

Vowel Reduction in Polish

by

Pawel Marcin Nowak

A.B. (Harvard University) 2001

M.A. (University of California, Berkeley) 2003

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John J. Ohala, Chair

Professor Ian Maddieson

Professor Alan Timberlake

Spring 2006



Abstract

Vowel Reduction in Polish

by

Pawel Marcin Nowak

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor John J. Ohala, Chair

It is well known that Polish vowels are heavily fronted and raised in the vicinity of consonants with a palatal component, but this phenomenon has not hitherto been analyzed in detail. In this work, we present a comprehensive analysis of the variability of Polish vowels, focusing on the role of consonantal context and its interaction with vowel duration. We also consider, however, other factors that are known to affect vowel articulation.

The study is based on data from four native speakers of Polish. Each subject was recorded reading an identical text twice, once at a normal speech rate and once at a fast speech rate, and contributed a short spontaneous narrative. The recordings were segmented with a semi-automatic amplitude-based algorithm and steady-states were automatically determined (defined as points at which the sum of the percentage rates of change of the first three formants was at its lowest). F1 and F2 of approximately 5000 tokens were measured at these points using a Linear Predictive Coding algorithm, and the

results were standardized. Each phonologically non-nasal vowel was classified according to a number of phonetic, phonological, morphological and syntactic criteria. The results were analyzed with multiple linear regression, with separate models constructed for different vowel classes in the height dimension (high, mid and low) and in the front-back dimension (front, central, back). In addition, a separate model was constructed for vowel duration.

Vowel duration and place of articulation of adjacent consonants were found to be very good predictors of formant variability in almost all cases, which yields strong support to Lindblom's (1963) target undershoot model of vowel reduction. As expected, the impact of consonants with a palatal component was particularly strong, but consonants with other places of articulation had a non-negligible effect as well. In all cases, the impact of the preceding consonants was stronger than that of following ones. Consistent with the findings of other researchers, we have found that vowel formants were influenced by several other factors (consonant manner, stress, content/function status of the word and speech style, among others). Word frequency, however, did not appear to play a significant role.

Table of contents

List of abbreviations	xii
Acknowledgements	xiv
1. Introduction	4
1.1. The undershoot model of vowel reduction	4
1.1.1. Lindblom (1963)	4
1.1.2. Undershoot vs. centralization	11
1.1.3. The role of duration	15
1.1.4. Summary and perspectives	25
1.2. Polish	26
1.2.1. The sound system of Polish	26
1.2.2. Vowel reduction in Polish	28
1.3. Outline of the dissertation	29
2. Methods	30
2.1. Speakers	31
2.2. Recording materials and procedures	32
2.3. Segmentation	33
2.4. Measurement procedures	35
2.5. Describing the data: phonology	38

2.5.1. Phonemic identity of the vowels	38
2.5.2. Consonantal context	41
2.5.3. Position in the word	42
2.6. Describing the data: morphology	44
2.7. Describing the data: syntax	45
2.8. Describing the data: word frequency	46
2.9. Summary of the variables recorded.	47
2.10. The statistical approach: an introduction	48
2.10.1. Multiple linear regression vs. analysis of variance	48
2.10.2. Principles of MLR: a brief summary	50
2.10.3. Automated search through predictor space and validation	56
2.10.4. Categorical variables	60
2.10.5. Statistical procedures: a summary	64
2.10.6. Assumptions	65
2.10.7. Correlation and causality	65
3. Summary statistics	68
3.1. Vowel phonemes	71
3.2. Subjects	73
3.3. Vowel duration	75
3.4. Speech rate	77
3.5. Comparative intensity	81

3.6. Word frequency	84
3.7. Preceding consonant place	87
3.8. Following consonant place	90
3.9. Preceding consonant manner	92
3.10. Following consonant manner	95
3.11. Preceding consonant voicing	97
3.12. Following consonant voicing	99
3.13. Number of following consonants	100
3.14. Speech style	102
3.15. Syllable position	104
3.16. Size of the phonological word	107
3.17. Postpausal phonological word	109
3.18. Prepausal phonological word	112
3.19. Clause-initial phonological word	115
3.20. Clause-final phonological word	116
3.21. Content vs. function words	118
3.22. Content vs. function syllables	120
3.23. Comparison of Euclidean distance scores	122
3.24. Conclusions	125

4. Vowel duration	127
4.1. Selecting the model	129
4.2. Categorical regression	135
4.3. Individual variables (MLR)	139
4.3.1. Intercept	140
4.3.2. Vowel	140
4.3.3. Speech rate	142
4.3.4. Syllable position	143
4.3.5. Clause-final phonological word	145
4.3.6. Following segment manner	147
4.3.7. Speech style	148
4.3.8. Preceding segment voicing	149
4.3.9. Number of following consonants	150
4.3.10. Comparative intensity	150
4.3.11. Preceding segment manner	152
4.3.12. Following segment place	153
4.3.13. Preceding segment place	155
4.3.14. Subject	156
4.3.15. Following segment voicing	157
4.3.16. Prepausal phonological word	157
4.3.17. Word content/function status	159
4.3.18. Possible interaction effects	159

4.4. Conclusions	160
5. First formant	163
5.1. Standardization	166
5.2. High vowels	168
5.2.1. Selecting the model	169
5.2.2. Categorical regression	173
5.2.3. Individual predictors (MLR)	176
5.2.3.1. Intercept	177
5.2.3.2. Following segment manner	178
5.2.3.3. Preceding segment manner	179
5.2.3.4. Preceding segment place	182
5.2.3.5. Following segment place	184
5.2.3.6. Comparative intensity	186
5.2.3.7. Prepausal phonological word	187
5.2.3.8. Speech rate	189
5.2.4. Summary	190
5.3. Mid vowels	190
5.3.1. Selecting the model	192
5.3.2. Categorical regression	196
5.3.3. Individual predictors (MLR)	199
5.3.3.1. Intercept	200

5.3.3.2. Vowel duration	200
5.3.3.3. Preceding segment place	201
5.3.3.4. Preceding segment manner	203
5.3.3.5. Following segment place	205
5.3.3.6. Following segment manner	207
5.3.3.7. Syllable position	211
5.3.3.8. Preceding segment voicing	214
5.3.3.9. Comparative intensity	215
5.3.3.10. Word content/function status	216
5.3.3.11. Speech style	216
5.3.3.12. Postpausal phonological word	217
5.3.3.13. Clause-final phonological word	219
5.3.3.14. Phonological word size	220
5.3.3.15. Vowel	220
5.3.4. Summary	229
5.4. The low vowel	231
5.4.1. Selecting the model	231
5.4.2. Categorical regression	235
5.4.3. Individual predictors (MLR)	238
5.4.3.1. Intercept	239
5.4.3.2. Vowel duration	239
5.4.3.3. Preceding segment manner	240

5.4.3.4. Following segment manner	242
5.4.3.5. Preceding segment place	244
5.4.3.6. Following segment place	245
5.4.3.7. Syllable position	247
5.4.3.8. Syllable content/function status	253
5.4.3.9. Preceding segment voicing	253
5.4.3.10. Speech style	254
5.4.3.11. Comparative intensity	255
5.4.3.12. Postpausal phonological word	256
5.4.3.13. Phonological word size	256
5.4.3.14. Number of following consonants	257
5.4.3.15. Following segment voicing	257
5.4.3.16. Clause-final phonological word	258
5.4.4. Summary	259
5.5. General discussion	260
5.5.1. Consonant place and duration	260
5.5.2. Other phonetic and phonological variables	264
5.5.3. Other relevant variables	266
5.5.4. Variables not selected into any model	267
5.5.5. Conclusions	268

6. Second formant	270
6.1. Standardization	273
6.2. Front vowels	274
6.2.1. Selecting the model	274
6.2.2. Categorical regression	278
6.2.3. Individual predictors	281
6.2.3.1. Intercept	281
6.2.3.2. Preceding segment place	282
6.2.3.3. Following segment place	283
6.2.3.4. Vowel duration	286
6.2.3.5. Speech rate	288
6.2.3.6. Other variables potentially interacting with place variables	290
6.2.3.7. Preceding segment manner	294
6.2.3.8. Following segment manner	297
6.2.3.9. Vowel	299
6.2.4. Summary	302
6.3. Central vowels	304
6.3.1. Selecting the model	304
6.3.2. Categorical regression	308
6.3.3. Individual predictors (MLR)	311
6.3.3.1. Intercept	311

6.3.3.2. Preceding segment place	312
6.3.3.3. Following segment place	314
6.3.3.4. Vowel duration	316
6.3.3.5. Postpausal phonological word	319
6.3.3.6. Word content/function	321
6.3.3.7. Other variables potentially interacting with the place variables	322
6.3.3.8. Preceding segment manner	324
6.3.3.9. Following segment manner	326
6.3.3.10. Vowel	329
6.3.4. Summary	331
6.4. Back vowels	333
6.4.1. Selecting the model	333
6.4.2. Categorical regression	337
6.4.3. Individual predictors (MLR)	340
6.4.3.1. Intercept	340
6.4.3.2. Preceding segment place	341
6.4.3.3. Following segment place	343
6.4.3.4. Vowel duration	346
6.4.3.5. Word frequency	348
6.4.3.6. Speech style	349
6.4.3.7. Preceding segment voicing	351

6.4.3.8. Following segment voicing	352
6.4.3.9. Prepausal phonological word	353
6.4.3.10. Postpausal phonological word	353
6.4.3.11. Clause-initial phonological word	354
6.4.3.12. Number of following consonants	354
6.4.3.13. Other variables potentially interacting with place variables	356
6.4.3.14. Preceding segment manner	357
6.4.3.15. Following segment manner	359
6.4.4. Summary	
6.5. General discussion	362
6.5.1. Consonant place and vowel duration	362
6.5.2. Other phonetic and phonological variables	367
6.5.3. Other relevant variables	368
6.5.4. Variables not selected into any model	369
6.5.5. Conclusions	370
7. General discussion	371
7.1. Vowel target undershoot in Polish	371
7.2. Stress	378
7.3. Non-phonetic and non-phonological variables	380
7.4. The effect of manner	381

7.5. Other phonetic and phonological variables	387
7.6. Conclusions	388
References	392
Appendix A	407
Appendix B	415

List of abbreviations

aff.	affricate
AIC	Akaike Information Criterion
alv.	alveolar
alveop.	alveopalatal
alvep.	alveopalatal
ant.	antepenultimate syllable
app.	approximant
clausef	clause-final phonological word
clausei	clause-initial phonological word
dent.	dental
dur.	vowel duration
Δ SED	change in the sum of Euclidean distances
encl.	enclitic
fcon	number of following consonants
freq	word frequency
fric.	fricative
fsm	following segment manner
fsp	following segment place
fsv	following segment voicing
int.	(comparative) intensity

lab.	labial
lat.	lateral
MLR	multiple linear regression
mono	monosyllabic word
nas.	nasal
pal.	palatal
palzd.	palatalized consonant
pen.	penultimate syllable
pos.	syllable position in the word
postp.	postpausal phonological word
pre.	preantepenultimate syllable
prep.	prepausal phonological word
procl.	proclitic
psm	preceding segment manner
psp	preceding segment place
psv	preceding segment voicing
pw size	phonological word size
scf	syllable content/function status
SED	sum of Euclidean distances
srate	speech rate
vel.	velar
wcf	word content/function status

Acknowledgements

I am immensely grateful to the members of my committee for their patience, helpfulness, wise advice and inspiration. Alan Timberlake asked just the right questions to help me clarify my thinking; I only regret that I have not had the opportunity to learn more from him during my years at Berkeley. Ian Maddieson was always available with his famous, vast knowledge in every relevant domain and his astounding attention to detail. John Ohala has been a source of inspiration ever since I accidentally came across one of his papers in my college days. The dissertation also benefited from discussions I had with Keith Johnson, Kenneth de Jong and Richard Wright. Of course, all the errors that undoubtedly remain are my sole responsibility.

Many other people had an important influence on my thinking about language and linguistics. I wish to acknowledge in particular the teaching and advice of Bert Vaux, Andrew Garrett, Sharon Inkelas and Juliette Blevins. I also learned a lot from my fellow graduate students and from my undergraduate students, all too numerous to mention.

In the Phonology Lab, Ron Sprouse's advice was invaluable, and not at all limited to technology and computers. Numerous fellow graduate students made it a stimulating and fun place to be around, among them Eurie Shin, Ryan Shosted, Anne Pycha, Julie Larson, Alan Yu and Reiko Kataoka.

I also owe a great deal to Belen Flores and Esther Weiss for all their help with all the administrative and bureaucratic issues and for prodding me to file on time so that I could drink champagne at the Commencement ceremony.

The dissertation would not have been possible without the generous help of my speakers, whose identities have to remain secret to keep the Committee for the Protection of Human Subjects happy.

My friends, in particular Bacci Willgohs, Paweł Rutkowski and Corey Yoquelet, have tried to keep me sane by reminding me, explicitly or otherwise, that work is not everything, and my family, especially my Parents, have been a constant source of support of all kinds throughout the years.

Thank you!

1 Introduction

Phonological descriptions of languages are frequently accompanied by vowel charts in which each vocalic phoneme is represented by a single point in a two-dimensional space.¹ Even the quickest glance at actual phonetic data reveals that this is a form of rather extreme idealization. Formant values of phonologically identical vowels vary widely by the token and it is more appropriate to identify vowel phonemes with regions of vowel space, although the problem here is that areas assigned to separate phonemes frequently overlap.

However, representing vocalic phonemes with single points in the vowel space is perfectly justified if it is assumed that each vowel has one specific *target* from which it deviates to a greater and lesser extent under the influence of *contextual factors*. This deviation is known as *vowel reduction* and a given vowel token may be said to be more or less reduced depending on its distance from the target.

Studying the mechanisms that govern vowel reduction is significant for a number of reasons. From the purely phonetic point of view, it has the potential to provide direct insights into the nature of coarticulation, a ubiquitous phenomenon with many implications for theories of speech perception and production. In addition, some recent studies have suggested that the degree of vowel reduction can be directly impacted by

¹ Allophones derived with categorical phonological rules might also be shown with separate points, but I eschew this complication at this point.

functional factors such as the speaker's concern for the listener's ability to comprehend the utterance. This makes vowel reduction a key issue in the ongoing debate about where to draw the boundary between phonetics and phonology (if indeed such a boundary is needed at all). This debate in turn has implications for the even more complex question of how to distinguish what is universal and what is language-specific.

This dissertation is an attempt to contribute to these debates by providing a comprehensive model of vowel reduction in one language, with a particular focus on the relative importance of the various factors that influence it. The language in question is Polish, and the work is based entirely on natural speech.

In the context of this work, "natural speech" means that variables are not experimentally controlled for (with very few exceptions) and that subjects perform tasks comparable to their normal patterns of language use. The advantages of such an approach are clear. Not only does it allow us to consider more variables, but it also enables us to get closer to our object of study – the natural linguistic behavior of human beings. However, such an approach also requires that the amount of data processed be far greater than in more tightly controlled experimental studies. Aside from time concerns, processing large amounts of data raises the problem of the degree of objectivity and consistency of the investigator's judgments necessary to transcribe and annotate it. One frequently-used method is cross validation by a number of researchers, but it was not possible to employ this option in this work. To address this issue in a different way, a significant background

concern in this work was defining and developing data-analytic procedures that are as objective and algorithmic as is feasible.

As far as Polish is concerned, the phonetic variability of its vowels has not hitherto been described in great detail. The language provides a particularly interesting ground for the undershoot model of vowel reduction, because it has a rather extensive set of consonantal place of articulation contrasts. It also has no phonological (i.e. neutralizing) vowel reduction, which is a two-fold advantage in this kind of study. Not only does it obviate the risk of confusing phonetic and phonological factors, but it also allows us to study significantly distressed or shortened vowels. (In a language with phonological vowel reduction, reduced vowel phonemes such as [ə] would typically be found in such positions). At the same time, it can be assumed that sound systems that display phonological vowel reduction originate in ones that do not, which means that studying the latter has clear implications for the former.

The remainder of this chapter discusses previous studies of vowel reduction in general and specifically in Polish and presents an outline of the work.

1.1 The undershoot model of vowel reduction.

1.1.1 Lindblom (1963)

The seminal paper for the study of vowel reduction was Lindblom's (1963) study. The experiment described in this study was based on the speech of a single male speaker of Stockholm Swedish, who recorded nonce CVC words involving the eight short (or lax) vowels of Swedish in three consonantal contexts (b_b, d_d, g_g). Each of the 24 CVC words thus obtained was produced in four different carrier phrases constructed so as to vary phrasal stress falling on the word in question. As a result, vowels were obtained with great variations in length (80-300 ms.). Lindblom summarized his results as follows: "the formant frequencies of all vowels are exposed to some change as vowel length decreases. This displacement is observed in the first, second and third formants, but is most conspicuous in the second. It appears to be a continuous monotonic function of vowel duration" (519).

Lindblom developed specific mathematical formulas predicting the values of the lowest two formants. According to these formulas, the observed formant values are a function of the target formant frequency of the vowel, the consonantal context and vowel duration. The basic idea is very simple: the shorter the vowel is, the more susceptible it is to consonantal influence. Therefore, the shorter it is, the more the observed formant

frequencies differ from the target formant frequencies. However, the precise relationship between the key elements of the model is rather complex.

The general formula for F2 takes the following form (Lindblom 1963:519, symbols slightly modified):

$$F2 = \alpha (F2i - F2t) * 2.73^{-\beta\tau} + F2t$$

where F2 represents the predicted steady-state formant values, τ represents vowel duration, α and β are consonantal constants (one per consonantal context, i.e. vowel-independent), F2t represents the target F2 frequency (one per vowel, i.e. consonant-independent) and F2i represents the F2 value observed at vowel onset. That last value is clearly an indirect representation of the acoustic target (or locus) of the preceding consonant, one that factors in the influence of the vowel on the consonant. Although the following consonant is not directly incorporated into the formula, it should be remembered that the formula is devised for perfectly symmetrical consonantal contexts.

It can be easily verified that the amount of duration-related undershoot in the second formant predicted by Lindblom's formula is directly related to the difference between the acoustic target of the consonant and the acoustic target of the vowel ($F2i - F2t$). It can therefore be said that Lindblom's formula indicates that the greater the acoustic distance between the vowel and the neighboring consonant is, the greater the deviation of the

vowel's second formant value from the target value will be in the same amount of time. In other words, the rate of formant change depends on the distance between the vowel and the consonant; it is the highest when the vowel and the consonant are the most distant acoustically. Of course, acoustic events reflect articulatory events. In particular, the rate of formant change reflects the speed of articulatory movements, and acoustic distances are correlated with differences in articulatory configurations. Lindblom's model thus suggests that the speed with which articulators move is directly correlated with the distance they have to travel.²

If we consider a hypothetical example of the vowels *e* and *o* produced with the same duration after a palatal consonant, the latter vowel can clearly be said to be more distant from the consonant in question, both acoustically and articulatorily. Lindblom's prediction in this situation is that there will be a greater difference between the observed and the target F2 frequency of *o* than between the observed and the target F2 frequency of *e*. The reverse might be the case after a different consonant, e.g. a velar. It should be noted that while this prediction is intuitively plausible, it is not logically necessary. It is perfectly possible to imagine a scenario under which the amount of undershoot is a function of a vowel duration variable and a set of consonantal constants only. In that case, *e* and *o* of the same duration would be predicted to differ from their respective targets by the same amount when pronounced after a palatal consonant (or any other

² We are not suggesting that the relationship between acoustic and articulatory events is straightforward. We will discuss the articulatory underpinnings of undershoot in depth in the final chapter of this work.

consonant, for that matter). In other words, under this scenario articulators move with the same speed regardless of the distance they have to travel.

Lindblom's general formula for the first formant is slightly different (1963:521, symbols slightly modified):

$$F1 = \alpha (375 - F1t) * 2.73^{-\beta t} + F1t$$

The variables and constants have the same meaning as in the case of the F2 formula, given the difference in formant numbering. The crucial observation is that there is no F1i value. Instead, the analogue of F2i – formant value at vowel onset – is a fixed value of 375 Hz. This value “refers to a frequency that, according to spectrographic measurements, the first formant comes close to almost instantaneously after the plosion during the transition from /b-/ , /d-/ or /g-/ to a following vowel” (521). This can be understood as indicating that the acoustic targets of labials, dentals and velars are approximately the same in the F1 domain, in contrast with the locus variability observed in the F2 domain.

The value of 375 Hz is not, of course, a phonetic universal. It has to be seen in the context of other formant values reported for the single Swedish speaker that Lindblom's study is based on. In fact, in Lindblom's data only non-high vowels have first formant values higher than 375 Hz. They range from 425-760 Hz. According to the formula, the

larger this value, the greater the amount that will be subtracted from the target F1 value of the vowel. This means that the more open the non-high vowel is, the greater the amount of F1-lowering in the same amount of time will be. Thus Lindblom's prediction is that the distance between the observed and the target F1 value will be greater in a low vowel such as *a* than in a mid vowel such as *e* with the same duration, in all consonantal contexts. Since, as was mentioned above, the value of 375 Hz in effect represents the acoustic target (locus) of the consonants considered by Lindblom, the F1 formula makes exactly the same prediction as the F2 formula: the amount of undershoot increases as the distance between the consonant and the vowel increases.

In the case of the second formant, Lindblom's formula predicts that target undershoot can result in formant values that are either higher or lower than the target F2 value, depending on whether the $F2_i$ value is higher or lower than the $F2_t$ value. In other words, the undershoot effect in the F2 domain does not have uniform directionality: whether F2-raising or F2-lowering is observed depends on the specific consonant-vowel combination. In the case of the first formant, we see that for all the vowels with target frequencies higher than 375 Hz – i.e. all the non-high vowels – the predicted effect is one of F1-lowering, regardless of the consonantal context. In other words, in the F1 domain, the undershoot effect has a uniform directionality for non-high vowels. This has the interesting consequence that although the undershoot effect may be more “conspicuous” in the F2 domain, as suggested by Lindblom, there is a sense in which it may be easier to describe in the F1 domain, because in the latter case we do not have to take into

consideration interactions between specific consonants and vowels. As suggested by Flemming (2006), this is also likely to be the reason why undershoot is more likely to be phonologized in the F1 domain than in the F2 domain. In the former case, duration-related variation in vowel formants can be abstracted away from the consonantal context; in the latter, it cannot.

In Lindblom's data, the three high vowels considered have F1 values slightly lower than 375 Hz, ranging from 325-370 Hz. The formula shown above predicts that in this case, decrease in duration should result in F1-raising. In other words, it predicts that the undershoot of high vowels should result in their becoming more open. This would suggest that the F1 acoustic targets (loci) of dentals, labials and velars are higher than the acoustic targets of high vowels. Given the fact that consonants require a higher degree of vocal tract constriction than vowels, this appears unlikely. In fact, Lindblom supplemented the F1 formula with a stipulation that the predicted F1 value will be equal to the target value for all vowels whose target F1 frequencies are lower than 375 Hz. In other words, he predicted *no* duration-dependent undershoot for high vowels. It is therefore interesting and somewhat puzzling to note that there are examples of such phenomena on the phonetic level. For instance, in Russian unstressed *i* and *u* are more open than their stressed counterparts (Alan Timberlake, p.c.). Flemming (2006) suggests that in such cases F1-raising may be due to vowel-to-vowel coarticulation with non-high vowels, but this issue clearly deserves further exploration.

Lindblom's (1963) formulas imply two major theoretical claims that are also made explicit in the discussion part of his paper. First, no special role is assigned to context-free centralization, which had been done in previous literature, for instance by Stevens and House (1963). On that analysis, vowels in "reducible" positions, e.g. unstressed, have a tendency to "degenerate" into schwa, regardless of the consonantal environment they are in. Lindblom argues instead that if reduction towards schwa is observed it is the result of assimilation to "schwa-like" elements in the environment (e.g. some vowels and "the rest position" of the vocal tract, 524). He also mentions, without endorsing but also without criticizing, Joos' (1948) view that centralization may be observed because the average effect of various consonants is centralizing; this view appears to be broadly in agreement with Lindblom's position, at least as far as F2 is concerned. Second, duration is seen as the only other factor that needs to be considered in addition to the segmental context, and the source of durational variation is irrelevant. Vowels are predicted to be reduced in the same way when shortened by destressing, increase in speech rate or any other factor, and none of these factors is predicted to have an additional influence on vowel spectra.

Subsequent research on vowel reduction has by and large focused on testing these two claims. The evidence for and against them will be now discussed in the two following sections. It may be worth pointing out that the two hypotheses are not entirely independent. Vowel reduction can only be expected to be inversely correlated with vowel duration if the underlying mechanism is contextual assimilation (undershoot). If

reduction is not caused by undershoot, but involves instead an attempt to approximate a point close to the center of the vowel space, in some consonantal contexts it could actually be expected to be *inhibited* by vowel shortening. However, there is no such necessary relationship in the opposite direction. Though extremely unlikely, it is possible that target undershoot is the cause of vowel reduction, but changes in vowel duration have little or no effect on its scope.

1.1.2 Undershoot vs. centralization

Delattre (1981) criticized Lindblom's view on the role of undershoot in vowel reduction, arguing that such coarticulatory influence is small. He analyzed acoustic and X-ray data from French, German, Spanish and English and concluded that each of these languages showed a different degree of vowel centralization, and that the way in which acoustic centralization was implemented articulatorily also differed by language. (Thus, in French there was little centralization of the tongue position but a significant degree of lip unrounding and unspreading, while in German the converse was observed). He also concluded that each language had a different "acoustic pole of attraction", i.e. that in each language the "center" of the vowel space towards which vowel formants moved was different. Delattre's results, while interesting, are seriously flawed. As pointed out by van Bergem (1993), he treated phonetic vowel reduction in languages such as French and Spanish on a par with phonological vowel reduction in English. Thus, the words recorded by his English subjects included such pairs as *disable~disability* and *abolish~abolition*, in

which a full vowel alternates with schwa. In his data, schwas seemed to preserve some information about the quality of the stressed vowel with which they alternated, suggesting that spelling pronunciation was involved.

Schouten and Pols (1979) studied spectral variation in a large corpus of Dutch CVC syllables recorded both as word lists and in a specially constructed passage. They concluded that consonantal influence on vowel midpoints was little different from random. However, consonantal influence did account for 3-9.5% of observed variance (more on F2 than on F1). While this is a small number in absolute terms, it has to be borne in mind that over 80% of variance was accounted for by the vowel's phonemic identity.

Koopmans-van Beinum (1980) observed that the vowels of Dutch speakers move closer to the center of the vowel space when they are unstressed and spoken in less formal speech conditions. She averaged over a number of consonantal contexts and she did not attempt to verify whether the overall centralization could be an average effect of differing consonantal contexts.

Nord (1986) analyzed data from four speakers of Swedish pronouncing bisyllabic content words constructed in such a way so that each vowel to be analyzed appeared in an invariant consonantal frame, with stress and word position independently varied. That consonantal frame always involved dental consonants, and Nord's conclusion was that

unstressed vowels in the initial syllable tended to coarticulate with them and move closer to the dental locus estimated at 350 Hz for F1 and 1650 Hz for F2. Final unstressed vowels, however, were reduced more towards the center of the vowel space (500/1500 Hz). The author ascribed this last result to coarticulation with the “relaxed position” of the vocal tract. His results are thus compatible with Lindblom’s theory, provided that it can be explained why final vowels were more susceptible to coarticulation with the rest position of the vocal tract than initial vowels. (The words were recorded in isolation, hence presumably the speakers’ vocal tract was in the rest position both before the start of the word and at the end of it). Perhaps it can be argued that anticipatory coarticulation in general tends to be stronger than perservatory coarticulation. Alternatively, it is possible that the consonants flanking the unstressed vowels in question had different coarticulatory effects in initial and final syllables, because they were pronounced in different ways. (This could be due to their location in the word or their location vis-à-vis the stressed vowels).

Van Son and Pols (1990, 1992) examined the realization of formant frequencies of Dutch vowels in a text read by a single speaker at two speech rates. The only statistically significant difference between the two speech rates was observed in F1. Curiously, for almost all the vowels the F1 value was *higher* in the fast speech rate, contrary to the predictions of the undershoot model. The authors suggested that the speaker may have spoken louder in the fast rate, which would result in more extreme jaw gestures. This, however, could not be verified, because the recordings had not been calibrated for

loudness. It should be pointed out that the amount of vowel shortening between conditions was only 15% in this study, resulting in a much smaller range of vowel durations than that examined by Lindblom (1963). It is conceivable that these between-rate differences should not be described as “vowel reduction” at all.

Fourakis (1991) examined the speech of eight Midwestern speakers of American English pronouncing the monophthongal vowels of their language in the frames [h_d] and [b_d]. Stress and speech rate were both independently varied, and the results were analyzed on a three-dimensional vowel space based on f_0 and the first three formants. It was found that the lack of stress and an increase in tempo tended to make vowels approach the neutral position in the three-dimensional vowel space, i.e. the value that would have been obtained for a uniform vocal tract. However, exactly the reverse happened for [ɪ] and [ɛ], which seems to clearly suggest that any putative “centralization”, if present at all, was overrun by contextual assimilation to the following [d], i.e. undershoot in Lindblom’s (1963) sense. These effects were not very robust statistically, but there was a statistically significant shrinkage of the overall vowel space brought about by changes in stress and tempo.

Van Bergem (1993) recorded 15 speakers pronouncing CVC syllables that happen to be existing content and function words in Dutch. He also found longer words in which these syllables appeared as stressed or unstressed, and he varied sentence accent. He analyzed

in detail the formant changes in one of the vowels, /ε/, and concluded that an undershoot account was preferable to a centralization account.

With the exception of the rather questionable data of Delattre (1981) and the data of VanSon and Pols (1990, 1992) whose validity is difficult to evaluate, the papers discussed in this section either support Lindblom's analysis of vowel reduction as target undershoot, or at least do not directly contradict it. They do not suggest that centralization should be considered a separate factor.

1.1.3 The role of duration

Fourakis (1991) and VanSon and Pols (1990, 1992) observed generally low correlation between vowel duration and the amount of acoustic vowel reduction. It will be recalled, however, that in the latter studies, the overall amount of durational variability of vowels was quite low (15% shortening in a fast speech rate). Other studies have generally observed that there is a relationship between a vowel's duration and its formant frequencies, but that one need also to consider other factors such as stress, speech style and lexical properties of words.

Harris (1978) examined data from three American English speakers pronouncing the nonsense words /əpɪpɪpə/ and /əpɪpɪbə/, with stress on either the penultimate or the antepenultimate syllable, pronounced at two speech rates. Unusually, the only formant

measured was F3, because F1 was obscured by background noise and F2 and F3 seemed to behave in a parallel manner. Harris concluded that stress and speech rate have different effects on vowel formants: the reductive influence of destressing is greater and more consistent across speakers than the influence of increased speech rate. She noted that stressed vowels spoken at the fast rate tended to have higher F3 values than the unstressed vowels spoken at the slow rate even when they were equal in duration. In addition, she pointed out that the length differences brought about by the voicing of the following consonant (in this case /p/ and /b/) do not seem to be correlated with formant frequency variation. On the basis of both the acoustic and electromyographic data, Harris concluded that stress involves “extra energy”, and suggested this explanation as an alternative to the duration-related part of Lindblom’s (1963) undershoot hypothesis.

Gay (1978) reported similar results. In his main experiment, four speakers of American English (trained phoneticians) recorded the nine monophthongal vowels of English in a [p_p] frame at two speech rates. Formant measurements made at the midpoint of the vowels showed that little variability was induced by speech rate changes. This raised the question of how the constancy of the acoustic target was achieved, given the reduction in duration. Two possibilities were entertained: increased coarticulation (articulator movement towards the vowel target begins closer to the consonantal closure) and increase in the speed of articulator movement after the release of the consonant. The examination of F2 onset frequency showed that at the faster rate it was closer to the target frequency of the vowel, thus supporting the first answer. The speed of movement did not

seem to change. In addition, a more limited corpus of [k_p_p] words with varied stress was created. It was found that there were no differences between vowels falling in the same stress condition at different speech rates other than duration. But unstressed vowels were consistently lower in amplitude and F0, as well as somewhat “reduced in color”, even when as a result of speech rate manipulation their duration was comparable to that of their stressed counterparts.

Nord (1986) found that Swedish speakers produced few spectral differences between stressed vowels, even when such vowels differed in duration. (Variations in duration were obtained by varying the position of the stressed syllable within the word). Unstressed vowels, however, tended to be reduced vis-à-vis the stressed vowels, even when their duration was comparable.

Fant et al. (1991) compared Swedish, French and English. They reported that the stressed and unstressed /e/ in French had exactly the same formant frequencies, despite the fact that unstressed vowels were shorter. While the authors were not explicit about this, it appears that the “expected” degree of reduction was observed in their Swedish and English data. The interest lies in the fact that the reported durations of unstressed vowels were roughly the same in all the languages studied (120-140 ms).

It appears then that there is a large body of evidence suggesting that changes in duration brought about by stress variation and changes in duration brought about by speech rate

variation do not result in similar changes in formant frequencies in several languages for which this problem was studied. A possible difference in the articulatory mechanisms underlying durational variations of the two types was suggested by Edwards et al. (1991), who analyzed the speech of four American English speakers using an optoelectronic tracking system. Their experimental material varied nuclear accent, phrasal position and speech rate. Accented syllables had longer opening and closing gestures with no change of peak velocity and with an increase in peak displacement for some of the subjects. Within the task-dynamic model, they interpreted this as indicating that durational changes related to accentuation result from changes in the *relative timing of adjacent gestures*; in other words, that the unaccented syllables are *truncated*.³ Syllables spoken at the slow rate, on the other hand, involved longer gestures with slower peak velocities, which they interpreted as a change in the *stiffness of gestures*. It should also be noted that Kuehn and Moll (1976), who used cineradiography, observed that different subjects achieved speech rate increases differently – some of them decreased articulatory velocity at the cost of displacement, which is consistent with the data of Edwards et al. (1991), while some of them increased it. (Cf. also the data of Gay 1978 discussed above). Thus at the articulatory level, the implementation of speech rate, but not stress, may be subject to individual variation. This is also supported by the electromyographic data of Tuller et al. (1982). The authors of that paper also suggested that the fact that the underlying mechanisms of durational lengthening in stressed syllables and in slow syllables are

³ Truncation is further discussed and modelled by Harrington et al. (1995), Carré and Chennoukh (1995) and Carré and Divenyi (2000).

different should not be seen as surprising, because in each case the goals of the speaker are different – achieving acoustic prominence in one case, slowing down in the other.

Koopmans-van Beinum (1980) analyzed speech of Dutch speakers in a variety of modes – vowels pronounced in isolation, isolated words, read text, retold story and free conversation. The observed correlations between duration and the degree of vowel reduction were not very high and they were significant for some of the vowels only. Discussing the discrepancy with the findings of Lindblom (1963), she noted that it arose because vowels tended to be more reduced than would be expected on the basis of duration alone in the “running speech” modes (retold story and free conversation). This raised the possibility that speech mode or style plays an independent role in vowel reduction, with the less “formal” modes allowing more reduction.

Another potentially related effect observed by some researchers is communicative intent, i.e. the idea that vowels pronounced “clearly” may show less reduction than other vowels, irrespective of their relative durations. Bernstein-Rattner (1985) examined a corpus of content and function monosyllables that were recorded spontaneously spoken by five English-speaking mothers to their children and to adult interviewers. Only words that could be matched for speaker, sentence position and intonational contour of the sentence were analyzed. She found that except for the “corner” vowels, the vowel space was more centralized in function words directed to adults, despite the fact that there were no

significant durational differences between the two registers. Most strikingly, the vowels [æ] and [ɪ] were clearly more centralized in adult-directed function words, despite the fact that their durations did not differ at all across the two registers. It is interesting to note that this effect was only observed in function words, which probably differed from content words in being unstressed and, as Ian Maddieson (p.c.) points out, more predictable from context.

Picheny et al. (1985) analyzed recordings of nonsense sentences spoken “conversationally” and “as clear as possible” by native speakers of American English. All five speakers increased vowel duration in the clear speech register, but the amount of the increase differed between the tense monophthongs (80 ms on average) and the lax vowels (20 ms on average).⁴ However, formant frequency differences were much greater in the case of the lax vowels, indicating a dissociation between reduction and duration. The authors suggested that this may have been due to the fact that the speech rate in the conversational mode was not fast enough to disturb the formant frequency patterns of the tense vowels at their midpoints.

Moon and Lindblom (1995) examined some American English vowels in a [w_l] frame, designed specifically to induce the greatest possible coarticulatory influence. They varied vowel duration not by varying stress or tempo, but by making use of the fact that vowels in English tend to shorten as the word in which they appear gets longer. Thus, they used

⁴ These values were not reported directly, but had to be read off from relatively small figures.

words of the form *will*, *willing*, *Willingham*. Five subjects pronounced these words in a “citation form” and then “as clearly as they could”; in addition, they recorded the vowels in the “null” context [h_d]. The main finding of the study was that though F2 differed from its null context value both in clear and in citation speech, in the latter the differences were greater. Having plotted F2 values against duration, the authors argued that these differences could not be reduced to durational variation, even though in clear speech the sounds were longer. There were vowels of approximately the same duration whose formant values nonetheless differed depending on whether they were “clear” or “citation” tokens. This effect, however, was not analyzed statistically, and speaker-dependent variation was considerable. In fact, the effect appeared to be really clear only in a handful of cases. In addition, some vowels in some of the subjects’ speech seemed to behave in a duration-independent way (in particular the “tense” vowels). The authors concluded that speaking clearly involves increased articulator speed that results in faster transitions between segments.

There are, however, several studies that question the existence of significant spectral variation between speech styles. Ladefoged et al. (1976) examined a range of American English vowels in a range of speech styles, ranging from very informal to the reading of frame sentences. They found that while there was a good deal of variation as such, “there was surprisingly little difference between styles” (230). Huang (1991) found little effect of speech style (read vs. retold story) on spectral properties of vowels, and suggested that the lack of such effects in American English as opposed to Dutch (cf. Koopmans-van

Beinum 1980) or British English (cf. Swaffield et al. 1961) may be best explained by sociolinguistic factors. Of interest is also the somewhat puzzling result of Summers et al. (1988), who compared speech produced in noise to speech produced in quiet conditions. They concluded that one could not claim that in the former the vowel space was expanded, i.e. vowels were less reduced. In fact, one of their two subjects appeared to centralize his vowels in the noisy condition. Despite that, the tokens recorded in the noisy condition were recognized more accurately in the perceptual part of the study, which the authors interpreted to be a consequence of increased amplitude and duration. Thus “speech in noise” was found to differ from “clear speech” in that it does not constrain vowel reduction, but it also results in improved intelligibility.

In recent years, there has been a growing interest in the impact of language use on linguistic structure. There is a large body of evidence that processes that can be described as “reduction” (i.e. failure to fully realize the target) are more common in more frequent words. For instance, Bybee (2000) presents evidence that final alveolar stops in English are deleted more often in more frequent words. According to Jurafsky et al. (2001), word frequency is correlated with different degrees of reduction because it provides a crude measure of word predictability. Analyzing English spontaneous speech, Jurafsky et al. argue that it is predictability that governs such reductive processes as word shortening, deletion and categorical vowel reduction (full vowel vs. schwa): the less predictable a word is, the less reduced it will be. Such behavior on the part of the speaker is readily explainable as an attempt to make the listener’s task easier (of course, not necessarily in

any way a conscious attempt to do so). Predictability, unfortunately, is very difficult to measure. Word frequency can capture it only in a very crude way, as many other complicating factors are involved (syntactic, semantic and pragmatic).

Several other researchers have presented evidence that variation in the scope of reduction that could be attributed to the speaker's paying attention to the listener's needs may be driven by more than predictability per se. Wright (2003) found that "hard" words had a more expanded vowel space than "easy" words, where "hard" words were those with higher neighborhood density and lower relative frequency. Thus it could be said that "hard" words were articulated more clearly. Interestingly, only the formant values of the "point" vowels seemed to be significantly affected, and not those of the mid vowels. Unfortunately, Wright did not report whether the vowels in those two types of words differed in duration.

It is also interesting to compare the results of Wright's (2003) study with those of Scarborough (2004). Investigating vowel nasalization rather than vowel reduction, she found that the degree of coarticulation was *greater* for the "hard" words. To account for this, she suggested that this type of coarticulation is actually beneficial to the listeners, because it enables them to clearly perceive the nasality of the following consonant. A parallel argument could conceivably be made for place-related coarticulation between a vowel and an adjacent consonant. It is thus possible that in a language other than English one could find the exact opposite of the phenomenon observed by Wright – i.e. vowel

space ostensibly shrunk in “hard” words. (“Ostensibly”, because that shrinkage would be a result of average consonantal influence, not a drive for centralization). In fact, as was mentioned above, one of Summers et al.’s (1988) subjects appeared to centralize his vowels in noise, with improvement rather than deterioration in their rate of recognition. What is problematic is the fact that both Wright (2003) and Summers et al. (1988) investigated American English. The issue clearly requires more research.

Finally, we need to consider views that do not question that there is a relatively straightforward correlation between the degree of vowel reduction and vowel duration, but question the causal nature of this relationship. This was the position espoused by Delattre (1981), who argued that Lindblom (1963) did not show that duration is the “primary cause” of vowel reduction, since it only reflects changes in stress and tempo. Van Bergem (1993) found generally high correlations between vowel durations and formant frequency displacements (though the result was less clear cut when data from individual speakers were considered separately). He nonetheless expressed his scepticism towards the view that vowel shortening *causes* vowel reduction; instead, he suggested that reduction and decreased duration *covary* as a result of tempo and style variation. Unfortunately, it appears that it may be very difficult to experimentally distinguish these two hypotheses.

1.1.4 Summary and perspectives.

Four decades after the publication of Lindblom (1963) paper, a large body of research has suggested that contrary to his claims, variations in speech rate and stress may not transform the articulation of vowels in the same way, even if they have comparable impact on vowel duration. In addition, evidence has been presented to the effect that factors such as speech style, communicative intent and lexical status of a given word may influence the spectral pattern of vowels independently of duration. At the same time, Lindblom's other major theoretical claim seems to have stood the test of time: no convincing evidence has been presented to the effect that in order to account for vowel reduction, one needs to invoke a context-free drive toward "centralization" in addition to coarticulatory influence of adjoining consonants.

One way to reconcile these findings is to assume that regardless of the factors that control the *scope* of spectral changes, the *cause* of these changes is always coarticulatory undershoot. Assessing the accuracy of this view is one of the major theoretical issues addressed in this dissertation. However, we consider it very likely that changes in vowel duration are in fact the most important influence on the degree of vowel reduction. Our primary focus will therefore be on determining the extent to which vowel variability in Polish can be captured by Lindblom's elegant and simple model of duration-driven undershoot.

1.2 Polish

1.2.1 The sound system of Polish: a brief introduction.

In this work, the generally accepted phonemic description of the Polish sound system (Stone 1990) is used in a slightly modified and simplified way. The consonantal system is illustrated in Table 1.1 below, and the vocalic system in Figure 1.1.

	Labial	Palatalized labial	Dental	Alveolar	Alveo-palatal	Palatal	Velar
Stop	p b	p ^j b ^j	t d			c ɟ	k g
Affricate			ts dz	tʂ dz _ɮ	tɕ dʑ		
Fricative	f v	f ^j v ^j	s z	ʂ z _ɮ	ɕ ʑ	ç	x
Nasal	m	m ^j	n		ɲ		
Trill				r			
Lateral				l			
Glide	w					j	

Table 1.1 The consonantal system of Polish assumed in this work (after Stone 1990).

i i u
 ε ɛ̃ ɔ ɔ̃
 a

Figure 1.1 The vocalic system of Polish assumed in this work (after Stone 1990).

The major difference between the system presented here and Stone (1990) is that the vowel [i̯] is treated as a separate phoneme, and not as a contextual variant of /i/ that

appears after non-palatalized consonants. From the perspective assumed in this work the key factor is that the difference between [i] and [i] is categorical. As such, it should be compared to the difference between, for example, [u] and [ɔ] rather than, for example, that between [u] produced in stressed and unstressed syllables. If [i] and [i] are indeed allophones, that relationship between them holds at a level of analysis so abstract so as to be irrelevant for the current work.

Moreover, many if not most other analyses of the Polish phonological system treat [i] as a separate phoneme (e.g. Rubach 1984, Sawicka 1995). Interestingly, they disagree as to its featural description: while Rubach (1984) treats it as a back vowel, Sawicka (1995) considers it to be front centralized. Examination of the data collected in this work, as reported later, suggested that “central” may in fact be the most appropriate classification and it is used here.

The other discrepancies between Stone (1990) and the system used here are comparatively minor. Thus bilabials and labiodentals are collapsed as one labial place of articulation. The consonant /ç/ is assumed to be phonemic, *pace* Stone, but in agreement with Sawicka (1995). In addition, Stone (1990) treats /l/ as dental rather than alveolar, which is clearly a mistake that is corrected here. Finally, obstruents classified as alveolar are strictly speaking post-alveolar or even retroflex, but this distinction was not added so as not to inflate the number of levels of the consonant place variables.

Further discussion of the phonological system of Polish will be presented in the following chapter, when the specific theoretical and classificational decisions made in the data analysis process are introduced.

1.2.2 Vowel reduction in Polish.

The language has no neutralizing phonological vowel reduction. It is also often asserted to have little or no phonetic vowel reduction. Thus Jassem (1962) observed no vowel reduction in unstressed syllables vis-à-vis stressed syllables. However, he compared stressed vowels to unstressed word-final vowels which, by his own account, tend not to differ greatly in duration. Sawicka (1995) claimed that vowel reduction of non-high vowels is rare and may occur only in phrase-final positions, as a result of the articulators slowing down and losing precision. (She also described vowel devoicing as fairly common in that position). She noted that high unstressed vowels are frequently deleted altogether, especially in posttonic position in words with antepenultimate stress (e.g. *prezydent* [ˈprɛzɪdɛnt] ~ [ˈprɛzdɛnt] “president”). More recently, Ramus et al. (2003) have argued that Polish is atypical as a stress-timed language precisely because it has no vowel reduction.

However, this opinion is largely due to the fact that these authors seem to equate vowel reduction with centralization. When vowel reduction is understood as a result of consonant-induced undershoot in Lindblom’s (1963) sense, it is clear that it is rather

common in Polish. Thus according to Sawicka (1995), all the vowels of Polish have fronted and raised allophones in the context of palatal and palatalized consonants, and in particular between two of them. In rapid speech, this may lead to the neutralization of the normally open-mid front unrounded vowel with the vowel /i/ (e.g. *pieniądze* [pʲɛ̃ɲɔndzɛ] ~ [pʲi̯ɲɔndzɛ] “money”). These facts can be accounted for very easily within the undershoot model. What is more, not all researchers agree that centralization is absent from Polish. Thus Rubach (1974), who analyzed recordings of radio speech aurally, claimed that vowel reduction to schwa is possible in rapid speech and can go as far as to loss of the vowel and creation of syllabic sonorants.

1.3 Outline of the dissertation.

In Chapter 2, the methods of data collection and analysis will be presented in detail. Chapter 3 will present general summary statistics about the data collected. Chapter 4 will present data concerning vowel duration. Chapters 5 and 6 will present analyses of the variability of the first and second formant, respectively. Chapter 7 will present a general discussion.

2 Methods

Research reported in this dissertation was based on natural and quasi-natural speech. The only variables that were experimentally controlled for at the time of data collection were speech style and (to a certain extent) speech rate. Other variables used to analyze the data were merely *observed* from the recordings. This put an emphasis on careful description of the obtained data, at all the levels that are potentially relevant (phonetic, phonological, morphological and syntactic). At the same time, an overarching concern was to make the description as objective and automatic as possible, that is to say, to limit to the absolute minimum the role of the investigator's subjective judgments.

A major consequence of this approach was that speech segments were by and large recorded according to their *phonological* rather than *phonetic* identity. That is to say, assimilations, deletions and insertions seen in the phonetic data were ignored, unless they resulted from regular phonological processes of Polish (according to Sawicka 1995). Crucially, this removed the need to rely on the often difficult decisions about the phonetic voicing, manner and place of consonantal segments. These features, in particular voicing and manner, are arguably continuous rather than categorical on the phonetic level, and any quantization that would have to be done to enable statistical analysis would have to be essentially arbitrary. Moreover, segments that are deleted, i.e. cannot be identified with a particular stretch of the sound signal, frequently leave traces of presence, whether

auditory, spectrographic or both. Thus, for example, a rhotic may be deleted, but the lowering of F3 may nonetheless be observed in the adjacent vowels.

Other variables were either assigned on the basis of preexisting sources (part of speech, frequency etc.) or were measured from the speech signal using algorithmic or nearly algorithmic procedures (duration, speech rate, intensity). The remainder of this chapter describes the process of data recording and annotation in detail. It also briefly discusses the principles of the statistical approach used in this work.

2.1 Speakers

The main part of this study is based on recordings obtained from four **subjects**. There were two women (W1, 44 years old and W2, 32 years old) and two men (M1, 27 years old and M2, 25 years old). Each speaker was originally from a different part of Poland, but all can be described as speakers of Standard Polish, with few dialectal features. All were college-educated and none was a professional speaker. Speaker M1 was fluent in English, while the others' competence in foreign languages was intermediate at best. Speaker M1 was also the only one who lived outside of Poland for any length of time (less than 3 years total) and had extensive linguistic training. None of the speakers reported any hearing or speech problems.

2.2 Recording materials and procedures

The text used in the recording was the final part of Krzysztof Zanussi's screenplay for the movie "Barwy ochronne" ("Camouflage"), which describes in detail a fight between two men (Zanussi 1978). The text was 514 words long, and the speakers were asked to read it twice: once at a normal pace and once as quickly as possible. Subsequently, they were asked to provide a short narrative on the topic of their choice. (Each speaker decided to tell a short story about recent events in their life.) The first part of the text, comprising 269 words, was used in the analysis, as well as a comparable initial part of each speaker's narrative.¹ In the remainder of this dissertation, the three types of recordings obtained will be referred to as *normal*, *fast* and *narrative* levels of the variable **style**.

The recordings were done in the subjects' homes, in conditions that were as quiet as was possible without using a sound-proofed room. A laptop computer was used, with the MobilePre USB audio interface with a sampling frequency of 44 kHz and the Shure SM10A unidirectional headworn microphone. The subjects expressed no concerns about the microphone and seemed to be comfortable wearing it.

¹ The fragments chosen were approximately the same length as the shortest narrative obtained.

2.3 Segmentation

Assigning boundaries to speech segments is a very complicated matter. Different approaches to the theory of speech production can conceivably lead to different decisions about where the boundaries between segments fall, and in many cases such boundaries have to be arbitrary anyway (e.g. between vowels and approximants). Rather than attempt to find the right way to segment speech, in this dissertation efforts were concentrated on finding a method that would be rapid and largely algorithmic, but could be overruled in the case of obvious errors. It was decided that the best solution would be to base the procedure on amplitude shifts in the speech signal, because it provides a criterion that can be applied to virtually any vowel-consonant boundary, which is not the case for spectrum-based criteria.

The problem with using amplitude for segmentation is that if it is calculated over a window that is too short, the measurement may become sensitive to individual glottal pulses. However, if the analysis window is too long, the amplitude-based procedure may fail to capture rapid changes that correspond to segment boundaries. A workable solution to this conundrum was found by using a relatively long analysis window (Gaussian with an effective duration of 10 ms.) with a very short timestep (1 ms.).

The aim was to locate points at which the shift in amplitude was large relative to the preceding and the following parts of the speech signal. For each window, the absolute

rate of change (absolute slope) of amplitude was estimated at the center of the window, by calculating the absolute difference between windows centered 5ms. before and 5ms. after the window in question. In the next stage, major and minor maxima in the slope of the amplitude contour were found. A major maximum was defined as a point where the slope of the amplitude contour was greater than in the preceding 20ms. and the following 20ms. In other words, such points could be found at most every 20ms.; in practice, they were found on average every 80ms.-100ms., which is not radically different from the average duration of segments in the data set. A minor minimum was defined as a point at which the slope of the amplitude contour was greater than in the preceding 5ms. or the following 5ms. In other words, such points could be found at most every 5ms.; in practice they were found on average every 14ms.-18ms. The investigator's task was limited to deciding whether a given major slope maximum corresponded to a real segment boundary, remove it if it did not and chose one of the minor slope minima as a boundary where necessary.

The procedure was remarkably consistent when boundaries between voiceless obstruents and vowels are concerned; in a great majority of cases, a major boundary was found within the first/last voicing period of the vowel. It is more difficult to evaluate the performance of the procedure in other situations, but its flexibility meant that in every case it was possible to obtain segmentation that appeared to be at least satisfactory.

The only real exception to this generalization concerned vowels adjoining glides or other vowels. In many cases, segmentation was possible, because clear and separate steady states (and transitions between them) could be seen in the formant tracks. Other cases, however, had to be treated as instances of coalescence. The vowels were identified, but they were not assigned to any particular phonemic category, and consequently they had to be excluded from the analysis. This included a substantial number of cases of coalescence of two identical vowels.

Examples of the operation of the segmentation algorithm are shown in Appendix A.

2.4 Measurements procedures

It was decided that measuring formant frequencies at vowel midpoints was sub-optimal, given that formant transitions into consonants were often asymmetrical duration-wise. (In particular, the transitions of alveopalatal consonants could be very long). Instead, a single steady state point was determined for each vowel using a Linear Predictive Coding algorithm (the Burg procedure, with 10 poles, a 25ms. window and resampling to 5500 Hz for women and 5000 Hz for men and a timestep of 1 ms., as implemented by Praat 4.3). For each window and each of the lowest three formants, the rate of change was calculated by finding the absolute difference between the preceding window and the following window. This value was then divided by the value of the formant at that point, which yielded a percentage rate of change. These percentage values for the three

formants were then summed up, and the steady state was assumed to correspond to the point where this summed up value was at its lowest. In addition, for each vowel a second point was estimated, corresponding to a similar minimum, but located within the middle 60% of the vowel.

Each vowel was then visually inspected, with the LPC formant trackings superimposed, to make sure that formant values would be collected in appropriate locations. In some cases, the procedure described above selected a point that was clearly located in the transitional part of the vowel. In about 2.8% of the cases, this could be dealt with by selecting the secondary steady-state, located within the middle part of the vowel. In a further 5.2% of cases, the steady-state location was adjusted by hand. Finally, in a small number of cases, involving primarily the high back vowel [u], conducting an LPC based measurement of formant values was deemed impossible and these tokens were excluded.

Examples of the steady-state location procedure are shown in Appendix B.

Formant frequency values and bandwidths were measured at these points using the same LPC parameters that were used to select steady-states. In addition, duration, average F0 and average intensity were recorded for each vowel. While **duration** could be considered directly in the models developed in the latter part of the study, this was not the case with f0 and intensity. As is well known, there is substantial variation in both intensity and f0

related to vowel height. To compensate for this, two new measures were devised: **comparative intensity** and **comparative f0**.

To calculate comparative intensity, all soundfiles were scaled to have the same average intensity. Then, average intensity values were calculated for each vowel of each subject. Next, the vowel's intensity value was divided by the average intensity value for the given subject/vowel combination, yielding a measure of intensity relative to the given vowel and subject. Next, this measurement was divided by the average of the same measurements obtained for the preceding and the following vowel. Thus a score greater than 1 indicated that the given vowel had higher intensity than the neighboring vowels, without considering differences due to vowel height. Conversely, a score lower than 1 indicated that the vowel's intensity was lower relative to the neighboring vowels.

Comparative f0 was calculated in a similar manner. However, it was not used in the main analyses reported in the subsequent chapters, because the pitch-finding algorithm of Praat 4.3 failed in a considerable number of cases. These included in particular many more heavily reduced vowels, i.e. those of great interest for the study.

The beginnings and the ends of pauses were determined with as much accuracy as was possible. When a perceivable pause was followed by a voiceless stop, the boundary was chosen so that the stretch assigned to the consonantal closure had a comparable duration to that of similar segments in that part of the waveform. Vowels appearing in

phonological words that followed a pause were marked as **postpausal**, while those appearing in phonological words preceding a pause were marked as **prepausal**. These two variables defined four different categories, since single phonological words appearing between pauses were both prepausal and postpausal, while many other phonological words were neither.

For the spontaneous narratives, inter-pause intervals were used to determine **rate of speech** in syllables per second. For the read texts, where the speakers produced few pauses, clauses were used to determine speech rate.

2.5 Describing the data: phonology

2.5.1 Phonemic identity of the vowels.

Each non-nasal **vowel** that could be clearly isolated from its neighboring segments was classified as one of the six oral vocalic phonemes of Polish (/i i ε a ɔ u/). In each case, it was assumed that the speaker had used the standard pronunciation of a given word. In other words, the actual phonetic pronunciation of the vowel in question was not considered in the determination of its phonemic status. According to the author's native intuitions about Polish, there were indeed no cases of categorical substitution of one phoneme for another among the oral vowels.

The so-called “nasal vowels” were not analyzed, but they nonetheless presented a problem. Polish has two orthographic symbols, “ę” and “ą”, which correspond to the modern reflexes of Common Slavic nasal vowels / \tilde{e} / and / \tilde{a} /.² Phonetically, they are not realized as nasal monophthongs, but typically sequences of a nasalized vowel followed by a nasal (before stops) or a nasalized glide (in other contexts). Their phonemic status is controversial – thus Stone (1990) treats them as separate phonemes, but Sawicka (1995) does not. According to her, the orthographic symbols in question correspond phonemically to combinations of non-nasal vowels and consonantal nasal elements. These phonemic nasal elements, however, cannot be identified with any of the nasal phonemes shown in Table 1.1, because in the word-final position, a fourfold contrast is possible (e.g. *ta* [t \tilde{a} w̃] “this” (instr.sg. fem.), *ton* [t \tilde{o} n] “tone”, *tom* [t \tilde{o} m] “volume”, *toń* [t \tilde{o} ɲ] “depth”). Sawicka suggests that this element should be identified as / η /. While not entirely unmotivated, this analysis is rather abstract, and it was not adopted here. Instead, word-final instances of [\tilde{a} w̃] and [\tilde{e} w̃] were not included in the main analysis with the phonemes [ɔ] and [ɛ], but were classified separately, as realizations of phonemic / \tilde{e} / and / \tilde{a} /.

As Sawicka notes, in non-word-final positions such four-fold contrast is not possible, unless an exaggerated spelling pronunciation is involved. In those contexts, there is no

² The correspondence is actually not straightforward, since several sound changes have occurred between Common Slavic and modern Polish.

difference between the pronunciation of the orthographic nasal vowels and orthographic vowel-nasal combinations. Thus, *kąta* “angle” (gen.sg.) and *konta* “(bank) accounts” (nom. pl.) are both pronounced as [kɔ̃nta], while the words *kęs* [kɛ̃w̃s] “bite” and *sens* [sɛ̃w̃s] “sense” rhyme. All those cases could either be treated as corresponding to underlying vowel-nasal sequences, in which case the vowels should be treated as all other instances of /ɛ/ and /ɔ/, or as examples of the phonemes /ɛ̃/ and /ɔ̃/, in which case they would not be analyzed together with oral vowels. The latter solution was chosen, partially because in the context of heavy nasalization that could be expected to be found there, the LPC measurements were likely to be rather unreliable.

The word-final orthographic “ę” is usually completely denasalized in the word-final position; in such cases, it was treated as an instance of the phoneme /ɛ/. Speaker M1 rather consistently realized nasal vowels (both orthographic “ę”/”ą” and vowel-nasal sequences) as the oral sequences [ɛw] and [ɔw]. Since it could not be determined beyond reasonable doubt that the [w] segment in these combinations is identical to the [w] representing the underlying glide /w/, these cases were excluded from the analysis.³

³ It is unclear what the source of this feature in this speaker’s speech is. Sawicka (1995) describes such pronunciation as characteristic of children and “uneducated” speakers; neither of these descriptions applies to this speaker. The speaker himself, who has linguistic training, believes that this pronunciation is a regional feature, but works on Polish dialectology do not confirm this (e.g. Dejna 1993).

2.5.2 Consonantal context

For each vowel, voicing, place and manner of the “phonologically preceding” and “phonologically following” consonant were recorded. As was implied before, this notion refers to the consonant that would be expected to precede or follow the vowel in question if there were no unexpected deletions or insertions and after all obligatory phonological rules have applied in the expected way.

The obligatory phonological rules of Polish include the following (Sawicka 1995): prepausal devoicing of obstruents, regressive voicing assimilation of obstruents and palatalization of underlyingly nonpalatal consonants before /i/ and /j/ (within a word). This means that the set of possible phonologically preceding and phonologically following consonants is somewhat richer than the set of underlying consonants shown in Table 1.1. It includes the segment [ɣ], which may arise from /x/ assimilating to a following voiceless obstruent, and palatalized dentals and alveolars (i.e. [tʲ] etc.).

The **place** of the preceding and the following segments was described as *labial*, *dental*, *alveolar*, *alveopalatal*, *palatalized*, *palatal* or *velar*. The category “palatalized” brought together palatalized labials and the rare cases of palatalized segments with a different major place of articulation (chiefly dentals). The **manner** of the preceding and the following segments was described as *stop*, *affricate*, *fricative*, *approximant*, *trill*, *lateral*

or *nasal*. The **voicing** of the preceding and the following consonants was described as *voiced* or *voiceless*.

Vowels that were phonologically not preceded by any consonants were excluded from the analysis. Otherwise, one would have to add a level called *none* to each of the abovementioned variables. This, however, would be a very heterogeneous collection of vowels adjacent to pauses and of vowels adjacent to other vowels and as such had little potential for explaining vocalic variance.

Finally, for each vowel the **number of consonants following** before another vowel within the same word was recorded. (More precisely, the number of consonants that would have followed without any deletions).

2.5.3 Position in the word

Boundaries between words and between phonological words were marked in the data. There is a difference between the former and the latter in Polish, because the language has a substantial number of clitics, i.e. unstressable words. Steffen-Batogowa (2000) summarized the controversies as to which Polish words should be considered clitics and on the basis of perceptual experiments suggested a fourfold division into non-clitics, semi-non-clitics, semi-clitics and clitics. In her scheme, the middle two groups comprised words with variable behavior and had a rather limited membership. In the current study,

they were disposed of, and the division was simplified into a binary one, which had the desirable effect of reducing the number of variables that were needed to be considered in the statistical analysis.

A vowel was *not* marked as appearing in a clitic if it appeared in one of the words classified by Steffen Batogowa as a non-clitic or a semi-non-clitic. These included all polysyllabic words, all nouns, adjectives, numerals, verbs, exclamations, indeterminate pronouns and nominative personal pronouns. All other monosyllables, classified as clitics and semi-clitics by Steffen-Batogowa, were classified as clitics in this study.

All non-clitic vowels in the dataset were described according to their “phonological” **position in the word**, i.e. ignoring any vowel deletions and insertions that may have occurred (as well as the presence of any clitics). The end of the word was used as the reference point. Due to their limited numbers, vowels appearing in the fifth syllable from the right edge of the word and beyond (i.e. beyond preantepenultimate) were grouped together as “other”.

A somewhat different procedure was needed for some combinations of monosyllables. As noted by Steffen-Batogowa (2000), the combinations of monosyllabic prepositions and monosyllabic personal pronouns, both of which are under normal circumstances clitics, act as complete words, i.e. receive stress on the first (i.e. penultimate) syllable. Similarly, when the clitic negative particle *nie* combines with a monosyllabic verb, it bears stress, as

though the whole combination was one word. Accordingly, in those circumstances vowels were numbered as if these combinations were single words.

Finally, for each vowel the **size of the phonological word** in which it appeared was recorded, in syllables. (Vowel deletions were not considered in this count, i.e. the measure reflects the “phonological” or underlying size of the phonological word).

2.6 Describing the data: morphology

Each vowel was classified according to the part of speech of the word in which it appeared (where non-obvious, these decisions were made on the basis of the *Cambridge-Klett Concise Polish Dictionary* and Grzegorzycowa et al. 1984). On this basis, the variable **word content/function** was created. Vowels appearing in nouns, adjectives, adverbs, verbs and numerals were marked as appearing in *content words*, while vowels appearing in auxiliary verbs, prepositions, pronouns, particles and conjunctions were marked as appearing in *function words*.

Using a similar binary distinction, Van Bergem (1993) showed that there was a significant difference in the degree of vowel reduction between content and function words in Dutch. However, the morphological structure of Polish is rather different from that of Dutch in one crucial respect: it is a much more inflectional language. As a consequence, grammatical information is often conveyed by endings that appear inside

items that are otherwise “content” words. It is therefore possible that an effect similar to that observed by Van Bergem for Dutch may pass unnoticed in Polish, unless a finer classification of the **syllable content/function** status is used. Vowels appearing in inflectional endings and in stem syllables of function words (as defined in the preceding paragraph) were classified as **function syllables**, all the other (i.e. stem syllables of content words) were classified as **content syllables**. (The decision whether a given syllable is part of an inflectional ending or not, which is not always obvious in Polish, was based on Grzegorzycykowa et al. 1984).

2.7 Describing the data: syntax

The text the speakers read and their transcribed spontaneous narratives were divided into clauses, i.e. syntactic units headed by a finite verb. In addition, units headed by adverbial participles were treated as separate clauses (cf. the discussion in Grzegorzycykowa 1996). The spontaneous narratives included numerous clause fragments and phrases appearing in appositions and as elements of lists; these were treated on a par with full clauses.

To examine the potential impact of syntactic structure on vowel reduction, it was recorded for each vowel whether it appeared in a phonological word that was **clause-initial** and/or **clause-final**. Combined together, these two variables described four categories, since a phonological word could also be both clause-initial and clause final in

the (rare) cases of one-word clauses, and neither clause-initial nor clause-final in the much more common case of clause-medial phonological words.

2.8 Describing the data: word frequency

Data about word **frequency** were obtained from the largest available corpus of Polish, provided by Polskie Wydawnictwo Naukowe PWN (<http://korpus.pwn.pl>). The corpus is based on 40 millions words found in a combination of books, newspaper articles, internet postings and recorded conversations. Token frequency rather than type frequency was used, since the highly inflectional character of Polish combined with frequent stem alternations means that a given vowel may only appear in one paradigmatic form of the lexeme.

One potential irresolvable problem stemming from the use of the PWN corpus was that it does not distinguish homophones. Thus for example the frequency count for the form *mamy* combines frequency information for the verb *mamy* “we have” and the noun *mamy* “moms”. This is unfortunate given that Gahl (2006) has shown that in the English Switchboard corpus low-frequency words are longer than their high-frequency homophones. However, homophony as such is likely to be much less common than in English.

2.9 Summary of the variables recorded.

Table 2.9.1 presents a summary of the variables recorded, describing the scale or levels used and referencing the section of this chapter in which they are defined. Where appropriate, the abbreviations which are used frequently in graphs in the following chapters are also shown.

Variable	Scale or levels	Sect.
subject	M1, M2, W1, W2	2.1
vowel	i i ε a o u	2.5.1
style	normal, fast, narrative	2.2
duration (dur)	continuous (in ms.)	2.4
comparative intensity (int)	continuous (unitless scale)	2.4
preceding segment place (psp)	labial, dental, alveolar, alveopalatal, palatalized, palatal, velar	2.5.2, 1.2.1
following segment place (fsp)	labial, dental, alveolar, alveopalatal, palatalized, palatal, velar	2.5.2, 1.2.1
preceding segment manner (psm)	stop, affricate, fricative, approximant, trill, lateral, nasal	2.5.2, 1.2.1
following segment manner (fsm)	stop, affricate, fricative, approximant, trill, lateral, nasal	2.5.2, 1.2.1
preceding segment voicing (psv)	voiced, voiceless	2.5.2, 1.2.1
following segment voicing (fsv)	voiced, voiceless	2.5.2, 1.2.1
number of consonants following in the word (fcon)	ordinal (consonants)	2.5.2
word position (pos)	(nonclitic) monosyllable, enclitic, final, penultimate, antepenultimate, preantepenultimate, other, proclitic	2.5.3
size of the phonological word (pw size)	ordinal (syllables)	2.5.3
word content/function status (wcf)	content, function	2.6
syllable content/function status (scf)	content, function	2.6
prepausal phonological word (prep)	yes, no	2.4
postpausal phonological word (postp)	yes, no	2.4
clause-initial phon. word (clausei)	yes, no	2.7
clause-final phon. word (clausef)	yes, no	2.7
speech rate (srate)	continuous (syllables/sec.)	2.8
word frequency (freq)	continuous	2.8

Table 2.9.1. Summary of the variables used in classifying data.

2.10 The statistical approach: an introduction.

The statistical analysis of the data in this work is based on multiple linear regression (MLR). Unless otherwise specified, the discussion of issues related to this technique is based on Cohen et al. (2003). This section introduces the principles of MLR and the specific way in which it is applied here.

2.10.1 Multiple linear regression vs. analysis of variance.

Cohen et al. (2003:4) note that in the behavioral sciences there lingers the belief that multiple linear regression techniques are “less scientifically respectable” than ANOVA-based techniques. This is, however, a misconception: both MLR and ANOVA are “special cases of the *general linear model* in mathematical statistics”; when all the independent variables are categorical and explicit cross-product variables are used in the MLR equation, MLR and ANOVA are mathematically equivalent. In fact, ANOVA was originally developed as a computationally simpler alternative to multiple regression, hence “any data analyzable by [ANOVA] may be analyzed by [MLR], whereas the reverse is not the case”.

The data considered in this dissertation are a clear illustration of the situation in which “the reverse is not the case”. ANOVA designs work best with approximately equal sample sizes in treatment groups (Gravetter and Wallnau 2000). There is no such

limitation with MLR.⁴ Since this study is based on samples of natural speech, no effort was made to keep sample sizes equal; in fact, given the number of variables that are considered, it is doubtful whether this would be even remotely possible.

MLR has a number of other advantages over ANOVA. First, continuous independent variables can be considered directly. Using a continuous variable like duration in an ANOVA analysis would require us to create a number of discrete levels (e.g. “short”, “medium” and “long”). This inevitably results in information loss, and in addition is very likely to be arbitrary. Second, unlike ANOVA, multiple linear regression provides a measure of effect size which is readily interpretable in real terms (the b-coefficient).⁵ Finally, in MLR there exist well developed techniques that enable the set of independent variables to be automatically trimmed down to include only the ones that are highly useful and as such can be expected to have predictive and explanatory power.

The only potentially problematic area in MLR is dealing with interactions. Unlike ANOVA, multiple linear regression does not automatically consider interaction effects. Interaction terms have to be added by the researcher or explored in different ways. It should be pointed out, however, that given the number of variables considered in this study, it would be absolutely impractical to consider all the possible interactions. The

⁴ In fact, as noted by Cohen et al. (2003:4) “standard computer programs now use a regression approach” in dealing with ANOVA designs with unequal sample sizes.

⁵ Eta squared and related measures express the proportion of variance accounted for in ANOVA (Cohen et al. 2003). They can therefore be interpreted as measures of relative effect size, but they do not provide an estimate of the effect size in real terms. For example, in our discussions of the b-coefficients for duration in chapters 5-7, we will interpret them as measures of the rate of change of formant frequencies. If an ANOVA design were used, we would not obtain a comparable measure.

precise way of dealing with them in this study will be explained in the subsequent chapters, when the actual data models are considered.

2.10.2 Principles of MLR: a brief summary.

A multiple linear regression analysis finds a formula of the type $y = \mathbf{b}_1\mathbf{x}_1 + \dots + \mathbf{b}_n\mathbf{x}_n + \mathbf{c} + \mathbf{e}$. In this formula, y represents the dependent variable, $\mathbf{x}_1 \dots \mathbf{x}_n$ represent independent variables (also known as predictors), $\mathbf{b}_1 \dots \mathbf{b}_n$ represent regression coefficients associated with the independent variables, \mathbf{c} represents the intercept (i.e. the value of y when all regression coefficients equal 0) and \mathbf{e} represents the error term. In other words, \mathbf{e} represents the difference between the observed value of y and the value predicted on the basis of the given combination of regression coefficients and independent variables. Multiple linear regression finds the optimal regression coefficients, i.e. the values with which the error term is minimized across all instances of y .

If the dependent variable and all the predictors use the same scale, the regression coefficients can be directly compared to illustrate the effect size, i.e. the relative importance of different predictors. However, this is unlikely to be the case. For instance, among the predictors used in this study are vowel duration, measured in milliseconds, and word frequency, measured in the number of occurrences of the word in the corpus. The \mathbf{b} coefficients assigned to these two predictors cannot thus be directly compared. One common way of circumventing the problem is to standardize all the variables. In this

case, if b_{dur} is 2, and b_{freq} is 1, the interpretation is that duration has a greater effect on the dependent variable than word frequency, because an increase of 1 standard deviation in duration measurement results in a change of 2 standard deviations in the dependent variable, while the change of 1 standard deviation in word frequency results only in a change of 1 standard deviation in the dependent variable. However, there is no obvious way of implementing this approach when some of the predictors are categorical, which is the case in this study. We will return to this problem shortly.

When b coefficients are calculated, they are accompanied by a **t-statistic** which evaluates their statistical significance. The null hypothesis is that the b is actually zero, i.e. that the predictor has no impact on the dependent variable. It should be stressed that significance of b and size of b are largely independent of each other. A b coefficient may be large, suggesting that the given predictor has a strong effect on the dependent variable, but it may actually not be statistically significant, because significance is a function of sample size (among other factors), but the size of b is not. Naturally enough, one does not want to attach importance to b coefficients that fail the significance test. However, an independent variable which fails the significance test may still be found to be a useful predictor of values of the dependent variable when proper validation methods are implemented, i.e. when the regression equation is used to predict values in a subset of data that was not used to construct the regression equation. In this situation, rejecting the non-significant independent variable is less obvious, because there is a strong possibility that it would have been found significant with a larger sample.

Residual standard error (RSE) is another useful value returned by MLR, because it provides direct information about how well the regression equation models the data in the units of the dependent variable. Thus, if we were to use MLR to predict formant values in Hertz, a RSE of 40 Hertz tells us that the values of y calculated by the MLR equation are likely to differ from the observed values by approximately ± 40 Hz. This enables us to evaluate the *practical* importance of the findings, i.e. the *practical* effectiveness of the model constructed. Moreover, RSE provides an unbiased estimate of the standard deviation of prediction errors in the population. In other words, if the regression equation were to be applied to the entire population from which the sample used to construct the equation was drawn, the standard deviation of errors (i.e. differences between predicted and obtained values) should be approximately equal to the RSE. This characteristic is useful in validating the model and will be discussed further in the following section.

Pearson's R, or the correlation coefficient, is a number between -1 and 1. An R of 0 means that there is no relationship between the dependent variable and the predictors, an R of -1 or 1 that there is a perfect correlation between them. The greater the deviation from 0, the stronger the relationship. The sign is not relevant for assessing the strength of the relationship, but only for its direction (direct vs. inverse correlation). The square of R, R^2 , can be interpreted as the proportion of variance in the dependent variable explained by a given combination of predictors.

Correlation strength can be assessed graphically in a direct way in linear regression with a single predictor by plotting the values of the dependent variable against the values of the predictor. Basically, the closer the points on the scatterplot are to the regression line, the stronger the relationship is. For example, figure 2.10.2.1 below illustrates the correlation between vowel duration and the first formant of the vowel *a* using the data from the current study. The R^2 value is 0.1879, indicating that 18.79% of variance in F1 values can be accounted for by variation in duration alone. The size of the correlation coefficient, R , is of course the square root or R^2 , i.e. 0.433. (We can see from the direction of the line on the chart that the relationship is direct rather than inverse, i.e. $R = 0.433$ rather than -0.433 .)

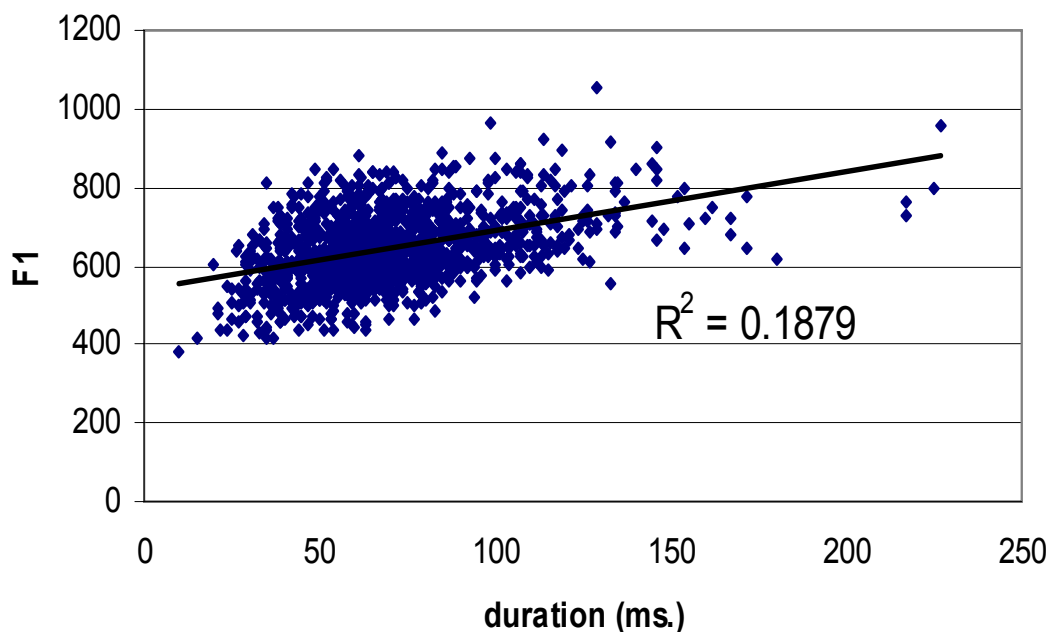


Figure 2.10.2.1 Correlation between vowel duration and the F1 values for the low vowel /a/.

When there are multiple predictors rather than one, a direct representation like that is not possible. However, MLR has the very interesting and useful property that the correlation coefficient between a dependent variable and the independent variables is equal to the correlation between observed values of the dependent variable and values predicted from the regression equation. The logic behind this equivalence is quite simple: predicted values are predicted on the basis of the independent variables alone, hence predicted values are perfectly correlated with the given set of predictors. It follows that the relationship between the observed and the predicted values of the dependent variable reflects the relationship between the dependent variable and the given set of predictors.

For example, figure 2.10.2.2 plots observed values of F1 of the low vowel /a/ vs. values predicted from a combination of five predictors: duration, place and manner of the following consonant and place and manner of the preceding consonant. R^2 of this relationship is equal to 0.325, indicating that $R=0.57$. These are also the values for the regression model used to obtain the predicted values, i.e. the values for the multiple correlation between F1 and the five predictors in question. Figure 2.10.2.2 thus illustrates indirectly but accurately the strength of the multiple linear relationship between the dependent and the predictors. As will be discussed in the following section, this property of regression models can be very useful in validating them.

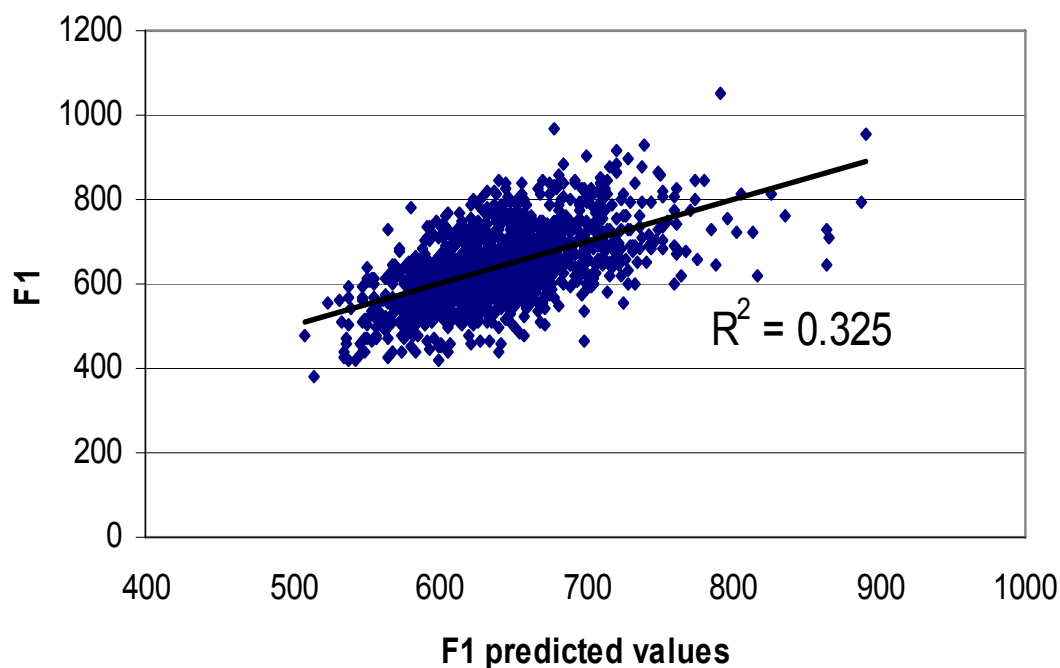


Figure 2.10.2.2 Correlation between observed F1 values for the low vowel /a/ and F1 values predicted from duration, preceding and following consonant place and preceding and following consonant manner.

When a predictor variable is dropped from the model, R^2 decreases. The amount of the decrease is known as **usefulness** of the variable. Estimating the importance of a predictor from usefulness is less preferable than comparing standardized regression coefficients, because usefulness reflects both the unique contribution of the variable to the regression model, as well as that which is shared with other variables. Usefulness can be also calculated from the square of **semipartial correlation** (also known as **part correlation**), i.e. correlation between the dependent variable and an independent variable when other independent variables are not allowed to vary.

Another useful value calculated by MLR analyses includes **partial correlation**, which differs from semipartial (part) correlation in that the variance attributable to other independent variables is removed from the dependent variable. (Thus, conceptually, partial correlation is related to the regression coefficients). Finally, one can calculate **zero-order correlations**, i.e. simple correlations between individual independent variables and the dependent variable.

2.10.3 Automated search through predictor space and validation.

When there are many potential predictors, it is not always clear which should be included in the model. For this reason, many automated methods of searching through predictor space have been developed. The method of choice in this work is based on the **Akaike Information Criterion AIC**), which estimates the log-likelihood of the models (Johnson 2006). Variables are added and/or removed from the model until a certain criterion is reached. The criterion assesses model fit against the number of predictors included in the model, i.e. there is a “penalty” for each variable added. The goal is to avoid the inclusion of variables with little explanatory power.

This is in turn motivated by the fact that regression models, in particular ones constructed using automated predictor search, may be overfitted, i.e. trained to fit the noise in the particular dataset rather than real relationships. (In fact, if the number of predictors were equal to the number of data points, the model would show perfect fit, even if there were

no true relationships in the data.) While the AIC partially alleviates the problem, more care is needed. One approach relies on evaluating the *practical* importance of a variable, in the light of the theory through which the data is analyzed. Thus we would be unlikely to want to pay much attention to a predictor whose contribution to multiple regression of formant values were 1Hz. But the best approach involves **validation**, i.e. assessing the success of the model in predicting values of the dependent variable beyond the data set used to construct the model.

Unfortunately, though validation is commonly advocated, there are no commonly agreed techniques beyond a few rules of thumb. For example, Schwab (2006) suggests defining a regression model, then splitting the sample into two subsamples and using the same predictors to construct regression models for each. The purely arbitrary criterion is that the R^2 values of the subsample models should not differ from the R^2 value of the main model by more than +/-5%. The main problem with this approach is that for each of the three models, separate regression coefficients are calculated, and these are by the very nature of multiple linear regression optimized for a given data set. It is thus possible that the three models will have actually different regression coefficients, but similar R^2 . Since regression coefficients are of the greatest theoretical interest as the best estimators of effect size, this approach is clearly unsatisfactory.

It seems that the best approach would involve constructing the model on one subset of the data (the **training set**), and then using the regression equation it provides to predict the

values of the dependent variable in a wholly separate data set (the **validation set**). The success of the model could then be assessed by comparing the observed and the predicted values in the validation set. The key question is how to find an appropriate criterion for doing that.

As has been discussed in the preceding section, the correlation between observed and predicted values of the dependent variable for a given data set is mathematically equivalent to the correlation coefficient (R) obtained by MLR for the same data set. If a model is successful, the correlation between predicted and observed values for the validation set should also be comparable to R obtained for the training set. It should be expected to be a little smaller, since, unlike the case with Pearson's R in the technical sense, it is based on regression coefficients calculated for a different subset of data.

In the remainder of this work, we will refer to the R -value obtained by correlating predicted and observed values in the validation set as " R " (and to its square as " R^2 "). This stresses the fact that these values only approximately reflect the relationship between the predictor variables and the observed values of the dependent variable in the validation set, because the correlation coefficients are optimized for the training set.

As will be shown when a concrete example is introduced, " R^2 " is not necessarily **maximized** by the inclusion of all the variables that maximize R^2 . In other words, some variables selected by the AIC method actually make predictions worse. To address this

issue, in this work new models were created on the training set with variables dropped one by one in the order that reversed the order in which they were selected by the AIC algorithm. Subsequently, new predictions were calculated for the validation set and “ R^2 ” was obtained. The collection of variables under which “ R^2 ” was maximized was then used to construct a **validated model** of the training set, though differences between R^2 and “ R^2 ” were also considered to assess how successful the validation was. In addition we compared RSE (residual standard error) with “RSE” (the standard deviation of differences between observed and predicted values in the validation set), since these values should also be similar if validation is successful. All the models discussed in this work were based on training sets comprising 75% of the data and validated on validation sets comprising 25% of the data, with random split between the two sets.

At this point, it may be appropriate to summarize the key terms discussed above in a tabular form, this is provided in table 2.10.3.1.

Term	Definition
training set	75% of the data, randomly selected
validation set	25% of the data, randomly selected
AIC model	MLR model of the training set including all variables selected by an automated AIC search
R^2	R^2 of the AIC model equal to R^2 of the correlation between predicted and observed values in the training set
“ R^2 ”	R^2 of the correlation between values observed in the validation set and values predicted for the validation set by the regression equation of the AIC model
RSE	residual standard error of the AIC model
“RSE”	standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the regression equation of the AIC model
maximized “ R^2 ”	maximum possible value of the correlation between observed and predicted values in the validation set using all variables from the AIC model or a subset thereof
validated model	MLR model of the training set including only the variables with which “ R^2 ” is maximized

Table 2.10.3.1 Summary of key terms introduced in this section.

The procedure defined here will be explained in more detail in the chapter 4 through an example (analysis of duration).

2.10.4 Categorical variables.

Multiple linear regression requires that independent variables be continuous. There are, however, two effective methods of using categorical variables in MLR. One of them requires that categorical variables be recoded as a series of dichotomous variables. The other, categorical regression, uses a technique known as optimal scaling to assign to the categorical variables numerical values that maximize model fit. The latter is less widely known and is currently only implemented in SPSS.

Dummy coding, which is one of the several possible coding schemes under the recoding approach, requires that one of the levels of a given categorical variable be selected as a reference level. For example, for the categorical variable *style*, we might select the level *normal* as the reference level. It is then effectively ignored in the fitting the regression line; in other words, the impact of *normal* speech style is reflected in the intercept only. For each of the other levels, a dichotomous variable is created; in the case at hand, these would be *fast* and *narrative*. All vowels spoken in the fast style are assigned the value of 1 on the dummy variable *fast* and 0 otherwise; conversely, all vowels spoken in the narrative style are assigned the value of 1 on the *narrative* dummy variable and 0 otherwise. The regression coefficients obtained for these variables can be interpreted directly as their contribution to the dependent variable in the units of the dependent variable, since the dummy variables are themselves unitless. Thus, in a hypothetical MLR of F1 in Hertz, the coefficients of -20 for *fast* and -30 for *narrative* would indicate that vowels spoken in these two styles are predicted to have F1 values 20Hz and 30Hz lower, respectively, than the vowels spoken in the *normal* style. All the dummy variables are entered into or dropped from the model as a block; however, t-values can be obtained for each dummy variable, indicating which levels of the categorical variable differ significantly from the reference level.

The overall regression model is not affected by which level of the categorical variable is chosen as the reference level, but this choice affects the ease of interpretation of the results. It makes most sense to select the level which could be expected to be the most

“neutral”, i.e. affect the dependent variable the least. In the current case, for most categorical variables, the level used was the one at which vowels could be expected to be the *least* reduced. However, the choice was irrelevant for categorical variables with two levels only and for some other variables no clear prediction could be made; in these cases the choice was random. The reference levels used with categorical variables in this work are summarized in table 2.10.4.1.

Categorical variable	Reference level
subject	m1
vowel	a
style	normal
duration (dur)	n/a (continuous variable)
comparative intensity (int)	n/a (continuous variable)
preceding segment place (psp)	dental
following segment place (fsp)	dental
preceding segment manner (psm)	stop
following segment manner (fsm)	stop
preceding segment voicing (psv)	voiceless
following segment voicing (fsv)	voiceless
number of consonants following in the word (fcon)	n/a (ordinal variable)
word position (pos)	penultimate
size of the phonological word (pw size)	n/a (ordinal variable)
word content/function status (wcf)	content words
syllable content/function status (scf)	content syllables
prepausal word (prep)	no
postpausal word (postp)	no
clause-initial word (clausei)	no
clause-final word	no
speech rate (srate)	n/a (continuous variable)
word frequency (freq)	n/a (continuous variable)

Table 2.10.4.1 Reference levels used in dummy coding the categorical variables.

However, dummy coding has one significant limitation. As was discussed before, the best way to estimate the relative contribution of different predictors is to compare their standardized regression coefficients. But with dummy coding, we do not obtain a single

regression coefficient for each categorical variable, but a number of them. This makes comparisons between different categorical variables difficult. To complicate matters further, it is basically impossible to compare regression coefficients associated with the different dummy variables with regression coefficients associated with continuous variables also used in the given regression model. This is simply because the required standardization cannot be effectively applied to dichotomies such as the dummy variables.

Categorical regression does not have this problem. It is based on a statistical technique of optimal scaling. It assigns to each level of a categorical variable a numerical value selected so as to maximize model fit. In addition, it standardizes all the data. As a result, a single regression coefficient can be calculated for each variable, continuous or categorical, and their sizes can be directly compared to each other to assess the relative importance of the variables.

However, categorical regression comes with problems of its own. One is that optimal scaling and standardization make it difficult to assess the differences between levels of the categorical variable in real terms. Another is that there are currently no implementations of categorical regression that allow for the predictor space to be searched automatically. Finally, given the fact that variables are scaled to maximize the fit, there is a danger of overfitting, i.e. discovering relationships where there are none. As was discussed above,

this could be alleviated with validation, but because variables are optimized for a particular data set, there is no obvious way of performing cross-sample validation.

It was decided that the best approach would combine the positive features of recoding and categorical regression. Thus, the procedure that was settled on for the models described in this work relied on selecting an optimal MLR model using the AIC and validation-based procedure described in the preceding section with dummy-coded categorical variables. The variables selected this way were then used to run categorical regression.

The procedure is only valid if regular MLR and categorical regression run on the same data set are truly equivalent. To assure this, usefulness (square semipartial correlation) of individual variables in the two models was compared. In addition, R^2 of the regular model and the categorical regression model were compared.

2.10.5 Statistical procedures: a summary.

The steps used in statistical analyses reported in the subsequent chapters are summarized in table 2.10.5.1. The analysis of duration in chapter 4 will provide an additional introduction and exposition through an actual example.

1	Select a combination of variables using the AIC procedure on the training set with dummy coding of categorical variables
2	Use the model constructed in step 1 to predict values of the dependent variable for the validation set
3	Drop the independent variables which do not maximize “R ² ” in the validation set
4	Use the independent variables left after step 3 to construct the validated model with the training set
5	Use the regression coefficients obtained for individual variables to explore the effect of categorical variables in detail
5	Use the independent variables left after step 3 to construct the validated categorical regression model with the training set
6	Compare the regular MLR model with the categorical regression model to assure their equivalence
7	Use the regression coefficients obtained in step 5 to assess the relative importance of independent variables

Table 2.10.5.1 Summary of steps used in statistical analyses.

2.10.6 Assumptions.

Multiple linear regression requires that a number of assumptions not be violated. These include linearity, normality and homoscedasticity. These assumptions are typically checked through a visual inspection of QQ normal plots and plots of fitted values vs. residuals. This procedure was followed here. Unless otherwise specified, it should be assumed that the assumptions were not violated.

2.10.7 Correlation and causality.

Association between variables is analyzed and tested statistically to draw inferences about causality. However, not even the strongest statistical relationship between two variables provides any evidence of a causal relationship between them in the absence of a well-defined theoretical framework. If x and y are strongly correlated, it may be because

x is influencing y, y is influencing x, both are being influenced by yet another variable or even because both are measures of the same property.

In this work, we assume that formant frequencies are likely to be influenced by the predictor variables we have selected rather than influence them. This is because formant frequencies provide indirect evidence about the shape of the vocal tract during the production of speech, which can be assumed to be a result of a combination of purposeful gestures by the speaker (reaching the target for the vowel, reaching the target for the next consonant) and physiological constraints affecting the movement of speech organs (such as duration of the interval in which the gesture is to be executed). This, of course, summarizes Lindblom's (1963) undershoot model of vowel reduction. However, it has to be remembered that the very existence of a causal relationship in this domain has been questioned: as was mentioned in Chapter 1, van Bergem (1993) suggested that changes in vowel duration do not *cause* changes in vowel spectra, but rather that they *covary* with them.

Results also need to be considered very carefully when we assess the potential impact of higher level variables. While it is extremely unlikely if not unimaginable that, for example, changes in vowel formants affect syntactic structure of the utterance, we cannot easily exclude the possibility that both are influenced independently by another variable we have not considered (e.g. the speaker's communicative intent).

In the following chapter, when the summary statistics about each predictor are presented, we will discuss the reasons why it may or may not have a direct causal effect on formant frequencies, and we will be returning to the issue of causality throughout the remainder of this work.

3 Summary statistics

In this chapter, we will present summary statistics of the data used in this work, considering in turn the *potential effect* of the predictors used. “Potential effect” means that at this point no statistical analysis of the data will be presented. Instead, graphical presentation of the data on the traditional F1-F2 vowel chart will be used to make the first assessment of the types of shifts that the vowel space undergoes in different conditions. When actual statistical analysis is applied in the subsequent chapters, many of these shifts will be validated as real effects of predictors on the independent variable, while others will not, suggesting in many cases that their presence in the graphical displays is an epiphenomenon caused by intercorrelations between different predictors.

Displaying the data graphically in the form of vowel charts works best when a small number of categories can be compared. For this reason, in this chapter the levels of most categorical and all ordinal predictors will be collapsed into two categories that could be expected to differ the most. (For instance, we will consider a binary division into “*palatal*” and “*nonpalatal*” consonants for the place variables instead of the seven place distinctions used elsewhere). For continuous variables, i.e. duration, speech rate, word frequency and comparative intensity, a different approach is needed to enable us to assess their potential effect using vowel charts. Accordingly, with these variables we will compare the outlying parts of the scale, i.e. the top and bottom 25% of tokens ranked

according to the given predictor. (For instance, vowels appearing in the 25% most frequent words with those appearing in the 25% least frequent words).

As will become apparent presently, not all shifts in formant frequencies necessarily make the overall vowel space substantially smaller. Thus, in addition to graphical displays, we will consider the sums of Euclidean distances between vowel formant means to assess the degree to which the vowel space has been shrunk or expanded in a given condition. Euclidean distance, applied to vowel spaces first by Liljencrants and Lindblom (1972), refers to the straight-line that can be drawn between any two vowels in a two-dimensional space. In other words, it is the hypotenuse of the right-angled triangle whose other sides delimit the distance between the two vowel points in question along the F1 and F2 axes. The greater the sum of Euclidean distances for a given condition is, the more dispersed the vowels are.

It should be stressed at this point that, although each number reported will be based on a substantial number of data points and as such will not be wholly unreliable, the Euclidean distance metric used here is not verified statistically. For this reason, it can only be used for assessment purposes. At the same time, the combination of graphical displays and Euclidean distance measurements will enable us to better formulate the questions that the statistical models reported in the following chapters are meant to answer.

It is also worth pointing out at this point that the phenomenon that may be reflected by a drop in the Euclidean distance score, i.e. diminished dispersion between the vowels, is often thought of as the defining characteristic of vowel reduction. However, in this work the term “vowel reduction” is used in a broader sense, to refer to all context-induced variability of vowels. To avoid confusion, we will use the term *vowel space shrinkage* to refer to the former phenomenon.

In addition, when each variable is introduced, we will briefly consider the theoretical reasons that may exist for its having an effect on vowel articulation. Given the lack of statistical verification, the discussion in this chapter will be entirely speculative.

All the formant measurements and Euclidean distance scores reported in this chapter will be provided in Hertz, without scaling to take into account the effect of differences in vocal tract size between subjects. As will be shown presently, all the subjects provided approximately equal number of tokens and, subject to exceptions noted, there are no reasons to expect that different independent variables have different distributions in the data collected from different speakers. This means that intra-speaker differences are likely to cancel each other out, which is sufficient for purely exploratory purposes.

3.1 Vowel phonemes

Using the methods and materials outlined in the preceding chapter, a total of 4923 vowel tokens were collected and used in this study. Given the uncontrolled nature of the material, there are large differences between the number of tokens of individual vowel phonemes.¹ The numbers in question are summarized in table 3.1.1. The table also shows the letters that the Polish alphabet uses for the vowels. These will be used from this point on for typographical convenience. Note that except for the use of the letter *y* for the central vowel /i/, the divergences from the International Phonetic Alphabet are minor.

Phoneme (IPA)	Polish alphabet	Number of tokens	Number of tokens (%)
i	i	316	6.4%
ɛ	e	1252	25.4%
a	a	1350	27.4%
ɔ	o	1156	23.5%
u	u	345	7%
ɨ	y	504	10.2%

Table 3.1.1 Breakdown of the data by vowel phoneme.

¹ These differences are likely to be motivated largely by differences in the underlying distribution of oral vowel phonemes in Polish, exacerbated by the fact that the LPC tracking algorithm had more problems with some vowels than others and some tokens of the former could not be analyzed (in particular *u*, as well as *i* in some heavily reduced contexts). No effort was made to equalize the number of tokens that fall in the different categories. To do so for all the levels of all the 22 independent variables considered would simply be impossible. To do so for the vowels only would be at least theoretically feasible. It would require, however, that many more words be considered, and the more frequent vowels occurring in them be rejected, which would disproportionately increase the amount of time needed to process the data. The annotation method used (multiple text tiers in Praat 4.3) was very efficient largely because it relied on providing a single phonemic tier (thus intervocalic consonants needed to be transcribed only once) and single annotations for higher-level domains spanning several vowels (words, prosodic words, clauses). It should be stressed at this point that equal sample sizes are *not* a necessary condition of multiple linear regression which will be used to analyze the data statistically in the following chapters.

Figure 3.1.1 presents shows *all* the data points used in this study in one two-dimensional vowel chart. As can be expected, the variation that can be seen is considerable: within the vowel space delimited by the most extreme data points, there are actually few areas that are not totally covered by the vowel tokens. As can also be expected, there is a considerable degree of overlap between areas belonging to individual vowel phonemes. Figure 3.1.2 presents the grand mean of individual vowel phonemes (across all four subjects).

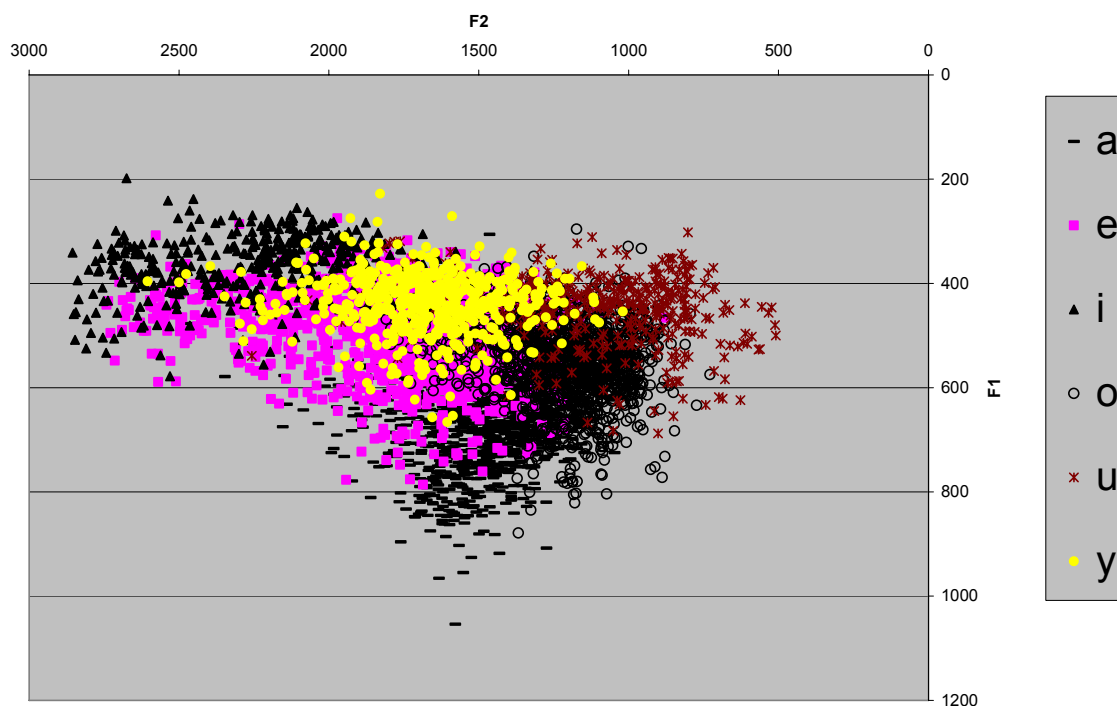


Figure 3.1.1 All the data points used in this study.

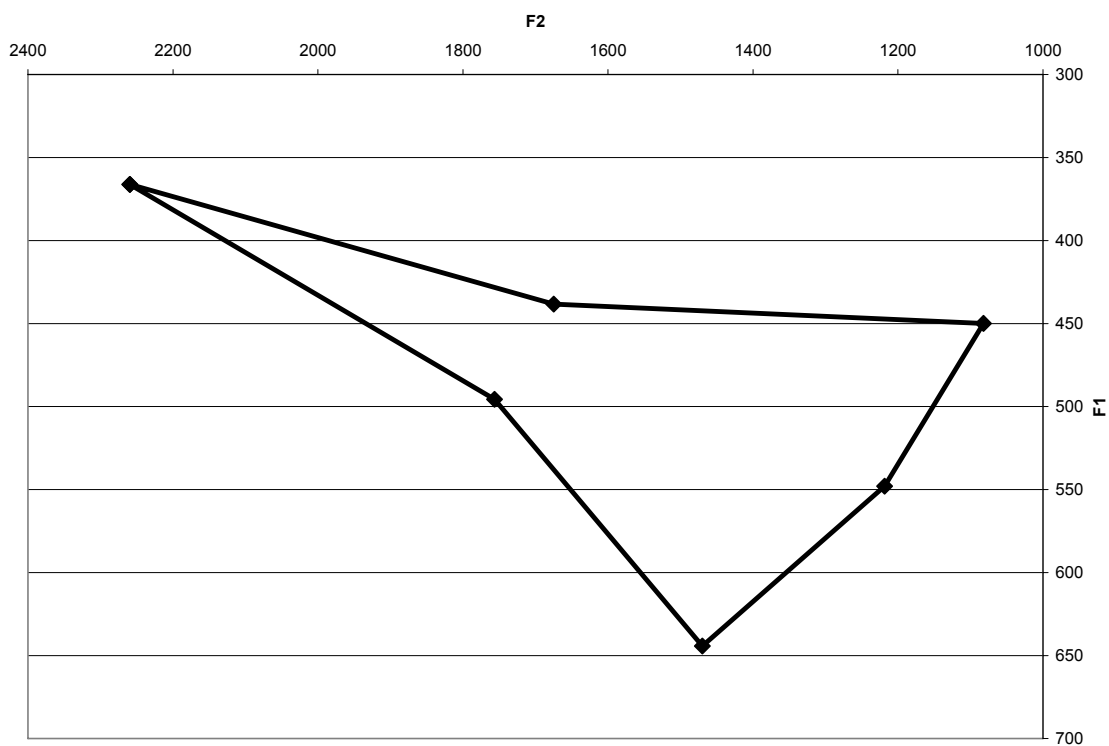


Figure 3.1.2 Grand means of vowel phonemes.

The mean sum of Euclidean distances for the entire data set is 15872.

3.2 Subjects

A large part of the variability seen in figure 3.1.1 is of course a result of differences between the sizes and perhaps shapes of the individual speakers' vocal tracts. These differences can be assessed from figure 3.2.1, where the vowel space of each speaker as well as the grand mean are shown separately. Unsurprisingly, the two male speakers have

lower formant frequencies and smaller vowel spaces, which is a natural consequence of the fact that male vocal tracts tend to be bigger.²

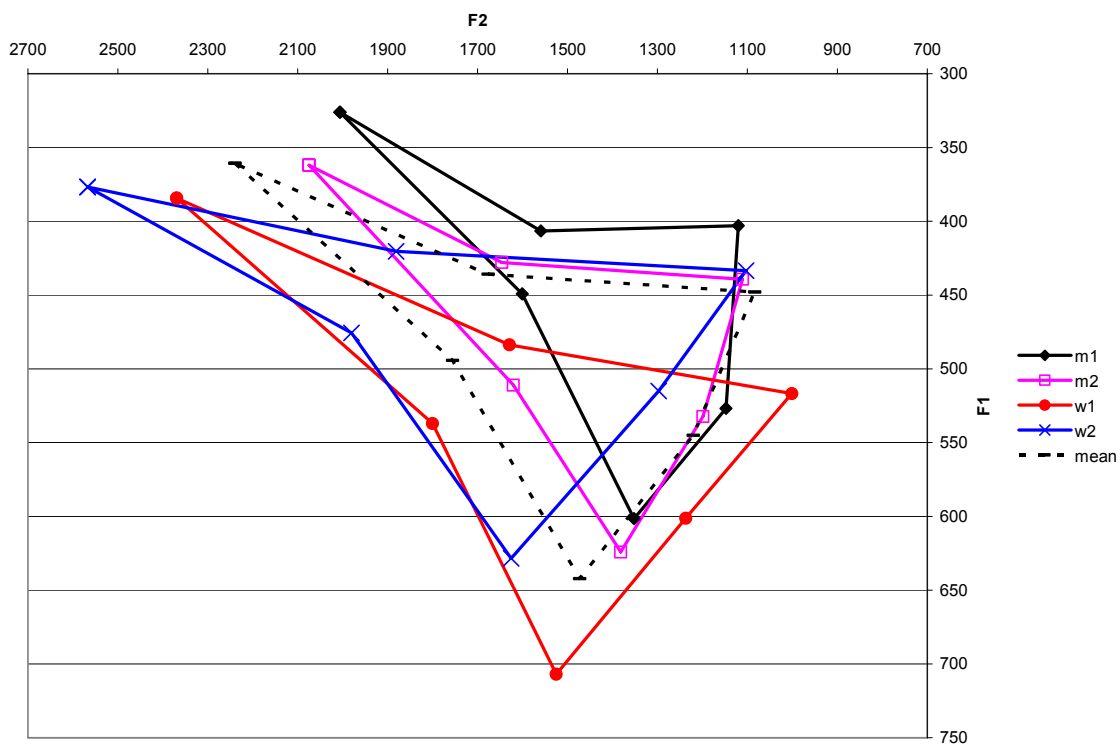


Figure 3.2.1 The vowel spaces of individual speakers.

Table 3.2.1 presents the number of tokens that each subject contributed to the data set.

Subject	Number of tokens	Number of tokens (%)
m1	1225	24.9%
m2	1175	23.9%
w1	1317	26.8%
w2	1206	24.5%

Table 3.2.1 Breakdown of the data by subject.

² Interestingly, for this set of speakers, the correlation may not hold *within* the sexes: speaker m2 is taller than speaker m1, and speaker w1 is taller than speaker w2, which is surprising in the light of the data in figure 3.2.1. (This assumes, of course, that height and vocal tract length are highly correlated).

3.3 Vowel duration

To illustrate the potential effect of vowel duration, this variable was standardized for each vowel, as otherwise it was likely to be obscured by intrinsic durational differences between vowels. To achieve this, within each vowel phoneme category, each duration score was converted to a z-score. Figure 3.3.1 compares the vowel spaces of the shortest 25% of vowel tokens (1231 tokens) with the longest 25% of vowel tokens, while table 3.3.1 presents the sums of Euclidean distances in these two subgroups, as well as percentage drop in the sum of Euclidean distances between long vowels and short vowels.

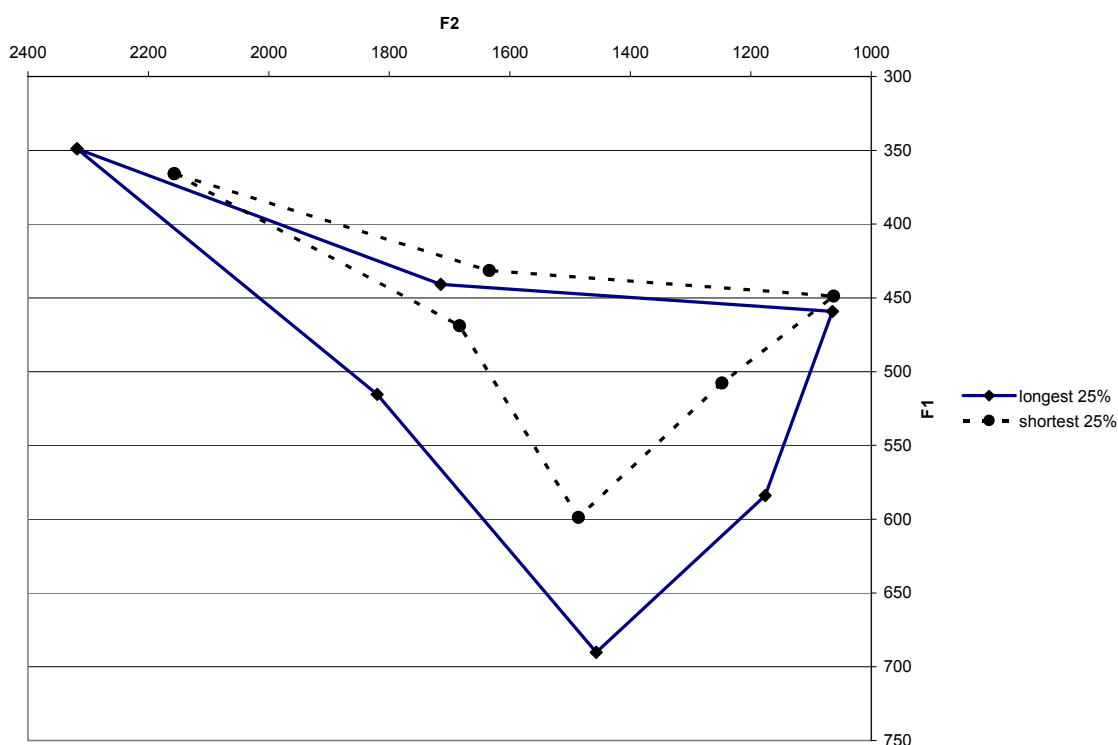


Figure 3.3.1 The potential effect of vowel duration.

Condition	SED	Δ SED
longest 25% of vowel tokens	17831	0%
shortest 25% of vowel tokens	14369	-19.4%

Table 3.3.1 Sums of Euclidean distances between vowel grand means in the longest and the shortest vowels (standardized). (SED – sum of Euclidean distances, Δ SED – percentage change in the sum of Euclidean distances).

Figure 3.3.1 shows a considerable difference between the longest and the shortest vowels. There is clearly a large shrinkage of the overall vowel space, confirmed also by the change in the sum of Euclidean distances. This potential effect can be decomposed into two factors. First, all the vowels except *i* are raised (i.e. have a lowered F1); nonhigh vowels are affected particularly strongly. Second, all the vowels except for *u* appear to be centralized in the F2 dimension (i.e. their F2 is lower or higher depending on the vowel's frontness). The F1 lowering effect is exactly what would be expected under Lindblom's (1963) undershoot model of vowel reduction: all consonants are likely to have lower F1 loci than most vowels, and the effect is likely to be greater the shorter the vowel is. (It has to be remembered, however, that Lindblom stipulated that there is no undershoot in high vowels). The source of the F2 effect is less obvious. One possibility is that the fronter vowels are more susceptible to the influence of F2-lowering consonants (e.g. labials) and the more back vowels are more susceptible to the influence of F2-raising consonants (e.g. palatals).

Factors affecting vowel duration are discussed at length in chapter 4.

3.4 Speech rate

Speech rate varied between 1.8 and 10.8 syllables per second. Figure 3.4.1 presents the breakdown of average speech rates by speech style and by subject. As is apparent, all subjects succeeded in making an appreciable difference between the normal reading and fast reading style, though the amount of the difference is not the same for all speakers (except for m2 and w2). This intra-subject difference appears to have been caused mostly by how quickly they were able to read in the fast reading style, rather than differences in the normal reading style. One interesting point to note is the fact that with the exception of speaker m1, all the subjects used the same speech rate for the normal reading style and the narrative style. (We will return to this point when we consider the potential effect of speech style in section 3.14).

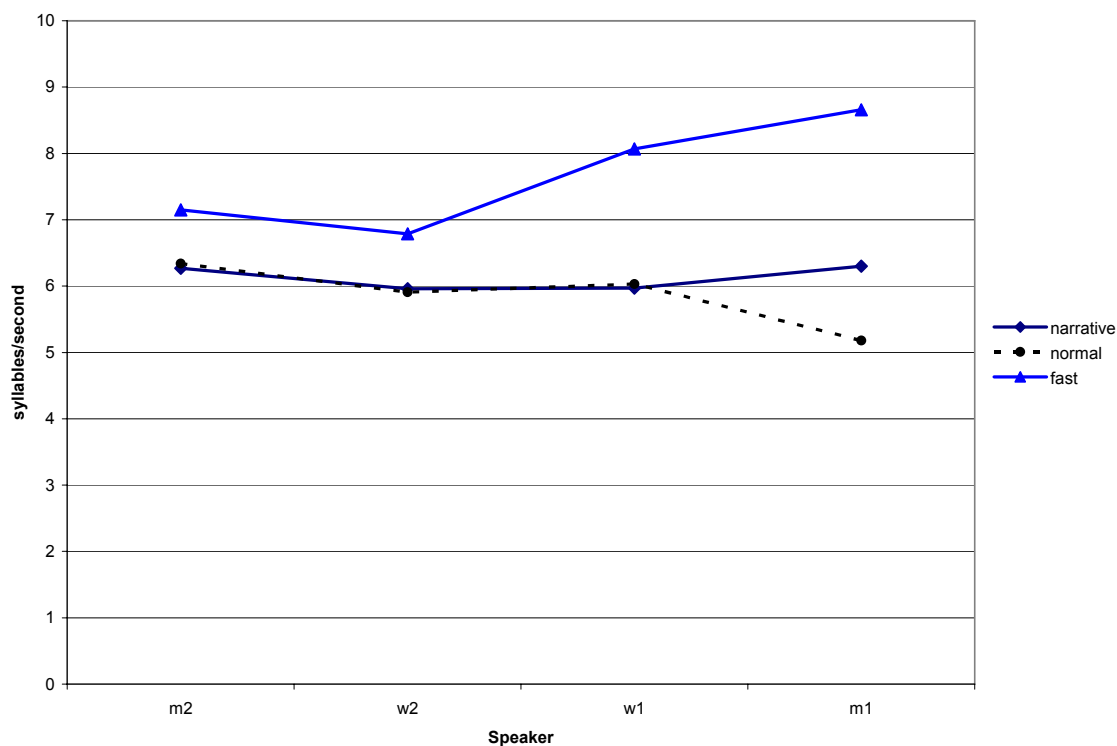


Figure 3.4.1 Breakdown of speech rate by subject and speech style.

Figure 3.4.2 shows the difference between the mean vowel spaces of the 25% of vowels that were spoken at the lowest rates and the 25% of vowels that were spoken at the highest rates. Table 3.4.1 shows the sums of Euclidean distances for these two categories.

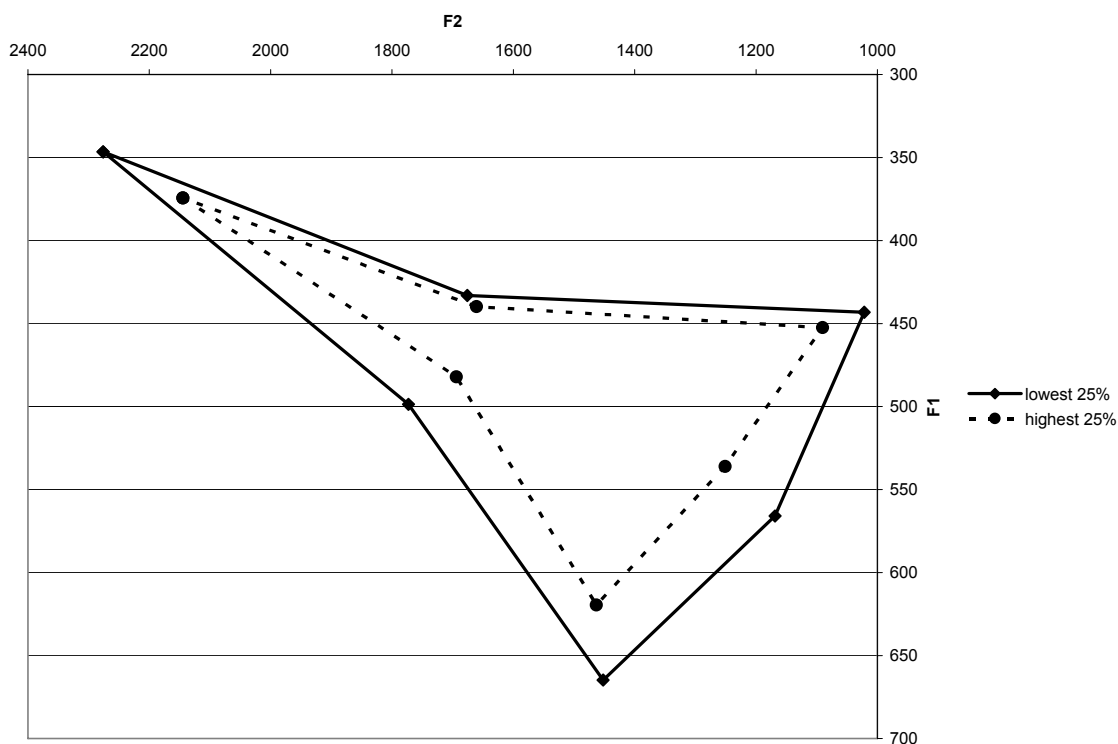


Figure 3.4.2 The potential effect of speech rate.

Condition	SED	Δ SED
25% of vowel tokens spoken at the lowest rates	17435	0%
25% of vowel tokens spoken at the highest rates	14181	-18.7%

Table 3.4.1 Sums of Euclidean distances between grand means of vowels spoken at the lowest and the highest speech rates.

The potential effect of speech rate appears to clearly be one of “classic” vowel space shrinkage. Of course, speech rate is very likely to be highly correlated with vowel duration and the question arises whether the former can be fully subsumed by the latter. To be more precise, the question is whether speech rate should be expected to have any additional effect on the spectral properties of vowel beyond the one it exerts as a likely determinant of vowel duration. As can be seen, the shifts in vowel space due to lower

duration (shown in the preceding section) and higher speech rate are not identical. However, at this point it is worth remembering that the data have not been subject to statistical analysis and the difference may be nothing more than an unintended consequence of the way data points have been sorted. In fact, this is likely to be the explanation for the fact that at the highest speech rates, high vowels appear to have a higher F1 than at the lowest speech rates, while the shortest high vowels (except *i*) have a lower F1 than the longest high vowels. As can be seen in figure 3.4.1, in his normal reading style, speaker m1 used a rate of speech that was lower than that of the other speakers. This means that his tokens are likely to be overrepresented among the low-rate tokens, which is in fact the case (503 out of 1231, i.e. 40.9%). As can be seen from figure 3.2.1, he also happens to be the speaker with the lowest overall formant values; his tokens may thus be pulling the means disproportionately down.

This reasoning, however, cannot explain why the speech rate effect is different from the duration effect in the non-high vowels, which the graphical displays suggest. Several explanations could be suggested for why that is the case. First, it may be pointed out that both the speech rate variable and the duration variable measure vowel duration with some degree of error, due in the former case to the fact that a higher level domain is used and in the latter to the fact that the semi-automatic segmentation algorithm used in this work is unlikely to have been perfect (leaving aside the intrinsic problems involved in segmenting vowels and consonants such as sonorants and voiced fricatives). Conceivably, these errors work in different directions with the two measures. (For

example, vowels spoken at the highest rates might turn out *not* to be the shortest vowels, because at higher speech rates consonantal constrictions may be reduced and a longer stretch may be assigned to the vowel). Alternatively, speech rate may indeed affect vowel articulation beyond its indirect effect through vowel duration. Since non-high vowels spoken at the highest rates appear to have *higher* F1 values than the shortest vowels, the explanation would have to be that at the highest speech rates speakers overcompensate by making their vocalic tongue gestures more extreme and more rapid. Finally, neither answer may be needed if the speech rate effect is not found to be significant by the statistical analyses presented in the following chapters.

3.5 Comparative intensity

Comparative intensity, i.e. the intensity of the vowel relative to the neighboring vowels with inter-vowel differences in inherent intensity subtracted away, varied between 0.752-1.38. Figure 3.5.1 compares the vowel spaces of the 25% of vowels with the lowest comparative intensity with the 25% of vowels with the highest score on this metric. Table 3.5.1 shows the sums of Euclidean distances in these two conditions.

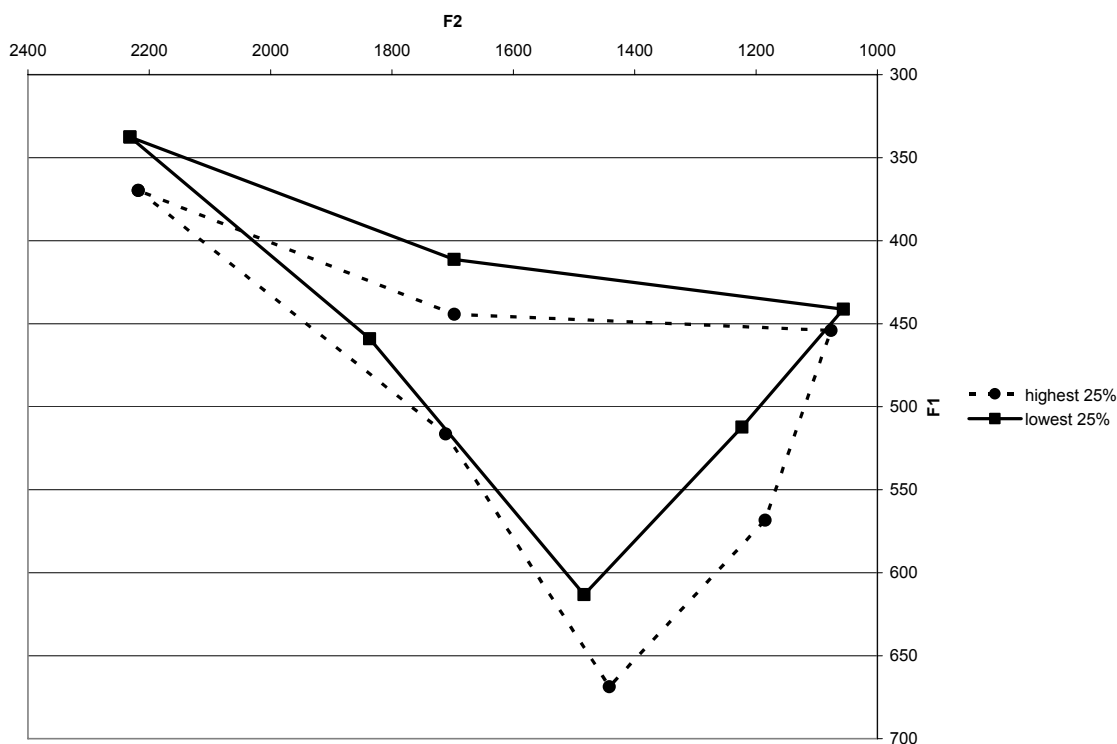


Figure 3.5.1 The potential effect of comparative intensity.

Condition	SED	Δ SED
25% of vowel tokens with the lowest CI score	16454	0%
25% of vowel tokens with the highest CI score	15958	-3%

Table 3.5.1 Sums of Euclidean distances between grand means of vowels spoken with the highest and the lowest comparative intensity.

The spectral effect appears to be due almost entirely to F1-raising in vowels with high comparative intensity (with the exception of the vowel *e*). This is, of course, hardly surprising: all other things being equal, less constricted vowels can be expected to have higher intensity. In fact, this is precisely the reason why raw intensity scores cannot be compared across vowels of different heights. The interesting question is, however, what drives the variation in intensity. When this variable was introduced, it was hoped that it

could provide information about how stressed or unstressed the vowel is, because intensity is one of the correlates of stress in Polish, albeit not a major one (Jassem 1962).³ This hope appears to have been justified. Table 3.5.2 compares the proportions of vowels that could be expected to be stressed on phonological grounds (vowels in penultimate syllables and non-clitic monosyllables, cf. section 3.15) in the two groups and in the entire dataset. Phonologically stressed vowels are clearly overrepresented in the high comparative intensity group and underrepresented in the other one.

Condition	Phonologically stressed vowel tokens	Phonologically stressed vowel tokens (%)
25% of vowel tokens with the lowest CI score	219/1231	17.8%
25% of vowel tokens with the highest CI score	755/1231	61.33%
All vowel tokens	1946/4923	39.5%

Table 3.5.1 Phonologically stressed vowel tokens in the two comparative intensity categories.

However, the differences in comparative intensity may not be wholly attributable to stress differences. The findings of Koopmans-van Beinum (1980) suggest that unstressed vowels could be expected to have a shrunken vowel space vis-à-vis stressed vowels. As table 3.5.1 shows, this is not the case here: in fact, the vowel space, as measured by the sum of Euclidean distances, is smaller in the high comparative intensity group than in the low comparative intensity group, albeit the difference is far smaller than those due to speech rate and duration. As will be shown in the section discussing the position of the syllable in the word (3.15), this is not the case when the phonologically stressed vowels

³ A major complication is that, as shown by Jassem (1962), the correlation between stress and intensity in Polish is only strong in words spoken with falling intonation, and the intonational contour of the utterances was not controlled for here.

are compared directly to the phonologically unstressed ones. This suggests that the ability of the comparative intensity variable to reflect stress-driven differences may be compromised by the existence of vowels that do have high comparative intensity but are not lexically stressed and vice-versa.⁴

3.6 Word frequency

Word frequency varied between 0-539,127 occurrences in the 40-million-words-large PWN corpus. Figure 3.6.1 compares the vowel spaces of the 25% most frequent words with the 25% least frequent words. Table 3.6.1 shows the sums of Euclidean distances for these two categories.

⁴ There exists the possibility that the high-intensity vowels that are not phonologically stressed reflect phrasal accent as opposed to lexical stress. On this scenario, we would have to assume that phrasal accent has an effect on vowel intensity, but not on vowel formant frequencies. It is unclear why it should be so. Van Bergem (1993) found that in Dutch both lexical stress and phrasal accent have independent effects on vowel reduction.

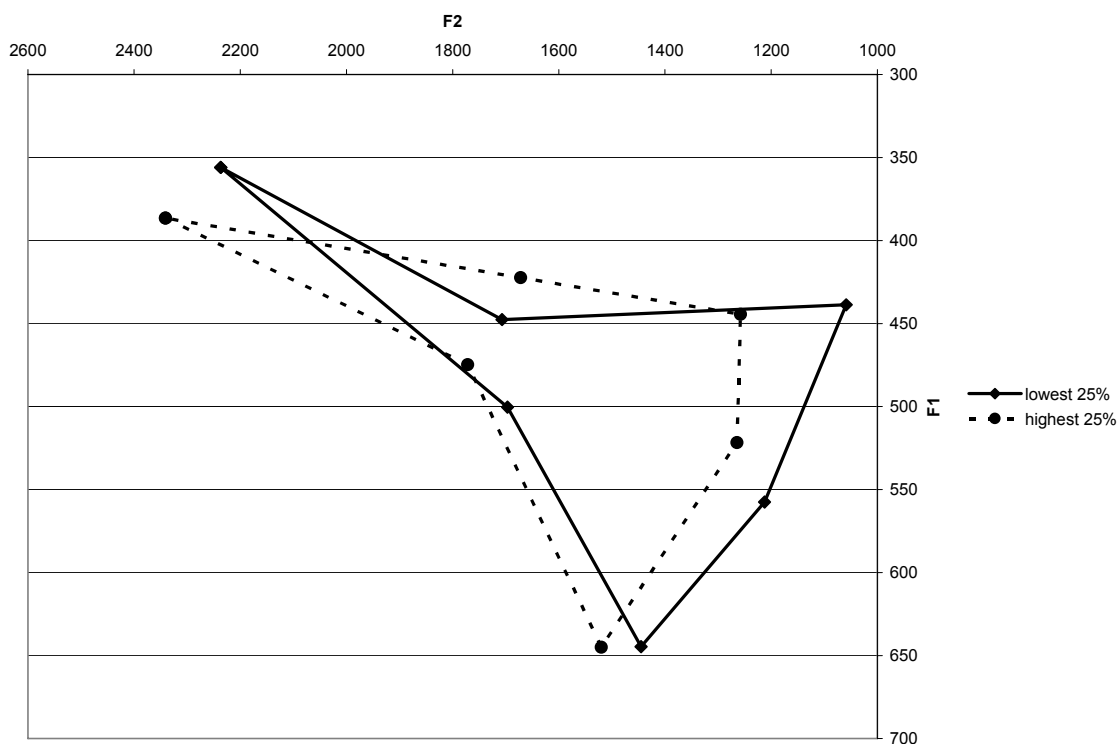


Figure 3.6.1 The potential effect of word frequency.

Condition	SED	Δ SED (%)
vowel tokens in the 25% least frequent words	16001	0%
vowel tokens in the 25% most frequent words	15028	-5.4%

Table 3.6.1 Sums of Euclidean distances between grand means of vowels appearing in the most and the least frequent words.

Reduction processes have been found to have a more limited scope in low-frequency words (Jurafsky et al. 2001). To a certain extent, this may be reflected by the Euclidean distance data in table 3.6.1. However, caution is needed. The approach used here to illustrate effects of independent variables is particularly ill-suited to capture a frequency effect. By the very nature of the word frequency metric, the high frequency group will contain vowel tokens from far fewer words than the low frequency group (although in

this case the problem is partially alleviated by the fact any low-frequency word appearing in the texts the subjects read is represented eight times in the data set, as each speaker read the text twice). As a result, the high-frequency group may be prone to accidentally reflect the specific properties of specific words rather than a frequency effect per se. The inconsistent picture shown by figure 3.6.1 indicates that this may have been the case here (F1 shifts in some vowels only, F2 changes with magnitude differing without a clear pattern). This suspicion will be further confirmed in the section discussing the potential effect of the content/function status of the word (3.21) and this issue will be discussed further there.

It should be noted at this point that the frequency effect on vowel reduction may be observed only to the extent that frequency is a somewhat crude index of *predictability*. According to Jurafsky et al. (2001), reduction can be expected to be more common in more predictable words because the speaker can assume that the listener will be able to recover information more easily. But word frequency is only one element of predictability. Word predictability is also a function of syntactic, semantic and discourse factors. To a certain extent, these factors may be reflected by other independent variables considered (in particular speech style, clause position of the word in the clause and the content/function status). But matters get more complicated when the predictability of a

given phoneme is to be considered.⁵ Thus any effect of word frequency on vowel spectra (or lack thereof) discovered in the data has to be treated with caution.

3.7 Preceding consonant place

Table 3.7.1 presents the frequency of the seven consonant place categories in the entire dataset. The greatest effect on vocalic spectra was expected to come from consonants with primary or secondary palatalization, i.e. alveopalatal, palatal and palatalized consonants. It is precisely in their context that traditional descriptions of Polish postulate the existence of fronted and raised allophones (e.g. Sawicka 1995). To capture this potential effect for the purposes of data presentation, in this chapter only these three place categories were collapsed into one category “*palatal*”, while the remaining place categories (dentals, alveolars, labials and velars) were collapsed into one category “*nonpalatal*”. (Quotation marks are used to avoid confusion with the term *palatal* used to describe a specific place of articulation). The frequency of these two consonant place categories is shown in table 3.7.2.

⁵ In this context, the findings of Wright (2003) and Scarborough (2004) are particularly interesting. Their findings indicate that reduction and coarticulation vary depending on the *difficulty* of individual lexical items, where difficulty is a function of relative word frequency and neighborhood density. Unfortunately, the computational burden involved with the estimation of word difficulty means that this factor cannot be considered in this work. The difficulties begin, though by no means end, with the fact that the commonly used metric of neighborhood density – one phoneme difference – aside from being crude in its own right is probably ill-suited to an inflectional language like Polish, in which words are longer than in an analytical language like English, and thus fewer near-homophones are likely to exist.

Place	Frequency	Frequency (%)
labial	1174	23.8%
dental	1316	26.7%
alveolar	892	18.1%
palatalized	371	7.5%
alveopalatal	461	9.4%
palatal	246	5%
velar	463	9.4%

Table 3.7.1 Breakdown of the data by place of articulation of the preceding consonant.

Place	Frequency	Frequency (%)
“palatal”	1078	21.9%
“nonpalatal”	3845	78.1%

Table 3.7.2 Breakdown of the data by place of articulation of the preceding consonant under a binary division. (“Palatal” = alveopalatal, palatal and palatalized consonants, “nonpalatal” = all the other consonants.)

The potential effect on vowel space of the preceding consonant place under the binary distinction is shown in figure 3.7.1, while table 3.7.3 shows the sums of Euclidean distances for these two categories. Both presentations differ from other displays, because of the restrictions the phonological system of Polish places on the distribution of the vowel *y*. This vowel never follows alveopalatal, palatalized and palatal consonants, and thus has not been shown on the “palatal” vowel space in figure 3.7.1. The Euclidean distances in table 3.7.3 have been calculated in both cases without including the vowel *y*, to allow for a meaningful comparison. It should be pointed out that the vowel *i* is also restricted phonologically as to what consonant it can follow. In a mirror image of the behavior of *y*, *i* can only be preceded by alveopalatals, palatals and palatalized segments. However, we have assumed after Sawicka (1995) that this restriction only applies word internally. Thus the cases of *i* in the “nonpalatal” context are those tokens of word-initial

i that were preceded by a word ending in a “nonpalatal” consonant, without a pause.

There are no parallel cases of *y*, since this vowel never occurs word-initially.

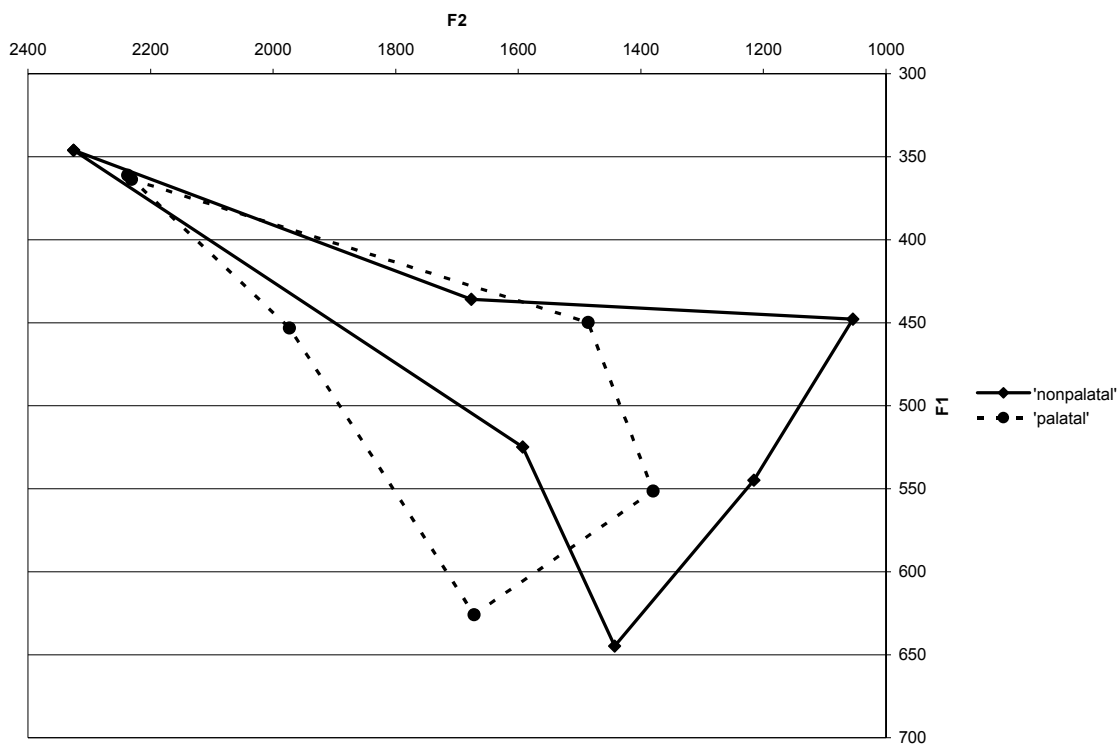


Figure 3.7.1 The potential effect of preceding consonant place under a binary division. Note that only five vowels are possible in the “palatal” condition. (“Palatal” = alveopalatal, palatal and palatalized consonants, “nonpalatal” = all the other consonants.)

Place	SED	Δ SED (%)
“nonpalatal”	12151	0%
“palatal”	9360	-22.97%

Table 3.7.3 Sums of Euclidean distances between grand means of vowels appearing after “palatal” and “nonpalatal” consonants. Only the vowels *i*, *e*, *a*, *o* and *u* were used in the calculations.

As can be seen, the presence of a preceding “palatal” consonant has a large effect on the shape of the vowel space. With the exception of the vowel *e*, the effect appears to be confined to the F2 domain: vowels preceded by a *palatal* consonant are much more front

(except for *i*). The vowel space is also dramatically shrunk; in fact, it is the greatest shrinkage observed in any condition.

3.8 Following consonant place

The same approach as in section 3.7 was taken to the investigation of the effect of the following consonant place. Tables 3.8.1 and 3.8.2 present summary statistics about the frequency of occurrence of the various place categories under the seven-way and the two-way division, respectively. Figure 3.8.1 illustrates graphically the transformation of vowel space under the binary division of consonants into “palatal” and “nonpalatal”, while table 3.8.3 shows the sums of the Euclidean distances for these two categories. It should be noted that in this case, there are no restrictions on possible vowel-consonant combinations, and all vowels were included in the calculation of Euclidean distances. However, to facilitate comparison of the preceding and the following consonant place effect, sums of Euclidean distances were also calculated for the limited five-vowel set (excluding *y*). These are shown in table 3.8.4.

Place	Frequency	Frequency (%)
labial	1286	26.1%
dental	1337	27.2%
alveolar	727	14.8%
palatalized	306	6.2%
alveopalatal	491	10%
palatal	204	4.1%
velar	572	11.6%

Table 3.8.1 Breakdown of the data by place of articulation of the following consonant.

Place	Frequency	Frequency (%)
“palatal”	1001	20.3%
“nonpalatal”	3922	79.7%

Table 3.8.2 Breakdown of the data by place of articulation of the following consonant under a binary division. (“Palatal” = alveopalatal, palatal and palatalized consonants, “nonpalatal” = all the other consonants.)

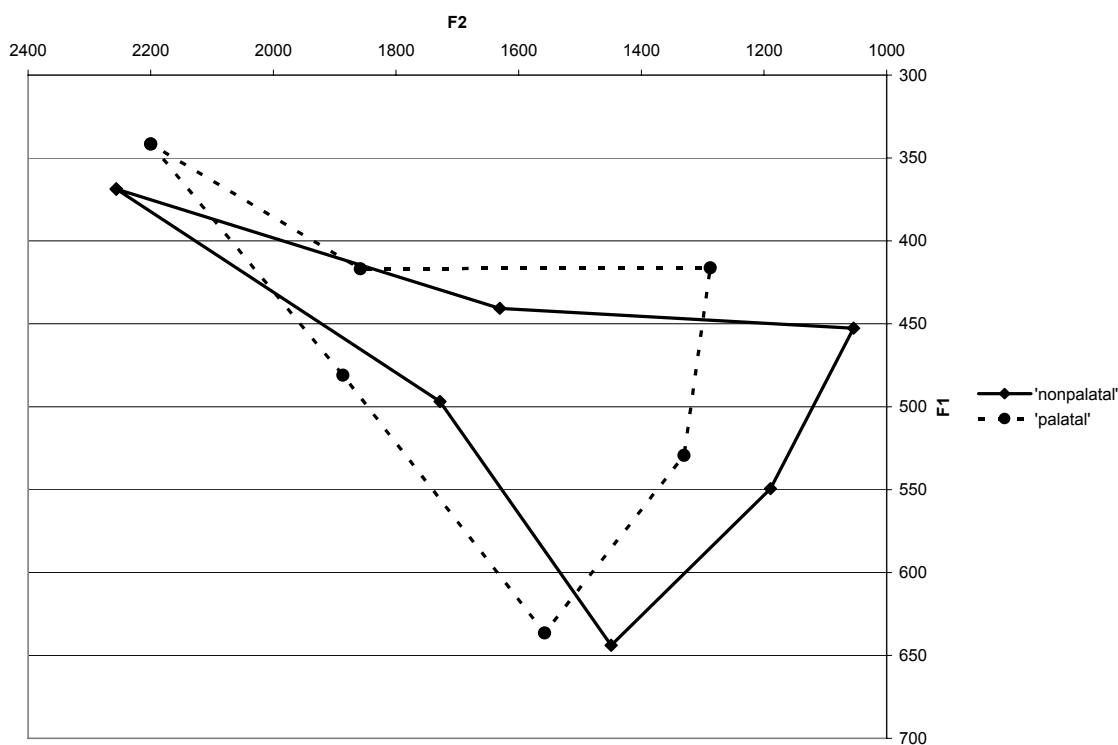


Figure 3.8.1 The potential effect of following consonant place under a binary division. (“Palatal” = alveopalatal, palatal and palatalized consonants, “nonpalatal” = all the other consonants.)

Place	SED	Δ SED (%)
“nonpalatal”	16290	0%
“palatal”	14008	-14%

Table 3.8.3 Sums of Euclidean distances between grand means of vowels appearing before “palatal” and “nonpalatal” consonants.

Place	SED	Δ SED (%)
“nonpalatal”	12197	0%
“palatal”	10023	-17.8%

Table 3.8.4 Sums of Euclidean distances between grand means of vowels appearing before “palatal” and “nonpalatal” consonants. Only the vowels *i*, *e*, *a*, *o* and *u* were used in the calculations.

From the graphical display, the effect of a following “palatal” consonant appears to be smaller than that of a preceding one. This is in line with impressionistic descriptions of the Polish vowel system (e.g. Sawicka 1995).⁶ The transformation is also somehow different, with a greater apparent effect in the F1 domain (with the exception of the vowel *e*). Judging from the Euclidean distance scores, the vowel space shrinks considerably, but less dramatically than in the case of a preceding “palatal” consonant.

3.9 Preceding consonant manner

Table 3.9.1 shows the frequency of the seven manner categories of the preceding categories in the entire data set. To explore the potential effect of the manner of the preceding consonant on the following vowel, in this chapter the levels of this variable were collapsed into two broad natural classes: *obstruents* (stops, fricatives and affricates) and *sonorants* (approximants, trills, laterals and nasals). The summary statistics for this binary division are shown in table 3.9.2. The potential effect of these two categories on the vowel space is illustrated in figure 3.9.1, while the sums of Euclidean distances are shown in table 3.9.3.

⁶ According to Sawicka (1995), fronting/raising is most common between “palatal” consonants”, rather common after a “palatal” consonant but before a “nonpalatal” one, and very rare in the reverse situation.

Manner	Frequency	Frequency (%)
stop	1618	32.9%
fricative	1229	25%
affricate	314	6.4%
approximant	338	6.9%
trill	327	6.6%
lateral	241	4.9%
nasal	856	17.4%

Table 3.9.1 Breakdown of the data by manner of articulation of the preceding consonant.

Manner	Frequency	Frequency (%)
sonorant	1762	35.8%
obstruent	3161	64.2%

Table 3.9.2 Breakdown of the data by manner of articulation of the preceding consonant under a binary division. (Obstruent = stop, fricative and affricate, sonorant = all the others.)

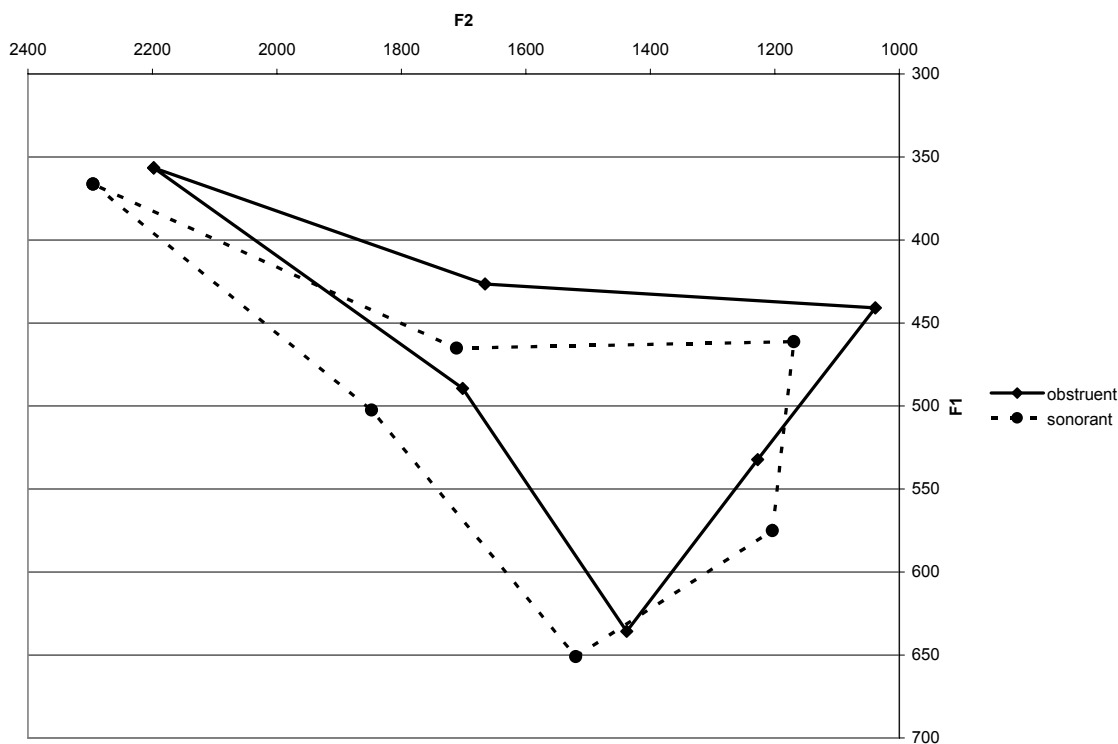


Figure 3.9.1 The potential effect of preceding consonant manner under a binary division. (Obstruent = stop, fricative and affricate, sonorant = all the others.)

Manner	SED	ΔSED (%)
sonorant	16263	0%
obstruent	15601	-4.1%

Table 3.9.3 Sums of Euclidean distances between grand means of vowels appearing after obstruents and after sonorants.

From the theoretical point of view, the expected effect of consonant manner divided binarily into obstruent and sonorant is unclear. From the perspective of the undershoot model, one could perhaps expect that sonorants might induce *less* coarticulatory pressure on the neighboring vowels, since they themselves require a more open oral constriction than obstruents (except for nasals). On the other hand, sonorant consonants are among the ones that have the greatest effect on vowel formants due to their acoustic properties (Johnson 2003). Judging from the data in table 3.9.3, the former tendency may be dominant, since the sum of Euclidean distances is greater in the sonorant condition, albeit the effect is small when compared with the one we have observed for example for vowel duration or speech rate.

Most likely, the data presented at this point show the limits of binary classification. It is very likely that the four types of sonorants have somewhat different effects on the vowel space and that these effects can only be assessed accurately when each manner class is treated separately, which will be the case in the statistical models presented in the following chapters.

3.10 Following consonant manner

The same approach as in the preceding section was used to explore the potential effect of the following consonant manner. Table 3.10.1 shows the frequency of the seven manner categories used in this work, while table 3.10.2 shows the frequency of the following consonants according to manner under the binary classification into obstruents and sonorants. Figure 3.10.1 illustrates the potential effect of the following consonant manner, and table 3.10.3 shows the sums of Euclidean distances in the two categories.

Manner	Frequency	Frequency (%)
stop	1455	29.6%
fricative	1527	31%
affricate	345	7%
approximant	508	10.3%
trill	249	5.1%
lateral	166	3.4%
nasal	673	13.7%

Table 3.10.1 Breakdown of the data by manner of articulation of the following consonant.

Manner	Frequency	Frequency (%)
sonorant	1596	32.4%
obstruent	3327	67.6%

Table 3.10.2 Breakdown of the data by manner of articulation of the following consonant under a binary division. (Obstruent = stop, fricative and affricate, sonorant = all the others.)

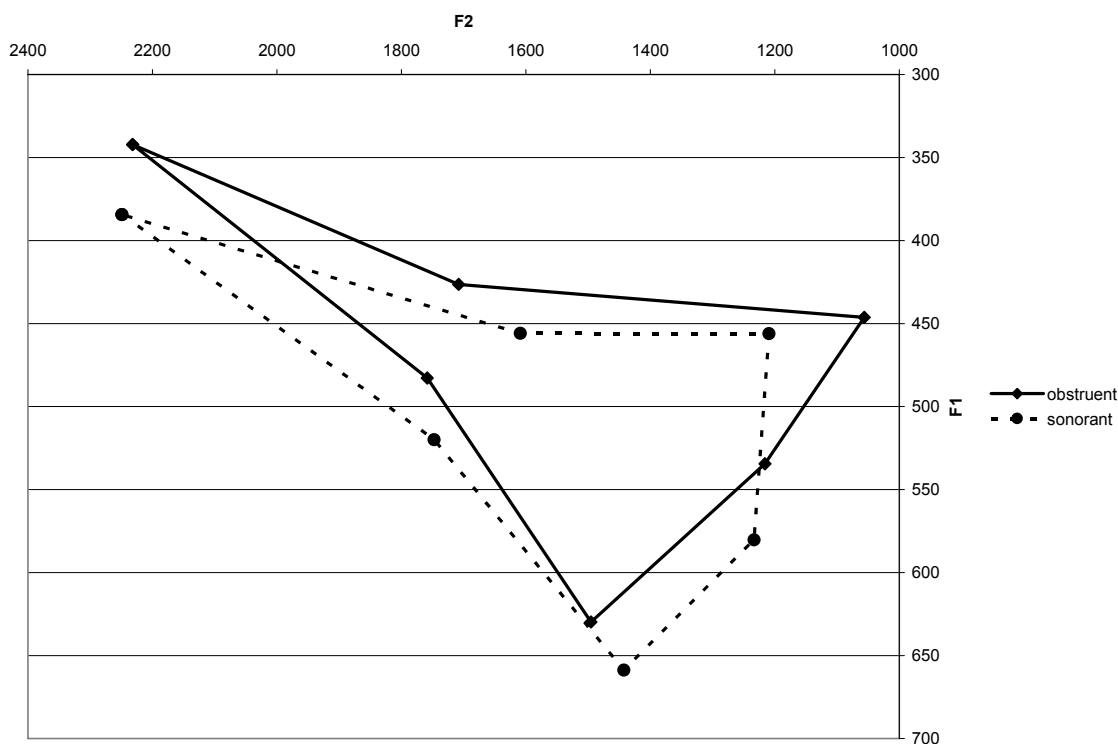


Figure 3.10.1 The potential effect of following consonant manner under a binary division. (Obstruent = stop, fricative and affricate, sonorant = all the others.)

Manner	SED	Δ SED (%)
sonorant	14715	-8.7%
obstruent	16123	0%

Table 3.10.3 Sums of Euclidean distances between grand means of vowels appearing before obstruents and after sonorants.

Graphically, the changes in vowel space due to the sonorant-obstruent distinction in the following consonant appear to be similar to those related to the preceding consonant manner, as discussed in the previous section. However, the examination of the Euclidean distance data reveals a different effect: in this case, the vowel space is actually shrunk in the sonorant condition. The difficulties in interpreting these potential effects parallel those involved with the interpretation of the potential effect of the preceding consonant

manner, as discussed in the preceding section. In this case also a more detailed discussion will be postponed until specific statistical models are analyzed in the following chapters.

3.11 Preceding consonant voicing

Table 3.11.1 presents summary statistics about voicing in consonants preceding the vowels examined. Figure 3.11.1 illustrates the potential effect of the voicing distinction on the vowel space, while table 3.11.2 shows the sums of Euclidean distances in these two conditions.

Voicing	Frequency	Frequency (%)
voiced	3200	65%
voiceless	1723	35%

Table 3.11.1 Breakdown of the data by voicing of the preceding consonant.

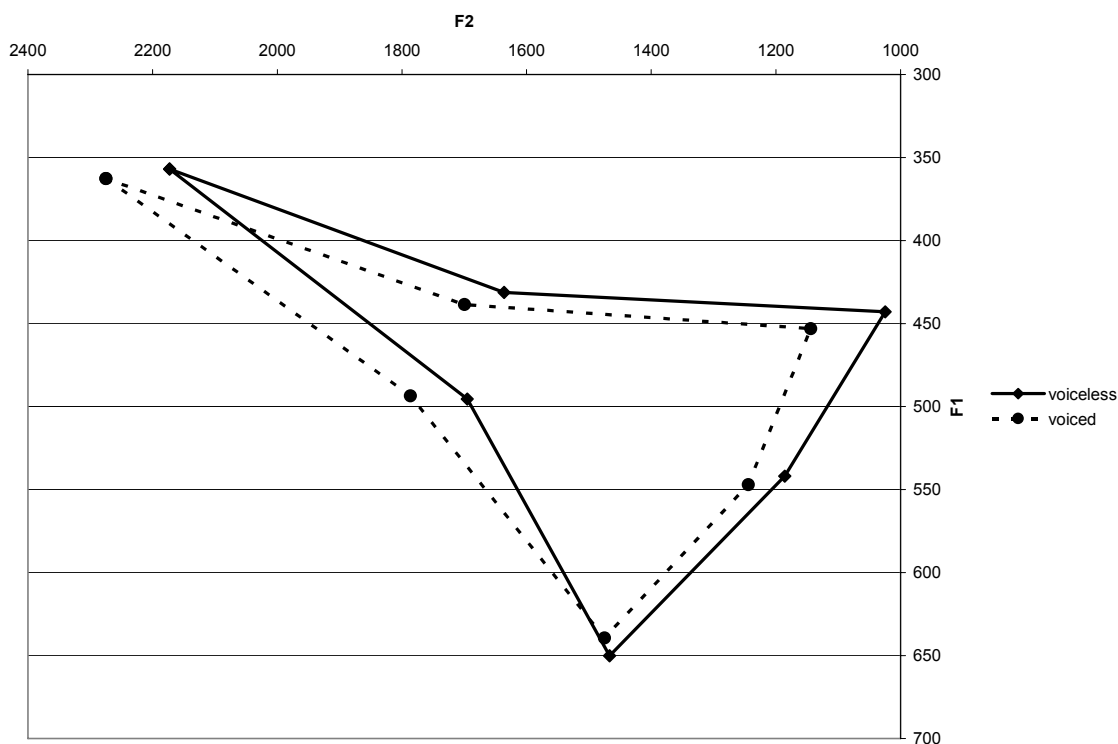


Figure 3.11.1 The potential effect of preceding consonant voicing.

Voicing	SED	Δ SED (%)
voiced	15679	0%
voiceless	15670	-0.0005%

Table 3.11.2 Sums of Euclidean distances between grand means of vowels appearing after voiced and voiceless consonants.

From the theoretical standpoint, no specific effect of preceding consonant voicing on the vowel space can be expected. From the graphic display, we see that vowels that follow voiced consonants tend to be somewhat fronter, but this may turn out to be an accident related, for example, to the different distribution of consonants with different places of articulation among the voiced and voiceless groups. The Euclidean distance data indicate that there is basically no shrinkage of the vowel space from one category to the other.

3.12 Following consonant voicing

Table 3.12.1 presents summary statistics about voicing in consonants following the vowels examined. Figure 3.12.1 illustrates the potential effect of the voicing distinction on the vowel space, while table 3.12.2 shows the sums of Euclidean distances in these two conditions.

Voicing	Frequency	Frequency (%)
voiced	3206	65.1%
voiceless	1717	34.9%

Table 3.12.1 Breakdown of the data by voicing of the following consonant.

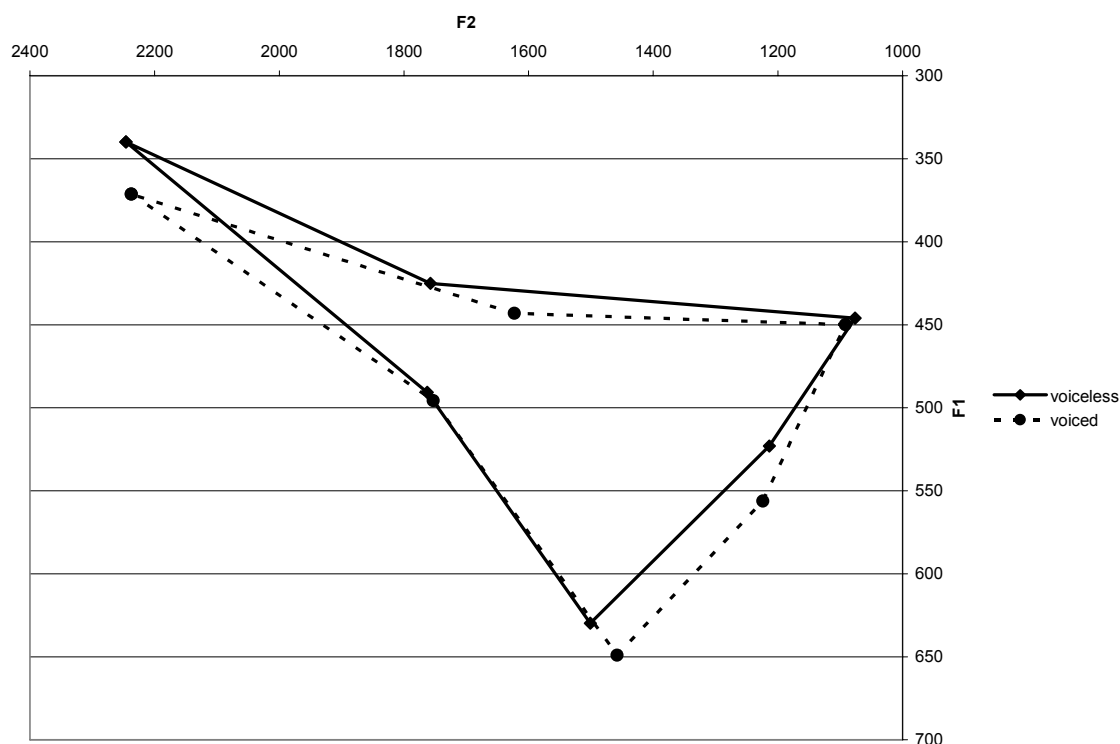


Figure 3.12.1 The potential effect of following consonant voicing.

Voicing	SED	Δ SED (%)
voiced	15676	-3.4%
voiceless	16228	0%

Table 3.12.2 Sums of Euclidean distances between grand means of vowels appearing before voiced and voiceless consonants.

As was the case with preceding consonant voicing, there are no specific expectations regarding the impact of the following consonant voicing status on vowel space. Some, but not all, vowels are slightly lower in the voiced condition, and the vowel space appears to be somehow shrunk in that condition. Whether there is any genuine effect of voicing must be verified by statistical analyses.

3.13 Number of following consonants

Summary statistics concerning the number of consonants following the vowel (before another vowel within the same word) are shown in table 3.13.1. To illustrate the potential effect of this factor, two categories were created: 0-1 consonants and 2 or more consonants. Frequencies within these two categories are shown in table 3.13.2. Their potential effect is shown graphically in figure 3.13.1, and the Euclidean distance data are provided in table 3.13.3.

Number	Frequency	Frequency (%)
0	1104	22.4%
1	2931	59.5%
2	783	15.9%
3	105	2.1%

Table 3.13.1 Breakdown of the data by number of following consonants.

Number	Frequency	Frequency (%)
0-1	4035	82%
2-3	1717	18%

Table 3.13.2 Breakdown of the data by number of following consonants (binary division).

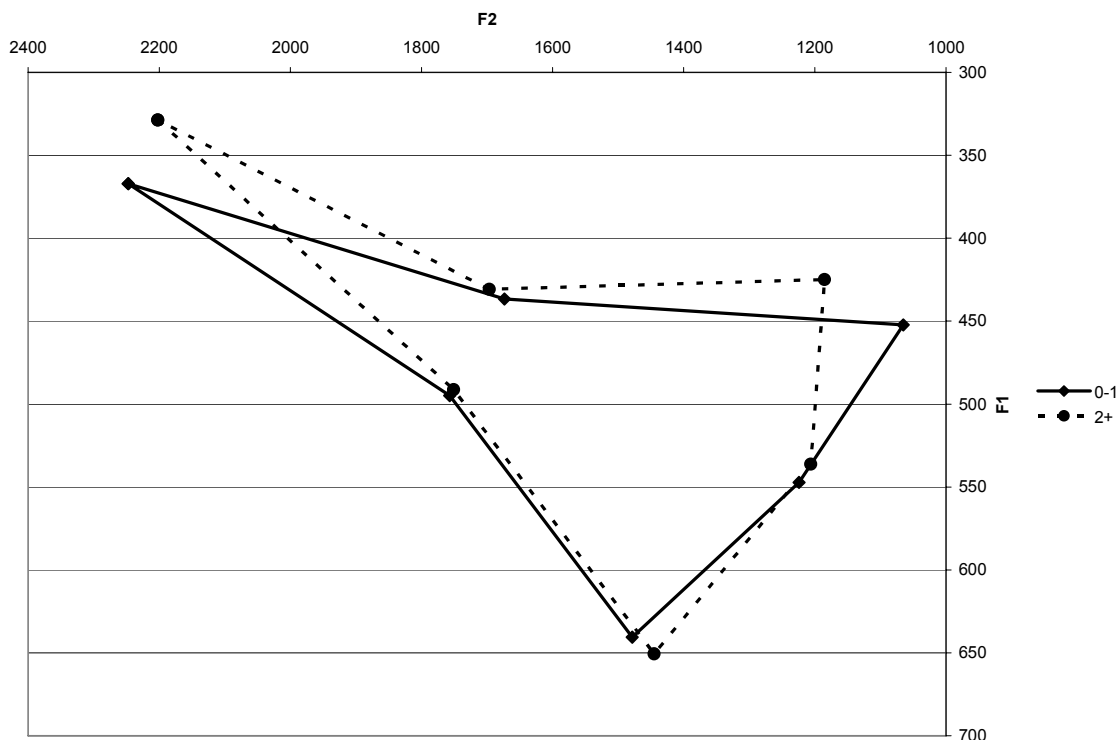


Figure 3.13.2 The potential effect of the number of following consonants.

Number	SED	Δ SED (%)
0-1	16066	0%
2-3	14936	-7%

Table 3.13.3 Sums of Euclidean distances between grand means of vowels appearing before 0-1 consonant and 2-3 consonants.

Because syllable division in Polish is a rather complicated matter on which there is no universal agreement (Sawicka 1995), in this work we have not considered syllable structure directly. To a certain extent, the potential impact of syllable structure could be

captured by the more objective metric of following consonant count. The higher the count, the more likely the syllable is to be a closed syllable. All other things being equal, if an effect were to be found, it could be expected to be one of increased reduction in closed syllables. The Euclidean distance data in table 3.13.3 indicate that this may be the case in Polish, but the graphical display in figure 3.13.1 shows that any effect is rather slight and confined to the vowels *i* and *u*. Moreover, since vowels could be expected to be shorter in closed syllables, the effect may be only an apparent one explainable by durational differences.

3.14 Speech style

Table 3.14.1 shows the breakdown of the data by speech style. Figure 3.14.1 presents the mean vowel spaces of all speakers combined in the three styles in a two-dimensional vowel space. Table 3.14.2 presents the sums of Euclidean distances in the three conditions.

Style	Number of tokens	Number of tokens (%)
normal reading	1737	35.3%
fast reading	1636	33.2%
narrative	1550	31.5%

Table 3.14.1 Breakdown of the data by speech style.

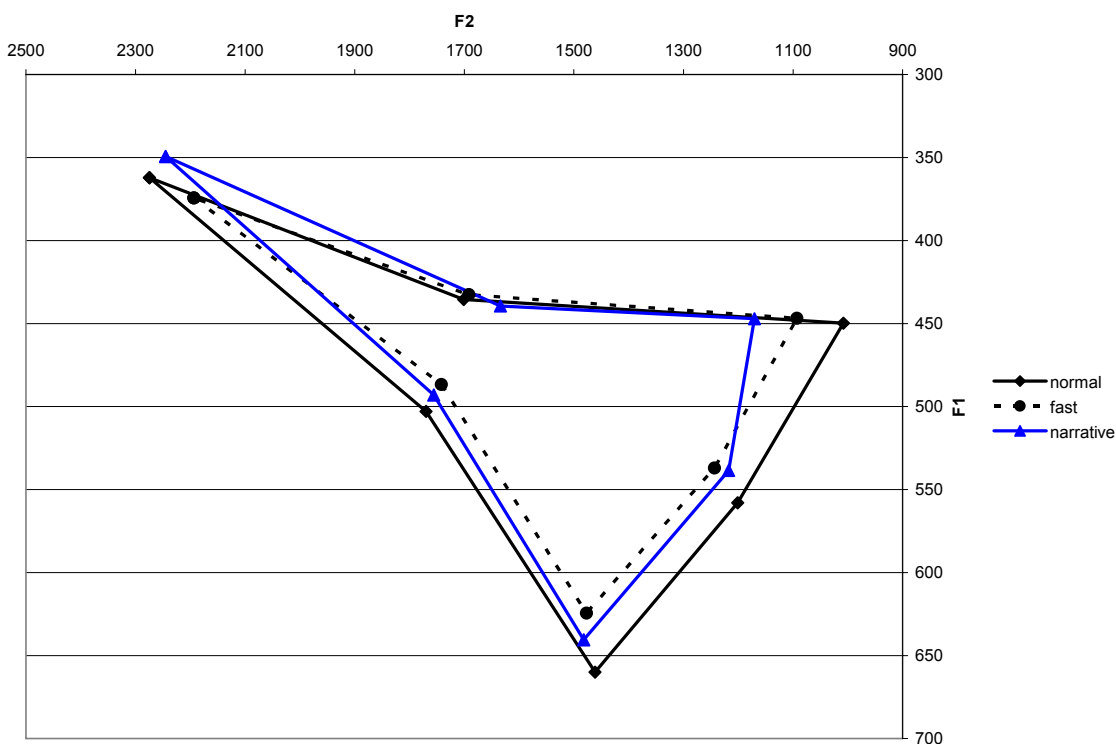


Figure 3.14.1 The potential effect of speech style.

Style	SED	Δ SED (%)
normal reading	17291	0%
fast reading	15034	-13.1%
narrative	15080	-12.8%

Table 3.14.2 Sums of Euclidean distances between vowel grand means in the three speech styles.

Both the graphical display in figure 3.14.1 and the changes in Euclidean distances suggest that there is a considerable difference between the normal reading style on the one hand, and the fast reading style and the narrative style on the other, but that the differences between the latter two styles are minor. Except for the high vowels, the overall effect appears to be shrinkage rather than unidirectional formant shift.

This potential effect is not surprising in the light of the findings of Koopmans-van Beinum (1980), who found that Dutch speakers had more expanded vowel spaces in more “formal” speech styles. This applies in particular to the difference between the slow reading style and the narrative style, while the shrinkage in the fast reading style may be more related to the speech rate increase that this task induced. To the extent that there is an independent speech style effect, it may reflect the fact that the scope of reduction is likely to be positively correlated with predictability (Jurafsky et al. (2001), cf. also the discussion in section 3.6). The text the subjects read in the two reading conditions was unfamiliar to them, while the narratives were personal stories told in an informal way to the investigator they were well acquainted with; it is likely that they assumed that in the latter case there was much more common ground. In addition, the very task of reading a story may have induced them to behave in a more formal way, i.e. use a more careful articulation.

3.15 Syllable position

Table 3.15.1 shows the distribution of vowel tokens in the entire dataset according to their position in the word. Since in Polish penultimate vowels and nonclitic monosyllables are expected to carry main lexical stress, in this chapter these two categories were collapsed into one category *stressed*, while all the other vowels were

collapsed into the category *unstressed*.⁷ The summary statistics for these two groups are shown in table 3.15.2. Figure 3.15.1 illustrates their potential effect on the vowel space, and table 3.15.3 shows the sums of the Euclidean distances in these two categories.

Position	Frequency	Frequency (%)
penultimate	1794	36.4%
monosyllable	152	3.1%
proclitic	338	6.9%
enclitic	158	3.2%
antepenultimate	805	16.4%
preantepenultimate	236	4.8%
final	1399	28.4%
other	41	0.8%

Table 3.15.1 Breakdown of the data by (phonological) position in the word..

Phonological stress	Frequency	Frequency (%)
stressed	1946	39.5%
unstressed	2977	60.5%

Table 3.15.2 Breakdown of the data by phonologically expected stress. (Stressed = penultimate and nonclitic monosyllables, unstressed = all the others.)

⁷ Secondary stress was not considered. It is a complex issue in Polish on which little is known for certain (see especially Steffen-Batogowa 2000 for a review).

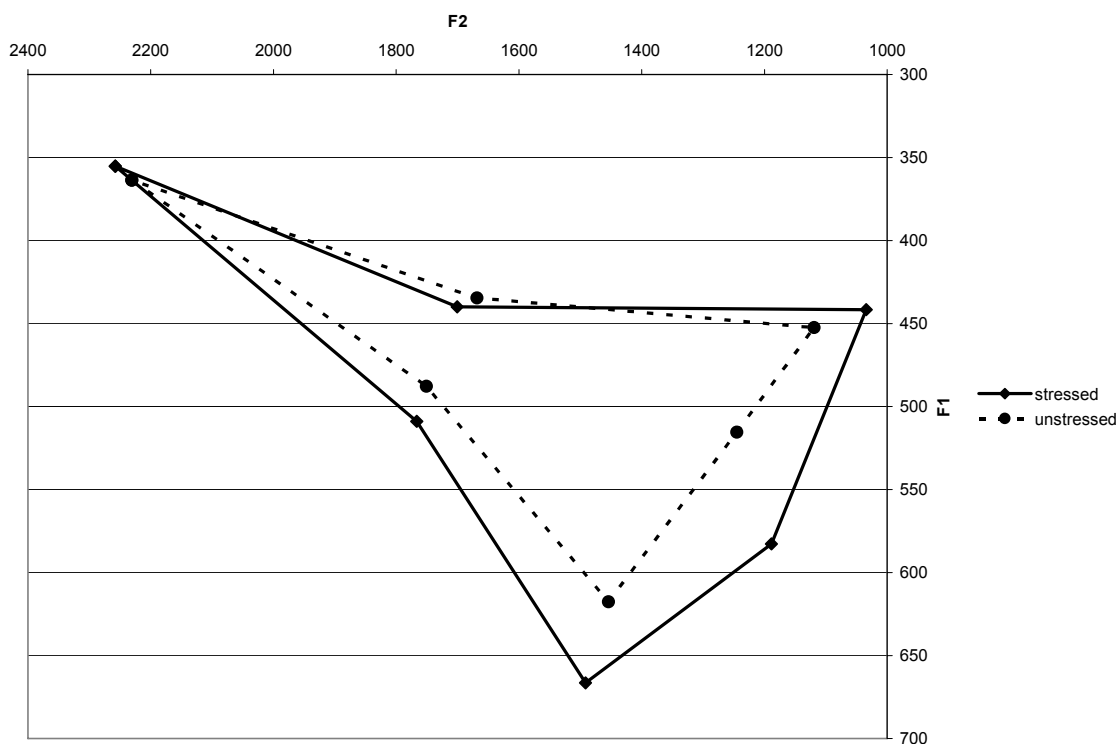


Figure 3.15.1 The potential effect of phonologically expected stress. (Stressed = penultimate and nonclitic monosyllables, unstressed = all the others.)

Phonological stress	SED	Δ SED (%)
stressed	16986	0%
unstressed	15131	-10.9%

Table 3.15.3 Sums of Euclidean distances between vowels that can be expected to be stressed and unstressed on phonological grounds.

Stressed vowels have been found to be less prone to reduction in several languages (Harris 1978, Nord 1986, van Bergem 1993). The data presented here appear to clearly confirm this for Polish: the vowel space of unstressed vowels is clearly shrunk, which can be seen in the graphical display and is also indicated by the Euclidean distance metric. The truly interesting question is whether this is a genuine independent effect, or whether it merely reflects durational differences between stressed and unstressed vowels. The

question is particularly important because there is evidence that in Polish the differences in duration between stressed and unstressed vowels are smaller than in some other languages. For example, Jassem (1962) estimated the duration ratio of stressed to unstressed vowels in Polish at 1.17:1 on average. For English, Crystal and House (1988b) estimated the same ratio at 2:1 for monophthongs and 1.5:1 for diphthongs. We will examine these issues carefully in the following chapters.

3.16 Size of the phonological word

Phonological words in the data varied between 1-9 syllables. The summary statistics about the distribution of vowel tokens in different categories are shown in table 3.16.1. For the purpose of exploration of the potential effect, the data were arbitrarily divided into shorter phonological words (1-3 syllables) and longer words (4 or more syllables). The summary statistics for these two groups are shown in table 3.16.2. Figure 3.16.1 illustrates their potential effect on the vowel space, and table 3.16.3 shows the sums of the Euclidean distances in these two categories.

Phonological word size	Frequency	Frequency (%)
1	76	1.5%
2	1322	26.9%
3	1878	38.1%
4	1089	22.1%
5	425	8.6%
6	108	2.2%
7	16	0.3%
8	4	0.08%
9	5	0.1%

Table 3.16.1 Breakdown of the data by size of the phonological word.

Phonological word size	Frequency	Frequency (%)
1-3	3276	66.5%
4-9	1647	33.5%

Table 3.16.2 Breakdown of the data by size of the phonological word under binary division.

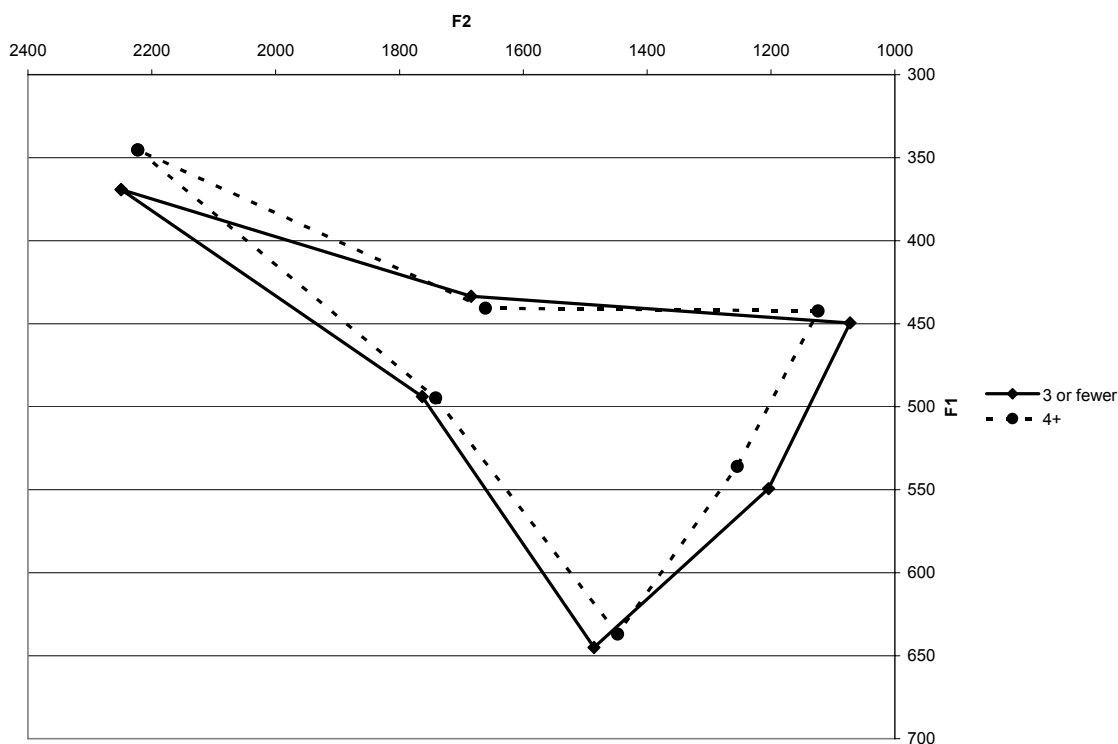


Figure 3.16.1 The potential effect of phonological word size.

Phonological word size	SED	Δ SED (%)
1-3	16221	0%
4-9	15044	-7.3%

Table 3.16.3 Sums of Euclidean distances between vowels in short and in long phonological words.

There are no theoretical reasons to expect that size of the phonological word per se has an impact on the degree of vowel reduction. However, vowels tend to be shorter in longer words; this effect was used by Moon and Lindblom (1994) to obtain English vowels of different durations. We could expect then vowels in longer phonological words to be more reduced, reflecting indirectly the effect of duration on vowel reduction. The data presented here indicate that this may indeed be the case, but this can only be verified through statistical analyses in the subsequent chapters.⁸

3.17 Postpausal phonological word

Table 3.17.1 presents the summary statistics concerning the distribution of vowel tokens in postpausal and non-postpausal phonological words in the entire dataset. Figure 3.17.1 presents the potential effect of these two conditions on the vowel space, and table 3.17.2 presents the sums of Euclidean distances.

Postpausal phonological word	Frequency	Frequency (%)
no	4102	83.3%
yes	821	16.7%

Table 3.17.1 Breakdown of the data by the position of the phonological word with regard to a preceding pause.

⁸ In addition, as Ian Maddieson points out (p.c.), longer phonological words may have more inflectional syllables, which could also result in an apparent effect of increased reduction.

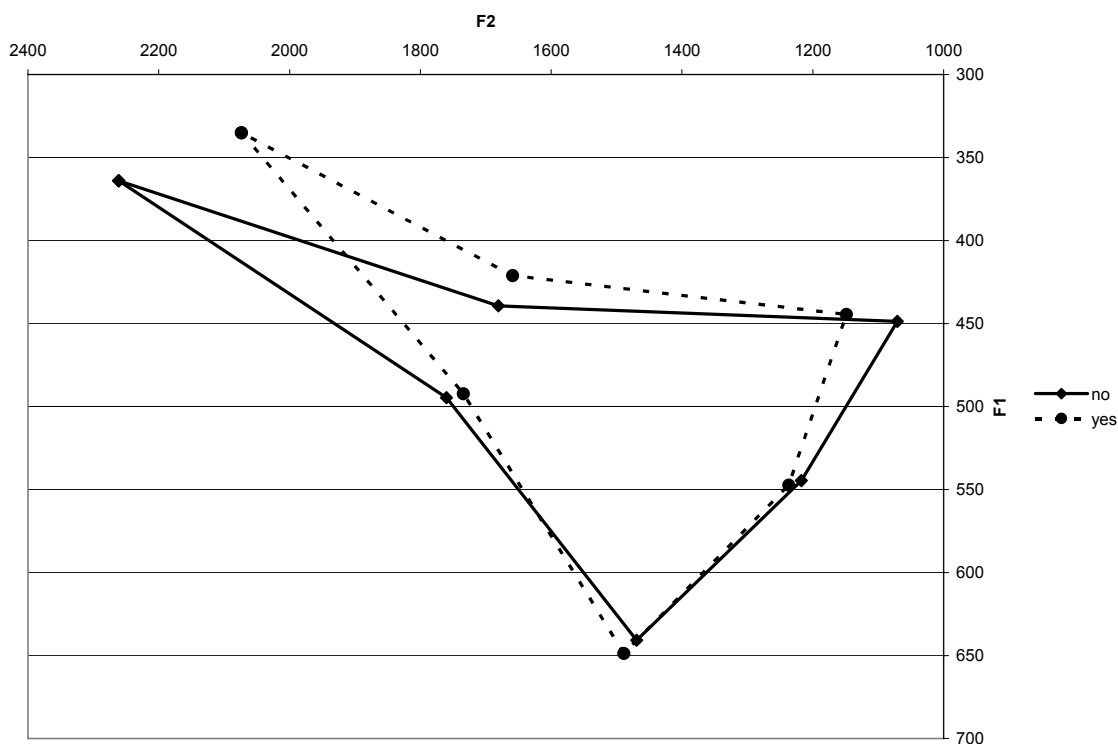


Figure 3.17.1 The potential effect of a pause preceding the phonological word.

Postpausal phonological word	SED	Δ SED (%)
no	16234	0%
yes	13554	-16.5%

Table 3.17.2 Sums of Euclidean distances between grand means of vowels appearing in postpausal and non-postpausal phonological words.

The postpausal variable was included for exploratory purposes (“data-mining”) rather than any specific theoretical reason. At the same time, the presence of a pause could conceivably correlate with changes in vowel space for several reasons. First, the vowels that are immediately adjacent to a pause may be subject to increased “coarticulation” with the rest position of the vocal tract (Nord 1986). The presence of such vowels in the postpausal dataset might thus skew the results for this group towards showing somewhat

greater centralization.⁹ Second, at least some pauses are likely to coincide with higher-level domains such as intonational phrases, clauses etc., and the phenomenon of domain-initial strengthening is well attested in a number of languages (Keating 1999, 2004). It has been examined in detail only for consonants and the possible predictions for vowels are not clear. On the one hand, the strengthening effect of a domain boundary on consonants manifests itself in longer and more extreme gestures. These would be likely to result in increased vowel target undershoot, unless of course the vowel in question were itself strengthened, i.e. articulated with more extreme gestures. In addition, postpausal segments, including vowels, may be lengthened (Cho 2004), complicating the matters even further. Finally, there is one more possibility that cannot be ignored and that is specific to our data set. Speakers made substantially fewer pauses in the fast reading condition than in the other two conditions, i.e. they made fewer pauses in the condition which appears to be correlated with greatest overall reduction (cf. section 3.14). For this reason, vowels from the fast reading style are underrepresented in the postpausal condition (162 out of 821, i.e. 19.7% vis-à-vis 33.3% in the entire dataset). This could skew the data towards displaying ostensibly less reduction in this condition.

The potential effect of postpausal condition on the vowel is quite large when the sums of Euclidean distances are considered and opposite in direction to what the account that relies on data skewing due to unequal distribution of tokens among styles would lead us

⁹ In this condition, as well as in the prepausal condition, clause-initial condition and clause-final condition, all vowels in the phonological word are treated together. Of course, effects such as “coarticulation” with the rest position of the vocal tract (and perhaps others) would show themselves primarily in the domain-edge adjacent vowel (immediately prepausal etc.) However, a strong effect should still be noticeable when we compare one group in which a certain proportion of vowels is affected with one in which none are.

to expect. Judging from the graphical display, the potential effect is confined to high vowels. It certainly does not provide a straightforward case of vowel space centralization, which makes the account in terms of increased influence by the rest position of the vocal tract unlikely. There remains the possibility that the effect is caused by stronger consonant articulation resulting from domain-initial strengthening, but it is unclear why only the high vowels should be affected. As is the case with other variables, only detailed statistical modeling will be able to tell us whether we are dealing with a genuine effect at all. Of course, there is the possibility that some or all of the effects discussed here are in fact present in the data, but the variables chosen are not sensitive enough to capture them.

3.18 Prepausal phonological word

Table 3.18.1 presents the summary statistics concerning the distribution of vowel tokens in prepausal and non-prepausal phonological words in the entire dataset. Figure 3.18.1 presents the potential effect of these two conditions on the vowel space, and table 3.18.2 presents the sums of Euclidean distances.

Prepausal phonological word	Frequency	Frequency (%)
no	4247	86.3%
yes	676	13.7%

3.18.1 Breakdown of the data by the position of the phonological word with regard to a following pause.

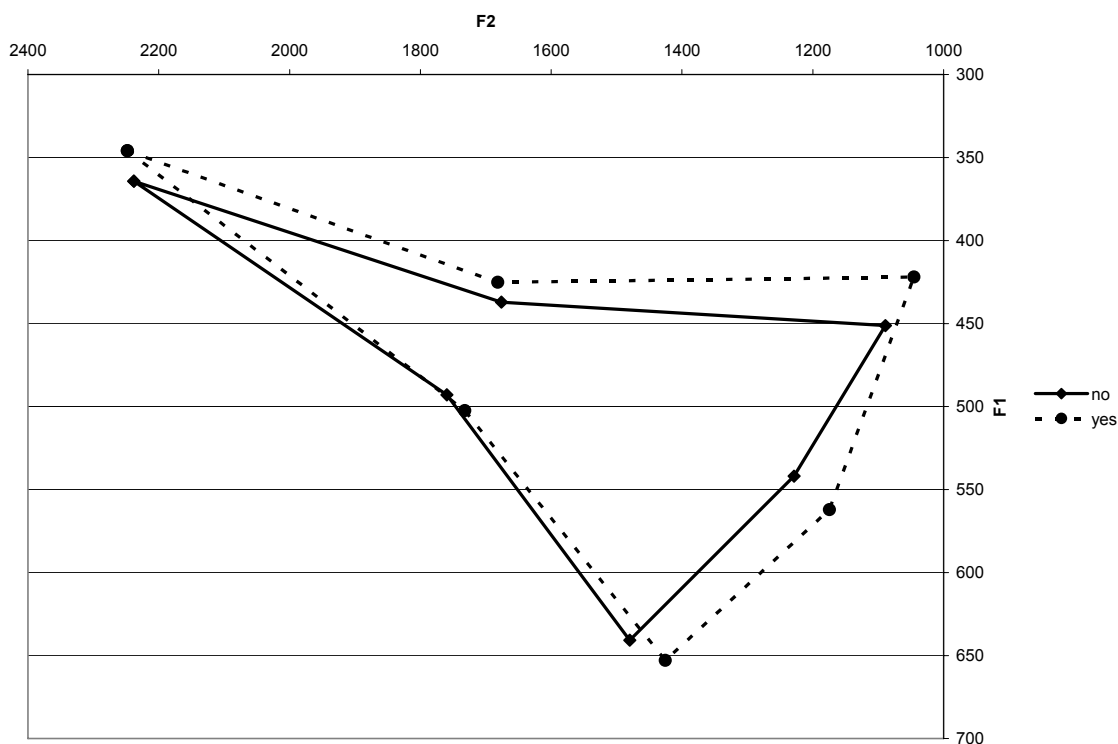


Figure 3.18.1 The potential effect of a pause following the phonological word.

Prepausal phonological word	SED	Δ SED (%)
no	15737	-6.1%
yes	16762	0%

Table 3.18.2 Sums of Euclidean distances between grand means of vowels appearing in prepausal and non-prepausal phonological words.

The potential sources for an effect of being in a prepausal phonological word on the vowel's quality are somehow clearer than is the case with the postpausal condition. Nord (1986) suggested that the anticipatory effect of the rest position of the vocal tract is greater than the perservatory one. We would thus be more likely to find increased centralization in the prepausal condition. As can be seen in the data presented here, this is clearly not the case: the vowel space appears to be expanded in the prepausal condition. It

is possible that what we observe is a durational effect, since domain-final lengthening is a well attested phenomenon (e.g. Wightman et al. 1992). It is also possible that this effect is compounded by the fact that vowels from the fast reading style, which are prone to undergo most reduction, are underrepresented in the prepausal dataset as well (140 out of 676, i.e. 20.7% vis-à-vis 33.2% in the entire dataset).

When the Euclidean distance data for the postpausal and the prepausal variables are taken together, another possible source of the observable effects emerges. As was noted in section 3.6, Jurafsky et al. (2001) have shown that predictability of information may influence the degree of reduction, with the material conveying the more predictable information being more prone to reduction. Given the fact that Polish is a language with a basic SVO word order (Stone 1990), old information may be expected to precede new information in the clause (although the effect is likely to be obscured by the fact that subjects, i.e. the prototypical topics, do not have to be overtly expressed). To the extent that pauses may be correlated with syntactic structure, we could expect more reduction after pauses and less before them. This appears indeed to be the case, with non-prepausal and non-postpausal words falling in between. Once again, whether this is a real effect requires statistical verification.

3.19 Clause-initial phonological word

Table 3.19.1 presents the summary statistics concerning the distribution of vowel tokens in clause-initial and non-clause initial phonological words in the entire dataset. Figure 3.19.1 presents the potential effect of these two conditions on the vowel space, and table 3.19.2 presents the sums of Euclidean distances.

Clause-initial phonological word	Frequency	Frequency (%)
no	3283	66.7%
yes	1640	33.3%

3.19.1 Breakdown of the data by the position of the phonological word with regard to clause beginning.

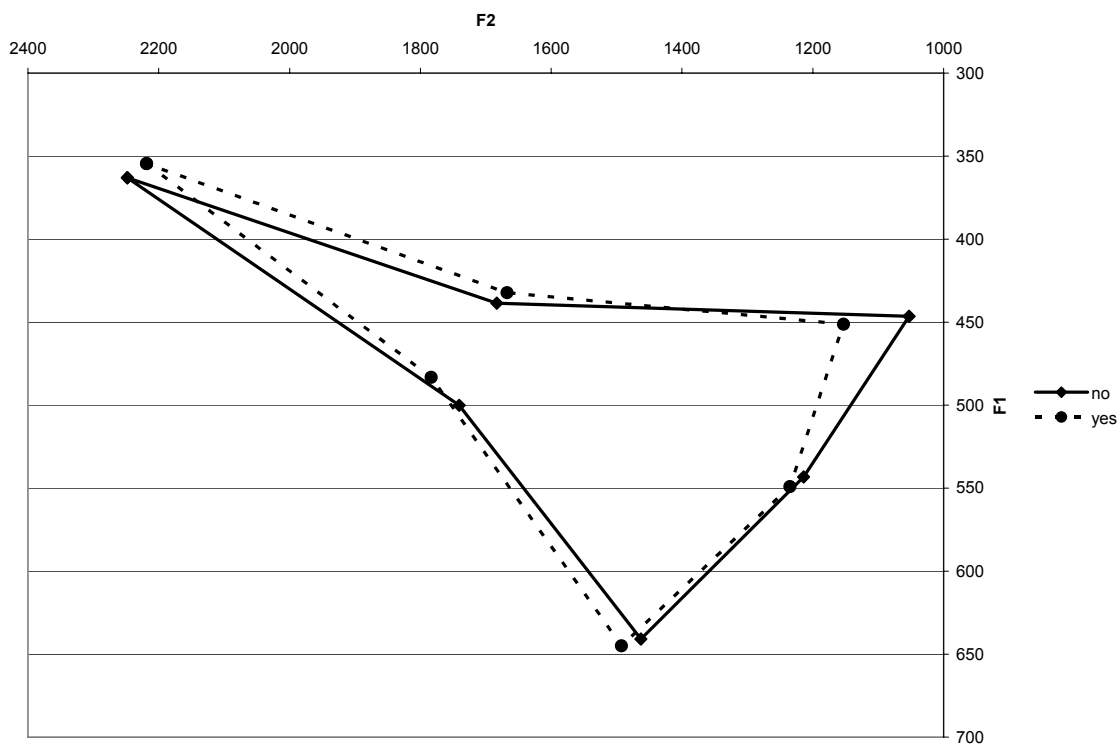


Figure 3.19.1 The potential effect of a preceding clause boundary.

Clause-initial phonological word	SED	Δ SED (%)
no	16216	0%
yes	15080	-7%

Table 3.19.2 Sums of Euclidean distances between grand means of vowels appearing in clause-initial and non-clause-initial phonological words.

In sections devoted to the pause-related variables (3.17-3.18), we have discussed the fact that the predictions related to domain-edge strengthening phenomena are unclear as far as vowels are concerned. This also applies to clause-initial words. A more useful approach may involve considerations of predictability, which we have invoked in sections 3.6 and 3.18. Given the fact that the beginning of the clause is more likely to present old (i.e. more predictable) information rather than new (i.e. less predictable) information, we could expect vowels in clause-initial phonological words to be more reduced. The Euclidean distance data suggest that this is the case. However, this potential effect is by no means clear, since what we see in figure 3.19.1 is a vowel space shift rather than shrinkage. The fact that the latter can be observed is almost certainly due to the fact that the vowel *u* is more fronted than others (for the other vowels the changes appear too small to be statistically significant).

3.20 Clause-final phonological word

Table 3.20.1 presents the summary statistics concerning the distribution of vowel tokens in clause-final and non-clause final phonological words in the entire dataset. Figure

3.20.1 presents the potential effect of these two conditions on the vowel space, and table 3.20.2 presents the sums of Euclidean distances.

Clause-final phonological word	Frequency	Frequency (%)
no	3294	66.9%
yes	1629	33.1%

Table 3.20.1 Breakdown of the data by the position of the phonological word with regard to clause ending.

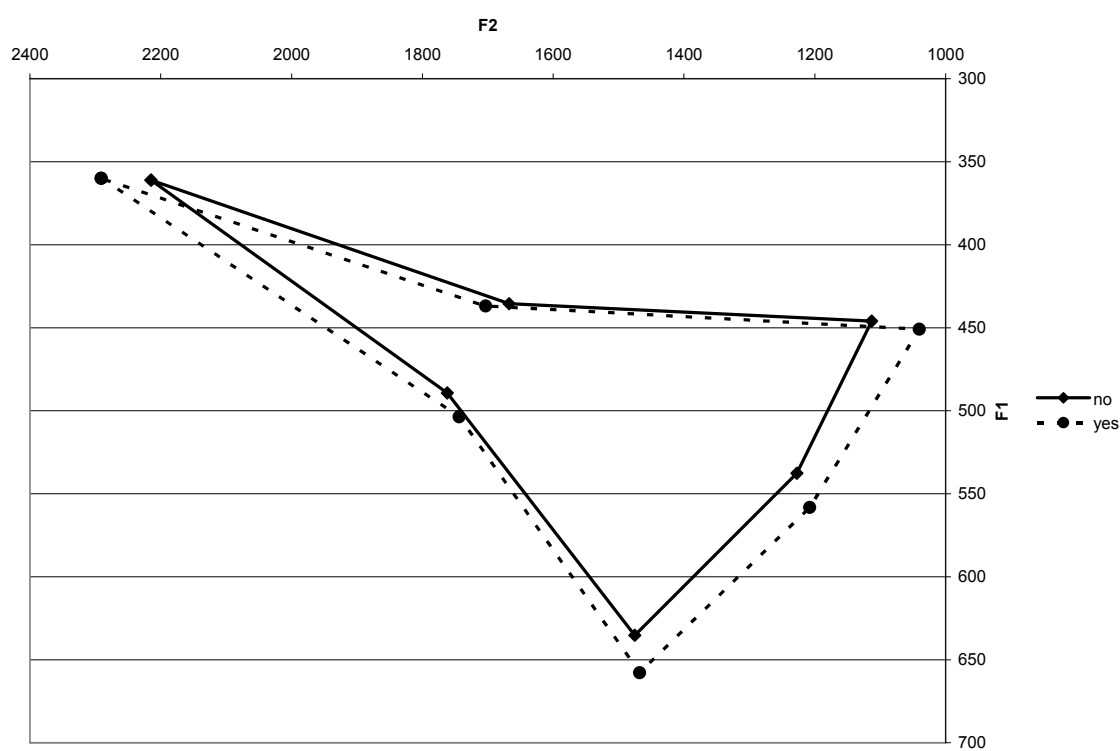


Figure 3.20.1 The potential effect of a following clause boundary.

Clause-final phonological word	SED	Δ SED (%)
no	15285	-9.8%
yes	16954	0%

Table 3.20.2 Sums of Euclidean distances between grand means of vowels appearing in clause-final and non-clause-final phonological words.

The phenomenon of domain-final lengthening, as well as considerations of information structure within a clause, both of which were discussed in the preceding sections, would lead us to expect that vowels should be less reduced in clause-final phonological words. To the extent that the effect observed here is verified statistically, this prediction is clearly confirmed both by the graphical display and by the Euclidean distance data.

3.21 Content vs. function words

Table 3.21.1 presents the summary statistics concerning the distribution of vowel tokens in content and function words in the entire dataset. Figure 3.21.1 presents the potential effect of these two conditions on the vowel space, and table 3.21.2 presents the sums of Euclidean distances.

Word	Frequency	Frequency (%)
content	3299	67%
function	924	33%

Table 3.21.1 Breakdown of the data by the content/function status of the word.

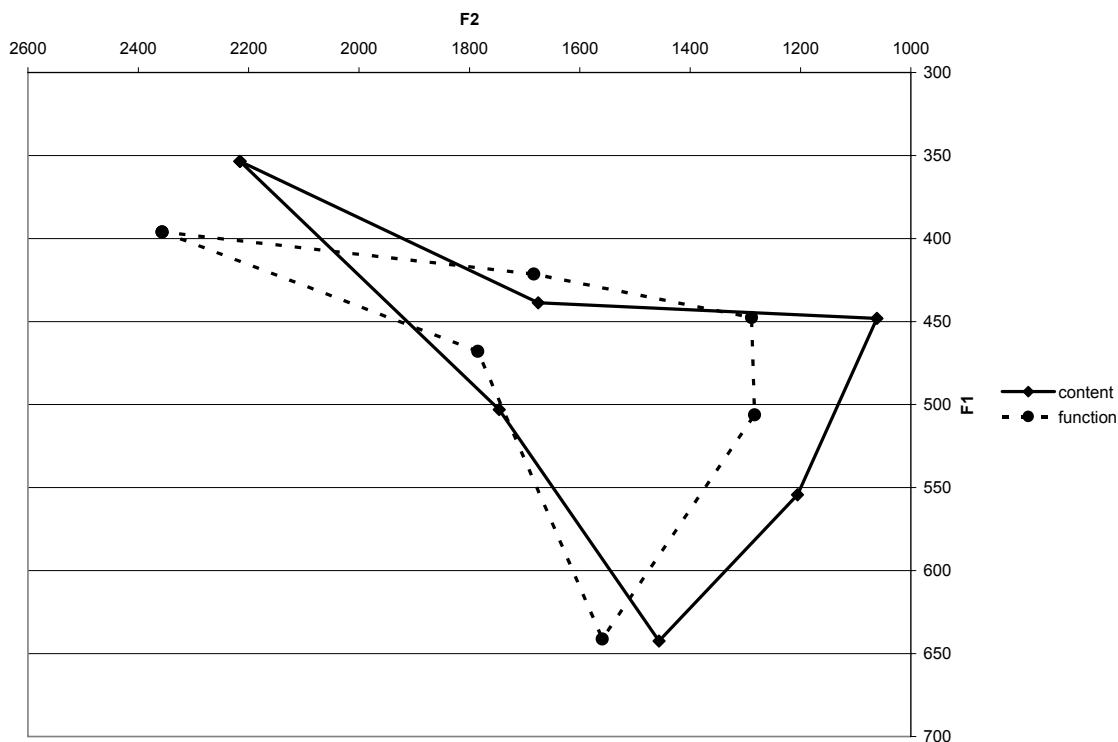


Figure 3.21.1 The potential effect of the word content/function distinction.

Word	SED	Δ SED (%)
content	15938	0%
function	14772	-7.3%

Table 3.21.2 Sums of Euclidean distances between grand means of vowels appearing in content and function words.

Vowels have been found to be less reduced in content words in other languages (e.g. van Bergem 1993). This tendency can be accounted for considerations of predictability of information (Jurafsky 2001, cf. also sections 3.6, 3.19, 3.20). All other things being equal, function words carry more redundant information and as such could be expected to reduce more. The Euclidean distance data confirm this prediction, but the somewhat idiosyncratic way in which the vowel space is transformed in figure 3.21.1 cannot be

easily accounted for. One possibility is that this merely reflects the fact that there are few function words that tend to be frequent. As such, the function word category may simply reflect accidental patterns – for instance, consonants of certain types may be more common among them. The fact that frequency-related considerations cannot be left aside when assessing the potential effect of the word content/function status becomes apparent when we compare figure 3.21.1 with figure 3.6.1. Graphically, the transformation of vowel space due to high-frequency of the word is almost identical to its transformation due to the vowel appearing in a function word.

3.22 Content vs. function syllables

Table 3.22.1 presents the summary statistics concerning the distribution of vowel tokens in content and function words in the entire dataset. Figure 3.22.1 presents the potential effect of these two conditions on the vowel space, and table 3.22.2 presents the sums of Euclidean distances.

Syllable	Frequency	Frequency (%)
content	2855	58%
function	2068	42%

Figure 3.22.1 Breakdown of the data by the content/function status of the syllable.

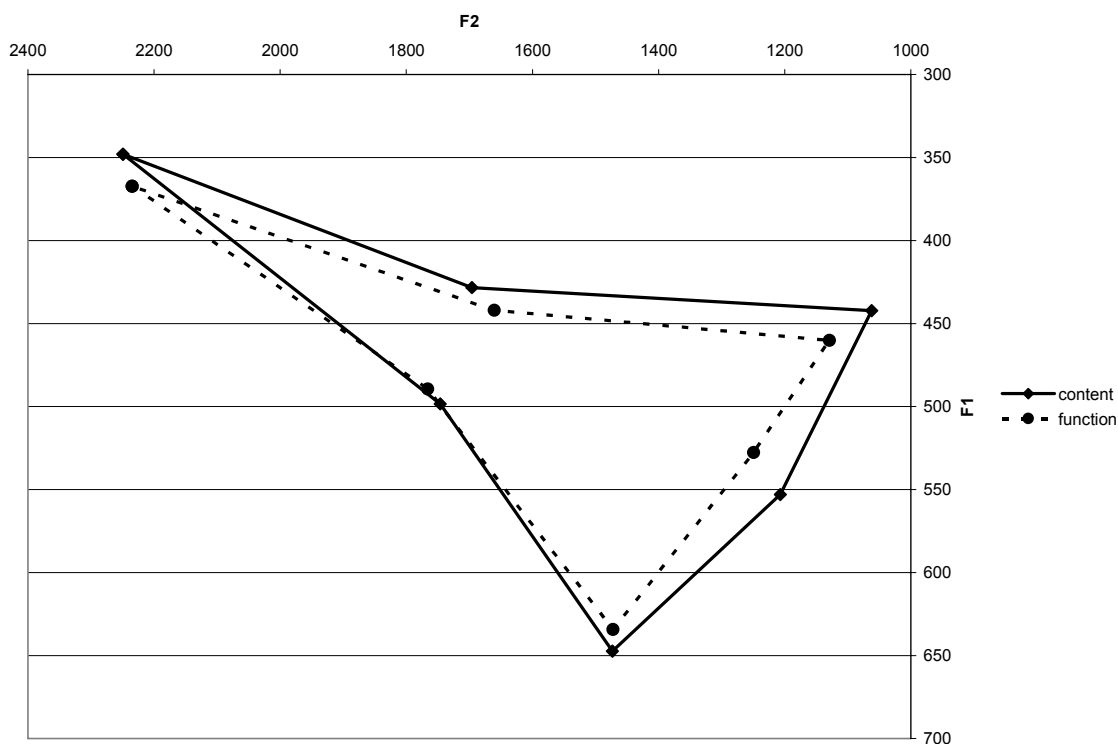


Figure 3.22.1 The potential effect of the syllable content/function distinction.

Word	SED	Δ SED (%)
content	16340	0%
function	15137	-7.4%

Table 3.22.2 Sums of Euclidean distances between grand means of vowels appearing in content and function syllables.

The reasons mentioned in the preceding section also apply to the content/function distinction between syllables: function syllables are expected to be more reduced due to their higher predictability than content syllables. This indeed appears to be the case, since the Euclidean distance data indicate that the vowel space is shrunk. Moreover, graphically, the transformation of the vowel space has the appearance of the “classic” centralizing shrinkage. However, caution is needed as the effect may be epiphenomenal.

Most importantly, many of the function syllables are inflectional endings; as such, they are likely to overrepresent the final, unstressed syllables, which are likely to be more prone to reduction anyway (3.15). As with any other variable, the existence of a true effect needs to be verified statistically.

3.23 Comparison of the Euclidean distance scores

To better evaluate the shrinkage of the vowel space in the different conditions, in this section we will compare the Euclidean distance scores across them. In order to enable a meaningful comparison with the preceding “palatal” condition, all the scores used in this section were recalculated using only the five vowels (*i*, *e*, *a*, *o* and *u*) that can appear in all consonantal contexts. Table 3.23.1 lists individual conditions ranked according to the Euclidean distance-measured size of the vowel space that is found in them. Table 3.23.2 presents the percentage shifts in the sums of Euclidean distances that happen within the individual variables, ranked according to the size of the percentage shift. (In this case, for the three level variable “style”, the shift was calculated as the percentage shift in the sum of Euclidean distances between the normal reading style and the average of the other two styles.)

Condition	SED
long duration	13219
low speech rate	12986
normal reading style	12689
stressed	12657
clause-final phonological word	12619
prepausal phonological word	12425
low comparative intensity	12238
“nonpalatal” following	12197
“nonpalatal” preceding	12151
content syllable	12148
sonorant preceding	12138
non-postpausal phonological word	12116
short phonological word	12091
non-clause-initial phonological word	12088
obstruent following	12008
<2 consonants following	12000
voiceless following	11979
content word	11859
low frequency word	11852
AVERAGE	11835
high comparative intensity	11782
non-prepausal phonological word	11744
voiced following	11741
voiceless preceding	11709
voiced preceding	11662
obstruent preceding	11608
non-clause-final word	11396
function syllable	11307
narrative style	11289
unstressed	11275
high frequency word	11254
clause-initial phonological word	11237
long phonological word	11222
fast reading style	11182
function word	11087
2+ consonants following	11048
sonorant following	11006
short duration	10814
high speech rate	10581
“palatal” following	10203
postpausal word	10029
“palatal” preceding	9360

Table 3.23.1 Comparison of the sums of Euclidean distances in different conditions. (SED= sum of Euclidean distances for the vowels *i, e, a, o* and *u*).

Variable	Δ SED (%)
preceding consonant place	22
speech rate	18.52
duration	18.19
non-postpausal vs. postpausal phonological word	17.23
following consonant place	16.35
speech style	11.45
position in the word	10.92
clause-final vs. non-clause final phonological word	9.69
following segment manner	8.34
number of following consonants	7.93
size of the phonological word	7.19
non-clause-initial vs. clause-initial phonological word	7.04
content vs. function syllable	6.92
content vs. function word	6.51
prepausal vs. non-prepausal phonological word	5.48
word frequency	5.05
preceding consonant manner	4.37
comparative intensity	3.73
following segment voicing	1.99
preceding segment voicing	0.40

Table 3.23.2 Percentage changes (Δ SED) in the sums of Euclidean distances within variables, calculated for the vowels *i, e, a, o, u*.

Bearing in mind that the effects have not been verified statistically, we can say that the conditions that are listed in the top half of table 3.23.1 – above the “average” row – seem to result in an expanded vowel space, while those listed in the bottom half seem to shrink it. In general, the picture that emerges is coherent and unsurprising. In particular, considering the very top of the ranking, we might say that the prototypical “unreduced” vowel is long, spoken at a low speech rate, read in the normal reading style and stressed. Most other conditions we find in this part of the table can also be accounted for by factors that were discussed in this chapter (e.g. preceding and following “nonpalatals”, being in a low-frequency content word etc.). For others, the cause of the effect is unclear (e.g. the effect of manner and voicing). The only condition whose presence in the top part of the

table is genuinely surprising is low comparative intensity, as the opposite effect could be expected. Of course, the bottom part of table 3.23.1 is essentially a mirror image of the top part and is thus equally unsurprising.

Essentially the same picture emerges when we consider the ranking of variables according to percentage shift in the sum of Euclidean distances presented in table 3.23.2. In general, the variables that would be expected to impact the vowel space the most appear on the top of the table (speech rate, duration) and those for which no strong effect could be expected appear at the bottom (e.g. the voicing variables). What is worth noting is that the comparative intensity variable is also very low in the table: this suggests that its rather surprising potential effect seen in table 3.23.1 may be epiphenomenal. In addition, the effect of preceding segment place is strikingly strong in comparison with others.

3.24 Conclusions

Any of the “potential effects” we have observed in this chapter could turn out to be epiphenomenal “when the statistical dust settles”. However, it will be somewhat surprising if at least some of the effects that appeared strong in this chapter do not turn out to be significant predictors of vowel reduction and, conversely, if many variables that did not appear important turn out to be crucial predictors. We now turn our attention to

statistical modeling, beginning with duration in the following chapter and continuing with each of the first two formants in the subsequent ones.

4 Vowel duration

In chapter 3 we have seen that vowel duration appears to have a strong effect on the shape of the vowel space. Of course, this potential effect remains to be verified statistically, but it will be hardly surprising if it does turn out to be a real effect. From the perspective of Lindblom's (1963) undershoot model, duration can be expected to be one of the crucial predictors of vowel reduction. As such, the factors that may influence it deserve to be analyzed separately and accordingly the current chapter is devoted to statistical modeling of vowel duration. In addition to being an interesting question in its own right, this analysis of duration will enable us to clarify our expectations regarding the impact of other predictors on formant frequencies. If a predictor that appears to have a strong influence on vowel formants (as discussed in chapter 3) is also found to strongly influence vowel duration, it will suggest that the former effect is epiphenomenal, i.e. that the predictor's apparent influence on vowel formants is a reflection of its influence on vowel duration rather than an independent effect in its own right. (Of course, whether this view is correct or not will be ultimately verified by statistical models of formant frequencies presented in the following chapters).

All the independent variables discussed before will be used as predictors in this chapter, except, of course, vowel duration, which will be the dependent variable.

As was discussed at the end of chapter 2, caution is needed whenever statistical correlations between variables are found, because the presence of a correlation does not in itself prove a causal relationship. One possibility that always has to be taken into consideration is that instead of the impact of x on y the correlation is in fact a result of y influencing x . This appears unlikely in the current dataset: most other variables are higher-level variables and it is extremely unlikely that they are affected by vowel duration rather than the other way round. For example, it is practically impossible to come up with a scenario according to which vowel duration affects word frequency or the choice of the neighboring consonants. In the few cases when difficulties concerning the directionality of a causal effect may arise, they will be discussed in the relevant sections. Likewise, we will note situations in which a correlation between x and y may not reflect a causal relationship between them at all: there is always the possibility that two variables covary because they are both influenced by yet another variable.

In addition to being an interesting question in its own right, the analysis of vowel duration in this chapter will serve two further functions. First, it will enable us to generally evaluate the effectiveness of the semi-automatic segmentation algorithm that was employed in this work. In general, if there are strong theoretical or empirical reasons (cross-linguistic ubiquity) to expect an effect of a certain predictor on vowel duration, the failure to find it in the data may indicate errors in the segmentation algorithm. Conversely, finding an unexpected effect for which no explanation can be found may indicate such errors as well. Of course, other possibilities abound and will be considered

when necessary. For example, the independent variables selected may not be sensitive enough to capture a certain effect, or an unexpected effect found may reflect a true but hitherto unexplored generalization about the durational properties of Polish vowels.

Second, the statistical modeling of vowel duration in this chapter will provide a practical illustration of the statistical procedures used in this work, which were described in a more abstract way in chapter 2.

4.1 Selecting the model

Following the methodology outlined in chapter 2, the entire dataset (4923 tokens) was randomly divided into a training set comprising 75% of vowel tokens (3692 tokens) and a validation set comprising 25% of vowel tokens (1231). The validation set was then used to select a model using the bidirectional stepwise search employing the Akaike Information Criterion (AIC), using raw duration scores in milliseconds as the dependent variable. With the exception of the variables clause-initial phonological word, postpausal phonological word and syllable content/function, all the independent variables were selected by the model. The 18 variables selected were then used to predict duration scores for vowel tokens in the validation set and “ R^2 ” (the correlation between observed and predicted scores in the validation set) was obtained. Independent variables were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “ R^2 ” was recalculated. It was found to be maximized when the

variables word frequency and phonological word size were also dropped from the model. Accordingly, the remaining 16 independent variables were used to calculate the validated model on the training set.

Table 4.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable from the model, with variables included in the validated model appearing above the thick line. Table 4.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 4.1.1 and 4.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

Independent variable	AIC rank	“R ² ”
vowel	1	n/a
style	2	0.1165
syllable position	3	0.1675
clause-final phonological word	4	0.2116
following segment manner	5	0.2466
preceding segment voicing	6	0.2705
comparative intensity	7	0.3009
speech rate	8	0.3127
subject	9	0.3262
preceding segment manner	10	0.3308
prepausal phonological word	11	0.3371
number of following consonants	12	0.3392
following segment place	13	0.3414
following segment voicing	14	0.3441
preceding segment place	15	0.3436
word content/function	16	0.3405
phonological word size	17	0.345
word frequency	18	0.3425
<i>full AIC model</i>	<i>n/a</i>	0.3426
<i>clause-initial phonological word</i>	<i>not selected</i>	<i>n/a</i>
<i>postpausal phonological word</i>	<i>not selected</i>	<i>n/a</i>
<i>syllable content/function</i>	<i>not selected</i>	<i>n/a</i>

Table 4.1.1 Variable selection for the duration model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R ²	“R ² ”	RSE	“RSE”
Value	0.3036	0.345	24.65	21.81
Difference	0.0414		-2.84	

Table 4.1.2 Duration models. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of duration in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

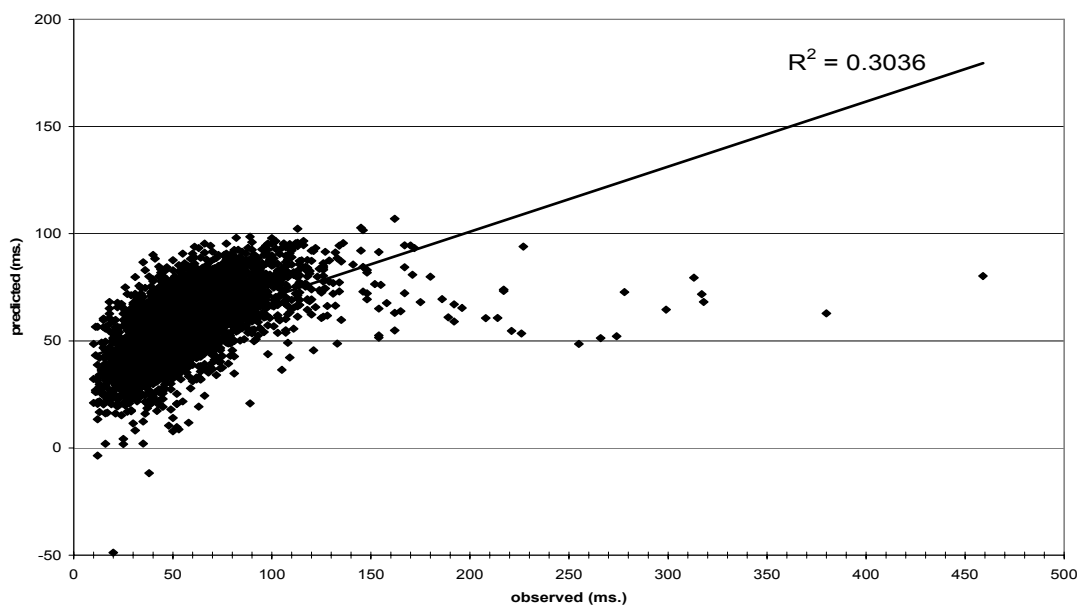


Figure 4.1.1 The validated model of duration. Correlation between duration values observed in the training set (horizontal axis) and values predicted for the training set by the 16 predictors included in the validation model (vertical axis). Note that unlike the observed values, predicted values can be negative.

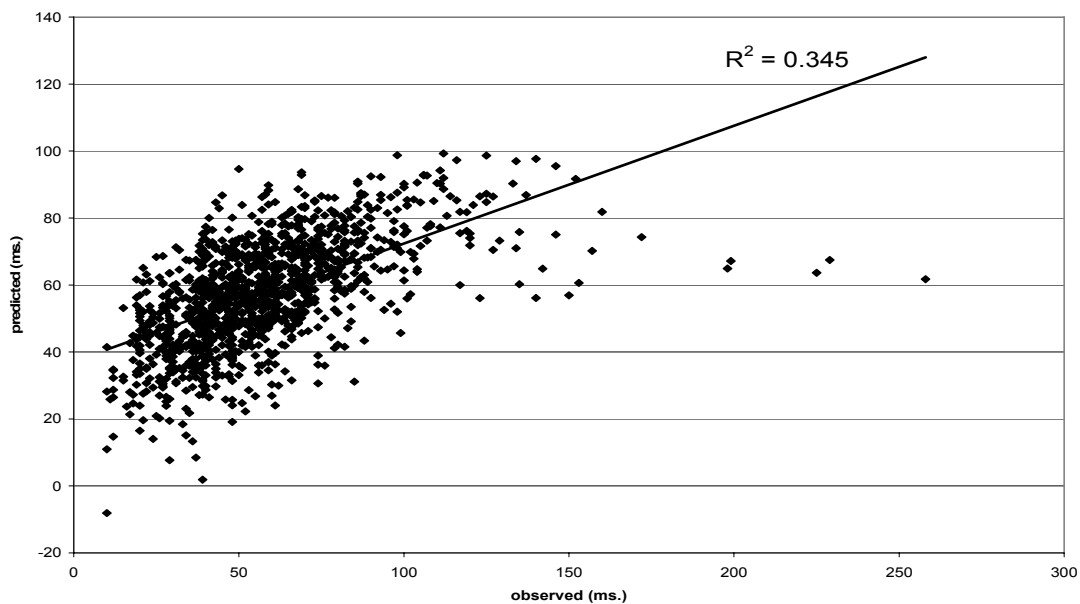


Figure 4.1.2 The validated model of duration. Correlation between duration values observed in the validation set (horizontal axis) and values predicted for the validation set by the 16 predictors included in the validation model (vertical axis). Note that unlike the observed values, predicted values can be negative.

Two questions need to be asked at this point, and the answers to both can only be tentative and orientational since there are few commonly agreed criteria against which they can be evaluated. The first is how successful the validation is. As was mentioned in chapter 2, Schwab (2006) suggests a difference of 0.05 or less between R^2 of models based on separate subsamples as a rule-of-thumb criterion of successful validation. This criterion is based on constructing separate models for the two subsamples, which means that regression coefficients are optimized for the given models. With the validation technique employed here, regression coefficients are optimized only for the training set, which means that the difference between the R^2 values could be expected to be even greater. In fact, it is 0.0414, i.e. it falls under the 0.05 criterion, suggesting that the validation is successful.¹ This is also confirmed by the fact that the difference between the RSE and “RSE” values is only 2.84. The residual standard error statistics estimate prediction error in the units of the dependent variable, thus we can interpret it as indicating that errors between obtained and predicted values in the validation set will tend to be within 3 ms. of the same errors in the training set, which is not a big difference.

The other question is whether the model itself is successful. R^2 values between 0.3036-0.345 indicate the proportion of the variance in the data explained by the model, and a correlation coefficient R between 0.55-0.587. RSE values indicate that the typical prediction errors are over 20 ms. All of these values indicate that the model is only

¹ Interestingly, the statistic is larger for the validation set, which is somewhat surprising. Since the validation set is smaller than the training set, it might be expected to be more susceptible to errors caused by outliers in the data. See also footnote 2.

moderately successful at best. This could indicate a number of problems. First, it could reflect errors of the semi-automatic segmentation algorithm. Second, it could reflect the impact of variables not included, including interactions between variables that are included. Finally, it could reflect the presence of important outliers in the data.

While the other two answers cannot be neglected and will be addressed to the extent possible further on in this chapter, the examination of figures 4.1.1 and 4.1.2 indicates that the last answer is not without merit. The graphical relationship is in general much stronger if only vowels under approximately 125 ms. are considered. While vowel tokens with observed values beyond that value do tend to have predicted values in the top part of the range, these rarely exceed 100 ms. The overall results appears to be that the regression lines are pulled down. In other words, the regression lines in both figures would be much steeper and data points would cluster around them more tightly if vowel tokens with values over 125 ms. were not considered. This indicates that vowels that could be called “super-long” in effect form a class of their own which is much less susceptible to the influence of context than “regular-length” vowels.²

To briefly test this effect, the AIC-stepwise search was conducted on the training set excluding the 80 vowels whose duration exceeded 125 ms. The R^2 value obtained was substantially higher (0.4517, indicating a correlation coefficient $R=0.65$). The variables

² Note that in the “super-long” vowels in the training set include some whose duration is over 300 ms., there are no such vowels in the validation set. This may account for the curious fact that predictions are more accurate in the validation set, as mentioned in footnote 1.

selected were in fact the same ones that were included in the final validated model, albeit selected in a slightly different order.³ Since there is no principled way of excluding vowels over 125 ms. long, the remaining discussion in this chapter will be based on the full data set model described in this section, but it has to be borne in mind that the specific regression coefficients discussed below could be somehow different if only the shorter vowels were considered.⁴

4.2 Categorical regression

The 16 variables that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.42, indicating correlation coefficient $R=0.64$. (It was thus substantially higher than the values obtained by MLR). Table 4.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple

³ The order was: vowel, style, syllable position, clause-final, following segment manner, preceding segment voicing, comparative intensity, speech rate, subject, preceding segment manner, prepausal, number of following consonants, following segment place, following segment voicing, preceding segment place and word content/function.

⁴ We could conceivably adopt the frequently-used method of ignoring tokens that fall beyond a specified range from the mean on a standardized scale, e.g. 2 standard deviations. However, this would very likely result in the elimination of a number of tokens on the low end on the scale. This would be an undesirable result, as the relationship appears to be strong there.

regression validated model, as discussed in the preceding section.⁵ All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	0	partial	part
vowel	1	0.065	0.339	0.337	0.395	0.329
speech rate	8	0.007	0.238	0.295	0.258	0.204
syllable position	3	0.034	0.220	0.278	0.262	0.208
clause final	4	0.014	0.151	0.175	0.175	0.136
following segment manner	5	0.010	0.150	0.159	0.183	0.142
style	2	0.040	0.116	0.214	0.128	0.098
preceding segment voicing	6	0.005	0.109	0.171	0.124	0.096
number of following consonants	12	0.004	0.107	0.003	0.125	0.096
comparative intensity	7	0.009	0.105	0.171	0.127	0.098
preceding segment manner	10	0.004	0.101	0.171	0.114	0.088
following segment place	13	0.005	0.094	0.073	0.117	0.090
preceding segment place	15	0.004	0.089	0.042	0.111	0.086
subject	9	0.006	0.085	0.062	0.110	0.084
following segment voicing	14	0.003	0.079	0.093	0.094	0.072
prepausal	11	0.005	0.078	0.176	0.091	0.070
word content/function	16	0.002	0.071	0.061	0.084	0.065

Table 4.2.1 Comparisons of MLR and categorical regression results, validated duration models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, *0* is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

In the first step, we will consider the question whether the MLR model and the categorical regression model are truly equivalent. This can be done by comparing the usefulness value of the MLR model with the square of the square of the part correlation coefficient of the categorical regression model. Mathematically, usefulness is equal to the squared part correlation coefficient. If the models are equivalent, the two values for the same variables should be similar. They cannot be expected to be exactly the same, in

⁵ Recall that “usefulness” is a technical term in MLR referring to the amount by which the R^2 value drops when a given variable is dropped from the model.

particular in the light of the higher R value of the categorical regression model, which can be attributed to the fact that categorical regression rescales all the variables in order to maximize the model fit. However, the relative values for different coefficients should be similar. Figure 4.2.1 compares the two values in a chart form.

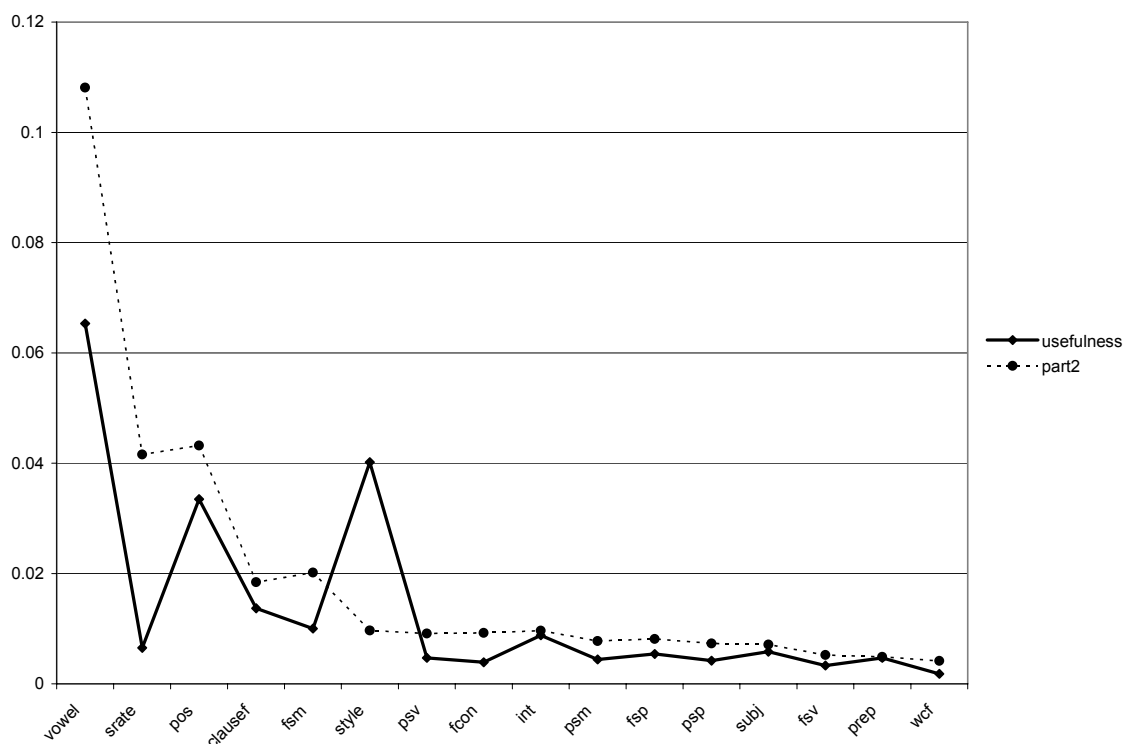


Figure 4.2.1 Duration. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

As can be seen, the relative values of MLR usefulness and CR squared partial coefficient are matched fairly well, with two apparent exceptions. MLR places the variable style much higher in the hierarchy than does CR; exactly the opposite is the case with the variable speech rate. Of course, given the fact that in this study one of the levels of the

variable style involved asking subjects to read as fast as possible, these two variables must be to a certain extent correlated (cf. section 3.4). It appears that the two ways of dealing with the categorical variable style – dummy coding in MLR and CR quantification – provide a different estimate of the relative importance of this variable vis-à-vis speech rate. It should be stressed that neither can be claimed to be more correct than the other. MLR returned low correlation coefficients, but was validated on a separate dataset; CR returned a better model fit, but was not validated. What this means is that the extent to which the speech style effect on duration is independent of speech rate has to remain somewhat indeterminate. This has to be borne in mind in the discussions that follow in this chapter.

Figure 4.2.2 shows graphically the relative contribution of independent variables to vowel duration estimated by the absolute values of the CR b-coefficients. The fact that different vowels have different intrinsic lengths is well known (Noteboom and Slis 1970); in this light, the fact that the greatest relative effect is that of the variable vowel is hardly surprising. For obvious reasons, the high-ranking of the speech rate variable is also unsurprising, subject to the caveat discussed in the immediately preceding paragraph. The impact of syllable position in the word, whose effect is very similar in size to that estimated for speech rate, can be accounted for to the extent that stressed vowels are known to be longer than unstressed vowels both in Polish and cross-linguistically (Jassem 1962, Crystal and House 1988). Two other variables also stand clearly above the rest: clause-final position of the phonological word and manner of the following segment.

Bearing in mind findings for other languages, this is also unsurprising (Crystal and House 1988).

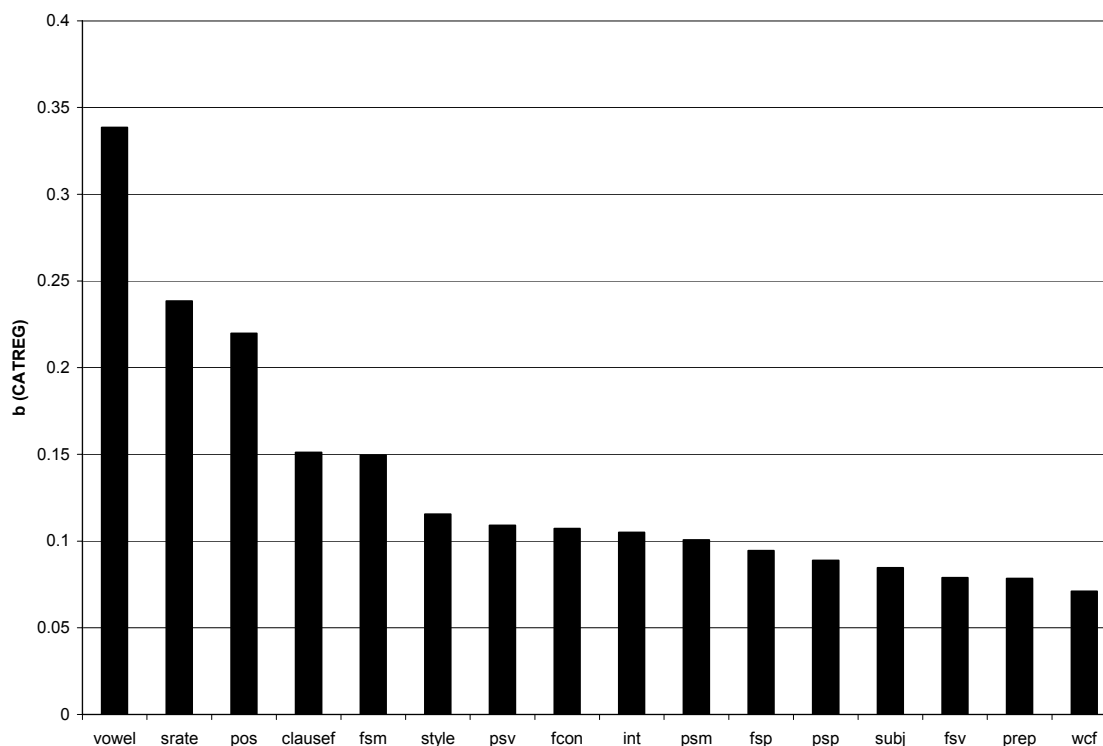


Figure 4.2.2 Relative impact of predictors on duration, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

4.3 Individual variables (MLR)

In this section, we will proceed to analyze each validated effect in detail, using regression coefficients for the individual dummy variables of the validated MLR model of the training set. For the dummy variables (i.e. the individual levels of the categorical variables) the MLR estimate can be interpreted directly as the amount in milliseconds

that being in a certain category adds to or subtracts from its estimated duration. There is no such obvious interpretation for continuous and categorical variables and in these cases the discussion will be somewhat more complicated.

4.3.1 Intercept

The intercept of a MLR equation is the estimated value of the dependent variable when all the regression coefficients equal 0. In the case of duration, this value is 14.2 ms. Given the choice of reference levels for categorical variables (cf. table 2.10.4.1 in chapter 2), this represents the vowel *a* spoken by speaker *m1* when the contribution of all the other variables equals 0. If all the variables were categorical and dummy-coded, this would represent the situation in which all the variables are at their reference levels. However, because the model includes continuous and ordinal variables, this is not the case. The intercept value thus is a mathematical construct with no intrinsic meaning beyond the role it plays in the regression equation.⁶

4.3.2 Vowel

Figure 4.3.2.1 presents the MLR-estimated regression coefficients for the levels of the variable vowel. The reference category is shown in capitals (in this case, the vowel *A*). Levels that are significantly different from the reference level are marked.

⁶ See in particular the discussion of comparative intensity in section 4.3.10.

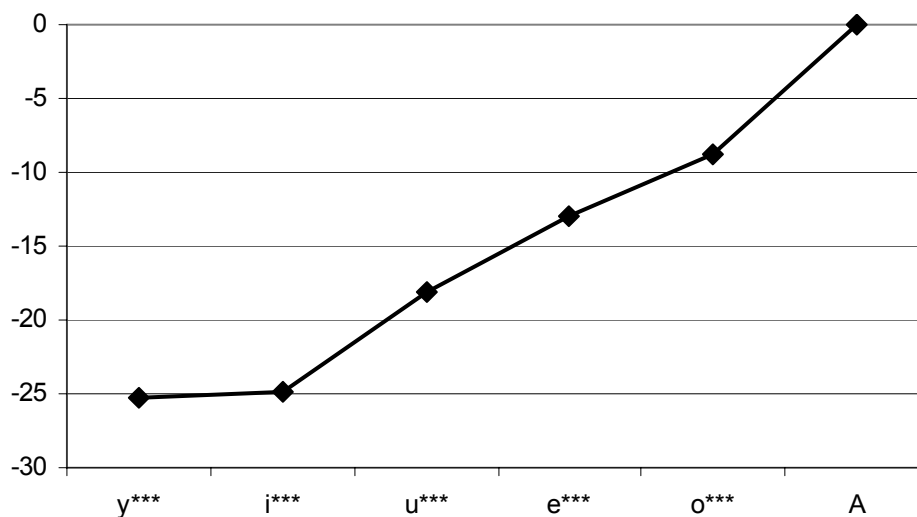


Figure 4.3.2.1 MLR-estimated effect of the individual levels of the variable vowel, in milliseconds, with *a* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

As can be seen, all the vowels are estimated to be significantly shorter than the vowel *a*, and the difference ranges from approximately 8 ms. in the case of *o* to approximately 25 ms. in the case of *y* and *i*. As can be seen from comparing figure 4.3.2 with figure 3.1.2 in chapter 3, the ranking of durational differences reflects almost perfectly the differences in vowel height. (The only exception is that the vowel *y*, which is estimated to be the shortest, is not the highest vowel). The fact that there is a correlation between vowel height and vowel duration is not surprising. It was attested in other languages, although its precise cause is unclear (cf. Neteboom and Slis 1970 and Ohala 1981).⁷ However, it also has to be remembered that the only differences that are tested for statistical

⁷ It may be worth pointing out that in their study of connected English speech, Crystal and House (1988) were only able to confirm the correlation for the extremes of the scale (i.e. high vs. low vowels).

significance in this model are these between *a* and all the other vowels individually, not those between any other combinations of vowels.

4.3.3 Speech rate

The regression coefficient for the continuous variable speech rate estimated by the validated MLR model was -0.93, statistically significant at $p < 0.0001$. As was discussed in chapter 2, speech rate varied between 1.8-10.8 syllables per second. This means that at the lowest speech rate, 1.67 ms. ($-0.93 * 1.8$) would be subtracted from the predicted value of the vowel, at the highest 10.04 ms. ($-0.93 * 10.04$). Thus the shift from the slowest to the highest speech rate predicts a decrease in vowel length of 8.37 ms. ($1.67 - 10.04$). It has to be remembered, however, that the MLR model estimated the contribution of speech rate as much lower than the CR model, likely ascribing the effect instead to speech style differences. Thus the true contribution of speech rate may, but does not have to be, underestimated here. This has to be borne in mind in particular when the value obtained is compared with the result seen in figure 4.2.2, which is based on categorical regression.

It is of course not surprising that there is an inverse relationship between speech rate and vowel duration. It is worth pointing out, however, that the causal underpinnings of this relationship are not, strictly speaking, obvious. It is possible to conceive of an analysis according to which speech rate is lowered by speakers deciding to produce longer and

longer vowels (i.e. making non-vowels longer is not part of the speakers' goal). The statistical methods employed here simply cannot invalidate this hypothesis. We assume, however, that it is more likely that speakers control the overall speech rate and when they change it, vowel duration changes as well.

4.3.4 Syllable position

Figure 4.3.4.1 presents the MLR-estimated regression coefficients for the levels of the variable syllable position. The reference category is shown in capitals (in this case, vowel in the penultimate syllable of a non-monosyllabic word). Levels that are significantly different from the reference level are marked.

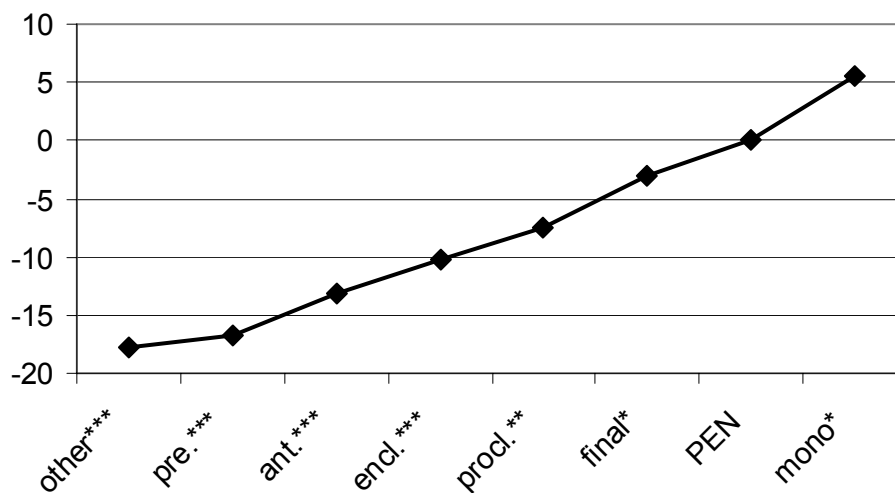


Figure 4.3.4.1 MLR-estimated effect of the individual levels of the variable syllable position, in milliseconds, with *penultimate syllable* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

It is a well-established fact that stressed vowels tend to be longer than unstressed ones (e.g. Lehiste 1970, Jassem 1962 specifically for Polish). This appears to be confirmed by the data in figure 4.3.4.1. As was discussed in section 3.15 in chapter 3, the vowels that could be expected to carry lexical stress are those in non-clitic monosyllables and in penultimate syllables of other words. As can be seen, vowel in all types of syllables are predicted to be significantly different in duration from vowels in the penultimate position. Since it is estimated that vowels in non-clitic monosyllables are predicted to be longer than vowels in penultimate syllables, we can extrapolate that the former are also likely to be significantly longer from all other vowels, which are predicted to be shorter. In other words, it appears that all vowels which can be expected to carry primary lexical stress are longer than those that cannot.

It is not surprising that the estimated difference between final vowels and penultimate vowels is very small (3 ms.). In fact in a more limited dataset, Jassem (1962) found no difference between these two types of vowels, which may reflect the more general phenomenon of domain-final lengthening (Wightman et al. 1992).⁸ He did find, however, a difference in length between vowels in penultimate syllables and in non-final, lexically unstressed syllables.

Two more interesting observations can be made. First, vowels in lexically unstressed syllables are predicted to be more and more shortened the further they are from the right

⁸ That is, the vowel is longer word-finally than could be expected on the basis of its stress status alone. Cf. also sections 4.3.5 and 4.3.16.

edge of the word. (This difference is quite large even when antepenultimate syllables are considered, i.e. 13.2 ms.) Second, vowels in clitics are longer than non-word-final unstressed vowels. This perhaps reflects the special status of the clitic domains as more than regular unstressed syllables but less than full words, although in the case of enclitics domain-final lengthening could also have played a role (with phonological word being the domain in question). It must be borne in mind, however, that for this model only differences between vowels in penultimate syllables and vowels in other syllables have been tested for statistical significance, and not the entire ranking.

4.3.5 Clause-final phonological word

Vowels appearing in clause-final phonological words were estimated to be 8.3 ms. longer than those that in other phonological words. There are two possible approaches to explaining this effect. On the one hand, it could be a reflection of the cross-linguistically common phenomenon of domain-final lengthening (Wightman et al. 1992). On the other, vowels in clause-final phonological words could be longer due to their greater information value in a typically SVO language such as Polish (cf. section 3.20).

On the former account, but not necessarily the latter, we could expect that more strictly domain-final vowels could be lengthened more. This can be analyzed as an interaction between syllable position in the word and clause-finality of the phonological word. To estimate this, the original variables syllable position and clause-final phonological word

were dropped from the model and new dummy variables introduced, corresponding to each combination of the levels of these two variables except for penultimate syllables in non-clause-final words, which served as the reference variable.⁹ We would expect to find that vowels in final syllables, enclitics and monosyllables, all of them from clause-final phonological words, are longer than the reference level.¹⁰ This was in fact largely confirmed. Vowels in final syllables and monosyllables of clause-final phonological words were estimated to be longer than penultimate syllables in non-clause-final phonological words (by 8.2 and 26 ms., respectively, both significant with $p < 0.0001$). Vowels in enclitics in clause-final phonological words were also found to be longer than the reference level, by 6.7 ms., but this effect was not significant. Most other categories displayed differences that were either negligible, or in the opposite direction. However, penultimate syllables in clause-final phonological words were significantly longer than those in non-clause-final phonological words (by 14.4 ms. at $p < 0.0001$). This indicates that either the effect of domain-final lengthening does apply to more than just the immediately domain-final vowels, or it could illustrate the information-value driven effect.

⁹ Such variables could also be added without dropping the original variables, but the results become very difficult to interpret in such a case, as estimations for the original variables are also changed. The procedure chosen is not inappropriate, as in fact what we are looking for is not a true interaction effect (extra contribution beyond the contributions of the original variables) but rather more fine-grained specification of the variables, in particular the variable clause-final.

¹⁰ Vowels in enclitics in clause-final phonological words are necessarily adjacent to the clause boundary. Vowels in monosyllables and final syllables in clause-final phonological words are adjacent to the clause boundary only if no enclitics follow; however, this is enough to set them apart from all the other vowels, which may never be adjacent to a clause-final boundary.

4.3.6 Following segment manner

Figure 4.3.6.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked.

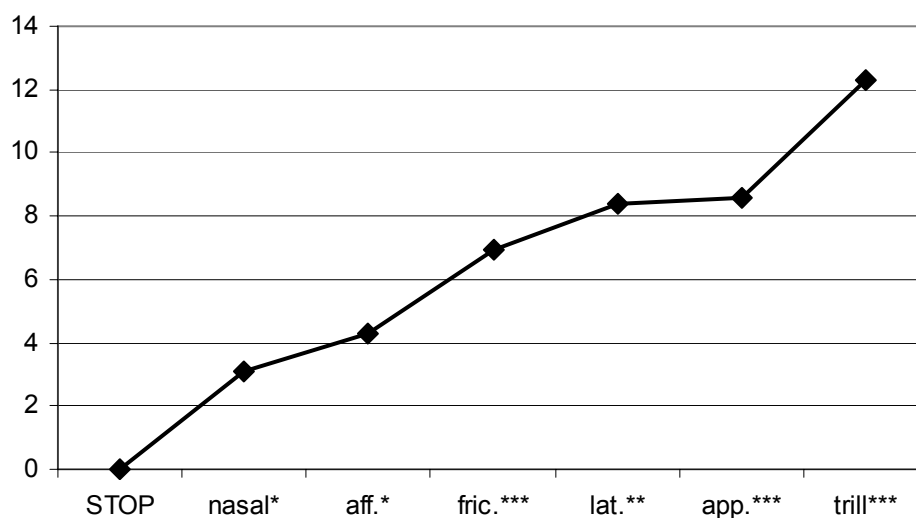


Figure 4.3.6.1 MLR-estimated effect of the individual levels of the variable following segment manner, in milliseconds, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

Vowels were found to be significantly shorter before stops than before consonants of all other manner categories. The small but significant differences between vowels preceding stops and vowels preceding nasals and affricates are somewhat surprising. Otherwise, the generalization may be that vowels are longer before segments with smaller degree of oral

constriction. This may be a genuine effect, but it may also be an effect of the amplitude-based segmentation criteria used in this work. Smaller oral constriction can be generally expected to correlate with more gradual amplitude loss between the vowel and the consonant, which could lead to a greater part of the vowel-consonant transition being assigned to the vowel.

4.3.7 Speech style

Figure 4.3.7.1 presents the MLR-estimated regression coefficients for the levels of the variable speech style. The reference category is shown in capitals (in this case, normal reading style). Levels that are significantly different from the reference level are marked.

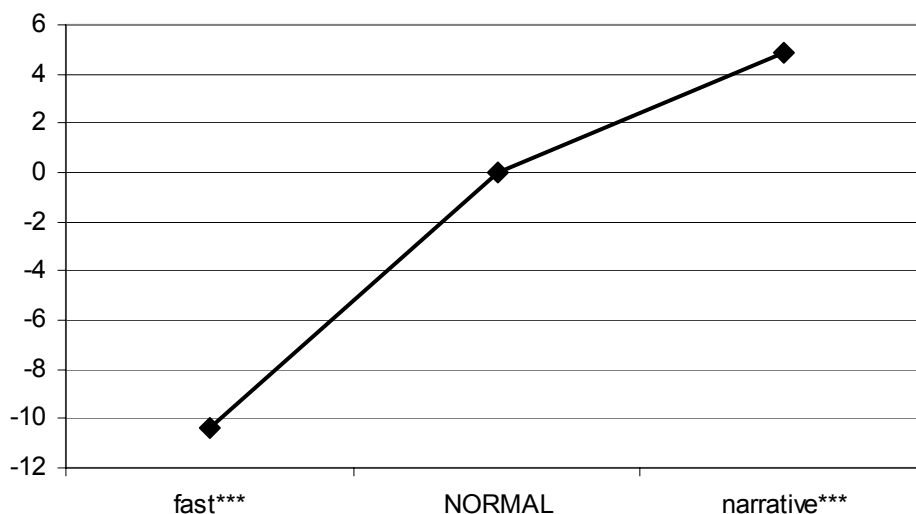


Figure 4.3.7.1 MLR-estimated effect of the individual levels of the variable speech style, in milliseconds, with *normal reading* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

Both the fast reading style and the narrative style are significantly different from the normal reading style. Given that these differences point in the opposite directions, it can also be inferred that they are likely to be significantly different from each other. Vowels in the fast reading style are predicted to be much shorter than vowels in the other two styles. Within this model, this is an effect that is independent of speech rate. It should be remembered at this point, however, that the categorical regression model appeared to ascribe the MLR-estimated speech style effect more to speech rate. Vowels in the narrative style are predicted to be somehow longer than vowels in the normal reading style. This effect is also independent of speech rate. This fact can be appreciated by considering figure 3.4.1 in chapter 3. We see there that three of the four speakers used the same speech rate in both styles, while one subject had in fact a *lower* speech rate in the normal reading condition.

4.3.8 Preceding segment voicing

Vowels were found to be significantly longer after voiced consonants (by 5.7 ms., $p < 0.0001$). The source of this effect is unclear. One important possibility is that this is an artifact of the measurement techniques used. The semi-automatic segmentation algorithm we employed tended to insert boundaries after the bursts of voiced stops, but at the beginning of voicing of voiceless stops. Any period of aspiration after voiceless stops was assigned to the consonant rather than the vowel; consequently, the duration of vowels after voiceless stops may have been underestimated. It is interesting to note that

Crystal and House (1988a) found that in English vowels are shorter after phonologically voiceless (i.e. aspirated) stops. That result may have exactly the same source as ours.

4.3.9 Number of following consonants

The MLR validated model estimated the regression coefficient for the ordinal variable number of following consonants at -3.56, significant at $p < 0.0001$. As was discussed in the preceding chapter, between 0-3 consonants followed vowels before another vowel in the same word. This indicates that the contribution of this variable varied between 0 ms. ($-3.56 * 0$) and -10.7 ms. ($-3.56 * 3$). In other words, vowels followed by three consonants (most likely to indicate a closed syllable) can be estimated to be 10.7 ms. shorter than vowels followed by no consonants (certainly an open syllable). The effect is not entirely surprising, since vowels have been found to be longer in open syllables in other languages (e.g. Navarro Tomás 1916, Maddieson 1985). However, it is worth pointing out that Jassem (1962) claimed that there is no such effect in Polish.

4.3.10 Comparative intensity

The MLR validated model estimated the regression coefficient for the continuous variable comparative intensity at 53.56, significant at $p < 0.0001$. As was discussed in the preceding chapter, comparative intensity varied between 0.752-1.38. This indicates that the contribution of this variable varied between 40.27 ms. ($53.56 * 0.752$) and 73.91 ms.

(53.56*1.38).¹¹ In other words, the vowel with the highest comparative intensity was estimated to be 33.64 ms. longer than the vowel with the lowest comparative intensity.

Unlike the other independent variables considered, comparative intensity cannot be thought as a cause of duration variation. If anything, the causal relationship may be the reverse: in longer vowels, there may be more time to achieve a greater opening of the vocal tract, which should correlate with higher comparative intensity. However, another possibility is that the contribution of the comparative intensity to the estimated duration score reflects the correlation of comparative intensity with a different variable which is not directly represented among the independent variables. One candidate is “non-lexical stress”, e.g. sentential accent and other kinds of emphasis. (In section 3.5 of chapter 3 we suggested that the potential effect of comparative intensity may be partially due to its correlation with syllable position, i.e. lexical stress. However, the regression coefficients reported here reflect the *unique* contribution of the variable, separate from that of the syllable position effect).¹²

¹¹ Recall that this is added to the intercept value of 14.2 ms., which is estimated using the unrealistic value of 0 for comparative intensity. As Ian Maddieson points out (p.c.), this result indicates that the intensity variable attracted to itself a considerable part of inherent vowel duration.

¹² One way to appreciate the difference is to examine the figures in table 4.2.1. The zero-order correlation coefficient of the comparative intensity variable, which is calculated without controlling for the other independent variables, is substantially higher than its regression coefficient (b).

4.3.11 Preceding segment manner

Figure 4.3.11.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked.

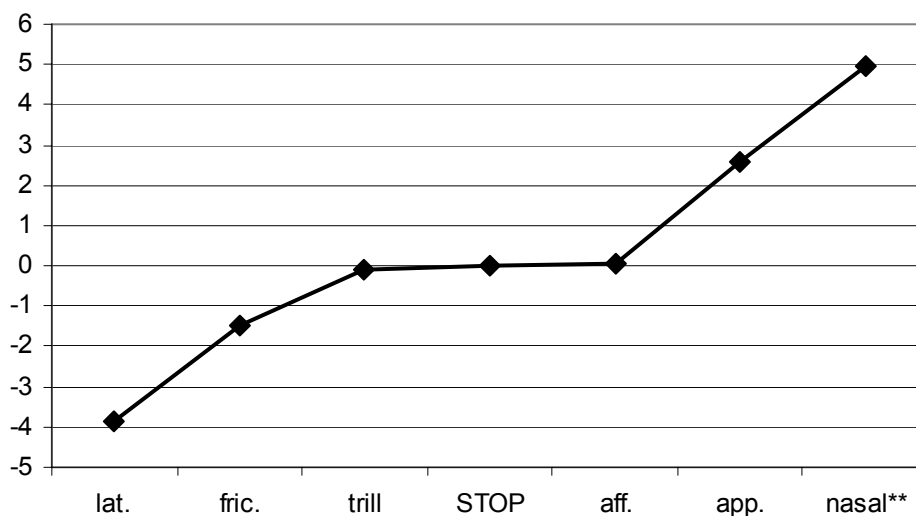


Figure 4.3.11.1 MLR-estimated effect of the individual levels of the variable preceding segment manner, in milliseconds, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

As can be seen, the differences attributable to the manner of the preceding consonant are small and do not form a clear pattern. In fact, the only difference that is statistically confirmed is between stops and nasals (5 ms.). Why there should be an effect like this is unclear. It may reflect properties of the semi-automatic intensity-based segmentation algorithm employed. Even though both stops and nasals involve a rapid oral closure, their

effects on the intensity contour are not the same; transitions between nasals and vowels are much more gradual (and thus the boundary is more difficult to locate with precision).

4.3.12 Following segment place

Figure 4.3.12.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dental). Levels that are significantly different from the reference level are marked.

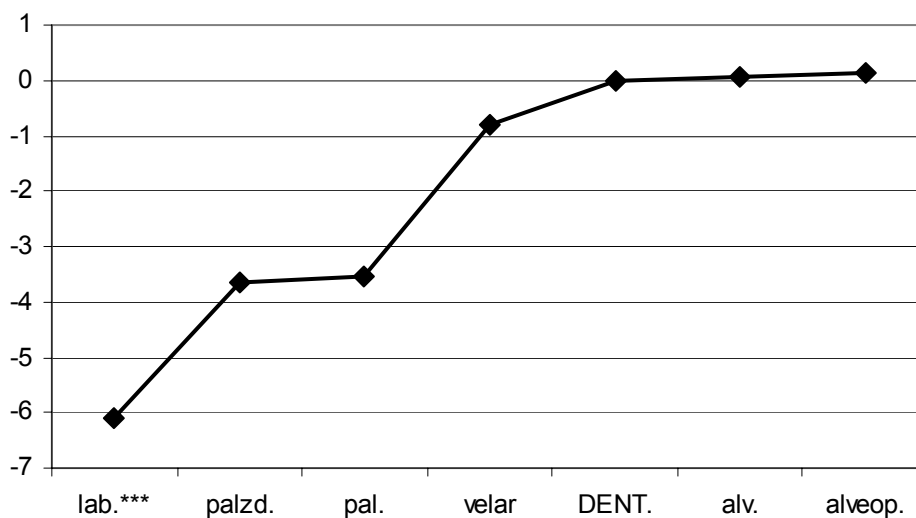


Figure 4.3.12.1 MLR-estimated effect of the individual levels of the variable following segment place, in milliseconds, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

The only difference induced by following segment place that is statistically significant is between labials and the reference category, dentals (6.1 ms.) This is a somewhat puzzling

result, as vowels have been found to be *longer* before bilabials in English (Crystal and House 1988), Palatalized and palatal consonants also condition somewhat shorter vowels, though not significantly so. If that effect was confirmed by with a larger sample of following palatals and palatalized consonants, it could perhaps reflect the fact that these consonants tend to have long transitions. Palatal transitions involve fairly close oral constriction, which entails lowered amplitude. As a result, the amplitude-based segmentation algorithm may have attributed the transition to the consonant rather than to the vowel. On the other hand, it is worth pointing out that most palatalized consonants are palatalized labials and that vowels before alveopalatals are not shorter than before dentals.

4.3.13 Preceding segment place

Figure 4.3.13.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this case, dental). Levels that are significantly different from the reference level are marked.

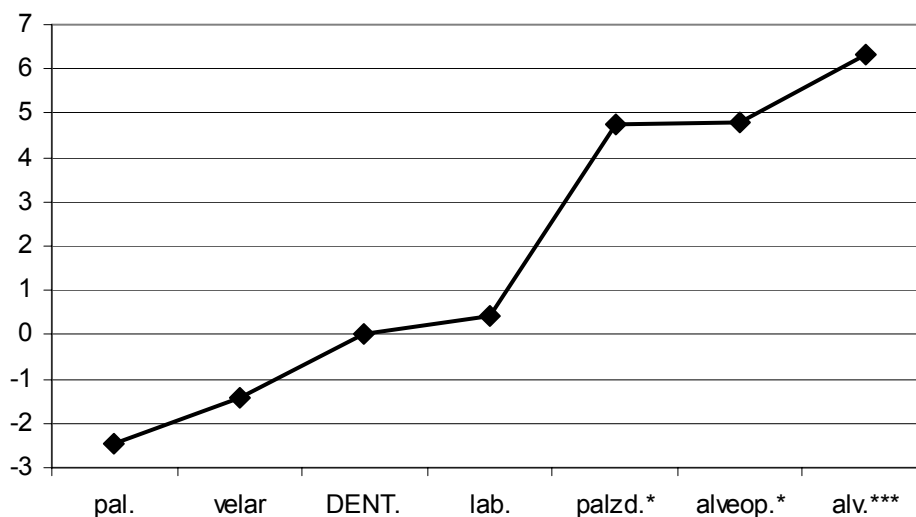


Figure 4.3.13.1 MLR-estimated effect (in milliseconds) of the individual levels of the variable preceding segment place, in milliseconds, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$).

When alveolars, alveopalatals and palatalized segments precede, they condition vowels to be significantly longer than they are after dentals. For the last two groups, it could be argued that the long palatal transitions need accommodation by increased vowel length, but this explanation cannot be easily extended to alveolars. What is more, vowels are actually shorter after palatal segments. This is genuinely puzzling and the effect of preceding segment place remains unexplained. It should also be noted that the place effects on duration are different when the following consonant is considered (cf. section 4.3.12).

4.3.14 Subject

Figure 4.3.14.1 presents the MLR-estimated regression coefficients for the levels of the variable subject. The reference category is shown in capitals (in this case, subject *m1*).

Levels that are significantly different from the reference level are marked.

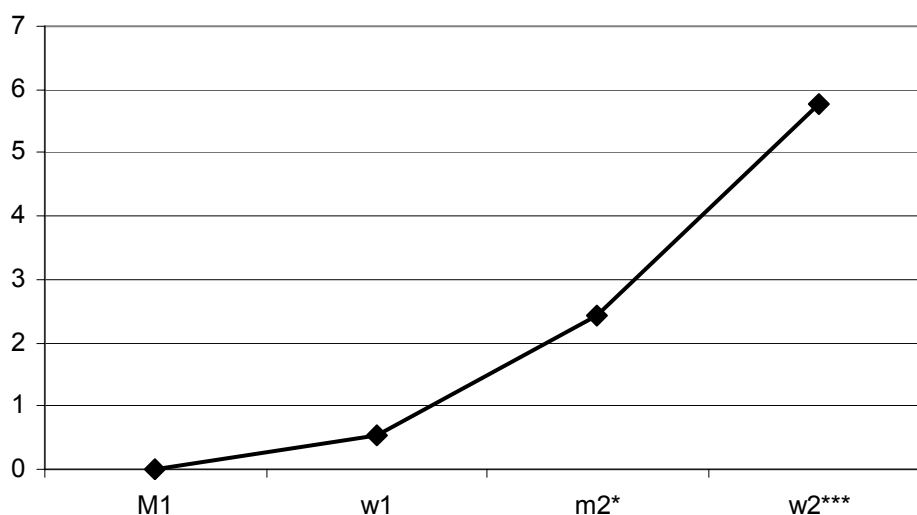


Figure 4.3.14.1 MLR-estimated effect of the individual levels of the variable subject, in milliseconds, with *m1* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$)

Figure 3.3.14.1 indicates that subjects *m2* and *w2* have significantly longer vowels than subject *m1* (by 2.4 and 5.8 ms.) This effect is of little linguistic interest and most likely reflects the subjects' idiosyncrasies.¹³ It does not seem that it can be ascribed to any

¹³ Alternatively, it may reflect different speaker-dependent errors of the segmentation algorithm.

known dialectal difference and it obviously does not reflect gender differences, since it is not aligned with gender differences among the subjects.

4.3.15 Following segment voicing

Vowels were found to be significantly longer before voiced consonants (by 4.5 ms., $p < 0.0001$). This type of effect is well attested in many languages, including English (cf. Crystal and House 1988). Lehiste (1977) claimed that it reflects the tendency for isochrony in VC groups, since voiced consonants tend to be shorter than voiceless consonants. Since consonant duration was not measured in this work, this claim cannot be verified here.

4.3.16 Prepausal phonological word

Vowels were found to be significantly longer in prepausal phonological words (by 6.7 ms., $p < 0.0001$). Problems with the interpretation of this effect parallel those raised by the clause-final effect (cf. 4.3.5). To the extent that pauses coincide with prosodic boundaries, the effect could be attributed to the phenomenon of domain-final lengthening (Wightman et al. 1992). Thus in this case also we would predict that vowels that are most likely to be immediately prepausal (vowels in monosyllables, word-final syllables and enclitics) will be lengthened more.

To test this possibility the same method was employed as in section 4.3.5. The original variables prepausal and syllable position were dropped from the model. Instead, dummy variables were added representing each combination of the levels of these two variables, except for penultimate syllable in non-prepausal phonological words (the reference level). The results closely paralleled those obtained for the similar investigation of the clause-final variable. Vowels in prepausal monosyllables were significantly longer than vowels in non-prepausal penultimate syllables (by 29.6 ms.). Vowels in prepausal enclitics and prepausal final syllables were also longer than the reference level, albeit not significantly so¹⁴ (in both cases by approximately 5.7 ms.). In addition, other unstressed vowels in prepausal words were actually significantly shorter than non-prepausal penultimates. This suggests that the prepausal lengthening effect is concentrated in the immediately prepausal syllables, which gives some support to the analysis in terms of domain-final lengthening effect. However, it needs to be pointed out that this is not the whole story: vowels in prepausal penultimate syllables were also significantly longer than vowels in non-prepausal penultimates, by 10.9 ms. (Recall that vowels in clause-final penultimates were also significantly longer than in non-clause-final penultimates). It appears then that the prepausal lengthening effect affects more than the immediately prepausal syllables; this could be analyzed either as a feature of the domain-final lengthening effect, or as suggesting that information value affects vowel length (given that prepausal words, likely to be clause-final, are more likely to carry new information).

¹⁴ The difference for the latter category was actually significant at $p < 0.1$, but this is below the lowest commonly accepted significance level.

It does not seem possible to distinguish between these two hypotheses given the current data.

4.3.17 Word content/function status

Vowels were found to be significantly longer in content words (by 4.5 ms., $p < 0.001$). This effect is not surprising and it is likely to reflect the lower predictability, and thus higher information value, of content words vis-à-vis function words (cf. Jurafsky et al. 2001).

4.3.18 Possible interaction effects

It is likely that the accuracy of the MLR model would improve if interaction effects were considered. For instance, Crystal and House (1988) found that in running English speech, vowels are lengthened before voiced consonants only in the prepausal condition. We have found evidence for both an independent prepausal lengthening effect and lengthening before voiced consonants, but in the light of their findings, it could be the case that vowels in prepausal phonological words are lengthened before voiced consonants to a greater degree than they are before voiceless consonants. We could also conceivably discover that durational effects differ before or after different combinations of voicing, place and manner (though Crystal and House (1988) did not confirm such effects in their data).

However, when a model includes 21 predictors, as is the case here, the number of potential interactions is huge. Even if we restrict them to those whose inclusion has sound theoretical motivation, adding interaction terms complicates the interpretation of main effects and their makes significance difficult to assess. This is particularly true for categorical variables, since new dummy variables have to be created for each combination of levels (except for the reference level for the interaction term). Adding many new variables poses the risk of artificially inflating the R^2 statistic without a genuine increase in the explanatory accuracy of the model. Since the examination of duration is *not* the main focus of this study, we will refrain from considering interaction effects here,¹⁵ but we will discuss them in our analysis of formant variability in the subsequent chapters.

4.4 Conclusions

Although the overall accuracy of the duration models presented here is not very high when the R^2 and RSE criteria are considered, the detailed examination of individual variables shows that they did capture the major contextual effects on vowel duration that have been observed in other languages. Most importantly, both MLR and CR confirmed strong effects of vowel height, speech rate, lexical stress, clause-final/prepausal position, speech style and syllable shape (as represented by the number of following consonants). In addition, a strong effect of comparative intensity was discovered, which likely reflects

¹⁵ The only exception are the considerations of interaction effects discussed in the sections on the clause-final phonological word effect (4.3.5) and the prepausal effect (4.3.16).

the impact of non-lexical accent on vowel duration. Furthermore, we have found minor but significant effects of preceding segment voicing, following segment voicing and place, and the content/function status of the word. In all of these cases, the findings are in agreement with what could be expected on theoretical grounds, on the basis of previous findings in other languages, or both.

The potential problems are of three types. First, there is the variance that the model failed to explain. This can reflect variables not considered (including interactions), measurement errors, genuine random variability in the data and, most likely, a combination of all three. Second, we found minor but significant effects which could not be fully accounted for, namely the impact of preceding and following segment manner and preceding segment place. These remain mysterious. Finally, we have not found evidence for two effects that have been seen in other languages: decrease in vowel duration as the length of the phonological word increases (Moon and Lindblom 1994) and the word frequency effect (Jurafsky et al. 2001, Gahl 2006). This is of course only a potential problem, because the fact that certain languages exhibit a certain effect does not guarantee that the same is the case in Polish

The last type of problem can in fact be due to the strictness of the validation criteria used. As can be seen from table 4.1.1, both word frequency and phonological word size were actually selected by the AIC criterion. When the original AIC model is considered, their effect was minuscule, but in the expected direction. Vowel duration was estimated to

decrease by approximately 1 ms. per every additional syllable in the phonological word and by 1.3 ms. per an increase in word frequency of 100,000. The fact that these two variables failed to pass validation may reflect the fact that these effects were not genuine, but may also be a function of the smaller size of the validation set.

It seems that the advantages of the validated model of duration outweigh its problems. Bearing in mind in particular the fact that even small effects were confirmed (e.g. the 4.5 ms. difference in vowel duration between content and function words), we take it to indicate that the semi-automatic segmentation algorithm employed in this work turned out to be fairly accurate. This in turn suggests that when in the following chapters we discover independent effects of predictors that are correlated with vowel duration (e.g. speech style and lexical stress) we can be reasonably certain that we are dealing with a genuine effect and not an effect that could be explained by duration if more precise segmentation were used. This does not mean, of course, that we will not attempt to exercise extreme caution when we consider the effects on vowel formants that will be found by the models discussed in the upcoming chapters.

5 First formant

In this chapter, we will present statistical models accounting for the variability of the first formant in our dataset. In our presentation, we will follow the statistical methodology explained in detail in chapter 2 and exemplified in our analysis of duration in chapter 4. Automatic Akaike Information Criterion search through predictor space will be conducted to select the optimal combination of independent variables. The regression coefficients obtained on a training set will be used to predict values for the validation set and variables which do not maximize the predictive power of the model will be dropped. The remaining variables will be used to construct a validated model on the training set. Regression coefficients obtained there will be used to investigate the effect of individual predictors in detail. In addition, we will use the same combination of variables to construct a categorical regression model of the data and assess in a more general way the relative impact of predictors.

Raw vowel formant scores cannot be analyzed directly in the same way raw duration scores can. Different speakers have different vocal tracts and, as a result, different vowel spaces (see figure 3.2.1 in chapter 3). Employing raw formant frequency scores in a regression analysis would result in highly significant results for the variable subject, which could obscure the role played by other, more linguistically interesting variables. For this reason, in our analysis of formant variability we will use standardized scores (see section 5.1 for details).

Another problem posed by formant variability analysis is that all vowels cannot be easily considered together. This is because some kinds of vowel reduction result in ostensible vowel space centralization (see, for instance, figure 3.4.2). If this is a real effect, it could not be captured by one general model applied to one formant: the effects of the lowering of high vowels and the raising of low vowels (or of the fronting of back vowels and the backing of front vowels) would essentially cancel each other out. For this reason, in this chapter we will construct three different models for the vowels of different heights (and three different models for the different degrees of vowel frontness in the analysis of the second formant in the following chapter). We will then compare the models in order to consider the patterns of first formant change in a more general way.

That said, the analysis of vowel reduction effects in the F1 domain is facilitated by the fact that consonants at the places of articulation we are considering here can be expected to have relatively uniform effects on the formants of different vowels. According to the perturbation theory of vowel acoustics, originally proposed by Chiba and Kajiyama (1941), a constriction at or near a velocity antinode (a point of maximum velocity and minimum pressure) of a natural resonance of the vocal tract lowers that resonance, while a constriction at or near a velocity node (a point of minimum velocity and maximum pressure) raises it. As shown in figure 5.1, the lowest natural resonance of the vocal tract (first formant) has an antinode at the lips and a node at the glottis. As can be seen, all the places of articulation that are relevant to Polish (labial through velar) are located closer to the lips than the glottis, i.e. closer to the velocity antinode. Accordingly, they can be

expected to lower the first formant. In perhaps simpler terms, it can be said that the “narrowing of the area function in the anterior half of the vocal tract (...) causes the first-formant frequency to decrease” (Stevens 1997:473).

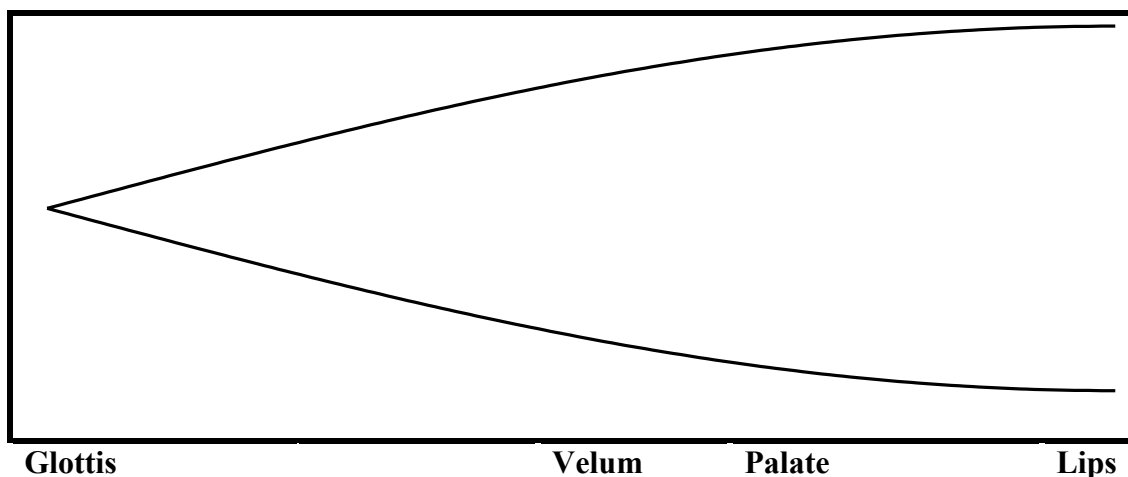


Figure 5.1 Locations of the velocity antinode (where the two sine waves are the furthest apart) and the velocity node (indicated by the intersections of the sine waves) in the first natural resonance of the vocal tract (F1). Based on Johnson’s (2003: 109) adaptation of Chiba and Kajiyama’s (1941) illustrations and on Ohala (2001).

In addition to considering the general effect, in the discussion of each of our models we will approach the question of potentially important interaction effects. Not all such effects can be considered: with twenty-two independent variables, even the first order interactions are numerous (231). We will focus primarily on interactions involving the key factors of the undershoot model of vowel reduction, i.e. duration and place. We will also consider potentially significant interactions involving other predictors that will turn out to have strong main effects. To analyze interactions, we will in each case add the interaction terms to the validated model to find out whether doing so significantly

increases R^2 in the training set. However, we will only accept the interaction effect as a real effect and analyze it in detail if it is validated, i.e. if it is found to increase the predictive power of the model when tested on the validation set. This will be verified by recalculating “ R^2 ” of the validation set for each significant interaction.

Throughout the chapter, we will loosely describe vowels as less or more reduced, where “more reduced” refers to increased influence of contextual factors. In other words, vowels will be taken to be “less reduced” when they appear to approach more closely their idealized target values. At the end of each section devoted to the three vowel classes, we will attempt to make these notions more precise.

5.1 Standardization

The F1 data were standardized for each intersection of the levels of the variables subject and vowel. That is to say, the entire dataset was divided into groups representing a single vowel phoneme spoken by a single speaker (twenty-four in total, since there were four speakers and six vowel phonemes). For each of these groups, the mean was calculated. It was then subtracted from each individual score, and the results were divided by the standard deviation for each group. The end product of this procedure are z-scores, which express the magnitude (in standard deviations) and the direction (by the sign) of the difference between a given result and the mean for the given group. For each group, the mean of z-scores is 0, and the standard deviation is 1.

The standardization procedure described above effectively eliminates differences between subjects that are due to their vocal tract sizes. Were the variable subject still to be found significant, this would indicate that the variability of vowels differs between speakers, i.e. that they used different vowel reduction strategies.

Table 5.1.1 presents the mean F1 values and standard deviations for each vowel of each subject that were used to standardize the scores. It also shows the mean standard deviations for each vowel (calculated by averaging the standard deviations of individual subjects).

vowel	m1		m2		w1		w2		average sd
	mean	sd	mean	sd	mean	sd	mean	sd	
i	326	39	362	42	384	48	377	63	48
y	406	40	428	43	484	60	420	49	48
u	403	37	439	48	517	58	433	47	48
e	449	66	511	54	537	83	476	62	66
o	527	68	532	51	601	82	515	68	67
a	601	78	624	46	707	78	628	101	76

Table 5.1.1 F1 means and standard deviations for individual subjects' individual vowels, used to standardize the data, as well as mean standard deviations for each vowel (averages of the subjects' standard deviations). All values are in Hertz.

It is worth pointing out that the three high vowels *i*, *y* and *u* have identical mean standard deviations and that the mid vowels *e* and *o* also pattern together in this respect. This suggests that they constitute genuine natural classes, although this way of calculating average standard deviations admittedly glosses over intra-speaker variability within each vowel category which can be seen in the data in table 5.1.1.

In the discussions that follow, we will be referring primarily to z-scores. However, in order to make it easier to directly assess the scope of a given effect in real terms, we will also informally reconvert z-scores to scores in Hertz, using average standard deviations for each class of vowels, calculated by averaging the average standard deviations for each vowel. These values are 48 Hz for the high vowels *i*, *y* and *u*, 66.5 Hz for the mid vowels *e* and *o* and 76 Hz for the low vowel *a*. Thus, we might say that a given variable lowers the F1 of high vowels by 1 standard deviation (sd), i.e. approximately 48 Hz. It has to be remembered that this last value will always be only an approximation, as in reality, due to the inter-speaker differences, one standard deviation does not mean exactly the same thing for different vowels of different subjects.

5.2 High vowels

High vowels have inherently low values of the first formant, and in general it is expected that vowel reduction will result in raising its value, i.e. in lowering the vowels in the two-dimensional vowel space. The prediction is somewhat less clear when consonantal influences are considered, since consonants of all the places of articulation relevant to our study involve constrictions that lower the first resonance of the vocal tract.

5.2.1 Selecting the model

There were 1165 high vowels *i*, *y* and *u* in the entire dataset (see section 3.1 for a detailed breakdown). Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (874) and a validation set, comprising 25% of tokens (291). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Of the twenty-two independent variables considered, twelve were selected into the model. They were then used to predict normalized F1 scores for vowel tokens in the validation set and “ R^2 ” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “ R^2 ” was recalculated. It was found to be maximized when five more variables were dropped from the model. Accordingly, the remaining 7 independent variables were used to construct a validated model of the training set.

Table 5.2.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 5.2.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 5.2.1.1 and 5.2.1.2 show the correlations between predicted and obtained values

of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

It should be noted that the variable vowel was not selected by the AIC search procedure, indicating that there are no important differences in the formant variability of the three different high vowels. In other words, with respect to vowel reduction they appear to behave like a true natural class. It will be shown later in the chapter that some differences between the vowels can nonetheless be discovered through the examination of interaction effects.

Independent variable	AIC rank	“R²”
following segment manner	1	<i>n/a</i>
preceding segment manner	2	0.1290
comparative intensity	3	0.2466
preceding segment place	4	0.2771
following segment place	5	0.2686
speech rate	6	0.2896
prepausal phonological word	7	0.2920
syllable content/function	8	0.2977
word content/function	9	0.2829
phonological word size	10	0.2852
number of following consonants	11	0.2875
clause-initial phonological word	12	0.2789
<i>full AIC model</i>	<i>n/a</i>	0.2794
preceding segment voicing	<i>not selected</i>	<i>n/a</i>
vowel	<i>not selected</i>	<i>n/a</i>
word frequency	<i>not selected</i>	<i>n/a</i>
duration	<i>not selected</i>	<i>n/a</i>
following segment voicing	<i>not selected</i>	<i>n/a</i>
clause-final phonological word	<i>not selected</i>	<i>n/a</i>
postpausal phonological word	<i>not selected</i>	<i>n/a</i>
speech style	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>
syllable position	<i>not selected</i>	<i>n/a</i>

Table 5.2.1.1 Variable selection for F1 high vowels model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R²	“R²”	RSE	“RSE”
Value	.279	.298	.823	.785
Difference	.019		-0.038	

Table 5.2.1.2 F1 in high vowels. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of F1 in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

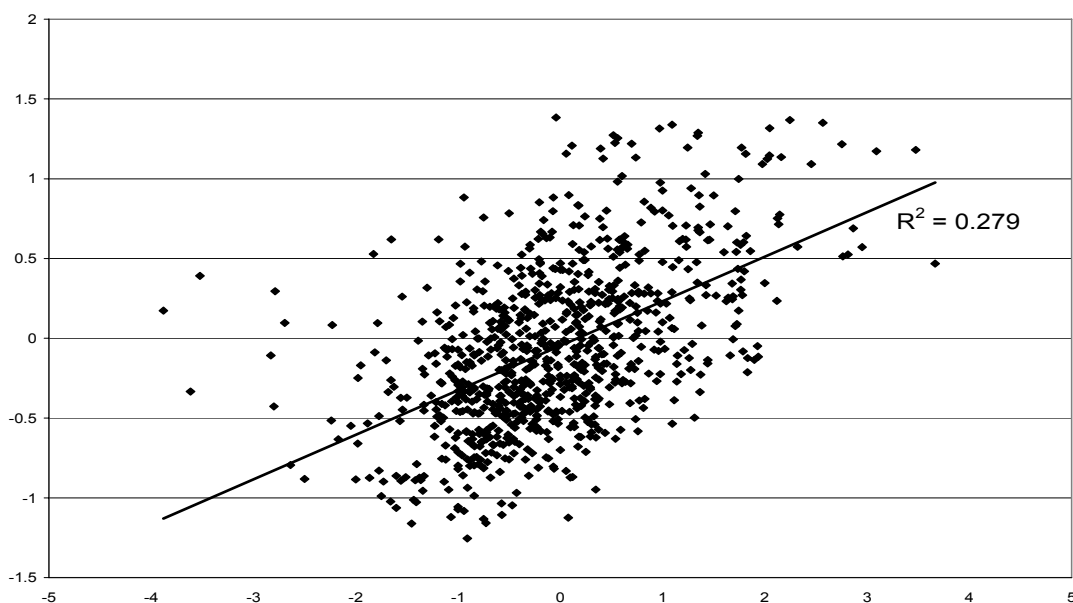


Figure 5.2.1.1 The validated model of F1 of high vowels. Correlation between F1 values observed in the training set (horizontal axis) and values predicted for the training set by the 7 predictors included in the validated model (vertical axis). All units are standard deviations.

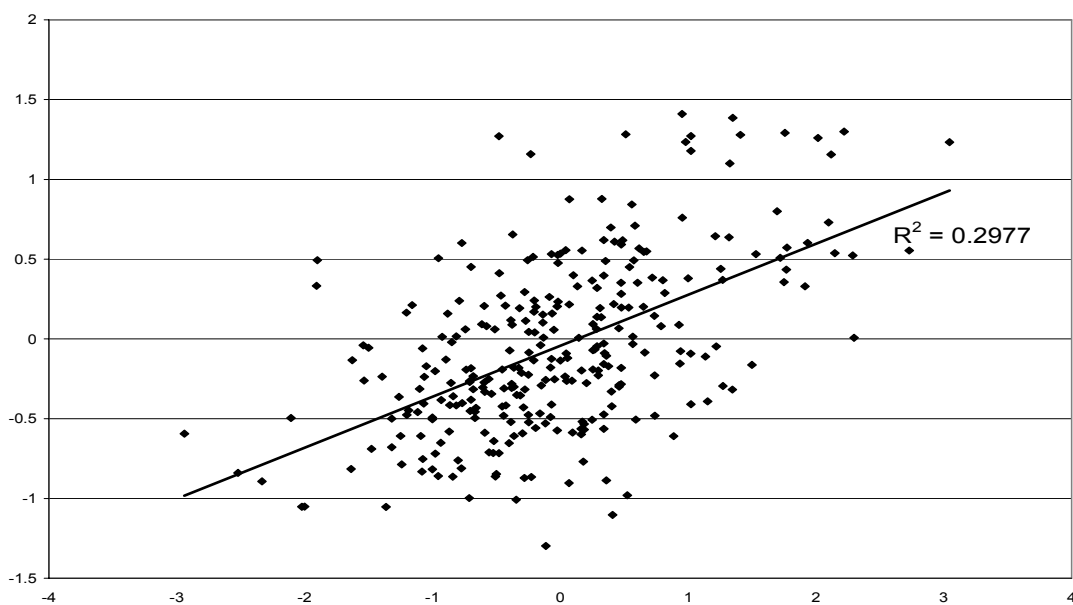


Figure 5.2.1.2 The validated model of F1 of high vowels. Correlation between F1n values observed in the validation set (horizontal axis) and values predicted for the validation set by the 7 predictors included in the validated model (vertical axis). All units are standard deviations.

As can be seen from the data in table 5.2.1.2, the validation of the model has been rather successful: differences between the equivalent key statistics of the training set and the validation set are small. This indicates that regression coefficients optimized for the training set fit equally well the subset of data that is contained in the validation set.

However, the overall predictive power of the model is limited, with less than one-third of variance accounted for and Pearson's correlation coefficient R equal to 0.528-0.546, indicating only a low-to-moderate relationship between the dependent variable and the set of predictors. As can be seen from figures 5.2.1.1 and 5.2.1.2, this is not due to the fact that the regression equation does not work equally well for the entire data range. (These figures should be compared with figures 4.1.1 and 4.1.2 in the preceding chapter, where it was clear that the relationship between vowel duration and its predictors was much stronger in one subset of the data). The weak performance of the model requires a different explanation and we will consider the possibilities further on in this chapter.

It is worth pointing out at this point that the F1 model of high vowels will turn out to be by far the least accurate statistical model considered in this work.

5.2.2 Categorical regression

The 7 variables that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of

different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.2905, indicating correlation coefficient $R=0.539$. The key statistics were thus very similar to the values obtained by MLR. Table 5.2.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	0	partial	part
following segment manner	1	0.0976	0.3476	0.3693	0.3748	0.3405
preceding segment manner	2	0.0869	0.3428	0.2883	0.3480	0.3126
preceding segment place	4	0.0200	0.1895	0.0307	0.1992	0.1712
following segment place	5	0.0153	0.1520	0.1531	0.1754	0.1501
comparative intensity	3	0.0122	0.1182	0.1814	0.1360	0.1157
prepausal	7	0.0064	0.0885	0.0900	0.1031	0.0873
speech rate	6	0.0038	0.0630	0.0754	0.0732	0.0618

Table 5.2.2.1 Comparisons of MLR and categorical regression results, validated F1 high vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). *Usefulness* refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, *0* is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 5.2.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. As can be seen, the values are similar, in particular if their relative values are considered. This indicates that in this case, multiple linear regression and categorical regression are closely matched and we are justified in treating the results obtained by both of these techniques as equivalent.

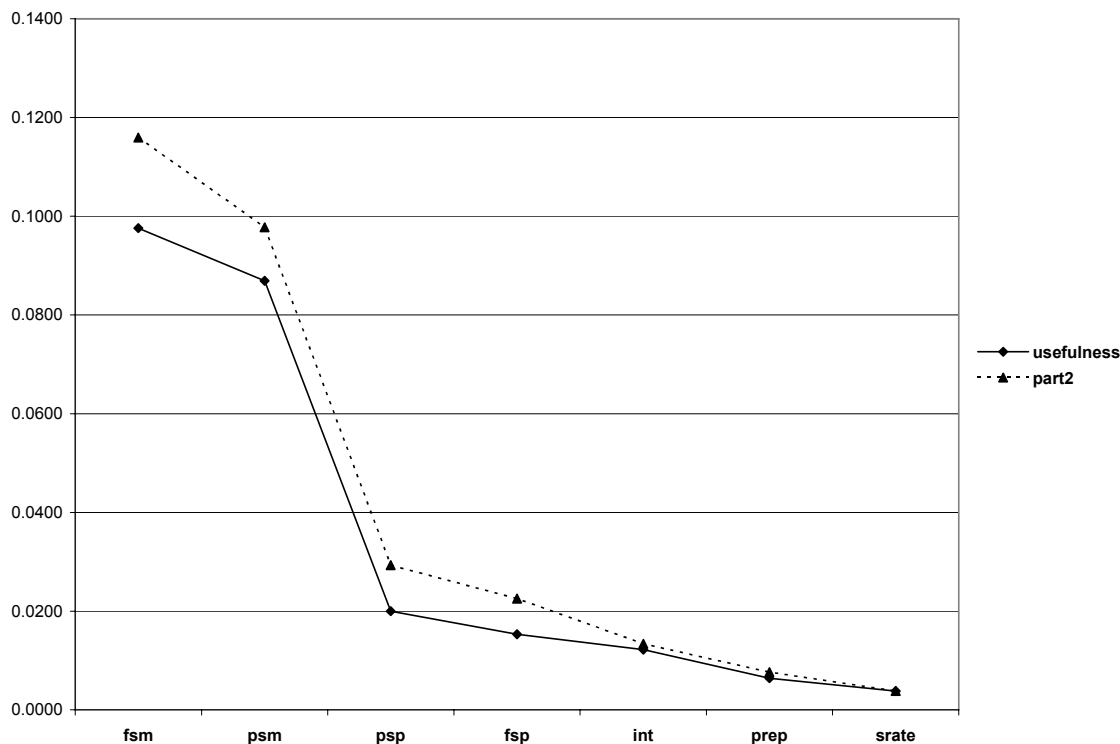


Figure 5.2.2.1 F1 of high vowels. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 5.2.2.2 shows graphically the relative contribution of independent variables to F1 of high vowels estimated by the absolute values of the CR regression coefficients. The manner variables emerge as far stronger than the place variables, but the fact that the four manner and place related variables emerge on top is hardly surprising. The other three variables in the model – comparative intensity, prepausal phonological word, and speech rate – are not entirely unexpected, although their role cannot be adequately assessed until we consider them in detail in the following section. What is surprising from the point of view of the undershoot model of vowel reduction is that vowel duration was not found to be an important predictor of vowel reduction.

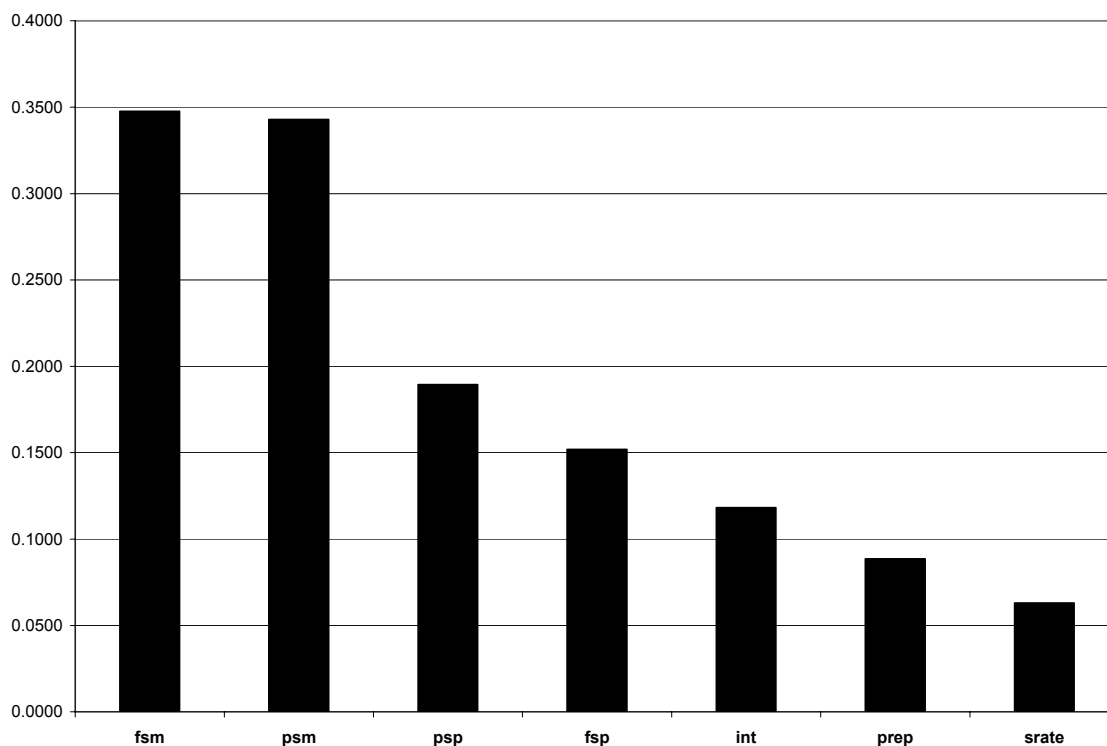


Figure 5.2.2.2 Relative impact of predictors on F1 of high vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

5.2.3 Individual predictors (MLR)

In this section, we will consider the impact of individual predictors in detail. The basic information will be provided in standard deviations, however, we will also provide Hertz estimates, obtained using the average standard deviation of high vowels (48 Hz). As was explained in section 5.1, such values need to be treated with great caution, as they may gloss over significant inter-speaker and inter-vowel variability.

All possible first-order interactions between the place and segment variables on the one hand and comparative intensity and speech rate on the other were examined. None of them significantly increased model fit. In addition, we considered interactions between the place and manner variables and duration, since the fact that this variable was not selected into the model could not be taken to mean definitely that it plays no role in the variability of the first formant of high vowels. Only the interaction between duration and following segment place significantly improved model fit (increase in R^2 by 0.0169), but it did not improve prediction (drop in “ R^2 ” by 0.0077). All the other interactions of place and manner variables with duration failed to significantly improve model fit. Finally, we considered the interactions of place and manner variables with the vowel variable, which also was not selected into the validated model, but could potentially be important through interaction with other variables. Two of the four interactions did significantly improve both model fit and prediction and are discussed in the relevant sections.

5.2.3.1 Intercept

The intercept of the MLR equation was -2.55. In other words, a high vowel was predicted to have an F1 lower by 2.55 standard deviations (122 Hz) from its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1, this is a purely mathematical construct with no inherent meanings, since some of the variables cannot have the value of 0 (in this case, speech rate and comparative intensity).

5.2.3.2 Following segment manner

Figure 5.2.3.2.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

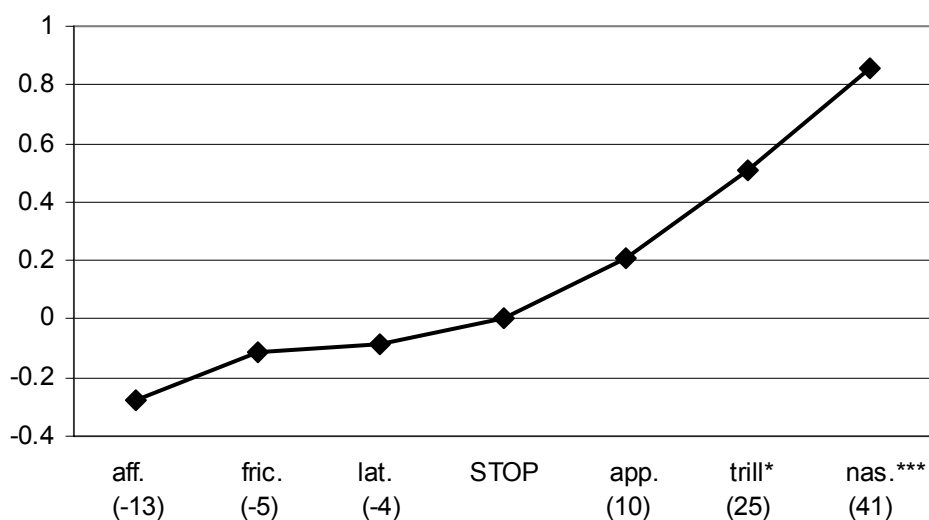


Figure 5.2.3.2.1 MLR-estimated effect of the individual levels of the variable following segment manner on F1 of high vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 48 Hz as the average value of standard deviation for the F1 of high vowels.

The largest statistically significant effect is the raising of the first formant in vowels followed by nasals. This can be interpreted as an expected effect of vowel nasalization (cf. Ohala 1975, Wright 1986). The first formant is also significantly raised in vowels preceding trills. This is not unexpected: as the data of Recasens (1991) and Recasens and

Pallares (1999) indicate, in comparison to other alveolar segments, the articulatory configuration for trills requires an active lowering of certain parts of the tongue (predorsum and mediodorsum). No other individual differences are significant vis-à-vis the reference category, stops. Two of them, however, are relatively large (approximants, raising of F1 by approximately 10 Hz) and affricates (lowering of F1 by approximately 13 Hz). If these effects were confirmed with a larger sample, the former could be attributed to the fact that approximants have a smaller degree of oral constriction than stops, and thus involve less elevation of tongue body which may affect the articulation of the high vowels. The latter effect, however, would be difficult to account for, as one would not expect following stops and affricates to have different effects.

The interaction between following segment manner and vowel was nearly significant (increase in R^2 by 0.0183, $p < 0.1$), but it did not improve prediction (drop in “ R^2 ” 0.0247).

5.2.3.3 Preceding segment manner

Figure 5.2.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

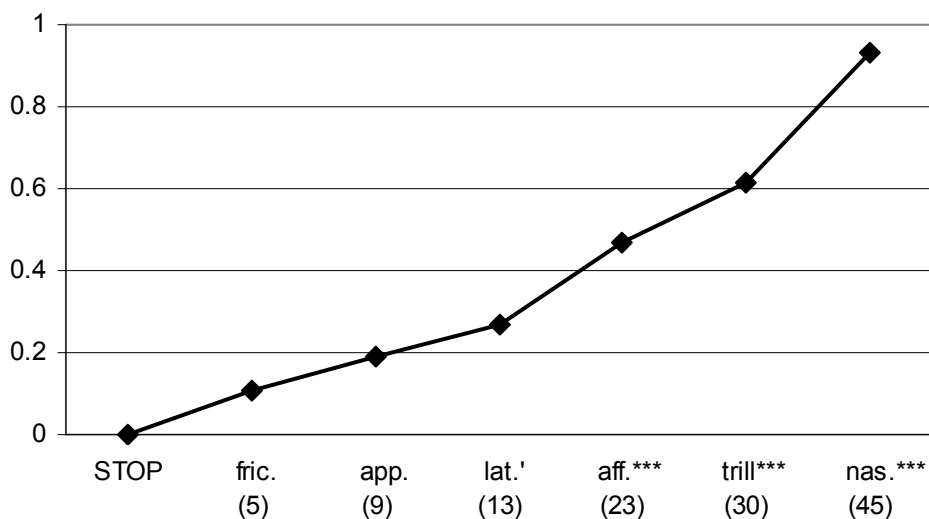


Figure 5.2.3.3.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F1 of high vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 48 Hz as the average value of standard deviation for the F1 of high vowels.

It is likely that the F1-raising effects of preceding nasals and trills can be accounted for in the same way as the effect of following consonants of this type in the preceding section. Accounting for the other effects, two of which are significant or nearly significant vis-à-vis the reference category, is less straightforward, but it can be pointed out that with the exception of nasals, there seems to be a general correlation between the degree to which F1 is raised after consonants of different manner classes and the degree of oral constriction they involve. This is not an unexpected effect from the perspective of the undershoot model of vowel duration, as more constricted consonants could be expected to be more compatible with lower F1 values. However, this reasoning may be somewhat problematic in the light of the fact that we do not see here exactly the same effect as we

did in the with the preceding segment manner condition. Moreover, the fact that there is a large (~18 Hz), but statistically untested difference after preceding fricatives and affricates is surprising.

A little more light can be shed on this issue by considering the interaction between preceding segment manner and the vowel variable. This interaction significantly improved model fit (increase in R^2 by 0.0205, $p < 0.01$) and improved prediction (increase in “ R^2 ” by 0.0059). To assess the role of this interaction in somewhat more detail, we calculated separate regression coefficients for the three levels of the variable preceding segment manner, i.e. built separate models for the three high vowels. This is a common approach to exploring interactions in multiple linear regression models (cf. Garson 2006). If there were no interaction, the regression coefficients should be similar. Figure 5.2.3.3.2 clearly shows that this is not the case. The vowels *i* and *u* appear to behave most similarly, although they differ in the scope of F1-raising induced by different manner categories (vis-à-vis stops). The vowel *y* differs from them in that no F1-raising vis-à-vis stops is observed after fricatives and affricates. Preceding nasals have the strongest F1-raising impact on all three vowels, but the size of the effect varies. It is the smallest on the most front vowel (*i*) and the largest on the most back vowel (*u*).

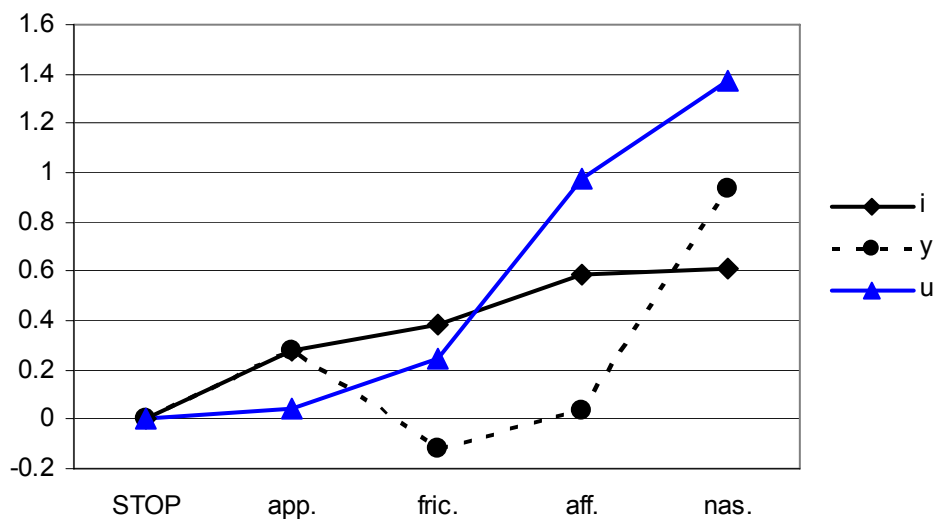


Figure 5.2.3.3.2 Interaction between vowel and preceding segment manner, F1 of high vowels. The three lines link regression coefficients calculated separately for the vowels *i*, *y* and *u*. All values are in standard deviations. All regression coefficients are relative to the reference category *stop* within the given group.

Not all preceding segment manner categories could be compared in figure 5.2.3.3.2, because in the training set there were no examples of laterals followed by *y* (a sequence not occurring in Polish) and no examples of a trill followed by *i*. The behavior of individual vowels differed in these contexts as well. More F1 raising (vis-à-vis stops) was observed for *u* following laterals than for *i*. The vowel *u* was found to have a raised F1 after trills (vis-à-vis stops), by approximately 60 Hz. The vowel *y* was found to actually have a lower F1 after trills than after stops (by approximately 11 Hz).

5.2.3.4 Preceding segment place

Figure 5.2.3.4.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this

case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

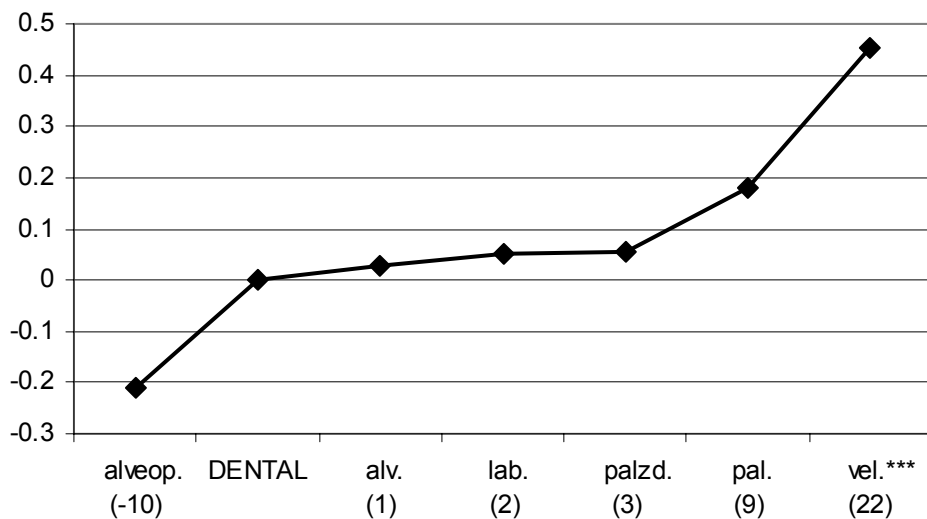


Figure 5.2.3.4.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F1 of high vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 48 Hz as the average value of standard deviation for the F1 of high vowels.

The only statistically significant effect vis-à-vis the reference place, dental, is the F1-raising impact of preceding velars. We also see that preceding alveopalatals condition lowered F1; should this result be verified statistically, it would not be surprising, as this is the typical effect of Polish alveopalatals on this formant (cf. Nowak 2006). What might be somewhat surprising is the fact that palatals and palatalized consonants do not seem to pattern in the same way – they raise F1 (non-significantly vis-à-vis the dentals).

The interaction between preceding segment place and vowel did not significantly improve model fit.

5.2.3.5 Following segment place

Figure 5.2.3.5.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

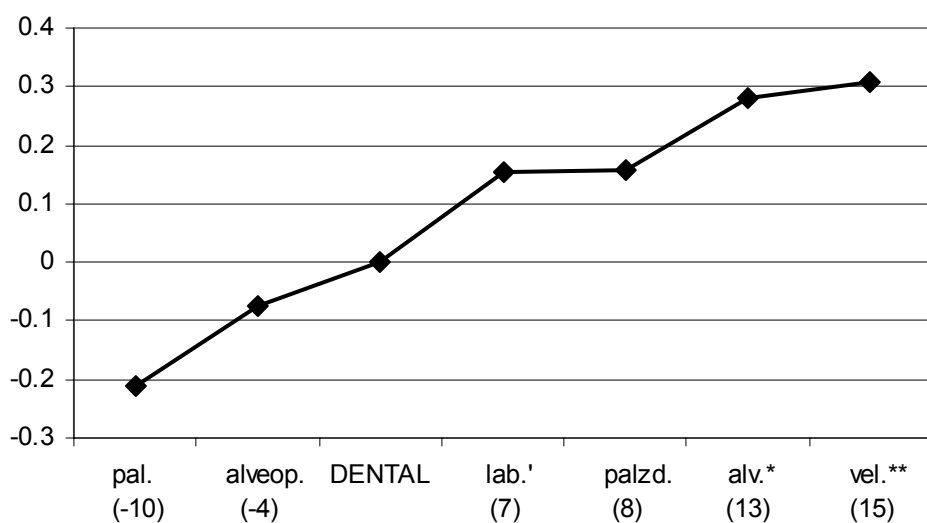


Figure 5.2.3.5.1 MLR-estimated effect of the individual levels of the variable following segment place on F1 of high vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 48 Hz as the average value of standard deviation for the F1 of high vowels.

Like the preceding velars discussed in section 5.2.3.4, following velars have a significant raising effect on the first formant (vis-à-vis dentals). Following alveolars also significantly raise F1 (vis-à-vis dentals), but in this case this must be juxtaposed with the lack of a similar effect in the context of preceding alveolars. No explanation is apparent for this difference. Palatals and alveopalatals lower F1 (non-significantly vis-à-vis dentals, but by interpolation very likely significantly vis-à-vis velars and alveolars). This is not a surprising effect, given the fact that such consonants require strong elevation of the tongue body and close oral constriction. Palatalized consonants, however, do not pattern in the same way. It is likely that in their case, the palatal effect is concentrated at their release.¹

More light can be shed on the role of following segment place by considering its interaction with the vowel variable. It was found to significantly improve model fit (increase in R^2 by 0.0212, $p < 0.01$) and it only minimally decreased prediction accuracy (drop in “ R^2 ” by 0.001). Regression coefficients calculated separately for the three high vowels are shown in figure 5.2.3.5.2. As can be seen, there are differences in the relative effect of following segment place (vis-à-vis dentals) on the three high vowels. Broadly speaking, the vowels *y* and *u* appear to behave in parallel (with the exception of following palatals), although the size of the observed effect differs. Surprisingly, the vowel *i* actually has higher F1 values in the context of alveopalatals than in the context of dentals; this has the effect of downplaying the scope of alveopalatal-induced F1-lowering

¹ Palatalized consonants, almost all of which are palatalized labials, can be realized phonetically as sequences of a “plain” consonant and the palatal glide [j] (e.g. Sawicka 1995).

when the high vowels are considered together as a class (cf. figure 5.2.3.5.1). We can also see that the general F1-raising effect of following velars is the strongest in the case of the vowel *u*, although it is observed for the other vowels as well.

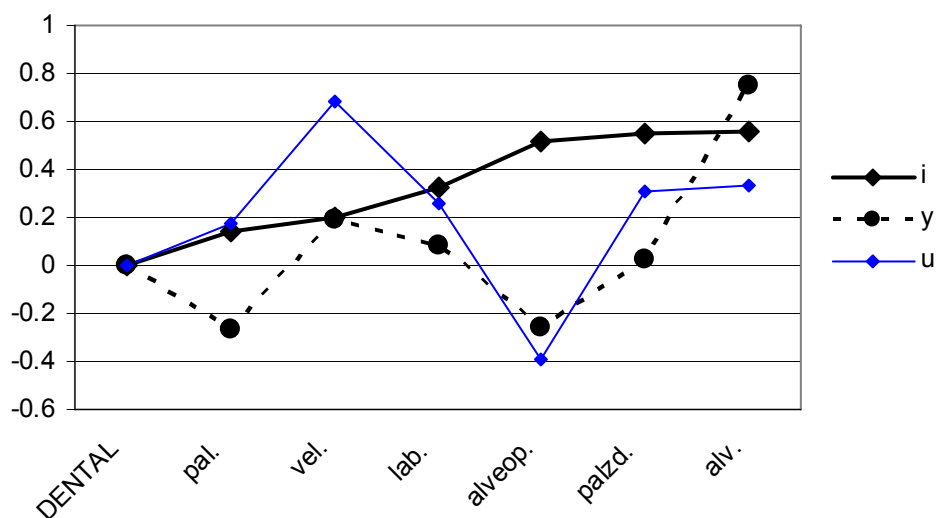


Figure 5.2.3.5.2 Interaction between vowel and following segment place, F1 of high vowels. The three lines link regression coefficients calculated separately for the vowels *i*, *y* and *u*. All regression coefficients are relative to the reference category *dentals* within the given group.

5.2.3.6 Comparative intensity

The MLR validated model estimated the regression coefficient for the continuous variable comparative intensity at 1.81, significant at $p < 0.0001$. Among the high vowels, comparative intensity varied between 0.761-1.302. This indicates that the contribution of this variable varied between adding 1.38 sd ($1.81 \cdot 0.761$) to the mean and adding 2.5 sd. ($1.81 \cdot 1.38$) to it. In other words, the vowel with the highest comparative intensity was

estimated to have the first formant 1.12 sd (~54 Hz) higher than the vowel with the lowest comparative intensity.

Comparative intensity is likely to correlate positively with the degree of opening of the vocal tract. In this context, it is not surprising that vowels with higher comparative intensity have a higher F1 than vowels with lower comparative intensity (cf. figure 3.5.1). It is more difficult to explain the underpinnings of the variation in comparative intensity among high vowels. We have suggested before that this variable may be correlated with non-lexical accent or other kinds of emphasis; it is unclear, however, whether this kind of reasoning is applicable to high vowels. This is because the F1 of high vowels is inherently low. There is thus likely to be a trade-off relationship between increasing the intensity of a high vowel on the one hand, and making it more “canonical” in spectral terms. It is unclear how this tension is resolved, and therefore whether the effect of comparative intensity on high vowels can be attributed to accentuation.

5.2.3.7 Prepausal phonological word

The F1 of high vowels was found to be lowered in prepausal phonological words, by 0.23 sd (~11 Hz). The source of this effect is not very clear. One possibility is related to the fact that pauses are likely to often correspond to the boundaries between syntactic domains. As was discussed in chapter 3, in an SVO language like Polish, new information is likely to come towards the end of the syntactic domain. From the

perspective of information value (cf. Jurafsky et al. 2001), we would expect that vowels carrying newer information should be less reduced. For high vowels, this is likely to indicate lower rather than higher F1, as found in the data here.

As was the case with duration, the interpretation of any potential general effect may be complicated by the phenomena affecting immediately domain-final vowels. In particular, Nord (1986) suggested that vowels in prepausal syllables are more susceptible to coarticulation with the rest position of the vocal tract, i.e. centralization. To informally explore this potential interaction effect, the original variable prepausal phonological word was dropped from the model. Instead, a new variable was added, defined as interaction between syllable position and being in the prepausal phonological word. Next, average effects (average of regression coefficients) were calculated separately for vowels in final and nonfinal prepausal syllables and final and non-final non-prepausal syllables.² High vowels in prepausal phonological words turned out to have a lower F1 than vowels in non-prepausal phonological words in both non-final and final syllables, but the difference was greater in the former (-0.39 sd. vs. -0.12 sd). This strongly suggests the presence of an interaction effect, although it has to be pointed out that it was not statistically verified within the current modeling paradigm.

² “Final” syllables here refers to syllables in enclitics, monosyllables and final syllables of non-monosyllables. The latter two categories strictly speaking comprise syllables that are only likely to be final in the phonological word – they are not if an enclitic follows.

5.2.3.8 Speech rate

The MLR validated model estimated the regression coefficient for the continuous variable speech rate at 0.0234, significant at $p < 0.05$. Among the high vowels, speech rate varied between 2.66-10.8 syllables per second. This indicates that the contribution of this variable varied between adding 0.06 sd ($0.0234 * 2.66$) to the intercept and adding 0.25 sd ($0.0234 * 10.8$) to it. In other words, the vowel spoken at the highest observed speech rate was estimated to have the first formant 0.19 sd (~ 9 Hz) higher than the vowel spoken at the lowest observed speech rate.

The fact that the variable speech rate was selected into the validated model of F1 for high vowels, but duration was not, might suggest that it served as a proxy for duration information. In other words, it is possible that the effect attributed by MLR to speech rate might in fact be an effect of duration. One reason why this may have happened in this case could be that vowel duration was measured less accurately for high vowels than for other vowels. Speech rate, which also indirectly measures vowel duration, could be subject to smaller error.

This is, however, unlikely. As was discussed in chapter 3, duration and speech rate appeared to have a different effect on high vowels when considered pre-statistically. In the case of the highest speech rate, the apparent effect was that of F1 raising (confirmed here), but the shortest high vowels appeared to have lower F1. In section 3.4 we

speculated about the possible reasons for this apparent dissociation between speech rate and duration in the case of high vowels. On the basis of the current results, nothing can be added to this discussion.

5.2.4 Summary

The strongest effects on the F1 of high vowels are those of preceding and following segment manner. Detailed analysis of regression coefficients showed that it is almost entirely the function of the lowering impact of nasals and trills. The place effects cannot be fully accounted for (in particular the lowering effect of velar consonants). The other effects observed also escaped a full explanation, but they were relatively small.

As was mentioned before, high vowels have inherently low values of the first formant. The lowest values of F1 were found in vowels appearing in non-nasal and non-trill contexts; as was discussed in the relevant sections, it is not surprising that less reduction is found in these contexts. The F1-lowering effects of lower speech rate, lower comparative intensity and non-prepausal phonological word can also be seen as indicative of decreased reduction, although little can be said for certain about the underlying causal mechanisms. The effects of place of neighboring consonants are complex and vary by the vowel.

As was noted earlier, the overall fit of the high vowel model was not very good. A partial answer to the question why this is the case may emerge from the examination of the separate models for the three high vowels we constructed to analyze interaction effects. The fit of the models for *i* and *y* was substantially higher than the fit for the overall model (R^2 equal to 0.37, 0.4 and 0.279, respectively). However, the separate model for *u* had the lowest fit (R^2 equal to 0.25). It is likely that the LPC tracking algorithm had the most problems resolving formant frequencies for *u*, because in that vowel the lowest two formants have the closest values. This could result in *u* having the greatest degree of error. It would not necessarily result in a significant result for the vowel variable, because the measurement error in question would be likely to be essentially random, i.e. uncorrelated with the predictors used. The only effect would be lowering the overall fit for high vowels, as observed in the data here.

5.3 Mid vowels

It is difficult to make strong predictions for the effect of vowel reduction on the height dimension of mid vowels, because their inherent values of the first formant are neither extremely low nor extremely high. The Polish mid vowels are usually defined as open mid (lower-mid), this suggests that contextual influences should generally result in the lowering rather raising of F1. However, this can only be verified statistically.

5.3.1 Selecting the model

There were 2408 mid vowels *o* and *e* in the entire dataset (see section 3.1 for a detailed breakdown). Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (1806) and a validation set, comprising 25% of tokens (602). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Of the twenty-two independent variables considered, fifteen were selected into the model. They were then used to predict normalized F1 scores for vowel tokens in the validation set and “R²” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “R²” was recalculated. It was found to be maximized when one more variable was dropped from the model (following segment voicing). Accordingly, the remaining 14 independent variables were used to construct a validated model of the training set.

Table 5.3.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “R²” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 5.3.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients.

Figures 5.3.1.1 and 5.3.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

It should be noted that in contrast with high vowels, the variable vowel was selected into the validated model, suggesting that there are some differences in the way the two mid vowels are impacted by contextual influences.

Independent variable	AIC rank	“R²”
duration	1	<i>n/a</i>
preceding segment place	2	0.1343
following segment manner	3	0.2480
preceding segment manner	4	0.2739
syllable position	5	0.3319
following segment place	6	0.3938
comparative intensity	7	0.4301
vowel	8	0.4523
word content/function	9	0.4699
preceding segment voicing	10	0.4873
style	11	0.4896
clause final phonological word	12	0.4979
postpausal phonological word	13	0.5055
phonological word size	14	0.5114
following segment voicing	15	0.5123
<i>full AIC model</i>	<i>n/a</i>	<i>0.5115</i>
number of following consonants	<i>not selected</i>	<i>n/a</i>
syllable content/function	<i>not selected</i>	<i>n/a</i>
word frequency	<i>not selected</i>	<i>n/a</i>
prepausal phonological word	<i>not selected</i>	<i>n/a</i>
speech rate	<i>not selected</i>	<i>n/a</i>
clause-initial phonological word	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>

Table 5.3.1.1 Variable selection for F1 mid vowels model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R²	“R²”	RSE	“RSE”
Value	0.5559	0.5123	0.6714	0.6758
Difference	-0.044		0.0044	

Table 5.3.1.2 F1 in mid vowels. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of duration in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

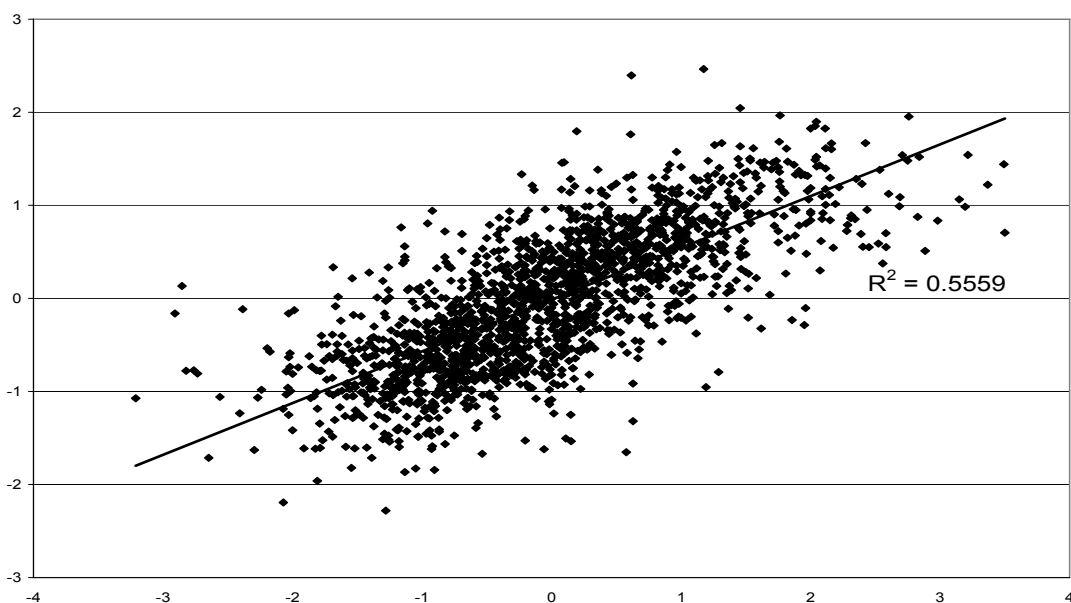


Figure 5.3.1.1 The validated model of F1 of mid vowels. Correlation between F1 values observed in the training set (horizontal axis) and values predicted for the training set by the 14 predictors included in the validated model (vertical axis). All units are standard deviations.

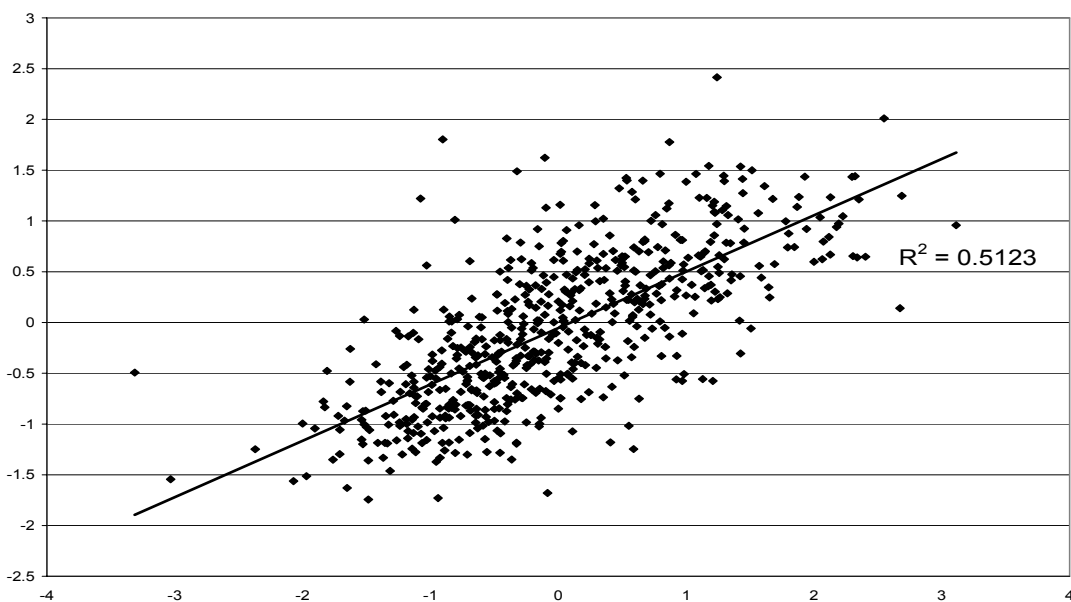


Figure 5.3.1.2 The validated model of F1 of mid vowels. Correlation between F1 values observed in the validation set (horizontal axis) and values predicted for the validation set by the 14 predictors included in the validated model (vertical axis). All units are standard deviations.

Table 5.3.1.2 indicates that the validation of the validated model is fairly successful; the R^2 and the RSE statistics for both the training set and the validation set are fairly similar. The overall predictive power of the model is moderately good. Judged from the R^2 values of 0.5123-0.5559, the model accounts for approximately twice as much variance in the data as did the model for high vowels. The R^2 values indicate multiple correlation coefficient R^2 between 0.716-0.746, which indicates a fairly strong linear relationship between the dependent variable and the predictors.

5.3.2 Categorical regression

The 14 variables that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.5786, indicating correlation coefficient $R=0.76$. The key statistics were thus very similar to the values obtained by MLR. Table 5.3.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	θ	partial	part
preceding segment place	2	0.1114	0.4888	0.2782	0.5024	0.3772
preceding segment manner	4	0.0734	0.3908	0.2268	0.4444	0.3220
duration	1	0.0516	0.2636	0.3841	0.3429	0.2369
following segment manner	3	0.0660	0.2628	0.2809	0.3684	0.2572
vowel	9	0.0154	0.1830	0.0019	0.2256	0.1503
syllable position	5	0.0237	0.1821	0.2518	0.2389	0.1597
following segment place	6	0.0248	0.1720	0.1632	0.2537	0.1703
preceding segment voicing	10	0.0111	0.1571	0.0101	0.2029	0.1345
comparative intensity	7	0.0136	0.1444	0.3207	0.2003	0.1327
word content/function	9	0.0090	0.1392	0.2535	0.1975	0.1308
style	11	0.0073	0.0941	0.1057	0.1401	0.0919
postpausal	13	0.0039	0.0668	0.0272	0.0990	0.0646
clause-final	12	0.0042	0.0626	0.1157	0.0903	0.0589
phonological word size	14	0.0011	0.0419	0.1502	0.0607	0.0395

Table 5.3.2.1 Comparisons of MLR and categorical regression results, validated F1 mid vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, θ is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 5.3.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. As can be seen, the values are very similar, in particular if their relative values are considered (which is indicated by the parallelism between the two lines). This indicates that in this case, multiple linear regression and categorical regression are closely matched and we are justified in treating the results obtained by both of these techniques as largely equivalent.

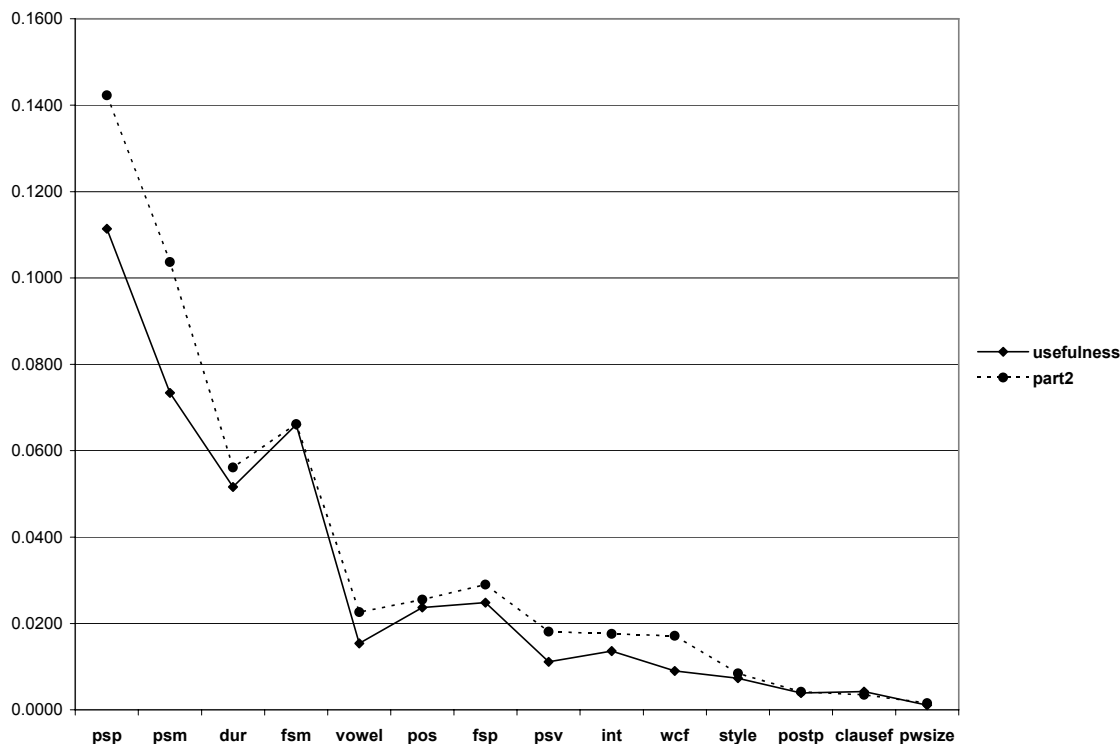


Figure 5.3.2.1 F1 of mid vowels. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 5.3.2.2 shows graphically the relative contribution of independent variables to F1 of mid vowels estimated by the absolute values of the CR b-coefficients. The variable preceding segment place emerges as the strongest, followed closely by preceding segment manner. Duration and following segment manner also appear to have fairly strong effects, the effect of the other variables is smaller. The picture that emerges from figure 5.3.2.2 thus strongly suggests that the variability of F1 of mid vowels may to a large extent be accounted for by the undershoot model of vowel reduction, but this needs to be verified by considering each variable with detail.

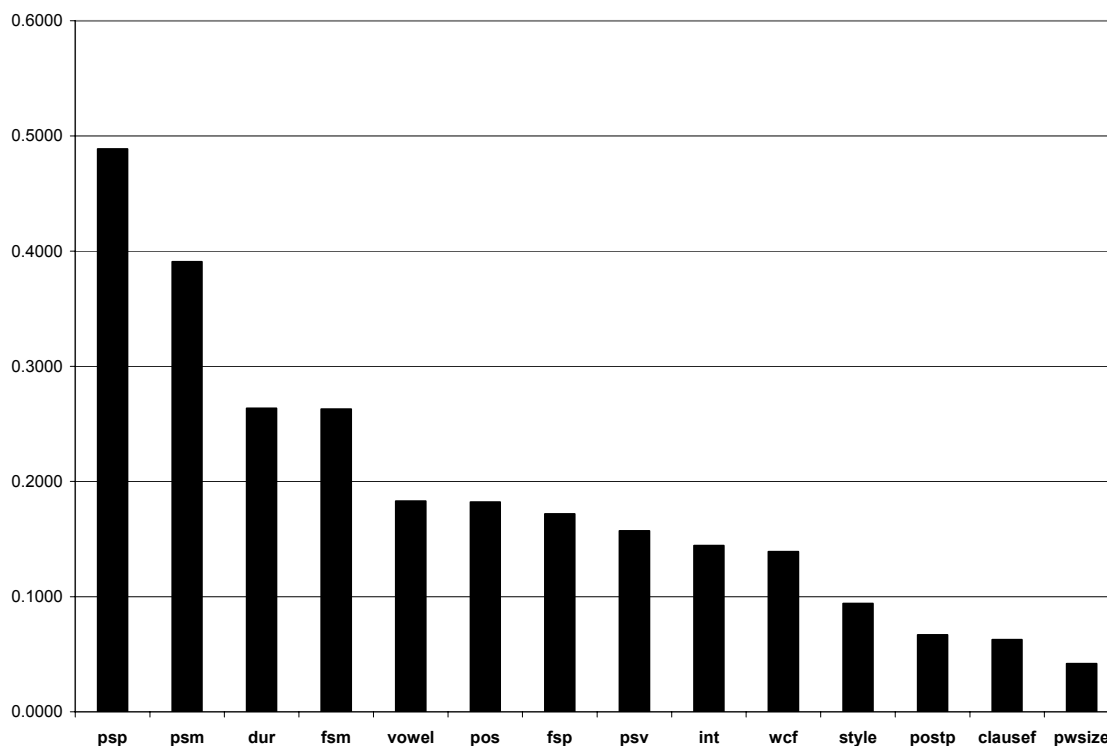


Figure 5.3.2.2 Relative impact of predictors on F1 of mid vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

5.3.3 Individual predictors (MLR)

In this section, we will consider the impact of individual predictors in detail. The basic information will be provided in standard deviations, but we will also provide Hertz estimates, obtained using the average standard deviation of high vowels (67 Hz). As was explained in section 5.1, such values need to be treated with great caution, as they may gloss over significant inter-speaker and inter-vowel variability.

5.3.3.1 Intercept

The intercept of the MLR equation was -2.3. In other words, a mid vowel was predicted to have an F1 lower by 2.3 standard deviations (154 Hz) from its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1, this is a purely mathematical construct with no inherent meaning, since some of the variables cannot have the value of 0 (in this case, vowel duration, comparative intensity and phonological word size).

5.3.3.2 Vowel duration

The MLR validated model estimated the regression coefficient for the continuous variable vowel duration at 0.0093, significant at $p < 0.0001$. Among the mid vowels, duration varied between 12-380 ms.. This indicates that the contribution of this variable varied between adding 0.11 sd ($0.0093 * 12$) to the intercept and adding 3.5 sd. ($0.0093 * 380$) to it. In other words, the longest vowel was estimated to have the first formant 3.4 sd (~ 228 Hz) higher than the shortest vowel. This means that the regression equation estimated F1 to be raised by 0.62 Hz for each additional millisecond of vowel duration ($228 / (380 - 12)$).

We suggested in the introduction to section 5.3 that contextual effects on mid vowels are likely to lower rather than raise their F1 value. Increased duration results in an F1

increase; this indicates that increasing the duration counteracts such contextual effects. Of course, this is exactly what is predicted by the undershoot model of vowel reduction.

Potential interaction effects between vowel duration and the place and manner variables, as well as the vowel variable, are discussed in the relevant sections.

5.3.3.3 Preceding segment place

Figure 5.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

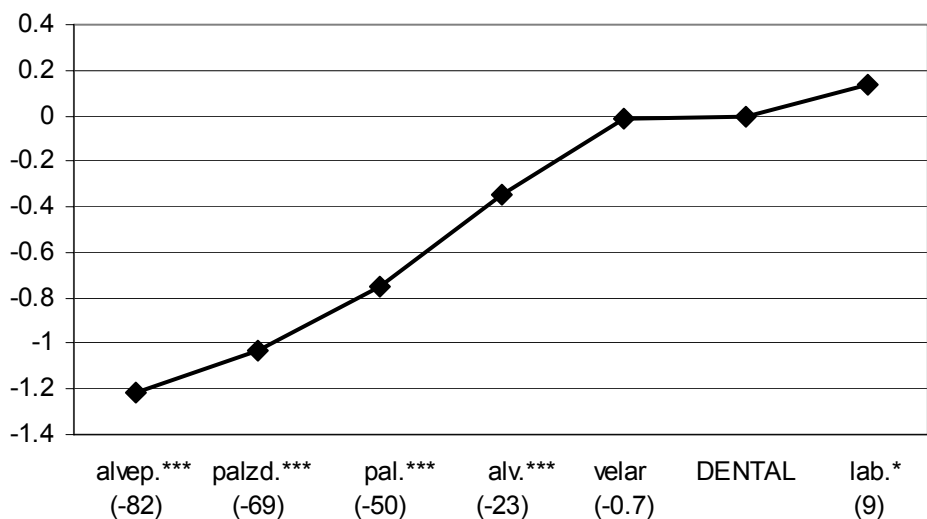


Figure 5.3.3.3.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F1 of mid vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

Categorical regression indicated that the variable preceding segment place was the strongest relative influence on the first formant of mid vowels. The strength of this variable can be additionally appreciated through the fact that all individual place categories, with the exception of velars, showed statistically significant differences vis-à-vis the dentals. In the case of preceding labials, the F1 was slightly higher than in the case of preceding dentals. For all other classes of preceding consonants, F1 was lower. What is particularly clear is that the three classes of consonants with the palatal component present exert a particularly strong lowering influence on F1 (between 50-82 Hz).

The inclusion of interaction term between preceding segment place and vowel duration significantly increased R^2 of the training set (from 0.5559 to 0.566, $p < 0.001$), but did not improve the predictive power of the model (drop in “ R^2 ” of the validation set from 0.5123 to 0.4491).

5.3.3.4 Preceding segment manner

Figure 5.3.3.4.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

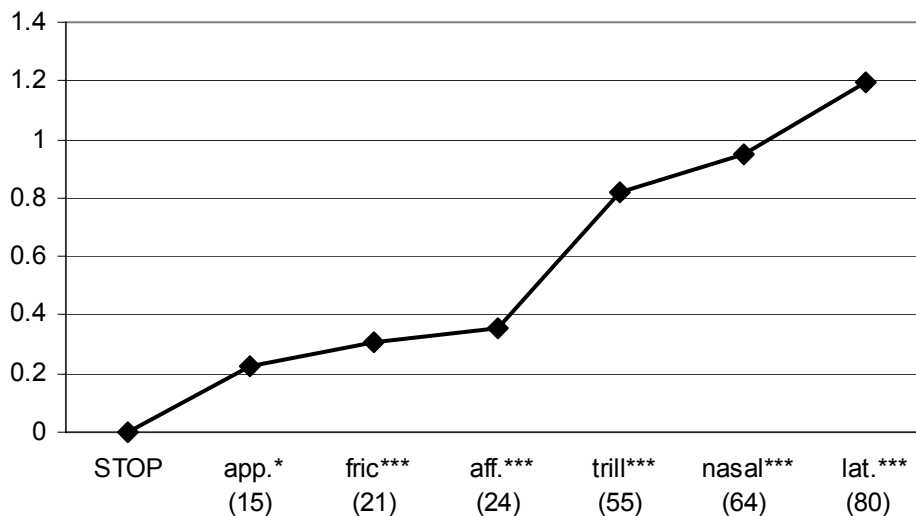


Figure 5.3.3.4.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F1 of mid vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown (‘***’ indicates $p < 0.001$, ‘**’ indicates $p < 0.01$, ‘*’ indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

Consonants of all manner classes strongly affect the F1 of mid vowels. F1 is predicted to be the lowest after stops and the highest after laterals. It seems that the difference between the effect of stops and the effects of approximants, fricatives, affricates and perhaps trills can be attributed to differences in the degree of oral constriction that these consonants require. The closer the constriction, the higher the tongue must be raised, which impedes the subsequent lowering required for open-mid vowels. In this sense, the abovementioned effects can be attributed to target undershoot. In addition, trills are known to require active lowering of medial parts of the tongue, in comparison with other alveolar consonants (Recasens 1991, Recasens and Pallares 1999). The effect of nasals and laterals observed in the data is likely to have a different explanation, namely one in terms of contextual nasalization and lateralization. These are also coarticulatory effects, but not directly related to there being conflicting requirements on the lingual articulator, although this may account for some part of the lateral effect. Nasalization can be expected to raise the first formant of mid vowels (cf. Ohala 1975 and Wright 1986). The predictions concerning lateralization are less clear, but it should be noted that there are known examples of vowel mergers in the height dimension in favor of the more open (i.e. higher F1) variant (e.g. Labov, Ash and Boberg 2006 for American English dialects).³ The effect of nasals mirrors the one observed for the effect of nasals on high vowels, while the effect of laterals is far stronger than it was in the case of high vowels.

³ It should be noted though that in the English dialects in question the neutralizations are triggered by a following lateral.

The inclusion of the interaction term between preceding segment manner and vowel duration resulted in almost significant increase in R^2 of the training set (from 0.5559 to 0.5586, $p < 0.1$), but it did not improve the predictive power of the model (drop in “ R^2 ” of the validation set from 0.5123 to 0.5082).

5.3.3.5 Following segment place

Figure 5.3.3.5.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

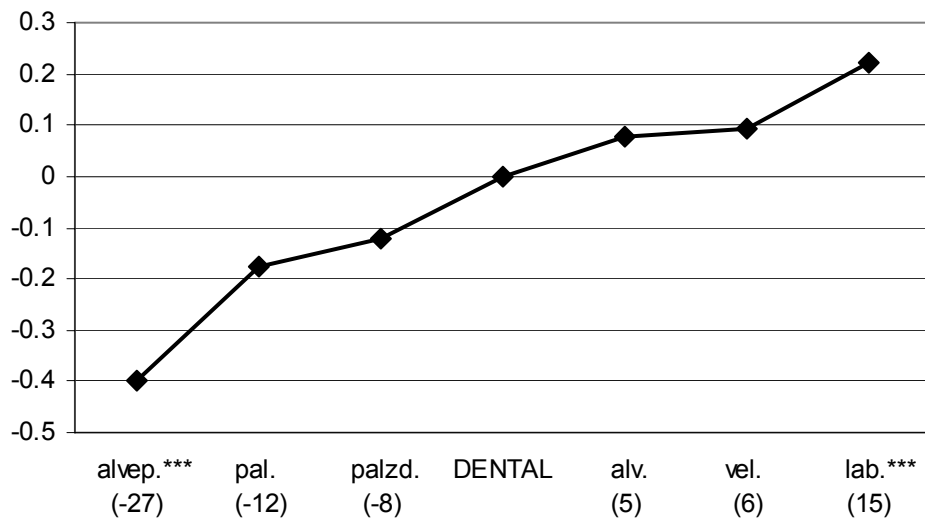


Figure 5.3.3.5.1 MLR-estimated effect of the individual levels of the variable following segment place on F1 of mid vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The abbreviations are as follows: lab. = labial, palzd. = palatalized, pal. = palatal, DENT. = dental, alv. = alveolar, alveop. = alveopalatal, vel. = velar. The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

While the effects are smaller in size, the general effect of following segment place is very similar to that of preceding segment place (section 5.3.3.2). In both cases labials are the only consonants that condition a significantly higher F1 values than dentals. In both cases, the lowest F1 values are predicted in the context of consonants with a palatal component, although in the case of following palatals and palatalized consonants the effect is not significant vis-à-vis dentals. In both cases, the differences between dentals, velars and alveolars are small (although the difference between preceding dentals and preceding alveolars is significant, but the one between the following ones is not).

The interaction between following segment place and vowel duration did not significantly improve model fit.

5.3.3.6 Following segment manner

Figure 5.3.3.6.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

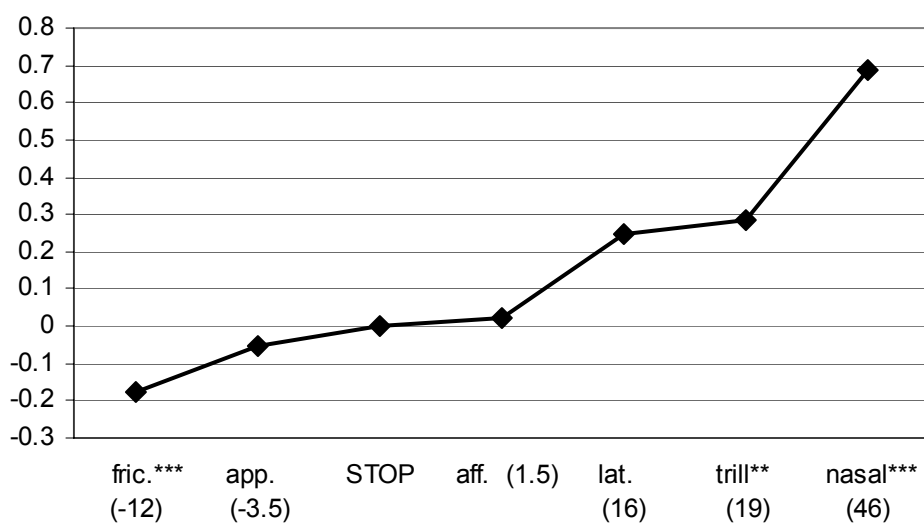


Figure 5.3.3.6.1 MLR-estimated effect of the individual levels of the variable following segment manner on F1 of mid vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

In general, the effect of the variable following segment manner resembles that of the variable preceding segment manner (section 5.3.3.4). Mid vowels are predicted to have highest F1 values in the context of trills, nasals and laterals, although in the case of following laterals the effect is much smaller and not statistically significant vis-à-vis stops. As for the remaining categories, only following fricatives are predicted to condition significantly lower F1 than following stops. This means that an explanation relating the manner effect to the degree of vocal tract opening is not appropriate in the case of following consonants. The precise source of the effect remains mysterious, but it is worth pointing out that such differences between the effect of preceding and following consonant manner parallel to a certain extent the differences observed for high vowels (see sections 5.2.3.2 and 5.2.3.3). For both high and mid vowels, the lowest F1 values occur after stops, i.e. after consonants with the greatest degree of oral closure, but the same potentially undershoot-related effect is not observed before stops.

To complicate matters further, the interaction between following segment manner and vowel duration turned out to be important. It both significantly increased the R^2 of the training set (from 0.5559 to 0.5593, $p < 0.05$) and increased the predictive power of the model (increase in “ R^2 ” of the validation set from 0.5123 to 0.5176). We will therefore consider it in detail.

Figure 5.3.3.6.2 provides a graphical illustration of this interaction effect. It compares the regression coefficients calculated for the validated model without the interaction term

(i.e. the coefficients shown in figure 5.2.3.3.6.1 above, called “general effect” here) with the same coefficients obtained when the interaction term was included in the validated model (called “main effect”). The importance of the interaction effect can be assessed from considering the degree to which the lines illustrating the general effect and the main effect are or are not parallel. In addition, the regression coefficients for the interaction terms are shown, multiplied by 100. This is an arbitrary value, chosen so as to make the interaction coefficients similar enough in scale to the general effect coefficients and the main effect coefficients to enable their presentation on one chart. Mathematically, the interaction coefficients multiplied by 100 indicate the increase or decrease in standardized F1 values as vowel duration increases by 100 ms. Note that if there were no interaction, the general effect and the main effect lines would be exactly the same, and the interaction would equal zero at all levels of the variable.

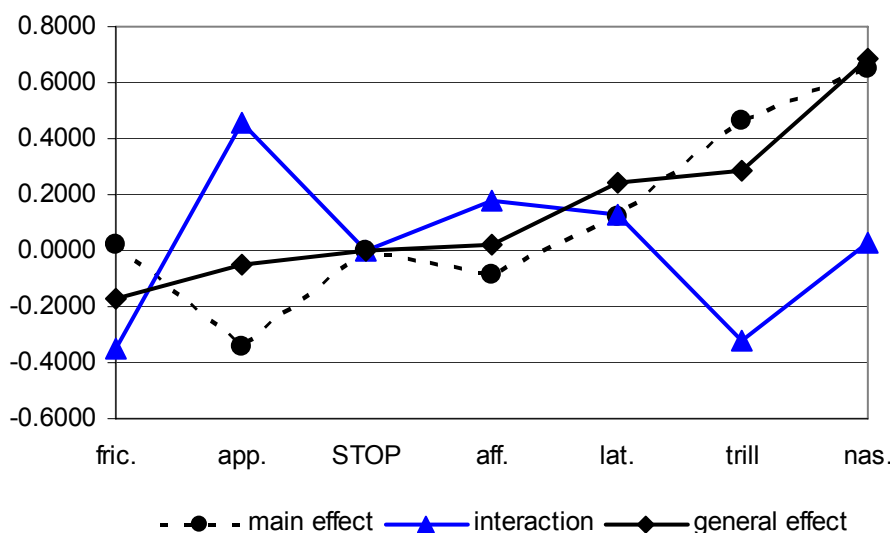


Figure 5.3.3.6.2 Interaction between vowel duration and following segment manner, F1 of mid vowels. “General effect” refers to regression coefficients of the validated model (i.e. the model with no interaction terms). “Main effect” refers to the regression coefficients obtained when the validated model is supplemented with the interaction variable. “Interaction” refers to regression coefficients of the interaction term multiplied by 100. All regression coefficients are calculated with reference to the reference category, shown in bold (*stop* for the general effect and the main effect, interaction of *stop* and *duration* for the interaction). All values are in standard deviations.

As can be seen, the general effect line and the main effect line are broadly parallel. The biggest difference be noticed at the approximant level. The regression coefficient calculated for approximants is much lower when interaction terms are included than when they are not. The explanation of this difference lies in the high positive value of the interaction term. What this indicates is that the presence of a following approximant has a strong lowering effect on F1 of the mid vowel, but this effect is very sensitive to vowel duration – it decreases as duration increases. The opposite can be observed for the effect of the trill. The estimated F1-raising effect of trills is greater when the interaction terms are included, but it decreases with the increase in vowel duration. Other effects are smaller, but can be interpreted in the same way. For example, when the interaction terms

are included, there is virtually no difference between vowels followed by stops and vowels followed by fricatives. However, the longer the vowel, the greater the F1-lowering effect of following fricatives vis-à-vis following stops.

It is notable that there does not appear to be any interaction with duration for the following nasals, which can be seen from the fact that the main effect coefficient and the general effect coefficient are almost identical and the interaction coefficient is close to 0.

5.3.3.7 Syllable position

Figure 5.3.3.7.1 presents the MLR-estimated regression coefficients for the levels of the variable syllable position. The reference category is shown in capitals (in this case, vowel in the penultimate syllable of a non-monosyllabic word). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

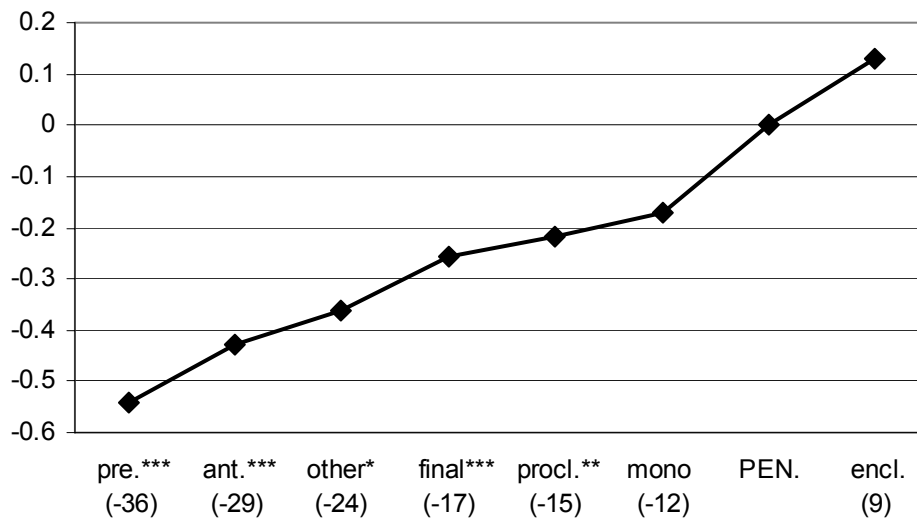


Figure 5.3.3.7.1 MLR-estimated effect of the individual levels of the variable syllable position on F1 of mid vowels, in standard deviations, with *penultimate syllable* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

Vowels in the two types of syllables that can be expected to carry lexical stress, i.e. monosyllables and penultimate syllables of other words, are not significantly different from each other. Vowels in enclitics are estimated to have a slightly higher F1, but the effect is not significant. Vowels in all other types of lexically-unstressed syllables are predicted to have lower F1 values. Assuming again that the target values of F1 of mid vowels in Polish tend to be higher rather than lower, it can be said that unstressed syllables are more prone to reduction than stressed syllables, which is the expected effect (cf. Nord 1986 for Swedish and Harris 1978 for English).

The effect discussed here may become even clearer if instead of considering each syllable position individually, we use a three way distinction between stressed syllables (monosyllables and penultimates), clitics (enclitics and proclitics) and unstressed syllables (all the others). This simplification results in a slightly reduced model fit (R^2 drops by 0.0062 and “ R^2 ” by 0.0059), but graphically the effect of lexical stress is clearer, as shown in figure 5.3.3.7.2.

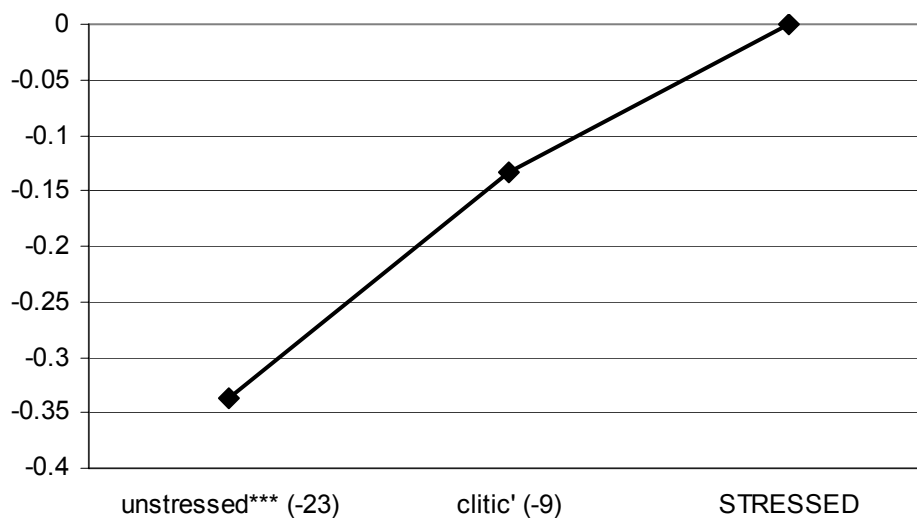


Figure 5.3.3.7.2 MLR-estimated effect of the individual levels of the supplementary variable stress on F1 of mid vowels, in standard deviations, with *stressed* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). Stressed vowels are vowels in penultimate syllables and non-clitic monosyllables, clitic vowels are vowels in enclitic and proclitics and unstressed vowels are all the other. The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

Using this simplified validation model, we considered interactions between stress and the preceding and following segment place. The former significantly improved model fit at the 0.01 level (increase in R^2 by 0.0055), but resulted in decreased prediction accuracy in

the validation set (drop in “R²” by 0.0021). The latter did not significantly improve model fit. We did find, however, an important interaction of stress and vowel. It will be discussed in section 5.3.3.15.

5.3.3.8 Preceding segment voicing

Mid vowels were predicted to have F1 value 0.3 sd (~20 Hz) higher after voiceless consonants than after voiced consonants; the effect was significant at $p < 0.001$. There are two possible explanations of this effect.

In section 4.3.8 it was shown that vowels were significantly longer after voiced consonants. It was also argued that this effect could be a result of the semi-automatic segmentation procedures used: boundaries between stops and vowels were typically found at the onset of voicing, which means that any aspiration accompanying voiceless stops did not count as a part of the vowel. As a result, the duration of vowels following voiceless stops was underestimated. Since duration F1 was found to be positively correlated with vowel duration, it is possible that the effect of preceding consonant voicing corrects this underestimation. However, we estimated in section 5.3.3.2 that the rate of change in F1 frequency was approximately 0.62 Hz per millisecond, and in chapter 4 the durational effect of preceding segment voicing was estimated at 5.7 ms. This suggests that amount of F1 correction explainable in this way is about 3.7 Hz

(0.64×5.7), which is much less than the 20 Hz effect of preceding segment voicing actually found.

In addition, John Ohala (personal communication) points out that segments with high airflow, such as voiceless fricatives and voiceless aspirated stops, tend to have a large glottal opening which may be assimilated by the neighboring vowels. This may result in an acoustic coupling between the supraglottal cavity and the subglottal cavity with effects mimicking those of nasalization (cf. Ohala and Ohala 1993 for a detailed account). The F1-raising observed for mid vowels after voiceless consonants is consistent with this explanation, although it is not fully clear whether the larger than normal glottal opening may indeed persist as far as the steady-state portion of the vowels.

5.3.3.9 Comparative intensity

The MLR validated model estimated the regression coefficient for the continuous variable comparative intensity at 2.32, significant at $p < 0.0001$. Among the mid vowels, comparative intensity varied between 0.752-1.38. This indicates that the contribution of this variable varied between adding 1.74 sd (2.32×0.752) to the intercept and adding 3.2 sd. (2.32×1.38) to it. In other words, the vowel with the highest comparative intensity was estimated to have the first formant 1.46 sd (~ 98 Hz) higher than the vowel with the lowest comparative intensity.

Comparative intensity is likely to correlate positively with the degree of opening of the vocal tract. In this context, it is not surprising that vowels with higher comparative intensity have a higher F1 than vowels with lower comparative intensity (cf. figure 3.5.1). Given that for open-mid vowels the canonical F1 can be expected to be higher rather than lower, this indicates that vowels with higher comparative intensity can be described as less reduced than vowels with lower comparative intensity. To the extent that this variable may correlate with accent and emphasis, this is not an unexpected result.

5.3.3.10 Word content/function status

Mid vowels were found to have a lower F1 value in function words than in content words, by 0.35 sd (~23 Hz); the effect was significant at $p < 0.001$. This can be interpreted as indicating that mid vowels are less reduced in content words, confirming the effect that has been found in other languages (e.g. van Bergem 1993).

5.3.3.11 Speech style

Figure 5.3.3.11.1 presents the MLR-estimated regression coefficients for the levels of the variable style. The reference category is shown in capitals (in this case, normal reading style). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

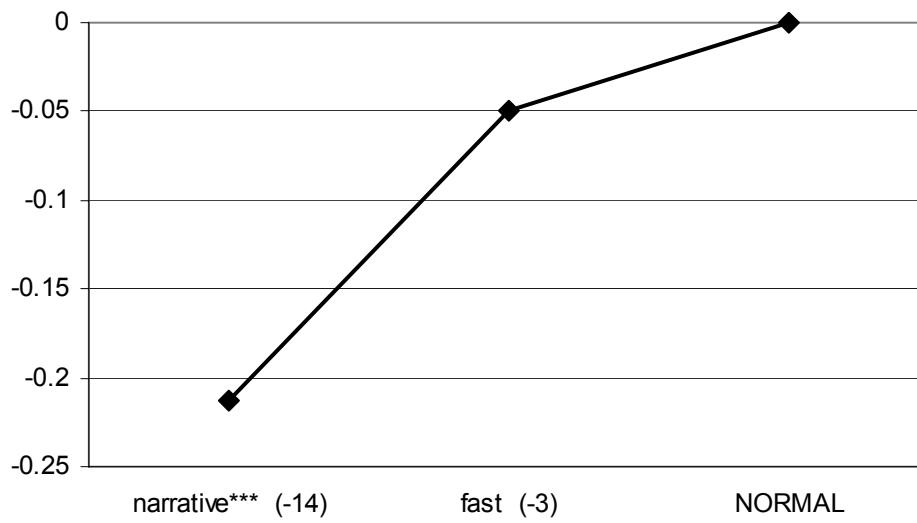


Figure 5.3.3.11.1 MLR-estimated effect of the individual levels of the variable style on F1 of mid vowels, in standard deviations, with *normal reading* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 67 Hz as the average value of standard deviation for the F1 of mid vowels.

As can be seen from figure 5.3.3.11.1, mid vowels had a significantly lower F1 in the narrative speech style. We interpret this as confirming Koopmans-van Beinum's (1980) finding for Dutch, according to which vowels are more reduced in less formal speech styles.

5.3.3.12 Postpausal phonological word

Vowels were found to have a higher F1 value in postpausal phonological words, by 0.18 sd (~ 12 Hz). The source of this effect is unclear. In chapter 3 we speculated that a potential effect of the vowel appearing in a postpausal phonological word would be to

make it more reduced in comparison to other vowels, since in an SVO language like Polish utterance-initial words are more likely to convey known, and hence predictable, information. This effect points in the opposite direction, as vowels in postpausal phonological words appear to be less reduced.

One possibility is that the effect observed is a reflection of the phenomenon known as domain-initial strengthening that has been observed for consonants (cf. Lavoie 2000) or other more phonetic effects that affect vowels that follow a pause. If this were the case, the F1 raising effect should be confined to immediately postpausal vowels. Unfortunately, it was impossible to verify this by investigating it as an interaction between the variables postpausal phonological word and syllable position, in the way paralleling our investigation of the prepausal variable (cf. section 4.3.16 in the analysis of duration). It turned out that only vowels in postpausal penultimates were sufficiently numerous in the mid vowel data for regression coefficients to be calculated when we dropped the original variables syllable position and postpausal phonological word from the model and added interaction terms. However, vowels in penultimate syllables in postpausal phonological words were found to have a significantly higher F1 than vowels in penultimate syllables in non-postpausal phonological words (by 0.16 sd, which is approximately 11 Hz, i.e. a value very similar for the general increase estimated for vowels in postpausal phonological words). In both groups, there was a similar proportion of penultimates that were not initial in phonological words (30.3% for postpausal phonological words and 36.4% in non-postpausal phonological words). This strongly

suggests that the raising effect of postpausal phonological words applies to vowels present in non-immediately postpausal syllables as well.

One further possibility is that postpausal phonological words are pronounced with a greater emphasis when the pauses coincide with paragraph and fragment boundaries. Anecdotically, that appeared to be the case, but this could not be verified in detail.

5.3.3.13 Clause-final phonological word

Vowels appearing in clause-final phonological words were found to have higher F1 values than vowels in non-clause final phonological words, by 0.15 sd (~10 Hz). It is possible that vowels in clause-final phonological words were less reduced because such words are more likely to convey new information in an SVO language like Polish. A brief investigation of the potential interaction effect between clause-finality and syllable position revealed that vowels in clause-final phonological words had higher F1 values in all comparable syllables; the difference was significant for penultimates. This suggests that the F1 raising effect in clause-final phonological words was present in the entire phonological word, and thus was not an effect of immediate adjacency to clause boundary.

5.3.3.14 Phonological word size

The MLR validated model estimated the regression coefficient for the ordinal variable phonological word size at -0.04, significant at $p < 0.05$. Among the mid vowels, phonological word size varied between 1-9 syllables. This indicates that the contribution of this variable varied between subtracting 0.04 sd ($-0.04 * 1$) from the intercept and subtracting 0.36 sd. ($-0.04 * 9$) from it. In other words, vowels in the shortest phonological words were estimated to have the first formant 0.32 sd (~ 21 Hz) higher than the vowel in the longest; that is, the F1 value was lowered by approximately 2.6 Hz with each additional syllable in the phonological word. It is not clear what the source of this effect is.

5.3.3.15 Vowel

If the vowel was *o*, it was predicted to have the standardized F1 value 0.45 sd lower than if it was *e*. Because the data were standardized by vowel and by subject, this does *not* mean that the *o* was predicted to have a lower F1 value than *e*. In fact, figure 3.1.2 showed that the mean value of F1 of *o* was actually approximately 50 Hz higher than that of *e*. What the vowel effect indicates instead is that there are differences between the ways in which reduction processes affect the two individual mid vowels. The way in which this can be investigated is through considering possible interaction effects

involving the vowel predictor. We will now consider the theoretically most interesting ones.

We will assess the importance of interaction terms in the usual way, by estimating the change in model fit (R^2) and in its predictive power (“ R^2 ”). To investigate the differences between vowels in more detail, we will use the approach to interactions that relies on fitting separate models for the two levels of the interacting variable and then comparing the regression coefficients (Garson 2006). Large differences between regression coefficients of the same predictor obtained for the model that considers the vowel *e* only and the model that considers the vowel *o* only will signal important interactions. It has to be remembered, however, that the effects shown can only be compared with regard to the reference levels for the two models. In other words, the fact that a higher regression coefficient may be obtained for one of the vowels does not necessarily indicate that the effect is greater for that vowel in absolute terms.

All first order interactions between vowel duration on the one hand and the place and manner variables on the other were found to result in both significantly better model fit and better predictions, as did the interaction between vowel and duration. Table 5.3.3.15.1 summarizes the impact of adding these interaction effects to the validated model.

Interaction	Change in R ²	Significance of the change in R ²	Change in “R ² ”
vowel and preceding segment place	0.0163	p<0.001	0.0356
vowel and following segment place	0.0045	p<0.01	0.002
vowel and preceding segment manner	0.0132	p<0.001	0.0094
vowel and following segment manner	0.0039	p<0.01	0.0046
vowel and duration	0.0096	P<0.001	0.0056

Table 5.3.3.15.1 Summary of the key statistics for interaction variables involving vowel duration.

Figure 5.3.3.15.1 illustrates the interaction of preceding segment place and vowel by displaying regression coefficients obtained separately for the vowels *e* and *o*. In this case, the variable palatal is not shown, as there were no cases of the vowel *o* in the training set. There is clearly a big difference in how the two mid vowels are affected by preceding consonants at different places of articulation. The front mid vowel appears to be much more susceptible to contextual F1 lowering by alveopalatals and palatalized consonants, as well as to a certain extent alveolars (vis-à-vis dentals). This agrees with previous impressionistic descriptions of Polish, according to which the vowel *e* is most strongly affected by neighboring consonants with a palatal element (cf. Sawicka 1995).

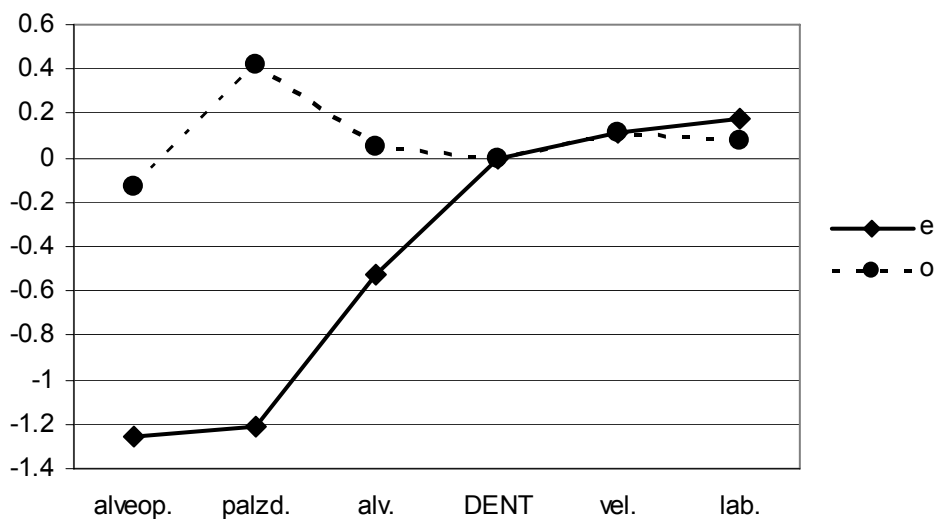


Figure 5.3.3.15.1 Interaction between vowel and preceding segment place. The two lines link regression coefficients calculated separately for the vowels *e* and *o*. All values are standard deviations.

Figure 5.3.3.15.2 illustrates the interaction of following segment place and vowel by displaying regression coefficients obtained separately for the vowels *e* and *o*. In this case also “palatal” consonants appear to affect the vowel *e* more than the vowel *o* (vis-à-vis dentals), but the differences are much smaller. Somewhat surprisingly in the context of the findings for preceding segment place reported above, alveolars appear to have a greater F1 lowering effect (vis-à-vis dentals) on *o* than they have on *e*.

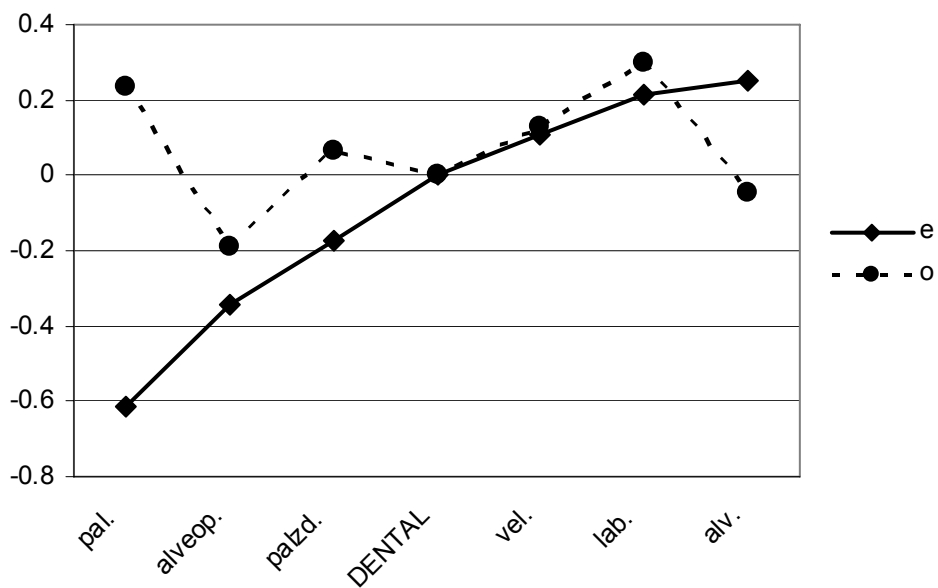


Figure 5.3.3.15.2 Interaction between vowel and following segment place. The two lines link regression coefficients calculated separately for the vowels *e* and *o*. All values are standard deviations.

Figure 5.3.3.15.3 illustrates the interaction of preceding segment manner and vowel by displaying regression coefficients obtained separately for the vowels *e* and *o*. In general, consonants of most manner categories appear to have a stronger F1 raising influence (vis-à-vis stops) on the vowel *e* than on the vowel *o*. The one clear exception is the preceding nasal, which affects the back mid vowel much more strongly than the front one.

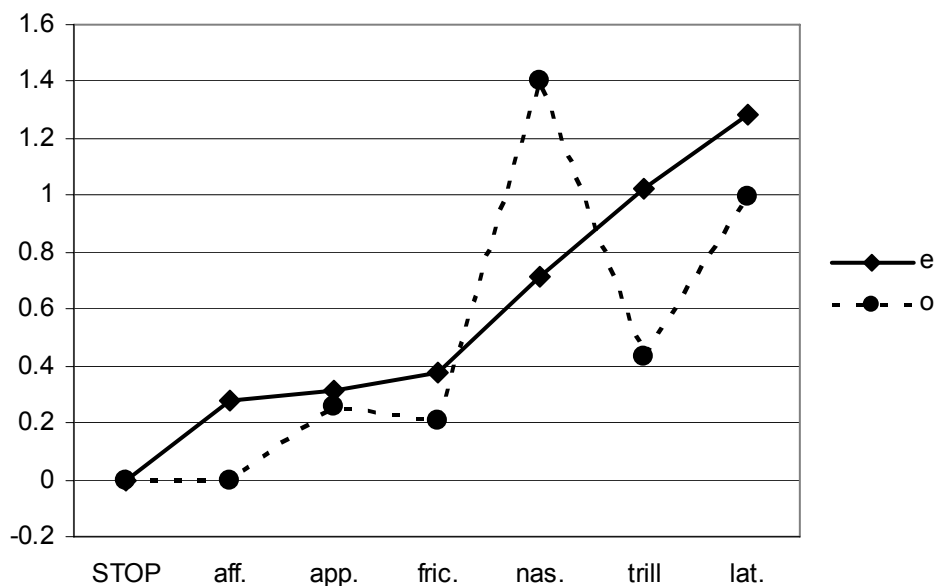


Figure 5.3.3.15.3 Interaction between vowel and preceding segment manner. The two lines link regression coefficients calculated separately for the vowels *e* and *o*. All values are standard deviations.

Figure 5.3.3.15.4 illustrates the interaction of following segment manner and vowel by displaying regression coefficients obtained separately for the vowels *e* and *o*. The differences in this case are not very large. The two biggest ones involve approximants and trills, which have a greater raising impact on the F1 of *e* than *o* (vis-à-vis stops). In a mirror image, fricatives appear to have a greater F1 *lowering* influence on *e* than on *o* (vis-à-vis stops).

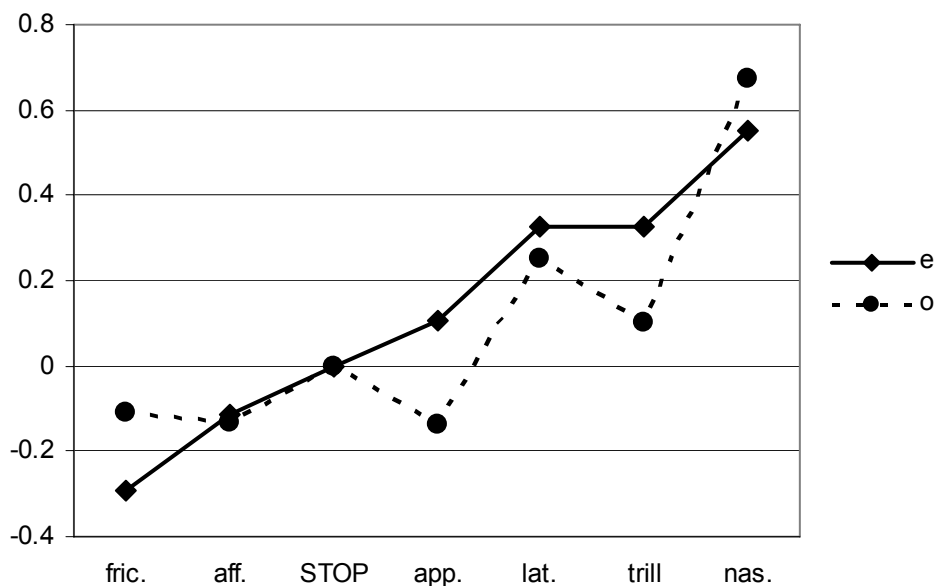


Figure 5.3.3.15.4 Interaction between vowel and following segment manner. The two lines link regression coefficients calculated separately for the vowels *e* and *o*. All values are standard deviations.

Summing up our investigation into the interaction of the place and manner interactions, it would appear that the vowel *e* undergoes more F1 lowering under the influence of the consonants with different places of articulation (vis-à-vis dentals), but more F1 raising as a result of being influenced by consonants with different manners of articulation (vis-à-vis stops, with the strong exception of preceding nasals). It would be tempting to conclude that this indicates that *e* is more prone to consonantal contextual influences than *o*. However, this could only be asserted if dentals and stops are truly appropriate as reference levels for the place and manner variables, respectively, i.e. the levels at which the vowels are closest to their idealized context-free realization. This, however, has not been independently established. It seems however that at least as far as the place variables are concerned, the lines linking regression coefficients for *o* do appear to be

generally “flatter” (and absolutely flat lines with all values equal zero would indicate *no* consonantal effects). However, there is no obvious way of verifying this intuition.

A different picture emerges when we consider the interaction between vowel and vowel duration. It should be recalled that the regression coefficient for duration for F1 of mid vowels was estimated at 0.0093. The regression coefficients calculated separately for *o* and *e* are 0.013 and 0.0067, respectively. The tokens of *o* in the validation set varied between 12-214 ms. in duration, which means that the contribution of the variable duration to the regression of *o* varied between 0.156 sd. (0.013×12 , approximately 10 Hz) and 2.782 sd. (0.013×214 , approximately 186 Hz). Thus the longest token of *o* was estimated to have F1 176 Hz higher than the shortest token of *o*, which indicates a rate of change equal to 0.87 Hz per milliseconds ($186 / (214 - 12)$). The tokens of *e* varied between 12-380 ms., indicating that the contribution of duration ranged between 0.08 sd. (0.0067×12 , approximately 5 Hz) and 2.55 sd. (0.0067×380 , approximately 171 Hz). Thus the difference between the longest and the shortest token of *e* was estimated to be 166 Hz, indicating that the rate of change was 0.45 Hz per millisecond ($171 / (380 - 12)$).⁴ Thus the vowel *o* appear to be more sensitive to changes in duration than the vowel *e*.

Similarly, *o* appears to be more sensitive to lexical stress. This is apparent from figure 5.3.3.15.5, which compares regression coefficients calculated separately for *o* and *e* for the three levels of the additional variable stress (cf. section 5.3.3.7). This interaction also

⁴ Note that the average of these two values of rate of change is 0.66 Hz/ms., very close to the 0.62 Hz/ms. we estimated for duration for mid vowels in general in section 5.3.3.2. This is of course not accidental.

significantly increased model fit (R^2 increase by 0.0047, $p < 0.001$) and increased the predictive power of the model (“ R^2 ” increase by 0.0115). (As in section 5.3.3.7, these values are relative to the supplementary model in which the original seven-level variable syllable position was substituted by the three-level variable stress).

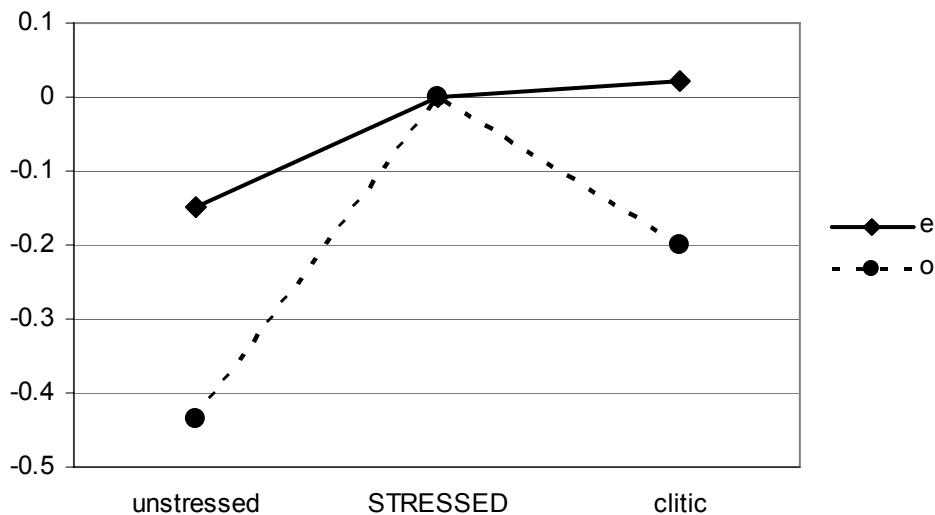


Figure 5.3.3.15.5 Interaction between vowel and stress. The two lines link regression coefficients calculated separately for the vowels *e* and *o*. All values are standard deviations.

The overall importance of interactions of the vowel variable with the other variables may be also assessed by adding all of the above-mentioned interaction terms to the original validated model (using the original variable syllable position rather than stress for consistency in this case). The global increase in R^2 is 0.0467, and in “ R^2 ” (i.e. prediction accuracy) 0.0416. Thus, we succeed in explaining 4.16-4.67% of variance more when we consider these interaction terms. This indicates that while the two mid vowels do behave in roughly similar ways as far as the variability of their first formant is concerned, there are considerable differences between them.

5.3.4 Summary

The first formant of mid vowels varies chiefly as a function of its consonantal context and vowel duration. Although the details concerning the individual variables are not always fully explainable, the overall picture seems to agree with what could be expected under the undershoot model of vowel reduction.

A particularly clear example of the target undershoot effect is the strongly F1-lowering influence of neighboring consonants with a palatal component. Given that the inherent F1 of open-mid vowels could be expected to be higher rather than lower, it appears possible to say that mid vowels are *less reduced* in the context of non-palatal consonants, labials in particular. Increased vowel length also limits reduction (i.e. the scope of F1-raising).

If the same reasoning is applied to the manner variables, we would have to say that vowels are less reduced in the context of nasal consonants. This is clearly counter-intuitive as it is also difficult to imagine that the idealized, context-free pronunciation of any phonologically non-nasal vowel is nasal. This suggests that the influence of neighboring nasals on mid vowels should be thought of as reductive, despite the fact that it results in F1-raising. This conclusion is further supported by the fact that cross-linguistically nasalization as a diachronic process results in the lowering of high and mid vowels (cf. Wright 1986 for a review of the literature). Likewise, the F1-raising influence of neighboring laterals should probably also be considered to be reductive. Lateral

consonants are known to result in vowel mergers in the height dimension, towards the lower phoneme (for example in English, cf. Labov, Ash and Boberg 2006).

Other factors that result in higher F1-values are lexical stress, high comparative intensity, normal reading style, appearance in a content word and in a clause-final phonological word. All of these effects would be independently predicted to correlate with decreased vowel reduction, as discussed in the relevant sections of this chapter and in chapter 3. F1 raising was also found in vowels following voiced consonants, vowels in postpausal phonological words and vowels in shorter phonological words. Where available, the potential explanations were discussed in the relevant sections.

Important interactions were found between the vowel variable and several other variables, suggesting that the two mid vowels do not behave in a fully uniform way. In particular, it appears that *e* is more strongly influenced by the neighboring consonants, while *o* is more sensitive to changes in duration. The picture is further complicated by interactions between other variables. Of course, we cannot preclude the possibility that when other interactions are concerned, including higher-order interactions, the picture would become even more complex.

5.4 The low vowel

Low vowels have inherently high values of the first formant. It is therefore expected that vowel reduction should uniformly result in F1-lowering.

5.4.1 Selecting the model

There were 1350 tokens of the low vowel *a* in the entire dataset. Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (1012) and a validation set, comprising 25% of tokens (338). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Since there was only one low vowel, the variable vowel was not included in the search. Of the twenty-one independent variables considered, sixteen were selected into the model. They were then used to predict normalized F1 scores for vowel tokens in the validation set and “R²” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “R²” was recalculated. It was found to be maximized when one more variable was dropped from the model (prepausal phonological word). Accordingly, the remaining 15 independent variables were used to construct a validated model of the training set.

Table 5.4.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 5.4.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 5.4.1.1 and 5.4.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

Independent variable	AIC rank	“R ² ”
duration	1	<i>n/a</i>
preceding segment manner	2	0.2471
syllable position	3	0.2943
following segment manner	4	0.3142
preceding segment place	5	0.3730
following segment place	6	0.4210
style	7	0.4140
syllable content/function	8	0.4469
preceding segment voicing	9	0.4562
comparative intensity	10	0.4684
postpausal phonological word	11	0.4938
clause-final phonological word	12	0.5025
phonological word size	13	0.5007
following segment voicing	14	0.5013
number of following consonants	15	0.5035
prepausal phonological word	16	0.5091
<i>full AIC model</i>	<i>n/a</i>	0.5070
word frequency	<i>not selected</i>	<i>n/a</i>
speech rate	<i>not selected</i>	<i>n/a</i>
word content/function	<i>not selected</i>	<i>n/a</i>
clause-initial phonological word	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>
vowel	<i>n/a</i>	<i>n/a</i>

Table 5.4.1.1 Variable selection for F1 low vowel model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R ²	“R ² ”	RSE	“RSE”
Value	0.5823	0.5091	0.6465	0.6851
Difference	-0.0732		-0.0386	

Table 5.4.1.2 F1 in the low vowel. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of duration in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

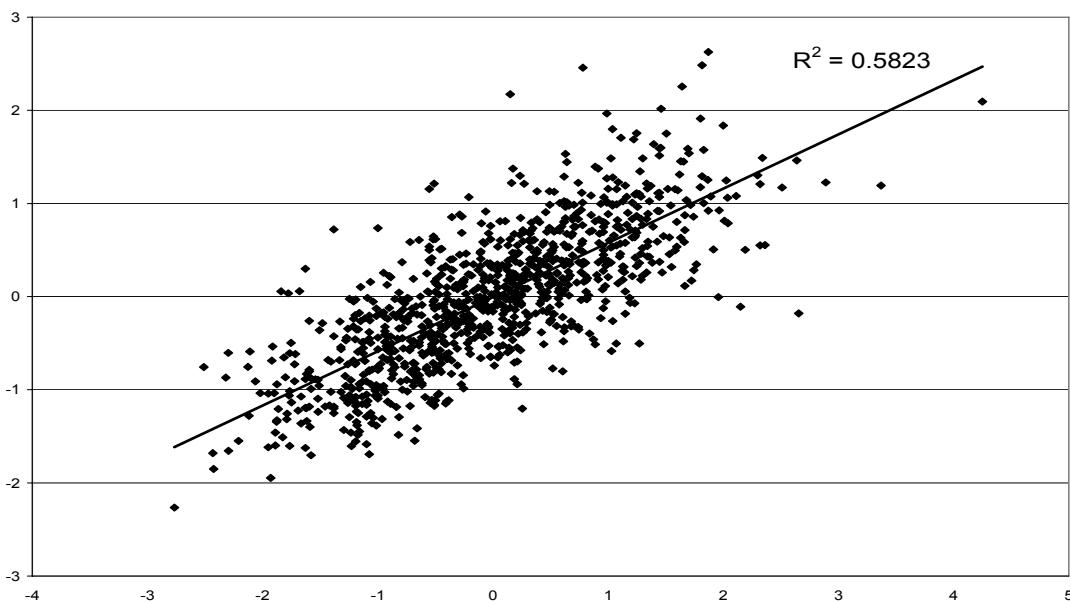


Figure 5.4.1.1 The validated model of F1 of the low vowel. Correlation between F1 values observed in the training set (horizontal axis) and values predicted for the training set by the 15 predictors included in the validated model (vertical axis). All values are standard deviations.

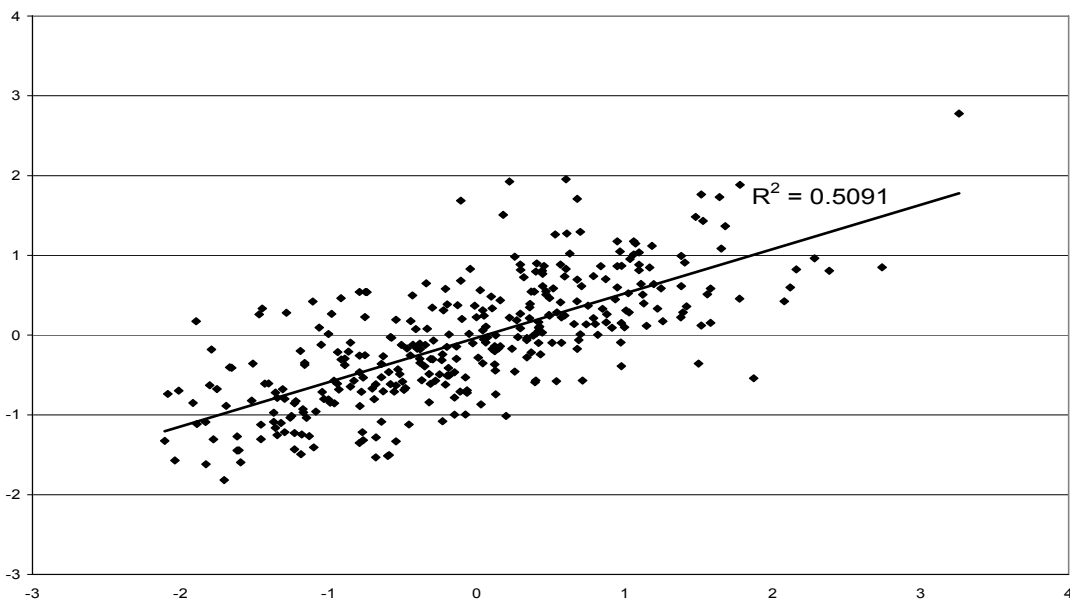


Figure 5.4.1.2 The validated model of F1 of low vowel. Correlation between F1 values observed in the validation set (horizontal axis) and values predicted for the validation set by the 15 predictors included in the validation model (vertical axis). All values are standard deviations.

As far as the accuracy of the model is concerned, results for the low vowel closely resemble those obtained for the mid vowels. The difference between R^2 and “ R^2 ” is somehow greater in this case, indicating that validation was less successful. However, the 0.073 difference observed is not very far from the 0.05 criterion suggested by Schwab (2006) for comparing models built separately on two subsets of data, i.e. with regression coefficients optimized separately for the two subsets. The validation method used here employs only one set of regression coefficients, optimized for the training set. This means that the “ R^2 ” value obtained for the validation set must be lower. In fact, when a separate model is constructed for the validation set, using the same set of predictors, but with regression coefficients optimized, the R^2 of the validation set 0.616, i.e. it is within 0.05 of the R^2 of the training set (0.5823, i.e. a difference of 0.034).

5.4.2 Categorical regression

The 15 variables that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.5822, indicating correlation coefficient $R=0.763$. The key statistics were thus virtually identical to the values obtained by MLR. Table 5.4.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular

multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	0	partial	part
duration	1	0.1076	0.4171	0.4887	0.5038	0.3770
preceding segment manner	2	0.0577	0.3535	0.1992	0.4188	0.2981
following segment manner	4	0.0559	0.3189	0.2775	0.3784	0.2642
preceding segment place	5	0.0386	0.2771	0.1940	0.3729	0.2597
syllable position	3	0.0497	0.2471	0.3308	0.3252	0.2223
following segment place	6	0.0259	0.1961	0.0473	0.2541	0.1698
syllable content/function	8	0.0115	0.1733	0.0836	0.2187	0.1448
preceding segment voicing	9	0.0103	0.1726	0.0488	0.2256	0.1497
style	7	0.0250	0.1711	0.1536	0.2489	0.1661
intensity	10	0.0098	0.1275	0.2554	0.1723	0.1131
postpausal	11	0.0068	0.0931	0.0857	0.1369	0.0893
phonological word size	13	0.0017	0.0780	0.0703	0.1151	0.0749
number of following cons.	15	0.0009	0.0723	0.0443	0.0977	0.0635
following segment voicing	14	0.0000	0.0677	0.0990	0.0935	0.0607
clause-final	12	0.0026	0.0563	0.1387	0.0824	0.0535

Table 5.4.2.1 Comparisons of MLR and categorical regression results, validated F1 low vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, *0* is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 5.4.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. As can be seen, the values are similar, in particular if their relative values are considered. This indicates that the MLR and the CR models are largely equivalent. The only considerable difference concerns the relative importance of preceding segment place and syllable position. The former is estimated as greater by categorical regression, the latter by multiple linear regression.

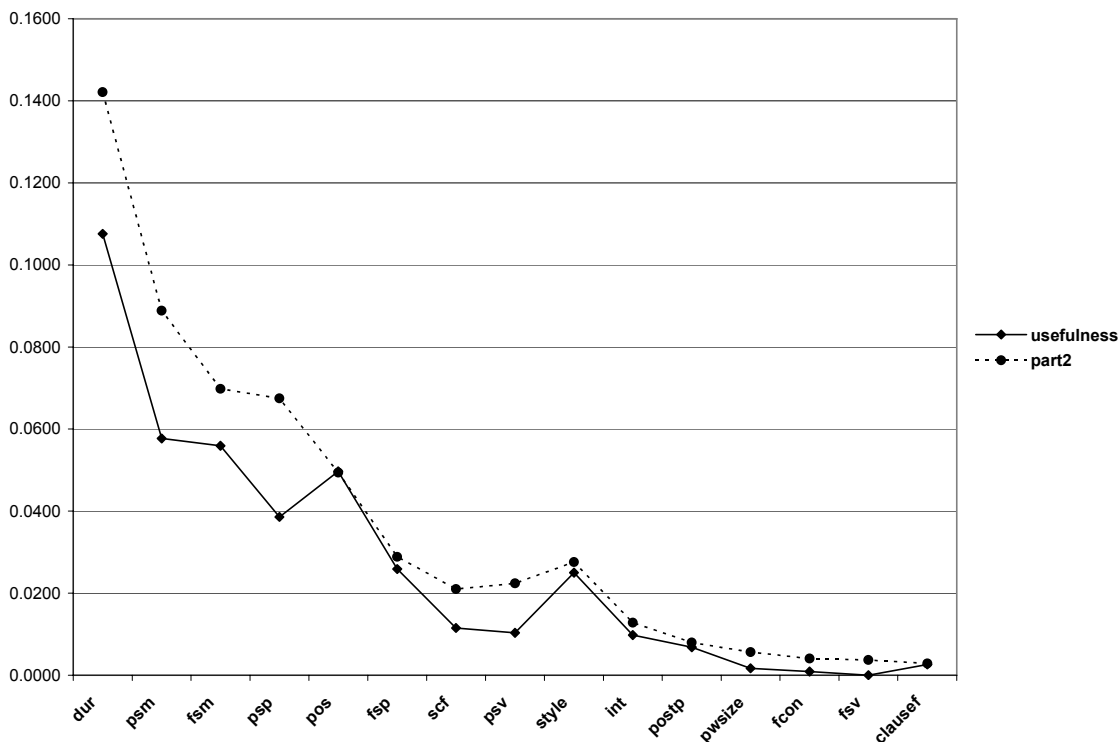


Figure 5.4.2.1 F1 of the low vowel. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 5.4.2.2 shows graphically the relative contribution of independent variables to F1 of high vowels estimated by the absolute values of the CR regression coefficients. In this case, duration appears to have the greatest relative effect, followed closely by the manner variables and preceding segment place. Syllable position and following segment place also appear high in the ranking. It should be noticed that with the exception of the vowel predictor, which was not considered in the case of the low vowels, the top variables for the low vowel and for the mid vowels (figure 5.3.2.2) are the same; the only difference lies in their ranking.

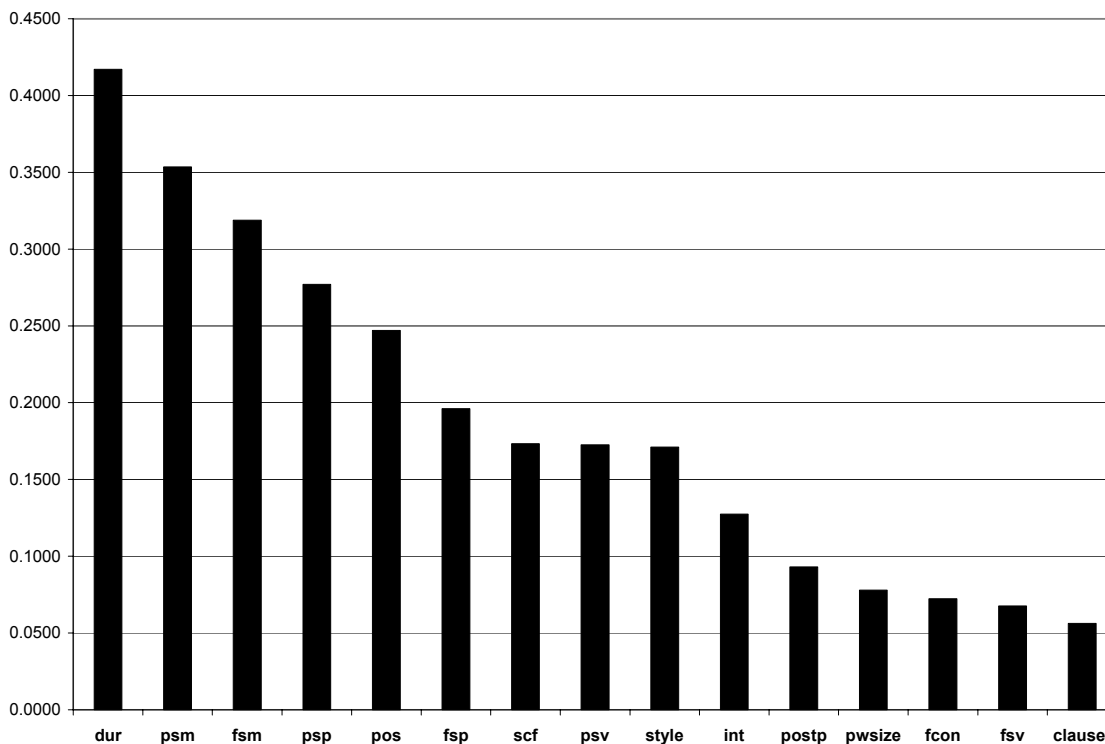


Figure 5.4.2.2 Relative impact of predictors on F1 of mid vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

5.4.3 Individual predictors (MLR)

In this section, we will consider the impact of individual predictors in detail. The basic information will be provided in standard deviations, but we will also provide Hertz estimates, obtained using the average standard deviation of high vowels (76 Hz). As was explained in section 5.1, such values need to be treated with great caution, as they may gloss over significant inter-speaker variability.

5.4.3.1 Intercept

The intercept of the MLR equation was -2.63. In other words, a mid vowel was predicted to have an F1 lower by 2.63 standard deviations (200 Hz) from its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1, this is a purely mathematical construct with no inherent meaning, since some of the variables cannot have the value of 0 (in this case, vowel duration, comparative intensity and phonological word size).

5.4.3.2 Vowel duration

The MLR validated model estimated the regression coefficient for the continuous variable vowel duration at 0.0155, significant at $p < 0.0001$. Among the mid vowels in the training set, duration varied between 10-225 ms.. This indicates that the contribution of this variable varied between adding 0.16 sd ($0.0155 * 10$) to the mean and adding 3.5 sd. ($0.0155 * 225$) to it. In other words, the longest vowel was estimated to have the first formant 3.3 sd (~ 254 Hz) higher than the shortest vowel. This means that the regression equation estimated F1 to be raised by 1.18 Hz for each additional millisecond of vowel duration ($254 / (225 - 10)$).

In section 5.3.3.15 we estimated the duration-dependent rate of change for the vowel *e* at 0.45 Hz per millisecond and for the vowel *o* at 0.87 Hz per millisecond. This leads to an

interesting generalization: the lower the vowel is, the more sensitive it is to changes in duration. (Recall that though both *e* and *o* are mid vowels, the former is phonetically higher than the latter, cf. figure 3.1.2). Of course, the fact that high vowels turned out not to be sensitive to changes in duration at all fits in this picture as well. We will return to this issue at the end of this chapter.

Potential interaction effects between vowel duration and the place and manner variables, are discussed in the relevant sections.

5.4.3.3 Preceding segment manner

Figure 5.4.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

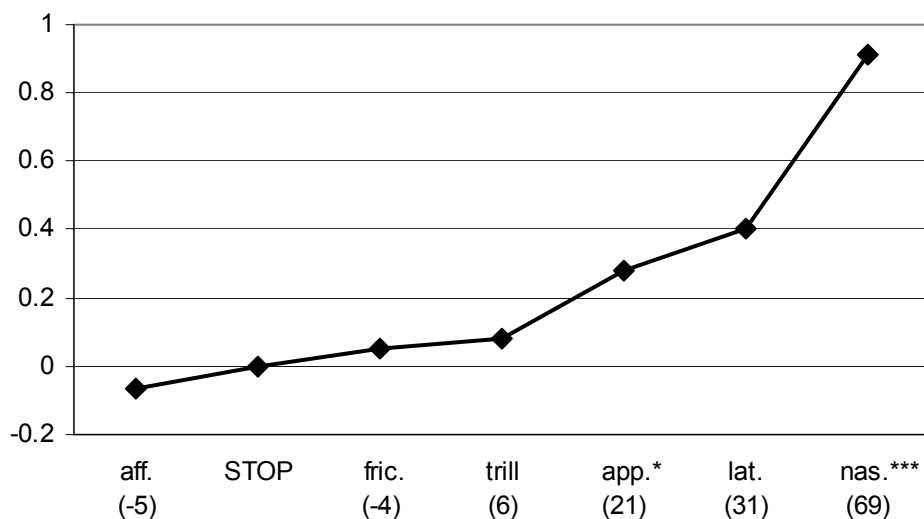


Figure 5.4.3.3.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F1 of the low vowel, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

The strongest effect of preceding segment manner is that of a preceding nasal consonant, which results in a strongly raised F1 vis-à-vis preceding stops. Taken at the face value, this could be taken to indicate that the low vowel is least reduced in the context of a preceding nasal. This is however, counterintuitive, as we would not expect a contextually nasalized vowel to most closely approach the idealized target. It seems that the influence of a preceding nasal should still be taken to be a reductive one, i.e. the F1-raising goes beyond what would be expected in a context-free realization of the vowel. This conflicts with some findings about the impact of nasalization on low vowels (Stevens 1998), but is in agreement with others (Wright 1986). It is likely that in contrast to the uniformly F1-raising effect on non-low vowels, the acoustic effect of nasalization on the first formant

of low vowels differs between languages and speakers; in our case, it appears to be F1-raising.⁵ The only other difference that is significant is between approximants and the reference category, stops. In this case, it is possible that the low vowel following approximants is less reduced than when it follows stops, because approximants involve a smaller oral constriction than do stops. This account could be perhaps extended to the (non-significant) F1-raising in the context of a preceding lateral. There are no significant differences between the impact of the other manner categories and the reference category, stops.

The inclusion of the interaction term between preceding segment manner and vowel duration did not significantly improve model fit.

5.4.3.4 Following segment manner

Figure 5.4.3.4.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

⁵ This does *not* imply that the low nasalized vowels in Polish become *perceptually* more open. Wright (1986) found nasalized mid vowels to be acoustically more open than nonnasalized ones (i.e. with higher F1 values), but perceptually less open. To explain this, he speculated that “it is possible that the position of the first nasal formant with respect to F1 influences perceived vowel height” (Wright 1986:57).

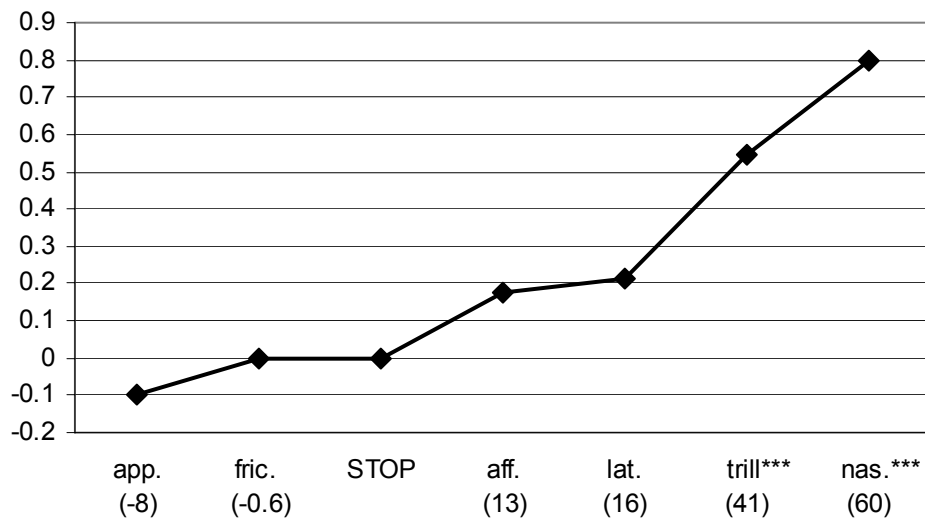


Figure 5.4.3.4.1 MLR-estimated effect of the individual levels of the variable following segment manner on F1 of the low vowel, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The abbreviations are as follows: aff. = affricate, fric. = fricative, lat. = lateral, app. = approximant, nas. = nasal. The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

The picture emerging from the examination of the effect of following segment manner resembles the one seen in the preceding section in the context of preceding segment manner. In particular, the remarks we made there with regard to nasals apply here as well. One major difference is that a following trill results in a strongly raised F1 (vis-à-vis dentals), but a preceding one does not. In addition, approximants are not significantly different from stops in this case. No clear explanation is apparent for these differences.

The inclusion of an interaction term between following segment manner and vowel duration significantly improved model fit (increase in R^2 by 0.012, $p < 0.001$), but failed validation (drop in “ R^2 ” by 0.01).

5.4.3.5 Preceding segment place

Figure 5.4.3.5.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. Each regression coefficient can be understood as the amount (in standard deviations) that is added to or subtracted from the value predicted for the reference category. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

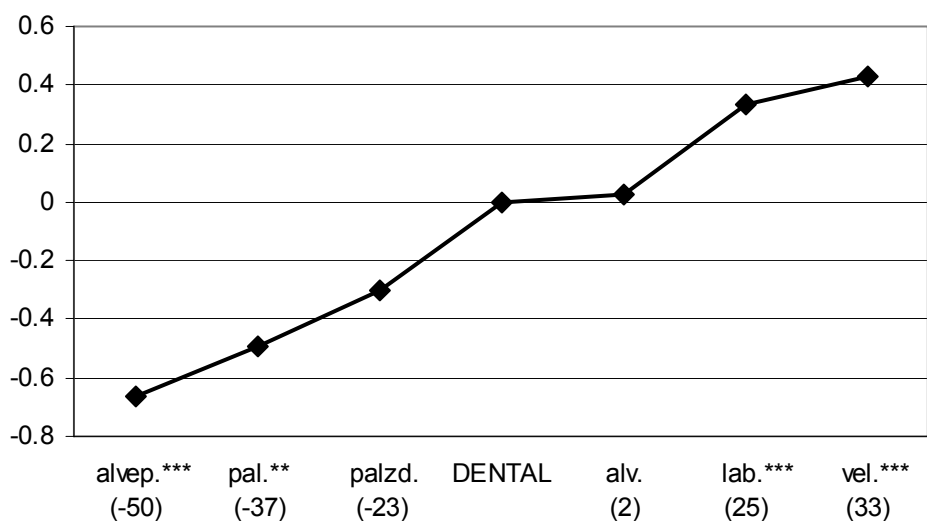


Figure 5.4.3.5.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F1 of the low vowel, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

The low vowel was estimated to have the lowest F1 values when it was preceded by alveopalatals, palatals and palatalized consonants; in the case of the first two categories,

that resulted in a significant difference vis-à-vis dentals. Thus it can be said that the low vowel is reduced the most in the context of consonants with a palatal component, which is not surprising and was also observed earlier for mid vowels. Preceding labials and velars represent the contexts in which the highest F1-values are predicted (significantly lower than in the context of preceding dentals). It can therefore be said that they represent the context least conducive to reduction.

It should be recalled at this point that our multiple linear regression analysis may have underestimated the size of the preceding segment place effect (cf. the comparison of MLR and CR in figure 5.4.2.1).

The inclusion of the interaction term between preceding segment place and vowel duration did not significantly improve model fit.

5.4.3.6 Following segment place

Figure 5.4.3.6.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. Each regression coefficient can be understood as the amount (in standard deviations) that is added to or subtracted from the value predicted for the reference category. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

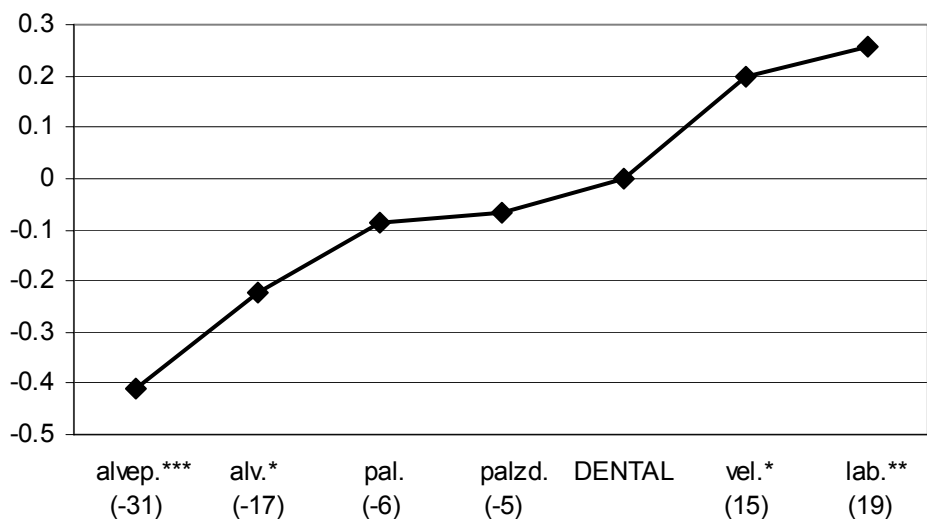


Figure 5.4.3.6.1 MLR-estimated effect of the individual levels of the variable following segment place on F1 of the low vowel, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

As was the case with mid vowels, the effects of following segment place on the low vowel are smaller in size than those of preceding segment place (section 5.4.3.6), but very similar qualitatively. One difference is that following alveolars, but not preceding alveolars, condition significantly lower F1 values than dentals. The source of this difference is unclear. Another difference is that the difference between following palatals and following dentals is not significant, although it points in the same direction (F1 lowering).

The inclusion of an interaction term between following segment place and vowel duration significantly improved model fit (increase in R^2 by 0.008, $p < 0.01$), but failed validation (drop in “ R^2 ” by 0.0172).

5.4.3.7 Syllable position

Figure 5.4.3.7.1 presents the MLR-estimated regression coefficients for the levels of the variable syllable position. Each regression coefficient can be understood as the amount (in standard deviations) that is added to or subtracted from the value predicted for the reference category. The reference category is shown in capitals (in this case, vowel in the penultimate syllable of a non-monosyllabic word). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

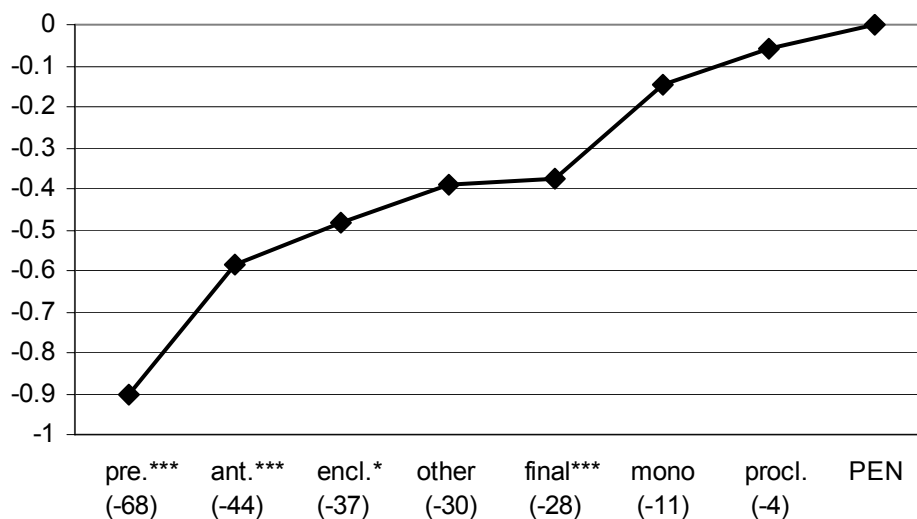


Figure 5.4.3.7.1 MLR-estimated effect of the individual levels of the variable syllable position on F1 of the low vowel, in standard deviations, with *penultimate syllable* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

Vowels in penultimate syllables are predicted to have the highest F1 values. Vowels in non-clitic monosyllables are not significantly different from them. Vowels in these two positions are expected to carry lexical stress, and the current data show that they are less reduced than unstressed vowels (with the exception of the fact that the difference between vowels in penultimate syllables on the one hand and vowels in proclitics and “other” syllables is not significant. These findings replicate the findings concerning the role of stress in vowel reduction in other languages (cf. Nord 1986 and Harris 1978). Except for details, they are also very similar to the findings for mid vowels reported earlier (cf. section 5.3.3.7).

As was the case with mid vowels, the effect in question becomes even clearer if instead of considering each syllable position individually, we use a three way distinction between stressed syllables (monosyllables and penultimates), clitics (enclitics and proclitics) and unstressed syllables (all the others). This simplification results in a slightly reduced model fit (R^2 and “ R^2 ” both drop by 0.01), but graphically the effect of lexical stress is clearer, as shown in figure 5.4.3.7.2. The picture that emerges is very similar to that obtained for mid vowels (cf. figure 5.3.3.7.2), with clitics displaying an intermediate degree of reduction between stressed and unstressed vowels, with non-significant differences from the stressed vowels. It should also be noticed that the stress-related variability of the low vowel matches more closely that of the vowel *o*, which is nearer to it in the F1 dimension. As was shown in figure 5.3.3.15.5, when the two mid vowels are considered separately, *e* is found to actually have slightly lower F1 values in clitics than in stressed syllables.

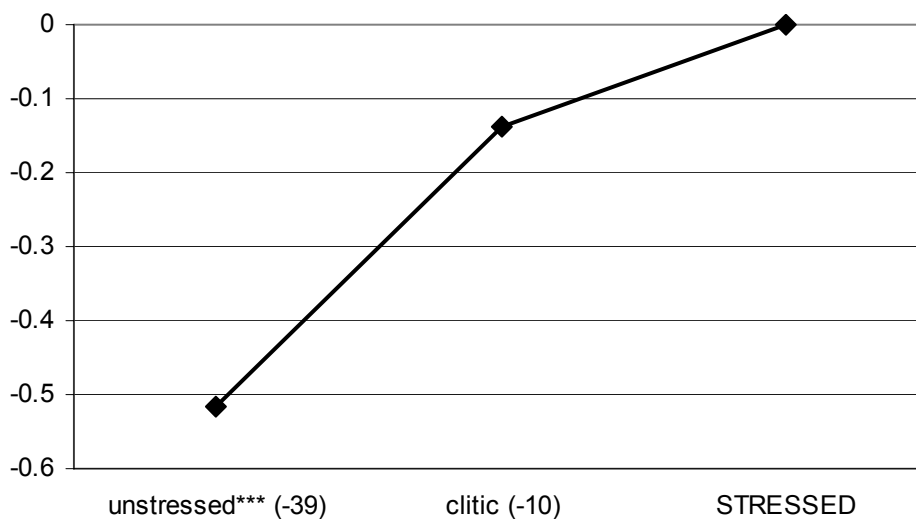


Figure 5.4.3.7.2 MLR-estimated effect of the individual levels of the supplementary variable stress on F1 of the low vowel, in standard deviations, with *stressed* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). Stressed vowels are vowels in penultimate syllables and non-clitic monosyllables, clitic vowels are vowels in enclitic and proclitics and unstressed vowels are all the other. The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of the low vowel.

Using this simplified validation model, we considered interactions between stress and the preceding and following segment place. Both significantly improved model fit at the 0.05 level (increases in R^2 by 0.008 and 0.01, respectively). The former improved prediction (increase in “ R^2 ” of the validation set by 0.005), while the latter decreased it minimally (drop in “ R^2 ” of the validation set 0.001). We will consider them both in detail. To do so, separate models were created for stressed and unstressed vowels (the number of vowels in clitics was very small, with many place categories missing). The source of the place-stress interaction can be assumed to lie in those categories for which regression coefficients differ the most. It has to be remembered that what can be examined in this

way is the difference between the differences between the place categories and the reference level (dentals) within the two stress categories. The fact that for a given level of the place variable a higher or lower standardized F1 value is predicted for one of the stress categories does *not* indicate such a difference in absolute terms.

Figure 5.4.3.7.2 shows the interaction between preceding segment place and stress. Clearly, the source of the interaction effect lies in the differences observed for the palatalized, palatal and velar categories. In the first two cases, stressed vowels undergo less F1-lowering (vis-à-vis dentals). In the case of velars, stressed vowels undergo less F1-raising (vis-à-vis dentals) than unstressed vowels.

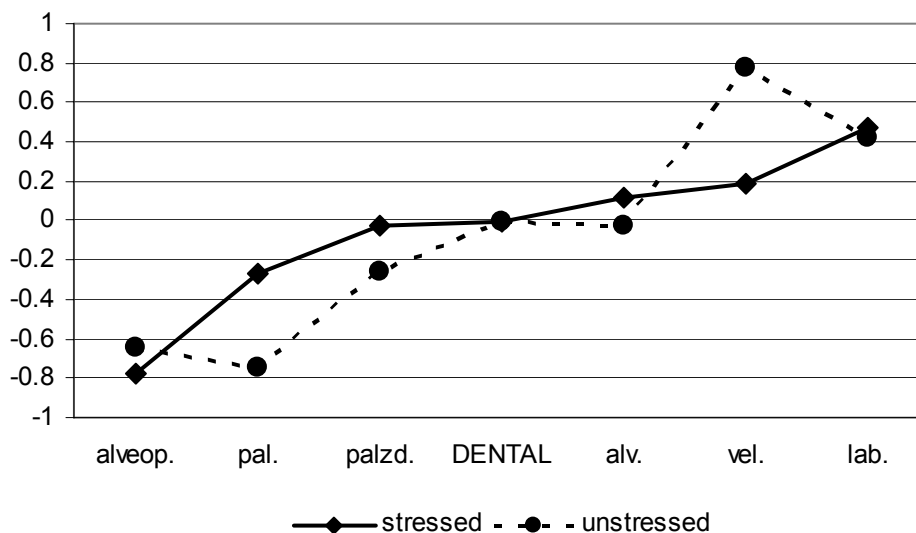


Figure 5.4.3.7.2 Interaction between stress and preceding segment place. The two lines link regression coefficients calculated separately for the stressed and the unstressed tokens of the low vowel *a*. All values are standard deviations.

Figure 5.4.3.7.3 shows the interaction between following segment place and stress. These results are difficult to interpret, but it is clear that while for stressed vowels alveopalatals and alveolars appear to pattern together against the other place categories (strong lowering of F1), for unstressed vowels the F1-lowering group (vis-à-vis other place categories) also includes dentals and palatalized consonants. One possibility that has to be borne in mind in this context is that stress might affect consonants as well as vowels in differential ways. Different realizations of the consonantal gestures in the stress conditions could result in different interactions with vowels.

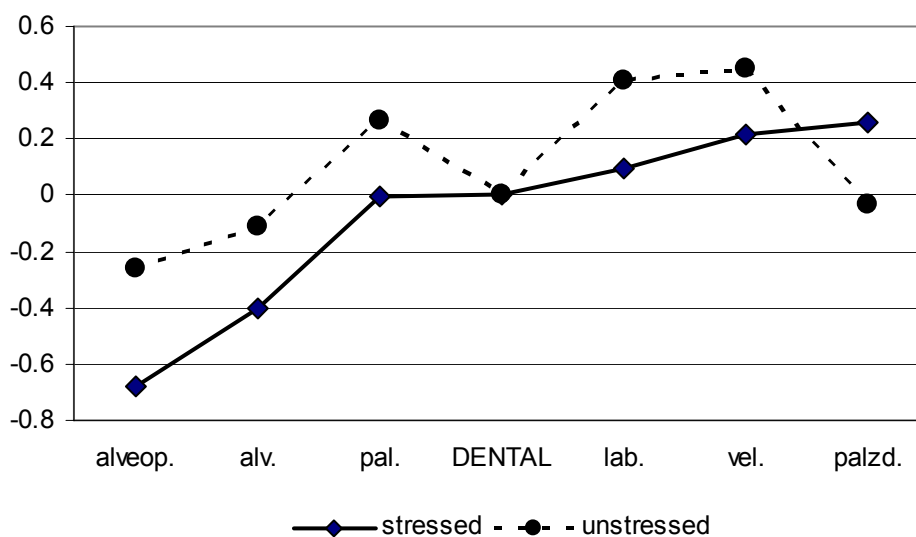


Figure 5.4.3.7.3 Interaction between stress and following segment place. The two lines link regression coefficients calculated separately for the stressed and the unstressed tokens of the low vowel *a*. All values are standard deviations.

5.4.3.8 Syllable content/function status

The low vowel was found to have a lower F1 value in function syllables than in content syllables, by 0.36 sd (~27 Hz); the effect was significant at $p < 0.001$. This can be interpreted as indicating that the low vowel is less reduced in content syllables, confirming the effect that has been found in other languages for the content/function distinction in words (e.g. van Bergem 1993 for Dutch). An interesting and puzzling question is why in this case it was the syllable-based variable that was found to be important, since the word-based one was also included in the predictor space and was actually selected in the case of mid vowels. Note that in both cases, the effects are similar in size (cf. section 5.3.3.10).

5.4.3.9 Preceding segment voicing

The low vowel was found to have F1 value 0.38 sd (~29 Hz) higher after voiceless consonants than after voiced consonants; the effect was significant at $p < 0.001$. A very similar effect was observed for mid vowels (cf. section 5.3.3.8). The same two possible explanations that were sketched in that context may be applicable in this case (underestimation of the duration of vowels after voiceless consonants and acoustic coupling between the vocal tract and the subglottal cavity due to the assimilation by the vowel of the large glottal opening typical of some voiceless segments).

5.4.3.10 Speech style

Figure 5.4.3.11.1 presents the MLR-estimated regression coefficients for the levels of the variable style. Each regression coefficient can be understood as the amount (in standard deviations) that is added to or subtracted from the value predicted for the reference category. The reference category is shown in capitals (in this case, normal reading style). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

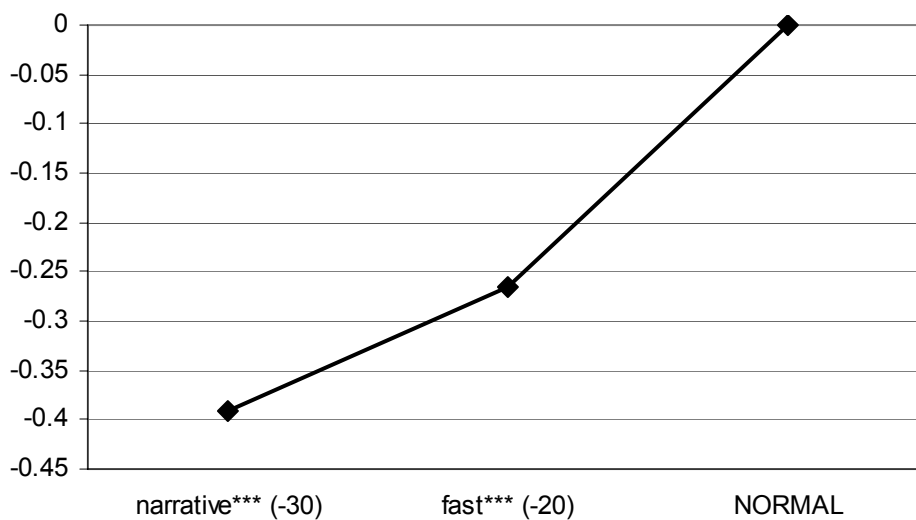


Figure 5.4.3.10.1 MLR-estimated effect of the individual levels of the variable style on F1 of the low vowel, in standard deviations, with *normal reading* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 76 Hz as the average value of standard deviation for the F1 of mid vowels.

The relative effect of style on F1 of low vowels is the same as on F1 of mid vowels (cf. 5.3.3.11). Highest F1 values are seen in the normal reading style, reflecting most limited reduction in what appears to be the most “formal” of the three styles (cf. Koopmans-van Beinum 1980), lower in the fast reading style and lowest still in the narrative style. However, for mid vowels the difference between the fast and the normal style was very small and non-significant; for the low vowel, the fast reading style is significantly different from the normal reading style and appears to pattern with the narrative style instead.

5.4.3.11 Comparative intensity

The MLR validated model estimated the regression coefficient for the continuous variable comparative intensity at 2.15, significant at $p < 0.0001$. For the low vowel in the training set, comparative intensity varied between 0.78-1.35. This indicates that the contribution of this variable varied between adding 1.68 sd ($2.15 * 0.78$) to the intercept and adding 2.9 sd. ($2.15 * 1.35$) to it. In other words, the vowel with the highest comparative intensity was estimated to have the first formant 1.22 sd (~93 Hz) higher than the vowel with the lowest comparative intensity.

Comparative intensity is likely to correlate positively with the degree of opening of the vocal tract. In this context, it is not surprising that vowels with higher comparative intensity have a higher F1 than vowels with lower comparative intensity (cf. figure 3.5.1).

For the low vowel, this indicates that vowels with more intensity are less reduced. To the extent that this variable may correlate with accent and emphasis, this is an expected result. It is notable that the size of the effect is almost identical to that observed for the mid vowels (cf. section 5.3.3.9).

5.4.3.12 Postpausal phonological word

The low vowel was found to have a higher F1 value in postpausal phonological words, by 0.22 sd (~ 17 Hz), significant at $p < 0.01$. This effect is almost identical in size to the one observed for mid vowels and its interpretation raises the same problems (cf. section 5.3.3.12).

5.4.3.13 Phonological word size

The MLR validated model estimated the regression coefficient for the ordinal variable phonological word size at -0.05. This effect was nearly significant ($p < 0.1$). Phonological word size varied between 1-9 syllables. This indicates that the contribution of this variable varied between subtracting 0.05 sd ($-0.05 * 1$) from the intercept and subtracting 0.45 sd ($-0.05 * 9$) from it. In other words, vowels in the shortest phonological words were estimated to have the first formant 0.4 sd (~ 30 Hz) higher than the vowel in the longest; that is, the F1 value was lowered by approximately 4.3 Hz with each additional syllable

in the phonological word. It is not clear what the source of this effect, but it resembles the one found for mid vowels (cf. section 5.3.3.14).

5.4.3.14 Number of following consonants

The MLR validated model estimated the regression coefficient for the ordinal variable number of following consonants at -0.07. Although this variable passed the AIC search and validation, the regression coefficient was not significantly different from 0 and has to be treated with caution. Between 0-3 consonants followed the low vowel in the same word before another vowel. This indicates that the contribution of this variable varied between subtracting nothing ($-0.07*0$) from the intercept and subtracting 0.21 sd. ($-0.07*3$) from it. In other words, vowels followed by no consonants had F1 0.21 sd. (approximately 16 Hz) higher than vowels followed by three consonants. Thus the F1 value was lowered by approximately 5.3 Hz with each additional following consonant. To the extent that the number of following consonants reflects syllable structure, this may indicate less reduction in open syllables.

5.4.3.15 Following segment voicing

The low vowel was found to have a higher F1 value when followed by voiceless consonants, by 0.1 sd (~ 8 Hz). This effect was not significant. The source of this effect is unclear. It may be relevant that the both mid and low vowels were found to have higher

F1 values after voiceless consonants (cf. sections 5.3.3.8 and 5.4.3.9). The account relating this effect to the vowel's assimilation of the large glottal opening of a neighboring voiceless consonant may be applicable in this case as well (cf. 5.3.3.8).

5.4.3.16 Clause-final phonological word

Vowels appearing in clause-final phonological words were found to have higher F1 values than vowels in non-clause final phonological words, by 0.12 sd (~9 Hz). This effect was significant at $p < 0.05$. It is possible that vowels in clause-final phonological words were less reduced because such words are more likely to convey new information in an SVO language like Polish. The same effect with very similar size was earlier found for mid vowels (cf. section 5.3.3.13).

5.4.4 Summary

The variation of the first formant of the low vowel *a* was found to be driven largely by the same factors as the same variation in mid vowels. In fact, virtually all the variables that were selected into the validated model of the mid vowels were also selected into the validated model of the low vowels. The only exception was the variable vowel, which was not considered in the case of the low vowel, and the variable word content/function. However, the validated model of F1 in the low vowel included the variable syllable content/function. In addition, it included two other variables that were not selected in the

case of mid vowels: number of following consonants and following segment voicing.⁶ The overall ranking of variables was very similar, many of the effects found were similar in size in the two models, and both models were equally different from the model for high vowels.

As was the case with mid vowels, the first formant of the low vowel varies mostly as a function of consonantal context and vowel duration. Long vowels in non-palatal contexts (and in particular in labial and velar contexts) have the highest F1 values, i.e. are the least reduced. For the low vowel, higher F1 values are also facilitated by lexical stress, high comparative intensity, normal reading style, content syllables, clause-final phonological words and more open syllables. With the exception of the last of these variables, the same effects were observed for the mid vowels.

When the manner variables are considered, the highest F1 values are observed after and before nasal consonants. The same effect was observed for mid vowels.

In another similarity to mid vowels, higher F1 values were found in vowels following voiced consonants, vowels in postpausal phonological words and vowels in shorter phonological words. For the low vowel, following voiceless consonants also condition higher F1 values. Where available, potential explanations for these effects were sketched in the appropriate sections.

⁶ The latter was actually selected by the AIC procedure for the mid vowels, but failed validation (cf. table 5.3.1.1).

5.5 General discussion

In this section we will attempt to synthesize the findings reported in this chapter. We will first consider the place variables and vowel duration, comparing their role to Lindblom's (1963) original model of vowel reduction. We will then in turn discuss other phonetic and phonological variables and finally nonphonetic variables.

5.5.1 Consonant place and duration

Lindblom's (1963) model of vowel duration was based on a study of a number of vowels spoken by a single male speaker of Swedish within a carefully controlled experimental design. All vowels were spoken in CVC words, in which both consonants were identical voiced stops *b*, *d* and *g*. The design of the current study is rather different. However, some comparisons can be made because the regression coefficients we obtain from multiple linear analysis present the *unique* contribution of variables to the model. This is essentially equivalent to controlling for other variables.

Our findings concerning the role of duration in Polish vowel reduction suggest that its importance is inversely related to vowel height. It seems to play no role for high vowels. For non-high vowels, its importance increases as the vowel's F1 gets higher. This can be illustrated by invoking our estimates of duration-related rate of change (cf. sections 5.3.3.2 and 5.4.3.2). As was reported earlier, it could be estimated at 0.45 Hz/ms. for *e*,

0.87 Hz/ms. for *o* and 1.18 Hz/ms. for *a*. (Recall that although both *e* and *o* are mid vowels, the former seems to have an inherently lower first formant).

The observation concerning the non-role of duration in the reduction of high vowels directly confirms Lindblom's (1963) claims (summarized in section 1.1). Comparing the effect of duration on non high vowels is much more difficult, as Lindblom's mathematical formulas are highly complex. The approach we took here was to plug in formant values into Lindblom's formulas and calculate the duration-related rates of change. We abstracted away from differences in the consonantal context by considering the average rate-of-change values.

Choosing the appropriate formant values to represent the three non-high vowels presented a difficulty, as Lindblom's formulas rely on explicit target formant frequencies. In this work, these are not directly known. Since there are reasons to assume that the F1 target frequencies of non-high vowels are relatively high, we decided to estimate them by using the grand means of vowel formant frequencies for a given vowel to which one average standard deviation was added. This yielded the values 559 Hz for *e*, 612 Hz for *o* and 716 Hz for *a*, averaged across all four speakers.

For each vowel, we plugged the target values estimated in this way into Lindblom's equations. For each consonantal context, we calculated two formant values, using the highest and the lowest duration value observed in our data for the given vowel. Next, we

calculated the rate of change by subtracting the latter value from the former and dividing it by the difference in duration between the longest and the shortest vowel. Finally, we averaged these values across consonantal context. In this way, we obtained rates of change that Lindblom's equations predict for three vowels with formant values comparable to the ones observed in the current study. Figure 5.5.1 compares these values with the rates of change we estimated for the Polish vowels using MLR regression coefficients for duration.

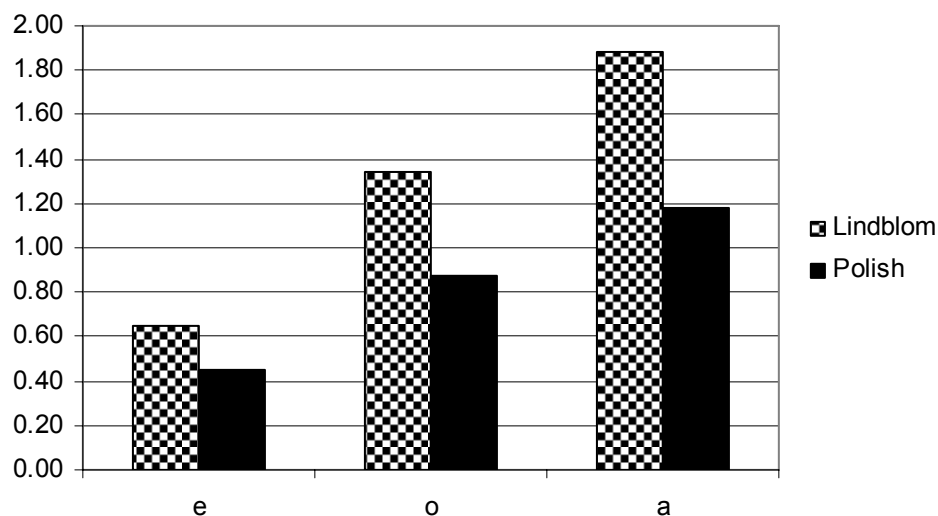


Figure 5.5.1 Duration related rates of change for non-high vowels estimated using the equations in Lindblom (1963) and MLR analysis of the Polish data. The values are in Hertz per millisecond. Refer to the text for details.

As can be seen, Lindblom's equations also predict that the duration-related rate of formant change increases as vowels become more open. In fact, considering the numerous differences between Lindblom's methodology and ours, the similarity of these

results is striking and can be taken as a general confirmation of Lindblom's finding on the role of duration in vowel reduction in the F1 domain.⁷

It is more difficult to compare consonant-place related differences observed in our study with Lindblom's model. Unlike our regression coefficients, Lindblom's equations do not provide an estimate of place effect independently of duration. Nonetheless, for each non-high vowel regardless of duration, Lindblom's equations predict that given the three different contexts, the highest F1 values will be found in vowels surrounded by bilabial consonants. Surrounding velars are predicted to condition F1 values slightly higher than dentals but lower than labials at short durations. At longer durations, they are predicted to condition F1 values slightly lower than dentals.

Table 5.5.1 presents the sums of the MLR-estimated regression coefficients for preceding and following labial consonants and for preceding and following velar consonants for mid and low vowels. These numbers thus represent the estimated differences between vowels surrounded by labial consonants and vowels surrounded by dental consonants (the reference category) on the one hand, and between vowels surrounded by velar consonants and vowels surrounded by dental consonants on the other. This is the closest

⁷ The fact that the rates of change estimated from Lindblom's equations are higher than the ones observed in this study may be due to many factors. One obvious possibility is that this is due to language difference between Polish and Swedish. It is also likely that there are individual differences between speakers. In fact, when Lindblom's equations are used to predict rates of change for individual speakers (using their estimated target formant values) it is clear that this is the case. However, in each individual case the overall ranking is the same.

approximation to Lindblom's experimental stimuli that is possible with our data. (For clarity, we use in this context Hertz values estimated from average standard deviations).

Context	Mid	Low
Labial (vs. dental)	24	44
Velar-(vs. dental)	5	48

Table 5.5.1 MLR-estimated differences between vowels surrounded by labial consonants and by dental consonants and between vowels surrounded by velar consonants and by dental consonants. The values are sums of regression coefficients obtained for preceding and following consonants with the given place of articulation, converted into Hertz using average values of standard deviations for given class of vowels.

As can be seen, in our data we have found the F1-raising effect of surrounding labials (vis-à-vis dentals) that is predicted by Lindblom's equations. We did not, however, replicate his findings with regard to velar consonants. It is possible that this is a result of language-specific differences in the articulation of these consonants between Polish and Swedish. Of course, we cannot compare the effect of other place categories, since Lindblom's study only considered dentals, labials and velars.

5.5.2 Other phonetic and phonological variables

The **manner** of neighboring consonants was found to be a strong predictor of vowel reduction. It was found to affect vowels of all heights. This effect seemed to be mostly due to the strongly F1-raising effect of sonorants of certain types (nasals, trills and laterals). The relationship of these effects to previous findings on other languages was discussed above in the relevant sections.

Comparative **intensity** was also found to be a good predictor of vowel reduction for vowels of all heights. In each case, a positive relationship was found between intensity and F1. Since both intensity and the first formant can be assumed to be directly correlated with the degree of vocal tract opening, this is not surprising. The most interesting question, however, remains: what was the cause of the variations in intensity? To the extent that increased intensity may have corresponded with accentuation and emphasis, it could be said to illustrate the limiting effect of such factors on the reduction of mid and low vowels, but the situation is less clear for high vowels (cf. section 5.2.3.6).

For the low and mid vowels, **syllable position** was found to be a good predictor of vowel reduction. In Polish it is related to lexical stress, and these vowels appeared to be less reduced in syllables that could be expected to carry lexical stress. No such effect was found for high vowels. These in turn were unique in showing an effect of **speech rate** which escapes full explanation (cf. section 5.2.3.8).

Other phonetic and phonological variables that were selected into at least one model included **number of following consonants**, **voicing** and **phonological word size**. They were found to affect either both the low vowel and the mid vowels, or just the former. In neither case did they turn out to be good predictors of vowel reduction in high vowels.

The **vowel** variable was selected into the model for mid vowels, suggesting that they do not fully behave as a class. There was also some evidence of vowel-differential behavior among the high vowels discovered through the analysis of interactions.

5.5.3 Other relevant variables

Vowels were found to be the least reduced in the most formal of the three speech **styles** used by the subjects, normal reading. This replicates Koopmans-van Beinum's (1980) findings for Dutch. The **content/function** distinction, which had also been found to be correlated with vowel reduction in other languages (e.g. van Bergem 1993) also played a role. However, in the case of mid vowels, it was the word-based predictor that was selected, and in the case of the low vowel, the syllable based one. No explanation is apparent for this difference. Mid and low vowels were also found to be less reduced in **clause-final** phonological words and **postpausal** phonological words. The former effect, but not the latter, could reflect syntactic structure.

None of the predictors mentioned in the preceding paragraph was useful in the case of high vowels. However, high vowels were found to have lower F1 values in **prepausal** phonological words. The source of this effect is not fully clear; it may, however, reflect more limited reduction in a specific syntactic context (cf. the discussion in section 5.2.3.7). This variable was not found useful in the case of mid and low vowels.

John Ohala (personal communication) points out that the pause-related and domain-related variables may be correlated with formant variability for another reason. It is well known that there is some correlation between pitch and the vertical position of the larynx (see Ohala 1972 for a review). The vertical position of the larynx is in turn correlated with the size of the vocal tract: the higher the larynx, the smaller the vocal tract. Decreasing the size of the vocal tract affects its resonances, making them higher. We could therefore expect some correlation between pitch and formant values. Pitch declination is a well known phenomenon (see Noteboom 1997 for a review). It leads us to expect higher pitch values domain-initially (clause-initially and postpausally) and lower pitch values domain-finally (clause-finally and prepausally). Consequently, we can also expect higher formant values in the former case and lower formant values in the latter. We have observed higher F1 values in postpausal mid and low vowels and lower F1 values in prepausal high vowels; these findings are compatible with this explanation. However, we have also observed higher F1 values in clause-final mid and low vowels, which is contrary to our predictions.

5.5.4 Variables not selected into any model

The variable **subject** was not selected into any model. This indicates both that the standardization procedure we used succeeded in eliminating vowel-formant variation due to differences between the individual subjects' vocal tracts, and that the subjects did not employ different vowel reduction strategies.

Word frequency was also not selected into any model. This may be surprising given the fact that frequency has been found to play a role in reductive processes in other languages (e.g. Jurafsky et al. 2001, Byebee 2000, Gahl 2006). However, Wright and Kilansky (2006) have recently argued that the role of frequency cannot be adequately considered without taking into account the more complex notion of “difficulty”, which is a function of frequency and neighborhood density. We did not investigate the latter factor in this work.

The only other variable that was not selected into any model was **clause-initial** phonological word.

5.5.5 Conclusions

To a large extent, our findings agree with Lindblom’s (1963) model of vowel undershoot. As predicted by Lindblom, no evidence for duration-dependent undershoot was found in high vowels. For other vowels, the role of duration was found to be closely related to vowel height, which also agrees with Lindblom’s predictions. Given that the F1 loci of consonants of all places of articulation are likely to be low, the general pattern appears to be that the role of duration increases as the distance between the locus frequency of the consonant and the target frequency of the vowel increases. In particular, the higher the F1 target frequency of the vowel, the higher is the duration-dependent rate of change of the first formant. To the extent that the first formant frequency conveys information about

articulator movement, this can be interpreted as indicating that the more open the vowel is, the faster is the required speed of articulator movement.

In addition to the place variables and duration, strong effects of manner were also observed. We also found smaller but non-negligible effects of such factors as stress, speech style and the content/function distinction (as well as some others). It is interesting to note that, like duration, these non-phonetic factors were also restricted to non-high vowels. In fact, in the case of high vowels, the fewest predictors were selected into the model and the accuracy of the model was the lowest.

6 Second formant

In this chapter, we will present statistical models accounting for the variability of the second formant in the vowels in our dataset. In constructing the models, we will follow the procedures outlined in chapter 2 and exemplified in the preceding chapters by our discussion of the models of vowel duration and F1. As was the case with F1, we will also employ standardized scores in our analysis of F2.

In the light of Lindblom's (1963) undershoot model of vowel reduction, analyzing the context-related variability of the second formant poses more problems than the analysis of the first formant. In the latter case, consonants at all the relevant places of articulation can be expected to lower the formant frequency of all non-high vowels, while little undershoot is predicted for high vowels. When the horizontal component of the two-dimensional vowel space is considered, the situation is different. This is because the impact of a given consonantal constriction depends on its location with regard to the velocity nodes and antinodes of the second natural resonance of the vocal tract (raising vs. lowering, respectively). As shown in figure 6.1, two nodes and two antinodes have to be considered in this case (cf. figure 5.1 for the simpler case of the first formant). Thus while some consonants can be expected to have a strong F2-raising effect because they are articulated with constrictions located close to the velocity nodes (e.g. those with a palatal component), others could be predicted to have an F2-lowering effect because of being located close to the velocity antinodes (e.g. labials).

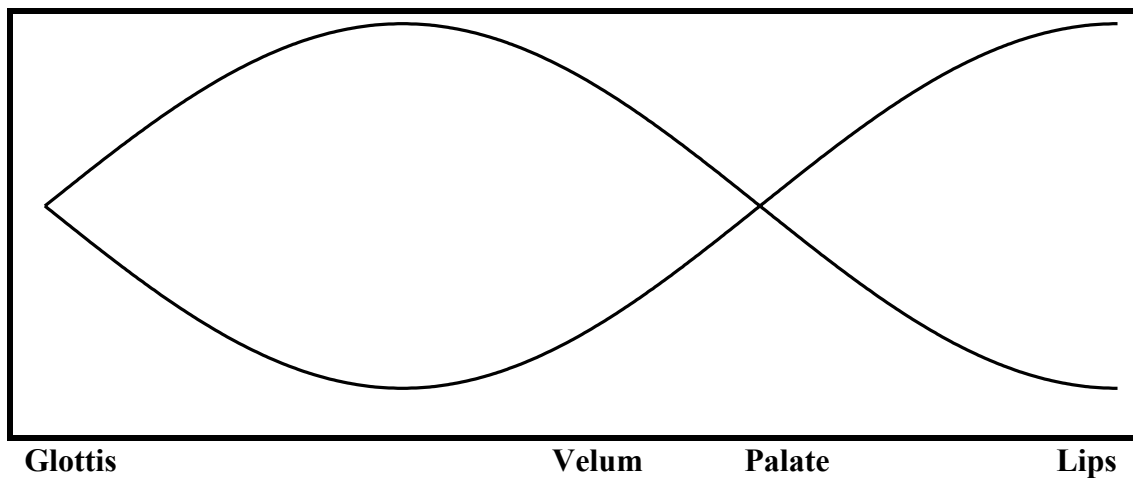


Figure 6.1 Locations of velocity antinodes (where the two sine waves are the furthest apart) and nodes (indicated by the intersections of the sine waves) in the second natural resonance of the vocal tract (F2). Based on Johnson's (2003:109) adaptation of Chiba and Kajiyama's (1941) illustrations and on Ohala (2001).

The consequence of this is that our models may not adequately capture the role of vowel duration. Within the undershoot model, increased duration inhibits vowel reduction by counteracting the consonantal influence. In the case of F1 of mid and low vowels, regression coefficients for the variable duration were positive, indicating an F1-raising effect counteracting the F1-lowering effects of neighboring consonants. For F2, we can predict that increased vowel duration will result in F2 increase in F2-lowering contexts, and F2 decrease in F2-raising contexts. As a result, a single regression coefficient for duration will, at best, partially capture the role of duration in some consonantal contexts. At worst, it will be meaningless.

In other words, to appreciate the role of duration in F2 variability we have to carefully consider the way it interacts with other variables, in particular the place variables. To this

end, for each model we will divide the preceding and the following segments into place-defined classes whose MLR-estimated effects on the second formant appear to be relatively homogeneous. We will then compute separate models for these different consonantal contexts and compare regression coefficients for duration obtained in this way. We will use this method regardless of whether we find that the interaction of duration and a place predictor significantly improves model fit and improves prediction of the given model. However, in the cases in which fit and prediction are not improved, we will not treat the results obtained as conclusive.

There are problems with the interpretation of such variables as lexical stress, speech style, frequency and the context/function distinction that parallel the problems related to the interpretation of the role of vowel duration. This is because at some level of these variables (e.g. stressed syllables) vowels would be predicted to be less affected by consonantal context. But what “less affected” means depends on what the precise place of articulation of the consonants in question is. We will therefore consider interactions of these variables and the place variables, even if these variables themselves are not selected into our models. If a given interaction is found to significantly improve model fit and to improve prediction, we will investigate it in detail by calculating separate regression coefficients for different consonantal contexts.

In addition, we will also consider potentially important interaction terms involving other variables that are selected into our models.

6.1 Standardization

The F2 data were standardized for each intersection of the levels of the variables subject and vowel. That is to say, the entire dataset was divided into groups representing a single vowel phoneme spoken by a single speaker (twenty-four in total, since there were four speakers and six vowel phonemes). For each of these groups, the mean was calculated. It was then subtracted from each individual score, and the results were divided by the standard deviation for each group. The end product of this procedure are z-scores, which express the magnitude (in standard deviations) and the direction (through the sign) of the difference between a given result and the mean for the given group. For each group, the mean of z-scores is 0, and the standard deviation is 1.

Table 6.1.1 presents the mean F2 values and standard deviations for each vowel of each subject that were used to standardize the scores. It also shows the mean standard deviations for each vowel (calculated by averaging the standard deviations of individual subjects).

vowel	m1		m2		w1		w2		average sd
	mean	sd	mean	sd	mean	sd	mean	sd	
i	2006	190	2075	173	2369	216	2567	200	195
e	1601	210	1620	203	1800	311	1980	327	263
y	1559	169	1646	198	1629	207	1881	244	204
a	1352	134	1381	151	1525	153	1626	209	162
o	1148	154	1199	157	1239	161	1296	208	170
u	1120	265	1111	273	1001	261	1103	288	272

Table 6.1.1 F2 means and standard deviations for individual subjects' individual vowels, used to standardize the data, as well as mean standard deviations for each vowel (averages of the subjects' standard deviations). All values are in Hertz.

In the discussions that follow, we will be referring primarily to z-scores. However, in order to make it easier to directly assess the scope of a given effect in real terms, we will also informally reconvert z-scores to scores in Hertz, using average standard deviations for each class of vowels, calculated by averaging the average standard deviations for each vowel. These values are 229 Hz for the front vowels *i* and *e*, 183 Hz for the central vowels *y* and *a* and 221 Hz for the back vowels *o* and *u*. Thus, we might say that a given variable lowers the F2 of front vowels by 1 standard deviation (sd), i.e. approximately 229 Hz. It has to be remembered that this last value will always be only a gross approximation, as in reality, due to the inter-speaker differences, one standard deviation does not mean exactly the same thing for different vowels of different subjects.

6.2 Front vowels

6.2.1 Selecting the model

There were 1568 front vowels *i* and *e* in the entire dataset (see section 3.1 for a detailed breakdown). Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (1176) and a validation set, comprising 25% of tokens (392). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Of the twenty-two independent variables considered, twelve were selected into the model. They were then used to predict normalized F2

scores for vowel tokens in the validation set and “ R^2 ” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “ R^2 ” was recalculated. It was found to be maximized when six more variables were dropped from the model. Accordingly, the remaining 6 independent variables were used to construct a validated model of the training set.

Table 6.2.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 6.2.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 6.2.1.1 and 6.2.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

Independent variable	AIC rank	“R²”
preceding segment place	1	<i>n/a</i>
following segment place	2	0.3097
preceding segment manner	3	0.3731
vowel	4	0.4737
following segment manner	5	0.5041
speech rate	6	0.5237
syllable position	7	0.5393
syllable content/function	8	0.5275
speech style	9	0.5312
word content/function	10	0.5336
prepausal phonological word	11	0.5359
word frequency	12	0.5338
<i>full AIC model</i>	<i>n/a</i>	<i>0.5297</i>
phonological word size	<i>not selected</i>	<i>n/a</i>
duration	<i>not selected</i>	<i>n/a</i>
comparative intensity	<i>not selected</i>	<i>n/a</i>
number of following consonants	<i>not selected</i>	<i>n/a</i>
preceding segment voicing	<i>not selected</i>	<i>n/a</i>
postpausal phonological word	<i>not selected</i>	<i>n/a</i>
following segment voicing	<i>not selected</i>	<i>n/a</i>
clause-initial phonological word	<i>not selected</i>	<i>n/a</i>
clause-final phonological word	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>

Table 6.2.1.1 Variable selection for F2 front vowels model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R²	“R²”	RSE	“RSE”
Value	0.5592	0.5393	0.6542	0.691
Difference	-0.02		0.037	

Table 6.2.1.2 F2 in front vowels. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of F1 in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

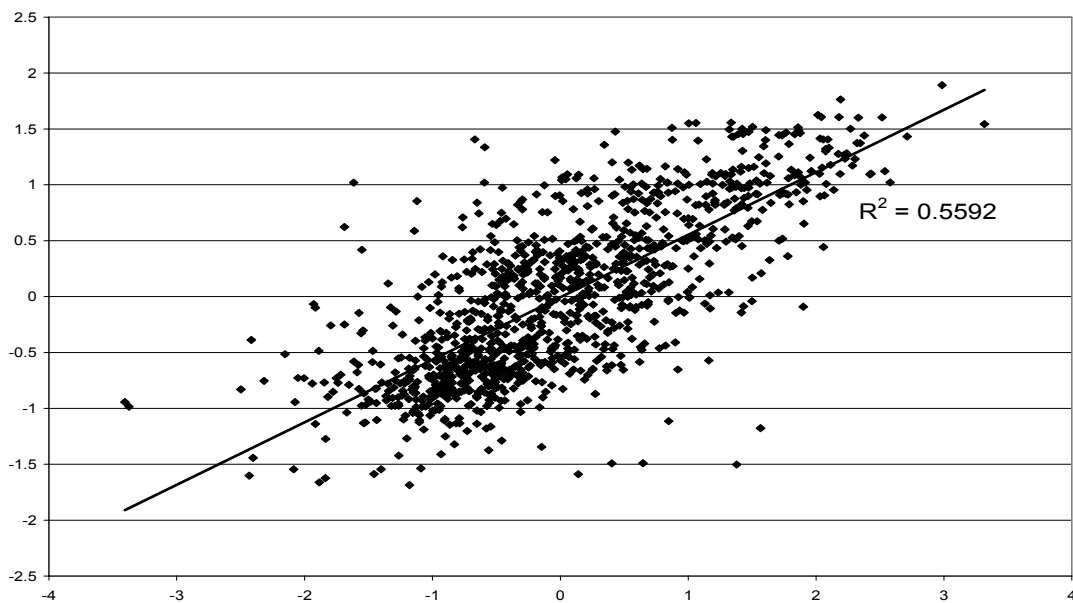


Figure 6.2.1.1 The validated model of F2 of front vowels. Correlation between F2 values observed in the training set (horizontal axis) and values predicted for the training set by the 6 predictors included in the validated model (vertical axis). All units are standard deviations.

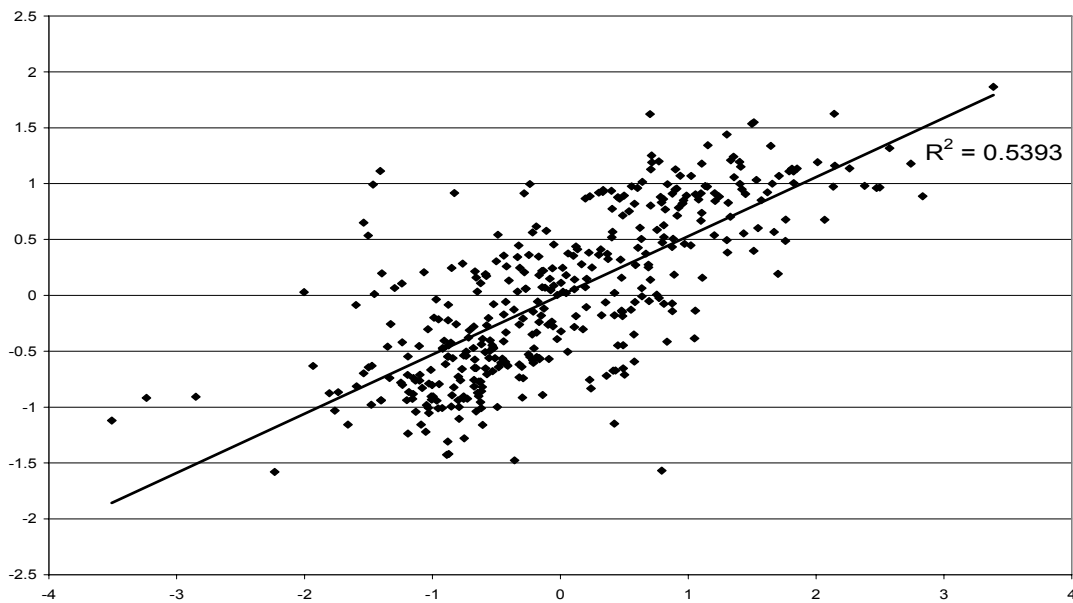


Figure 6.2.1.2 The validated model of F2 of front vowels. Correlation between F2 values observed in the validation set (horizontal axis) and values predicted for the validation set by the 6 predictors included in the validated model (vertical axis). All units are standard deviations.

As can be seen from the data reported above, the validation of the model was rather successful, with small differences between the corresponding statistics for the training set and the validation set. The R^2 values of 0.5393-0.5592 indicate that the correlation coefficient R is between 0.73-0.75, which suggests a fairly strong linear relationship between the dependent variable and the predictors.

6.2.2 Categorical regression

The 6 predictors that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.5567, indicating correlation coefficient $R=0.746$. The key statistics were thus very similar to the values obtained by MLR. Table 6.2.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	θ	partial	part
preceding segment place	1	0.2246	0.653	0.575	0.647	0.565
following segment place	2	0.0842	0.331	0.200	0.428	0.315
preceding segment manner	3	0.058	0.230	0.371	0.314	0.220
vowel	4	0.0298	0.200	0.029	0.255	0.176
following segment manner	5	0.0221	0.195	0.152	0.266	0.184
speech rate	6	0.0092	0.08	0.071	0.118	0.079

Table 6.2.2.1 Comparisons of MLR and categorical regression results, validated F2 front vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, θ is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 6.2.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. The relative values of these two indicators appear to be similar enough to justify treating MLR and categorical regression as equivalent.

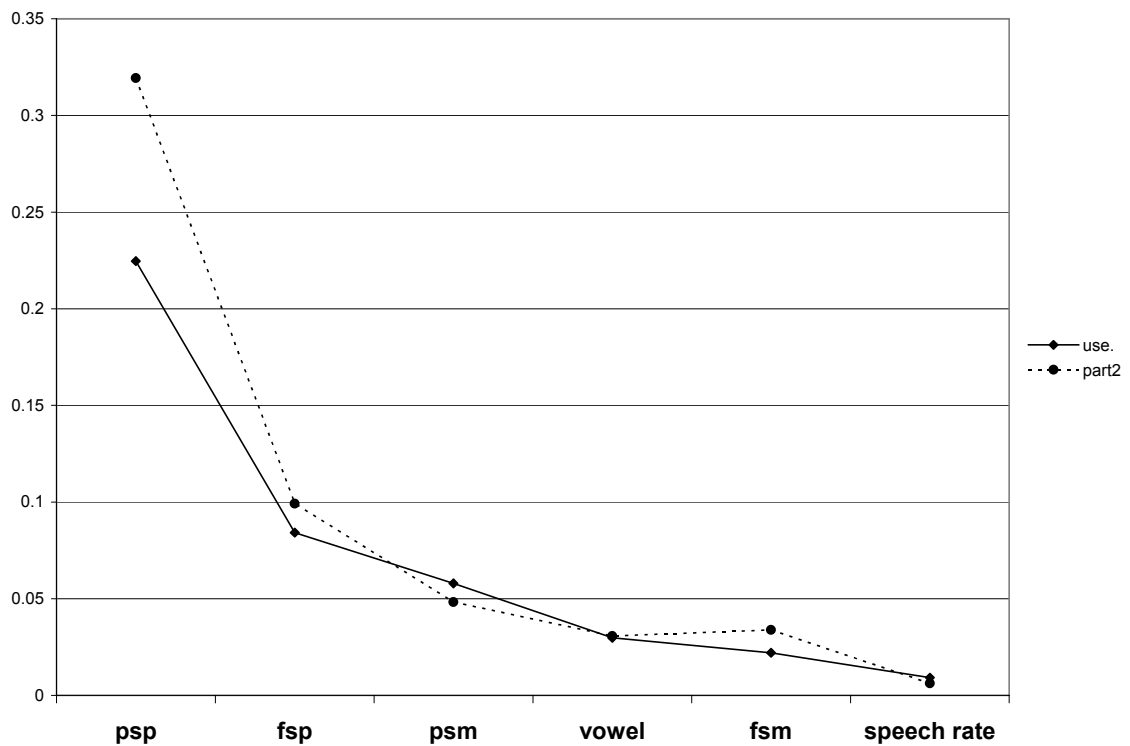


Figure 6.2.2.1 F2 of front vowels. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 6.2.2.2 shows graphically the relative contribution of independent variables to F2 of front vowels estimated by the absolute values of the CR regression coefficients. As can be seen, preceding segment place emerges as by far the strongest predictor of F2 in front vowels. Its impact is approximately twice as strong as that of following segment place, which in turn has a slightly stronger influence than the manner variables and the vowel variable. The last predictor, speech rate, does not have a very strong impact.

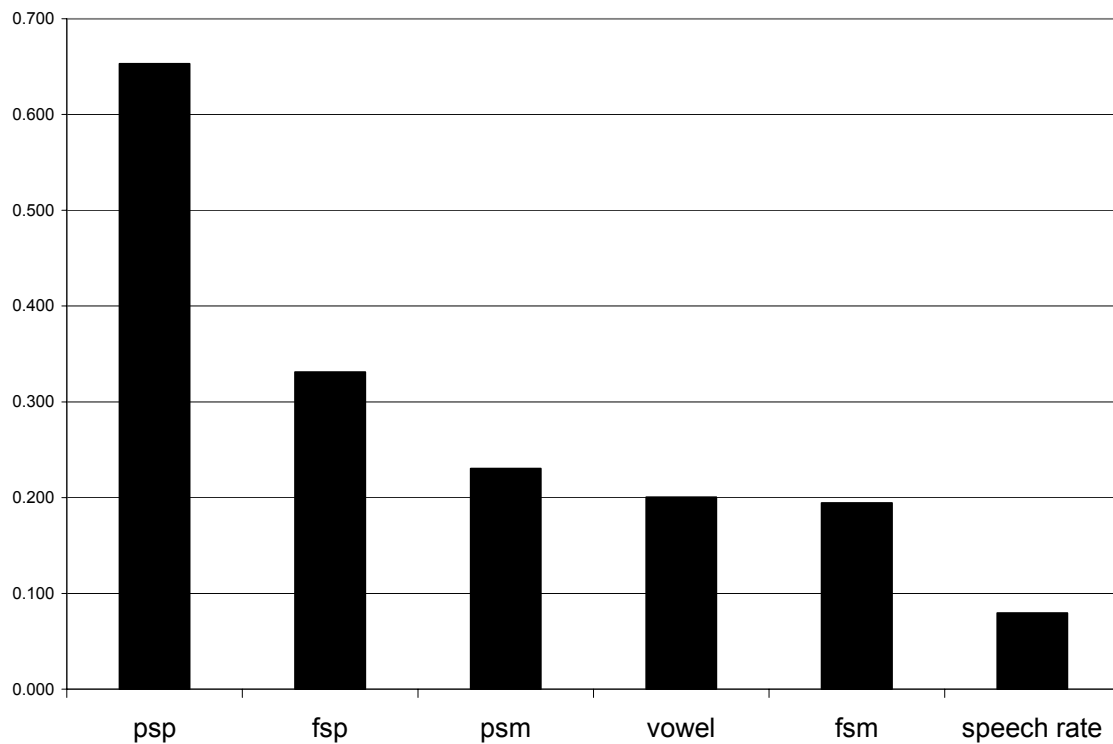


Figure 6.2.2.2 Relative impact of predictors on F2 of front vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

6.2.3 Individual predictors (MLR)

6.2.3.1 Intercept

The intercept of the MLR equation was -0.155. In other words, a front vowel was predicted to have an F2 lower by 0.155 standard deviations (35 Hz) from its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1,

this is a purely mathematical construct with no inherent meaning, since some of the variables cannot have the value of 0 (in this case, speech rate).

6.2.3.2 Preceding segment place

Figure 6.2.3.2.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

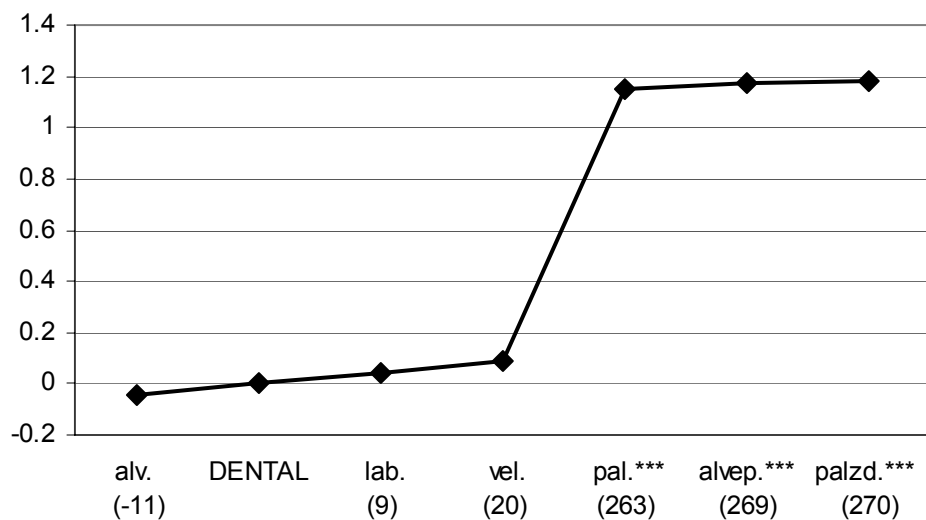


Figure 6.2.3.2.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F2 of front vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 229 Hz as the average value of standard deviation for the F2 of front vowels.

In terms of their effect on the following vowel, figure 6.2.3.2.1 clearly shows that consonants fall into two distinct categories. On the one hand, there are only small differences between vowels following alveolar, dental, labial and velar consonants. On the other hand, all consonants with a palatal component condition F2 that is significantly higher than that observed after dentals, by more than 260 Hz.

We will make use of this patterning of preceding consonants in our investigation of potential interaction effects between the preceding segment place variable and other variables. To this end, we will compute separate regression models for front vowels following alveolars, dental, labials and velars on the one hand, and the “palatal” consonants on the other. We will refer to the former group as *non-fronting* and to the latter as *fronting*. It should be noted that these are merely names of convenience, reflecting the relative properties of the consonants. That is to say, we are not claiming that consonants in the former group have a fronting influence vis-à-vis the idealized target F2 frequency of the vowel, while the latter do not.

6.2.3.3 Following segment place

Figure 6.2.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

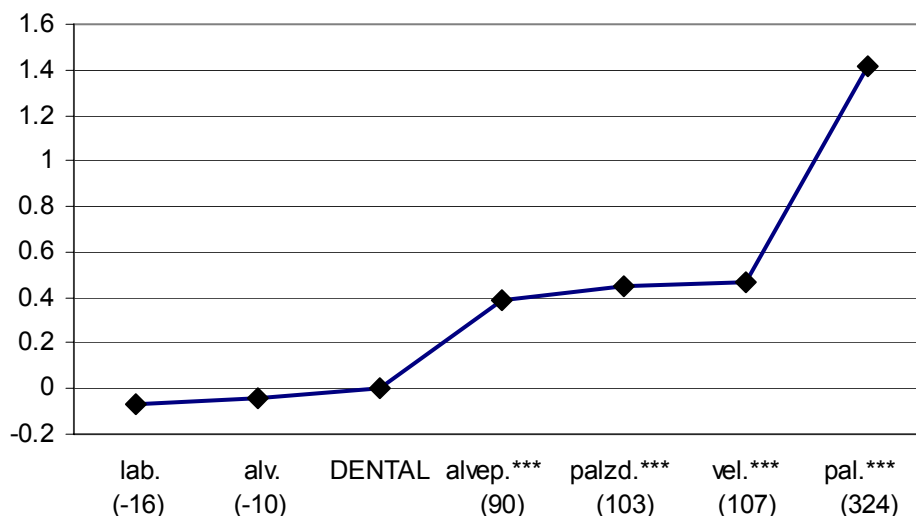


Figure 6.2.3.3.1 MLR-estimated effect of the individual levels of the variable following segment place on F2 of front vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 229 Hz as the average value of standard deviation for the F2 of front vowels.

The effect of following consonant place on front vowels is slightly different than the effect of preceding consonant place discussed earlier. All consonants with a palatal component condition significantly higher F2 (vis-à-vis dentals) in this case as well, but the effect is only comparable in size in the case of palatals; with alveopalatals and palatalized consonants it is much smaller. This is not surprising in the case of palatalized consonants, as it is likely that in their case palatalization is stronger at the moment of release than at the moment of closure. (As was mentioned earlier, it is often asserted that palatalized consonants of Polish can be realized phonetically as sequences of non-palatal consonant and the palatal glide, e.g. Sawicka 1995). It is not clear why there should be a

similar difference for alveopalatals, but it is consistent with the findings of Nowak (2006), showing that perceptually, the vowel plays a greater role in facilitating the perception of a preceding rather than following alveopalatal.

One other important difference between the preceding and following consonant effect is that F2 of front vowels is significantly higher before velars (*vis-à-vis* before dentals) but not after velars (*vis-à-vis* after dentals). This difference appears to indicate that there are differences in CV and VC coarticulation patterns involving velar consonants. One possibility is that coarticulation is stronger with a preceding velar, resulting in the relative backing of the vowel in the vowel space. With following velars, coarticulation is less strong and as a result, there is no backing of the front vowel in the vowel space. It should be noted, however, that the effect is somehow puzzling, given the well-known fact that the constriction tends to be moved forward during the articulation of a velar consonant (Houde 1968).

To analyze interactions in depth, we will compute separate regression coefficients for front vowels appearing before *fronting* consonants (alveopalatals, palatalized, velars and palatals) and *non-fronting* ones (dentals, labials and alveolars). As was explained at the end of the preceding section, these names should be thought of names of convenience rather than ones that express specific theoretical claims about the target frequencies of front vowels.

6.2.3.4 Vowel duration

As can be seen from table 6.2.1.1, vowel duration not only failed to be selected into the validated model, but in fact was not selected into the AIC model. However, as was argued in the introduction to this chapter, in the case of the second formant, it is important to consider interaction effects, because consonants with different places of articulation can have effects with opposite directionality.

In the case of front vowels, neither the interaction of preceding segment place and duration nor of following segment place and duration resulted in a significant improvement of model fit. As a result, our investigation into the potential differences in the duration effect in the context of different types of consonants should be treated as informal and purely exploratory.

Table 6.2.3.4.1 presents the regression coefficients for the variable duration calculated separately for four models (vowels preceded/followed by fronting/non-fronting consonants; for the definition of fronting and non-fronting see sections 6.2.3.2 and 6.2.3.3). The models were calculated using all the predictors included in the validated model plus duration.

Condition	fronting	non-fronting
preceding consonant	0.00035 (0.08 Hz/ ms)	0.0018 (0.4 Hz/ms)
following consonant	-0.0018 (-0.4 Hz/ms)	0.0009 (0.21 Hz/ms)

Table 6.2.3.4.1 Regression coefficients for the variable duration in different consonantal contexts, F2 of front vowels. Fronting and non-fronting consonants for the preceding and the following segment condition are defined in sections 6.2.3.2 and 6.2.3.3, respectively. The values are in standard deviations and can be interpreted as change in standardized F2 value per each additional millisecond of vowel duration. The values in brackets were estimated using 229 Hz as the average value of standard deviation of F2 for front vowels.

If we consider only the following consonants, the regression coefficients reported in table 6.2.3.4.1 are consistent with the undershoot model of vowel reduction. For the fronting consonants, the sign of the coefficient is negative, i.e. increasing the duration of the vowel results in less vowel fronting. For the non-fronting consonants, the coefficient is positive, which indicates that increasing the duration of the vowel results in fronting it. On this interpretation, the target F2 value of the front vowels would be estimated to be somehow lower than the one observed before the fronting consonants, and somewhat higher than the one observed before the non-fronting consonants. Intuitively, this appears to be plausible, especially for the vowel *e*.

The regression coefficients obtained for the preceding consonants cannot be reconciled with the ones obtained for following consonant. Both coefficients are positive, indicating that under the undershoot model of vowel duration, the target F2 frequency of front vowel would be expected to be higher than the one observed in any consonantal context. This does not appear to be very likely. As we have noted before, Polish vowels, and the front vowel *e* in particular, are impressionistically fronted in palatal contexts (e.g. Sawicka 1995).

To sum up, the data illustrating the interaction of following consonant place and vowel duration are consistent with the undershoot model of vowel reduction. They do not present, however, conclusive evidence to this effect, since the interaction did not significantly improve model fit. The data illustrating the interaction of preceding segment place and vowel duration do not appear to be consistent with the undershoot model. This difference between the following and preceding consonant place effects may reflect a genuine difference in coarticulatory mechanisms or it may simply be a function of the fact that these interaction effects are non-significant. In either case, vowel duration does not appear to play a major role in the variability of F2 of front vowels. This can also be appreciated by noting that the rates of change in Hz/ms estimated from regression coefficients, as shown in table 6.2.3.4.1, are all smaller than the F1 rates of change estimated in chapter 5 for non-high vowels.

6.2.3.5 Speech rate

The MLR validated model estimated the regression coefficient for the continuous variable speech rate at -0.073, significant at $p < 0.001$. Among the front vowels, speech rate varied between 1.78-10.8 syllables per second. This indicates that the contribution of this variable varied between subtracting 0.13 sd ($-0.073 * 1.78$) from the intercept and subtracting 0.79 sd. ($-0.073 * 10.8$) from it. In other words, the vowel spoken at the highest observed speech rate was estimated to have the second formant 0.66 sd (~ 151 Hz) lower than the vowel spoken at the lowest observed speech rate.

The higher the speech rate, the less front the front vowel is. On the face of it, this could be summarized as indicating increased centralization at higher speech rates. It is not easy, however, to account for such an effect within the undershoot model of vowel reduction. Given that the average consonantal influences appear to increase rather than decrease the estimated F2 of front vowels, and that increased speech rate should be correlated with decreased vowel duration, we would predict that any general effect of speech rate should point in the opposite direction, i.e. that front vowels should be more front at higher speech rates. A different explanation for the observed effect must be sought. One purely speculative possibility is that the fronting consonants themselves are reduced at higher speech rates, i.e. are not realized with the same amount of tongue body fronting. This would cause them to have a more limited fronting influence on front vowels.

To a certain extent, this appears to be confirmed by the investigation of the interaction between preceding consonant place and speech rate. Adding this interaction to the validated model resulted in a nearly significant increase in model fit (increase in R^2 by 0.0046, $p < 0.1$) and improved the predictive accuracy of the model (increase in “ R^2 ” of the validation set by 0.009). The regression coefficient obtained for preceding fronting consonants was -0.11, indicating a decrease in F2 frequency of approximately 27 Hz for each additional syllable per millisecond in speech rate. The regression coefficient for non-fronting consonants was also negative, but smaller in size (-0.04, indicating a decrease in F2 of approximately 10 Hz per each additional syllable per second). Thus in

the case of preceding non-fronting consonants, the speech rate effect could be consistent with the undershoot mechanism.

The interaction of speech rate and following segment place failed to significantly improve model fit.

6.2.3.6 Other variables potentially interacting with place variables

Predictors reflecting differences in style, syllable position, word frequency and the content/function distinction were not selected into the validated model. However, as was argued above, their relative importance could be obscured by their non-uniform effect in different consonant place environments. We have therefore explored the potential interactions between these variables and the place variables. The results are summarized in table 6.2.3.6.1. As can be seen, four interactions fulfilled our criteria, i.e. resulted in significantly increased model fit and increased prediction. They are analyzed in detail.

Predictor	preceding segment place	following segment place
style	0.0141, p<0.001, -0.0041	0.0144, p<0.001, 0.006
stress	0.0247, p<0.001, 0.0132	0.011, p<0.05, -0.005
syllable content/function	0.0106, p<0.001, -0.0056	0.0125, p<0.001, 0.009
word content/function	0.0176, p<0.001, -0.0064	0.0026, n/s
word frequency	0.041, p<0.001, -0.0037	0.008, p<0.01, 0.0154

Table 6.2.3.6.1 Exploration of interactions between the place variables (columns) and other selected variables (rows) for F2 of front vowels. Each cell shows in turn change in model fit (i.e. increase in R^2 of the training set) resulting from the addition of the interaction term, significance of the R^2 increase and change in the predictive power of the model (change in “ R^2 ” of the validation set). Changes in predictive power were not calculated for interaction that did not significantly improve model fit. Cells are in bold if a given interaction both significantly improves model fit and improves prediction. The variable stress is the simplified version of the variable syllable position, cf. section 5.3.3.7.

Figure 6.2.3.6.1 illustrates the interaction between preceding segment place and stress by showing separately the regression coefficients obtained for front vowels following the fronting consonants and the non-fronting consonants. The differences are somewhat puzzling, as they point in the opposite direction from what could be expected on theoretical grounds. Assuming that it is likely that vowel reduction is decreased in stressed syllables (which was confirmed for F1 in the preceding chapters), we would expect that in the fronting contexts, vowels in stressed syllables should have lower F2 values than other vowels. In the non-fronting contexts, it would appear more likely that front vowels would have higher F2 values in stressed syllables, though here the predictions are somewhat less clear. In fact we see in both cases that the reverse is the case.

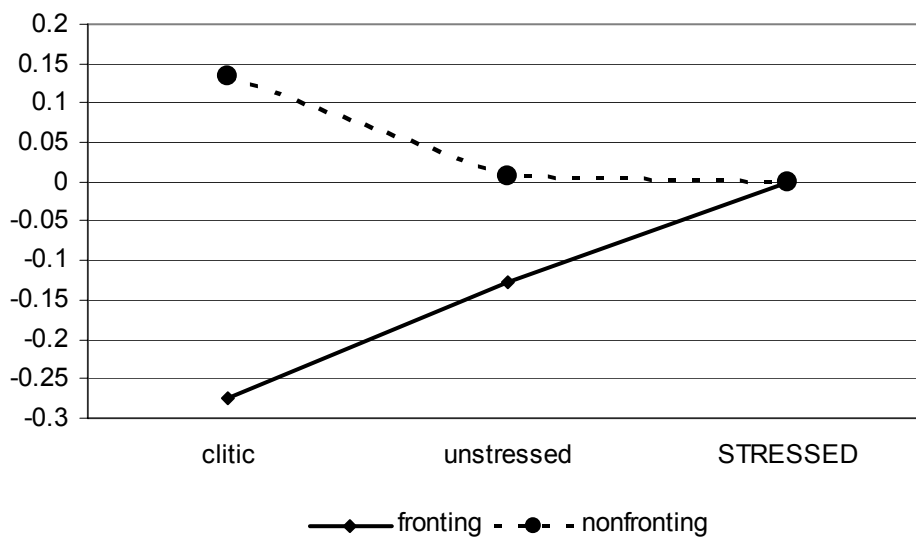


Figure 6.2.3.6.1 Interaction between stress and preceding segment place. The two lines link regression coefficients calculated separately for the two different classes of consonants. All values are standard deviations.

In fact, the expected picture emerges when we consider interaction between stress and following segment place. (As indicated in table 6.2.3.6.1, this interaction did significantly improve model fit, but it slightly reduced the predictive power of the model. It should therefore be treated with caution.) Figure 6.2.3.6.2 shows the regression coefficients calculated separately for preceding fronting and non-fronting consonants.

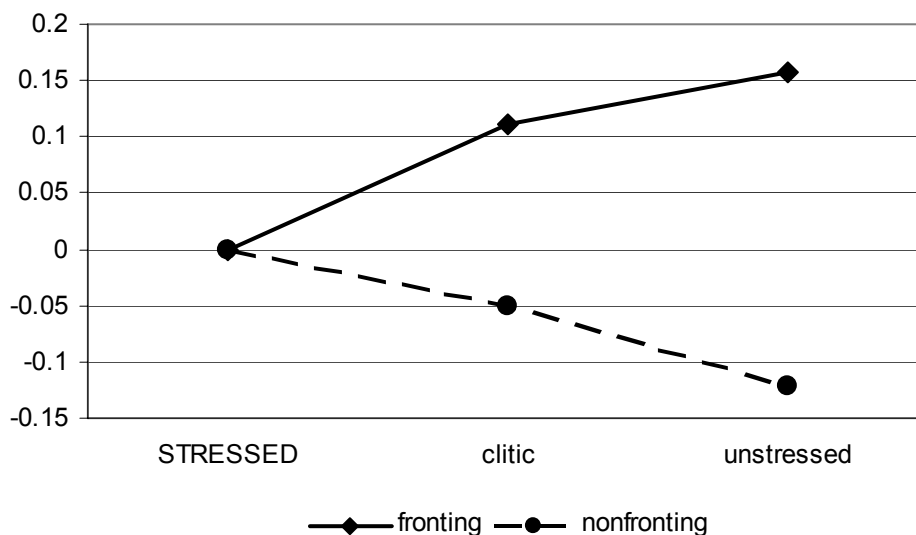


Figure 6.2.3.6.2 Interaction between stress and following segment place. The two lines link regression coefficients calculated separately for the two different classes of consonants. All values are standard deviations.

Thus it can be cautiously concluded that stress does not reduce coarticulation between the vowel and the preceding consonant; in fact the opposite may be the case. As far as the following consonants are considered, however, stress does appear to limit coarticulation. It is worth pointing out at this point that a similar picture seemed to emerge from our consideration of the interactions of vowel duration and the place variables (cf. section 6.2.3.4).

Figure 6.2.3.6.3 illustrates the interaction of following segment place and style. Based on the findings of other researchers for other languages (e.g. Koopmans-van Beinum 1980), as well as our findings concerning the first formant, as reported in the preceding chapter, we could expect vowels to be more reduced in the less formal speech styles, in particular in the narrative style. Figure 6.2.3.6.3 suggests this is the case: we see that front vowels in the fronting contexts have lower F2 values in the normal reading style than in the two other styles, while in the non-fronting contexts, they have higher F2 values in the normal reading style.

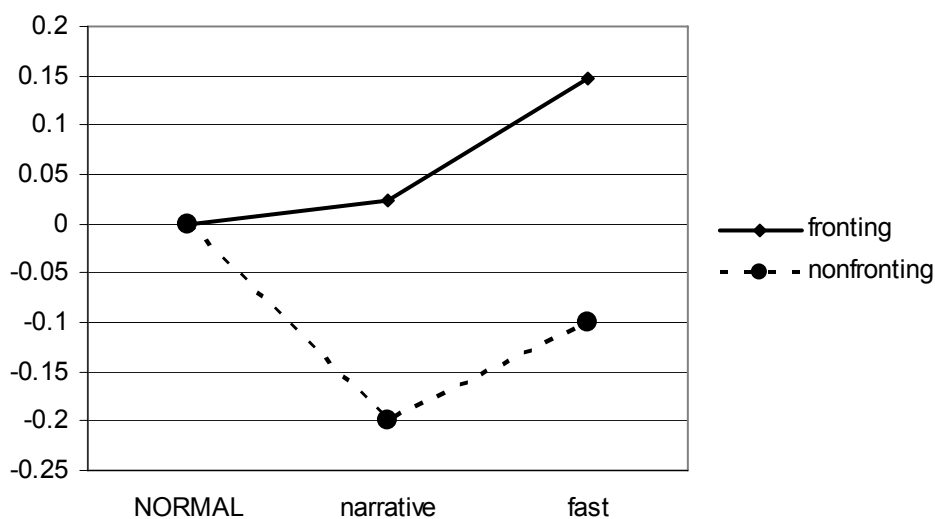


Figure 6.2.3.6.3 Interaction between style and following segment place. The two lines link regression coefficients calculated separately for the two different classes of consonants. All values are in standard deviations.

Vowels in function syllables were found to have lower F2-values than vowels in content syllables when followed by fronting consonants (by 0.025 sd, i.e. approximately 6 Hz) and higher F2 values when followed by non-fronting consonants (by 0.011 sd, i.e.

approximately 2 Hz). Since we could expect vowels to be more reduced in function syllables (cf. van Bergem 1993 and chapter 5), this effect is unexpected and difficult to explain. However, it is very small in size.

The regression coefficients of the variable word frequency, calculated separately for vowels following fronting consonants and non-fronting consonants were very small. They indicated that in both cases, F2 decreases as word frequency increases. In the case of following fronting consonants, it decreases by approximately 1 Hz for each additional 100,000 in word frequency. In the case of non-fronting consonants, the decrease is by approximately 6 Hz per each additional 100,000 in word frequency. The latter effect may be consistent with the expectation that there should be more coarticulation in more frequent words, but the former cannot.

6.2.3.7 Preceding segment manner

Figure 6.2.3.7.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

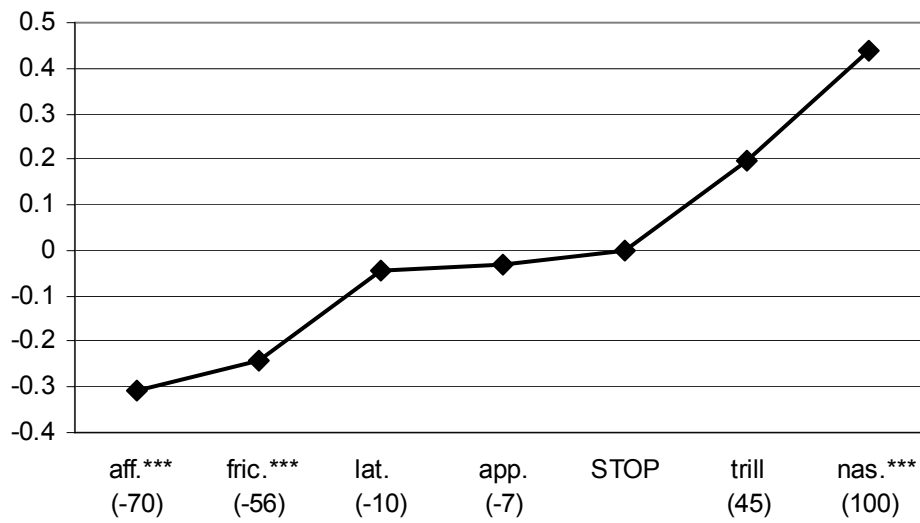


Figure 6.2.3.7.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F2 of front vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 229 Hz as the average value of standard deviation for the F2 of front vowels.

Front vowels were found to have significantly higher values of the second formant after nasals (*vis-à-vis* stop) and significantly lower F2 values after fricatives and affricates (*vis-à-vis* stops). There are no clear reasons for these differences. One possibility is that consonants with different manners have different coarticulatory requirements. In this case, more light could be shed on the effect in question by investigating a potential interaction of preceding segment place and manner.

Adding to the validated model an interaction term between preceding segment place and manner significantly improved model fit (increase in R^2 by 0.021, $p < 0.001$) and prediction (increase in " R^2 " by 0.008). Figure 6.2.3.7.2 shows the regression coefficients

calculated separately for the fronting and the non-fronting consonants. The F2 lowering effect of preceding fricatives and affricates can be observed in both groups of consonants, although it appears to be stronger for the non-fronting consonants. The greatest difference can be seen for laterals and nasals. Non-fronting laterals and nasals appear to have little influence on the F2 of the following vowel. However, it is precisely after nasals and laterals that the F2-raising influence of the fronting places of articulation is most pronounced. The precise reasons for this difference, however, remain unclear.

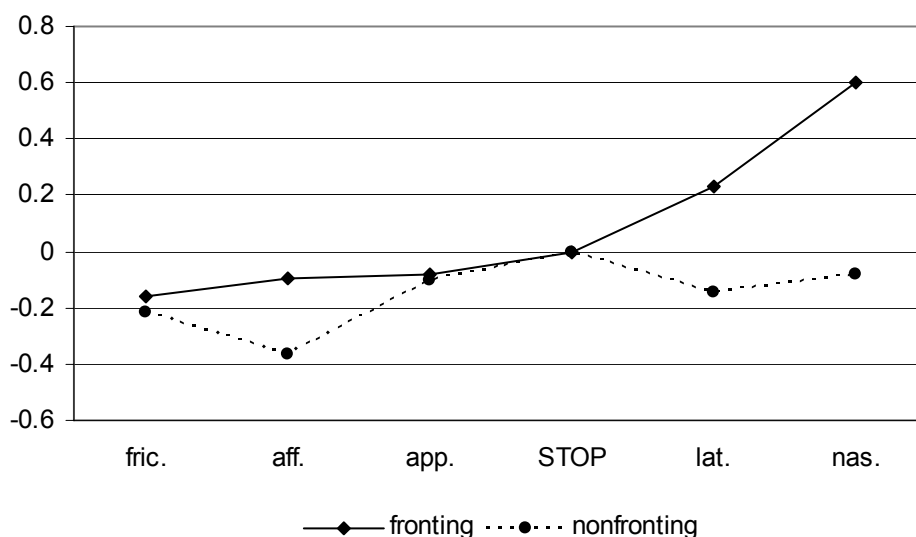


Figure 6.2.3.7.2 Interaction between preceding segment place and preceding segment manner. The two lines link regression coefficients calculated separately for the two different classes of consonants. Trills are not included, as there were no fronting trills in the data. All values are standard deviations.

6.2.3.8 Following segment manner

Figure 6.2.3.8.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

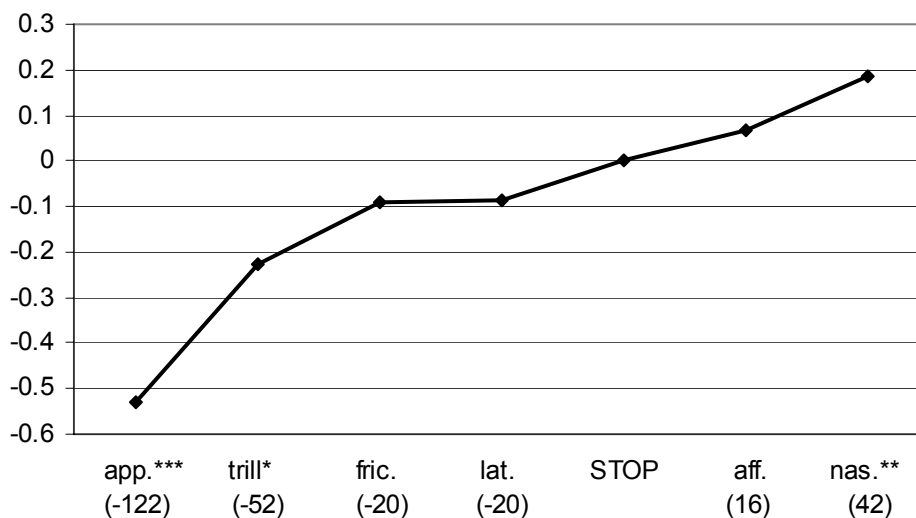


Figure 6.2.3.8.1 MLR-estimated effect of the individual levels of the variable following segment manner on F2 of front vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 229 Hz as the average value of standard deviation for the F2 of front vowels.

Like preceding nasals, following nasals condition significantly higher F2 values in front vowels (vis-à-vis stops). In addition, F2 is significantly lower before approximants and trills. This stands in contrast to the results of the preceding segment manner variable

reported above. In this case also there is the possibility that the manner variable interacts with the place variable.

Adding to the validated model an interaction term between following segment place and manner significantly improved model fit (increase in R^2 by 0.016, $p < 0.001$) and prediction (increase in “ R^2 ” by 0.005). Figure 6.2.3.8.2 shows the regression coefficients calculated separately for the fronting and the non-fronting consonants. With the exception of approximants, there are very few manner-related differences between vowels preceding fronting consonants. F2 is predicted to be somewhat higher, which likely indicates less reduction, before non-fronting consonants that are affricates, nasals or laterals (as opposed to stops). The source of this effect is unclear. The biggest difference, however, can be seen for approximants. The highest F2 values were found in the context of fronting approximants (i.e. the palatal glide) and the lowest in the context of non-fronting approximants (i.e. the labiovelar glide). This seems to indicate that coarticulation between a vowel and a following approximant is greater than between a vowel and a following consonant of another type. A comparison of figures 6.2.3.7.2 and 6.2.3.8.2 reveals that this is not the case for preceding approximants.

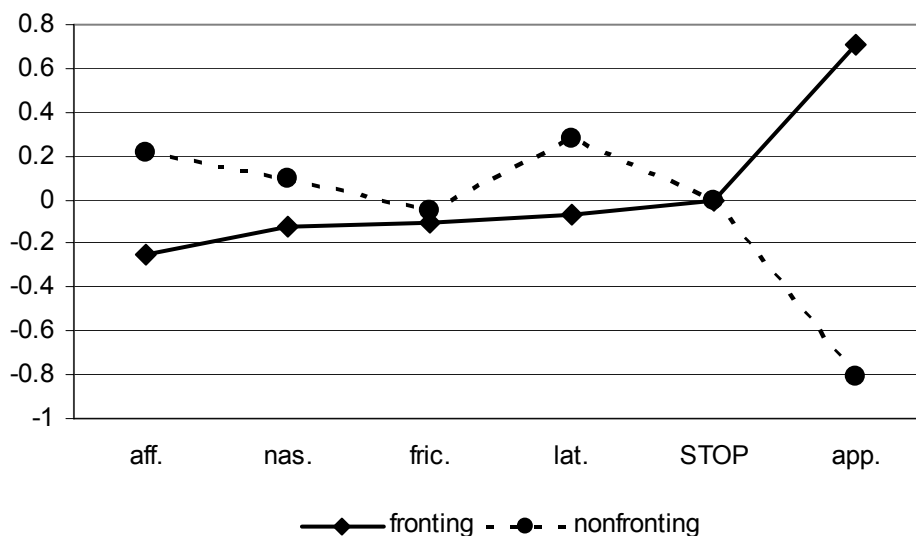


Figure 6.2.3.8.2 Interaction between following segment place and following segment manner. The two lines link regression coefficients calculated separately for the two different classes of consonants. Trills are not included, as there were no fronting trills in the data. All values are standard deviations.

6.2.3.9 Vowel

The front vowel was estimated to have standardized F2 0.57 sd. (~131 Hz) lower if it was *i* rather than *e*, significant at $p < 0.001$. Since the data were standardized by vowel and by speaker, this does not indicate that the second formant values of *i* were lower than those of *e*. As was shown in figure 3.1.2, the reverse was the case. What the vowel term indicates is that the two front vowels interacted with other variables in somewhat different ways.

Both the interaction between preceding segment place and vowel and between following segment place and vowel significantly increased model fit (increases in R^2 by 0.028 and

0.024, respectively, both significant at $p < 0.001$) and improved the predictive power of the model (increases in R^2 of the validation set by 0.011 and 0.02, respectively). The former interaction is shown in figure 6.2.3.9.1, the latter in figure 6.2.3.9.2.

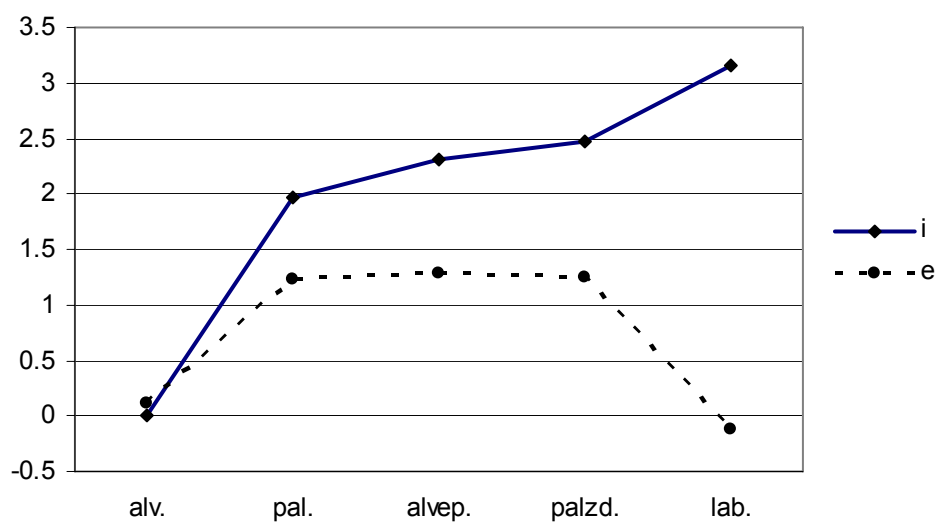


Figure 6.2.3.9.1 Interaction between vowel and preceding segment place. The two lines link regression coefficients calculated separately for the two different front vowels. There were no instances of *i* preceded by dentals or velars, and these places are not included. All values are standard deviations.

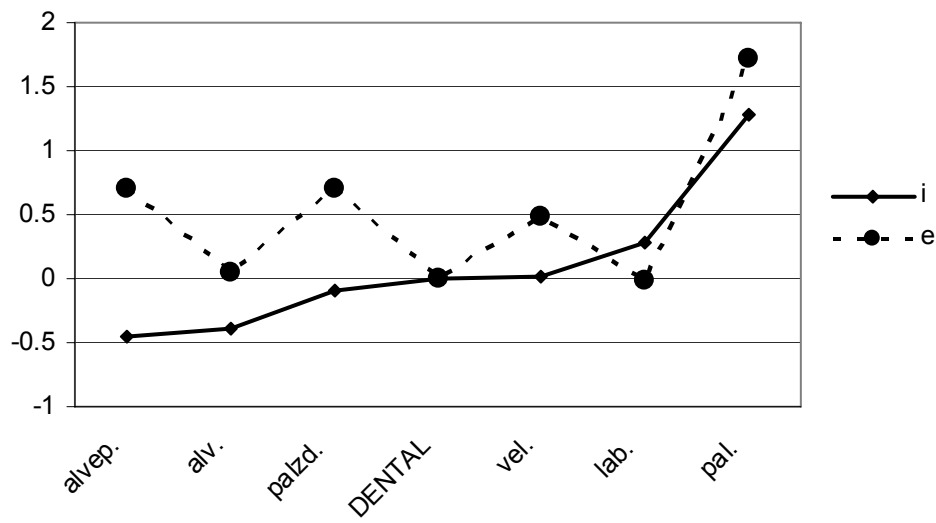


Figure 6.2.3.9.2 Interaction between vowel and following segment place. The two lines link regression coefficients calculated separately for the two different vowels. All values are standard deviations.

Figure 6.2.3.9.1 shows that the preceding place effect for the vowel *e* is exactly what would be expected from the consideration of the general preceding place effect (cf. section 6.2.3.2). That is, there are no differences between labials and alveolars on the one hand, and between palatals, alveopalatals and palatalized consonants on the other. For *i*, the picture looks different and is somehow surprising. In particular, F2 of *i* is estimated to be higher after labials than after consonants with other places of articulation. It has to be remembered, however, that non-palatal and non-palatalized consonants can only precede *i* in Polish across a word boundary. This means that coarticulatory processes may be subject to different constraints in these specific cases than in the more regular instances of consonant-vowel adjacency inside a word. It also means that there are few examples of *i* preceded by non-palatal and non-palatalized consonants. In fact, the estimates shown in figure 6.2.3.9.1 are based on only two cases of preceding alveolars and only five cases of

preceding labials. As a result, the estimated regression coefficients may not be very reliable.

In figure 6.2.3.9.2, we see that the regression coefficients estimated separately for *e* follow the general following place pattern discussed in section 6.2.3.3. In particular, higher F2 values are found before palatals, palatalized consonants and alveopalatals. However, with the exception of high F2 values before palatals, the vowel *i* does not behave in the same way. In particular, lower F2 values are predicted before alveopalatals than before dentals. This is genuinely puzzling and no obvious answer is apparent. One possibility is that following dentals, velars and labials coarticulate strongly with the preceding *i*, but alveopalatals (and alveolars) do not. However, at this time it has to remain a mere speculation.

One important consequence of the interaction shown in figure 6.2.3.9.2 is that the general effect of following consonant place discussed before in section 6.2.3.3 clearly underestimates the amount of F2-raising that *e* undergoes when it precedes alveopalatals.

6.2.4 Summary

The variation of front vowels in the F2 domain is mostly the function of preceding and following consonant place. Duration is not selected as an important predictor of vowel reduction. Its interaction with following segment place is consistent with the undershoot

model, but it does not significantly improve model fit. These non-significant interaction effects suggest that the target formant frequencies of the front vowels are lower than those observed before consonants with the palatal component and velars, but higher than those observed before dentals, labials and alveolars.

Stress, speech style, word frequency and the content/function distinction are not important predictors of vowel reduction per se. The interaction of style and following segment place indicates the expected effect of greater reduction in unstressed syllables. The interaction of stress and following segment place shows the expected increased reduction in unstressed syllables, but it fails validation. Other interactions involving these variables and the place variables either cannot be validated or yield unexpected results. The speech rate effect is likewise unexpected and difficult to account for.

The manner variables have a relatively strong impact on F2 frequency of front vowels as well. The results are complex and differ for preceding and following consonants. There is no clear explanation for them.

The vowel term was significant, indicating that the two front vowels behave differently. In particular, there is some evidence suggesting that the vowel *e* is more affected by neighboring consonants than the vowel *i*.

6.3 Central vowels

6.3.1 Selecting the model

There were 1854 tokens of the central vowels *a* and *y* in the entire dataset (see section 3.1 for a detailed breakdown). Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (1391) and a validation set, comprising 25% of tokens (463). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Of the twenty-two independent variables considered, thirteen were selected into the model. They were then used to predict normalized F2 scores for vowel tokens in the validation set and “ R^2 ” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “ R^2 ” was recalculated. It was found to be maximized when five more variables were dropped from the model. Accordingly, the remaining 8 independent variables were used to construct a validated model of the training set.

Table 6.3.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 6.3.1.2 compares the statistical

data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 6.3.1.1 and 6.3.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

Independent variable	AIC rank	“R ² ”
preceding segment place	1	<i>n/a</i>
following segment place	2	0.1667
following segment manner	3	0.3754
preceding segment manner	4	0.4228
vowel	5	0.4354
postpausal phonological word	6	0.4456
vowel duration	7	0.4512
word content/function	8	0.4547
word frequency	9	0.4583
phonological word size	10	0.4549
preceding segment voicing	11	0.4512
clause-final phonological word	12	0.4507
number of following consonants	13	0.4474
<i>full AIC model</i>	<i>n/a</i>	<i>0.4463</i>
prepausal phonological word	<i>not selected</i>	<i>n/a</i>
following segment voicing	<i>not selected</i>	<i>n/a</i>
speech rate	<i>not selected</i>	<i>n/a</i>
syllable content/function	<i>not selected</i>	<i>n/a</i>
comparative intensity	<i>not selected</i>	<i>n/a</i>
clause-initial phonological word	<i>not selected</i>	<i>n/a</i>
speech style	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>
syllable position	<i>not selected</i>	<i>n/a</i>

Table 6.3.1.1 Variable selection for F2 central vowels model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R^2	" R^2 "	RSE	"RSE"
Value	0.5351	0.4583	0.703	0.6962
Difference	0.077		-0.0068	

Table 6.3.1.2 F2 in central vowels. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R^2 = squared multiple correlation coefficient of the validated model of the training set; " R^2 " = squared correlation coefficient between observed values of F1 in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; "RSE" = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

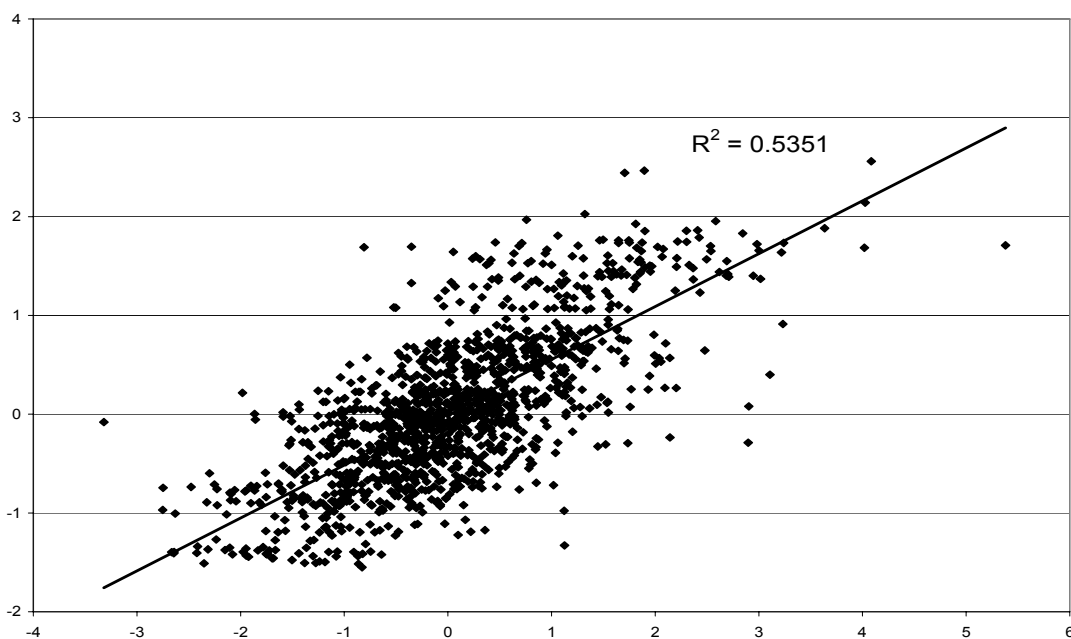


Figure 6.3.1.1 The validated model of F2 of central vowels. Correlation between F2 values observed in the training set (horizontal axis) and values predicted for the training set by the 8 predictors included in the validated model (vertical axis). All units are standard deviations.

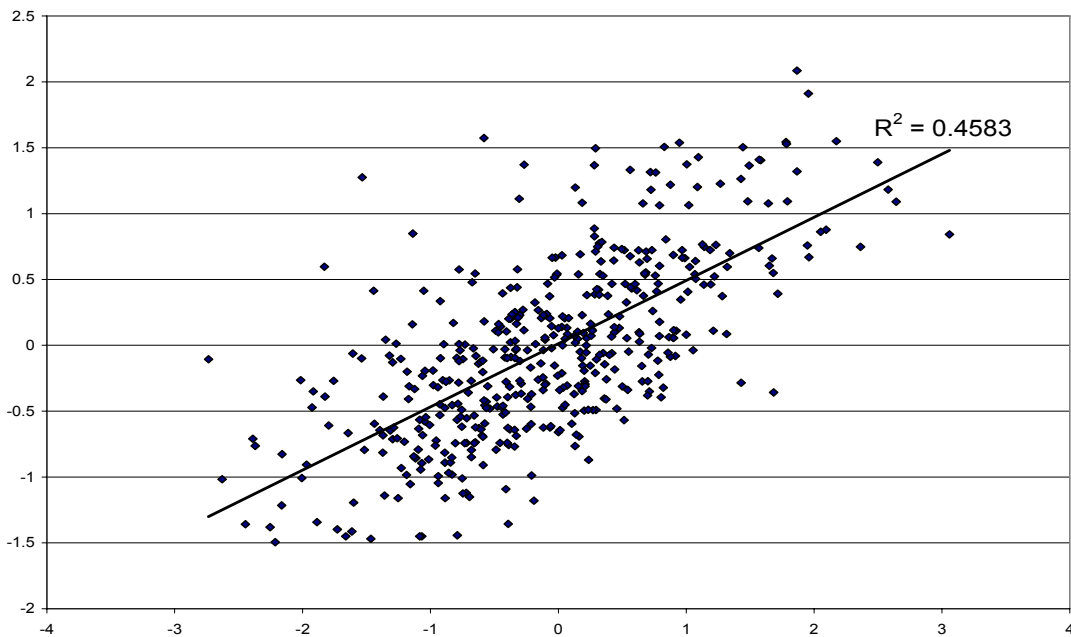


Figure 6.3.1.2 The validated model of F2 of central vowels. Correlation between F2 values observed in the validation set (horizontal axis) and values predicted for the validation set by the 8 predictors included in the validated model (vertical axis). All units are standard deviations.

The validated model is moderately accurate. R^2 values between 0.4583-0.5351 indicate a multiple correlation coefficient R between 0.68-0.73, i.e. a fairly strong linear relationship between the dependent variable and the predictors. As far as model validation is concerned, the RSE and “RSE” values differ only minimally, but the difference between R^2 and “ R^2 ” is fairly large (0.077). However, this difference is not very far from the 0.05 criterion suggested by Schwab (2006) for comparing models built separately on two subsets of data, i.e. with regression coefficients optimized separately for the two subsets. The validation method used here employs only one set of regression coefficients, optimized for the training set. This means that the “ R^2 ” value obtained for the validation set must be lower. In fact, when a separate model is constructed for the

validation set, using the same set of predictors, but with regression coefficients optimized, the R^2 of the validation set 0.4912, i.e. it is within 0.05 of the R^2 of the training set (0.5351, i.e. a difference of 0.044).

6.3.2 Categorical regression

The eight predictors that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.527, indicating correlation coefficient $R=0.726$. The key statistics were thus very similar to the values obtained by MLR. Table 6.3.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC rank	usefulness	b	θ	partial	part
preceding segment place	1	0.1728	0.515	0.517	0.568	0.475
following segment place	2	0.1521	0.440	0.459	0.533	0.433
preceding segment manner	4	0.0174	0.170	0.027	0.228	0.161
following segment manner	3	0.0157	0.153	0.238	0.213	0.150
vowel	5	0.0044	0.109	0.010	0.135	0.094
postpausal PW	6	0.0041	0.078	0.119	0.111	0.077
vowel duration	7	0.0024	0.050	0.026	0.063	0.043
word content/function	8	0.0019	0.033	0.164	0.046	0.032

Table 6.3.2.1 Comparisons of MLR and categorical regression results, validated F2 central vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, θ is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 6.3.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. The relative values of these two indicators are very similar, which justifies treating MLR and categorical regression as equivalent.

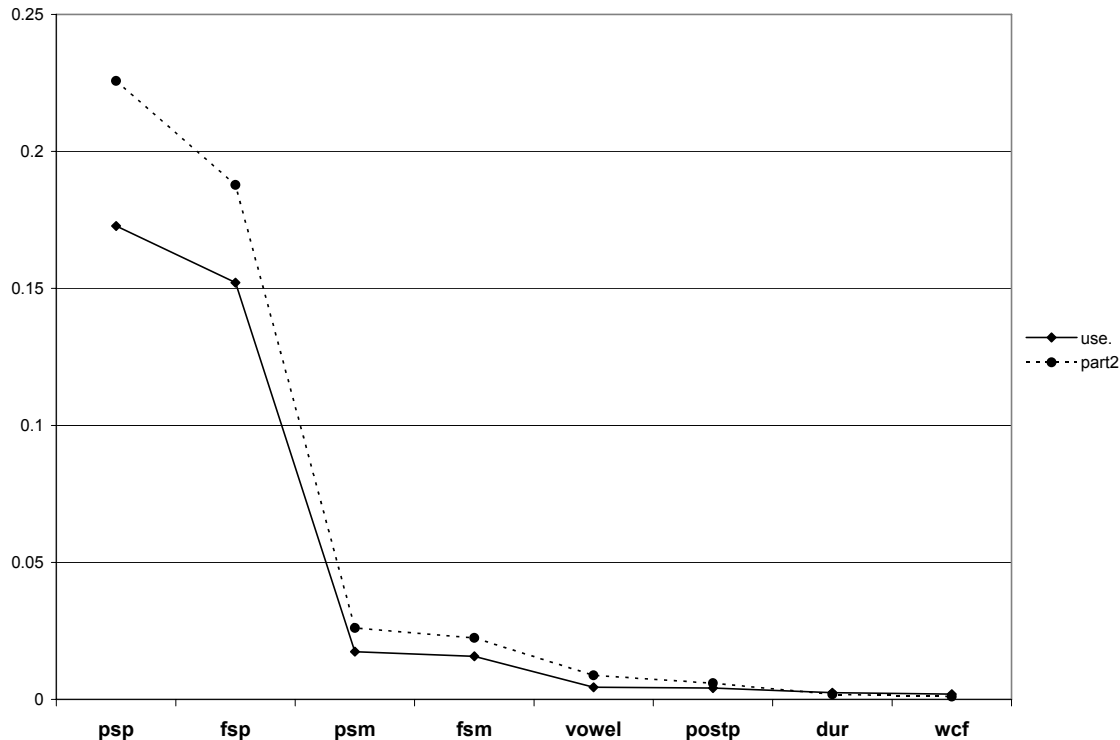


Figure 6.3.2.1 F2 of central vowels. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 6.3.2.2 shows graphically the relative contribution of independent variables to F2 of central vowels estimated by the absolute values of the CR regression coefficients. The place variables are by far the strongest predictors, followed by the manner variables. For both manner and place, the impact of the preceding consonant is slightly greater than that of the following consonant. Other variables have a smaller impact.

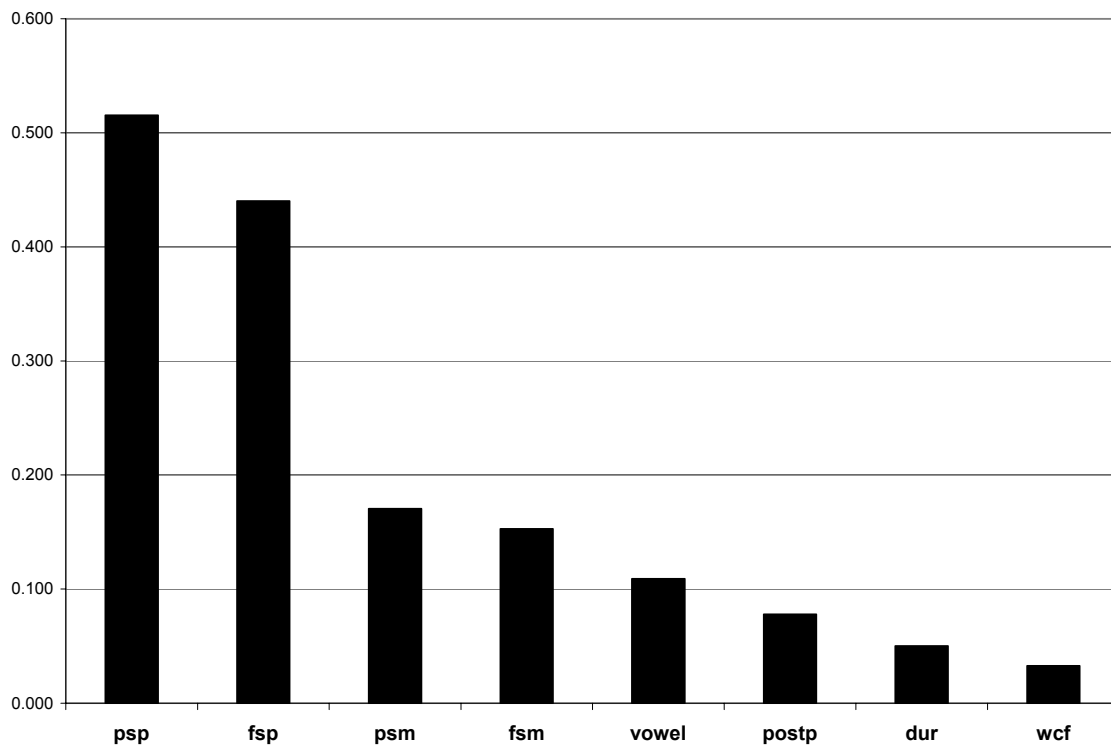


Figure 6.3.2.2 Relative impact of predictors on F2 of central vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

6.3.3 Individual predictors (MLR)

6.3.3.1 Intercept

The intercept of the MLR equation was 0.124. In other words, a central vowel was predicted to have an F2 higher by 0.124 standard deviations (23 Hz) than its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1,

this is a purely mathematical construct with no inherent meanings since some of the variables cannot have the value of 0 (in this case, duration).

6.3.3.2 Preceding segment place

Figure 6.3.3.2.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

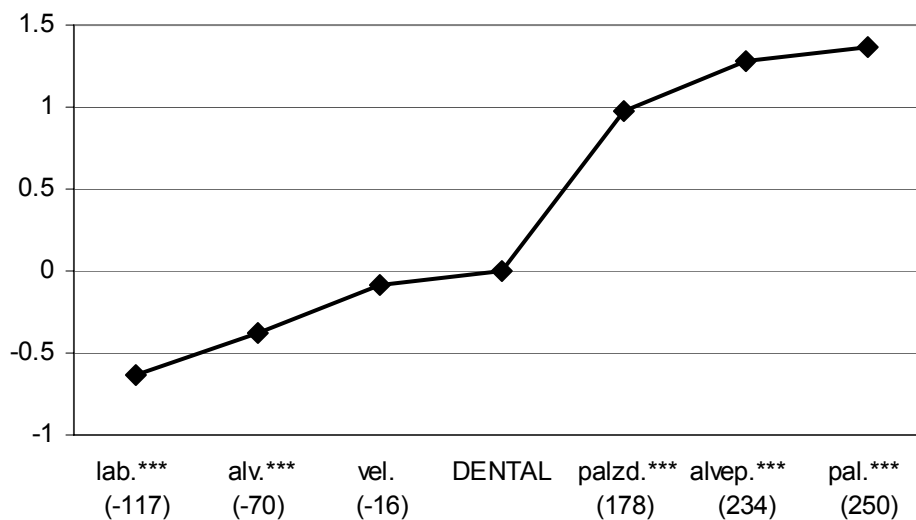


Figure 6.3.3.2.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F2 of central vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 183 Hz as the average value of standard deviation for the F2 of central vowels.

As far as the effect of preceding consonant place is considered, central vowels seem to fall into three categories. Vowels preceded by labials and alveolars have significantly lower F2 values than vowels preceded by dentals. Vowels preceded by velars are not significantly different from vowels preceded by dentals. Vowels preceded by consonants with a palatal component have F2 values significantly higher than vowels preceded by dentals.

In absolute terms, the F2-raising effects of preceding alveopalatals, palatals and palatalized consonants (*vis-à-vis* dentals) on central vowels are slightly smaller than those observed for front vowels. The greatest preceding consonant F2-lowering effect (*vis-à-vis* dentals) on central vowels, that of labials, is much greater than the greatest effect observed on front vowels, that of alveolars (117 Hz vs. 13 Hz). As a result, with dentals as the reference category, the variability of F2 in central vowels that can be attributed to preceding consonants is 367 Hz (from -117 Hz for labials to +250 Hz for palatals). The same variability for front vowels is just 283 Hz (from -13 Hz for alveolars to +270 for palatalized consonants). This is despite the fact that the overall variability in central vowels is lower than in front vowels (standard deviations of 183 Hz and 229 Hz, respectively).

In our investigation of interaction effects involving preceding consonant place further on in this chapter, we will refer to labials and alveolars as *backing* consonants, to velars and dentals as *neutral* consonants and to palatals, alveopalatals and palatalized consonants as

fronting consonants. In this case also these are merely names of convenience, reflecting the *relative* properties of the consonants. That is to say, we are not claiming that consonants in the fronting group have a fronting influence vis-à-vis the idealized target F2 frequency of the vowel, those in the backing group have a backing influence, and those in the neutral group have little or no influence. While this is not unlikely, it cannot be confirmed within our current paradigm.

6.3.3.3 Following segment place

Figure 6.3.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

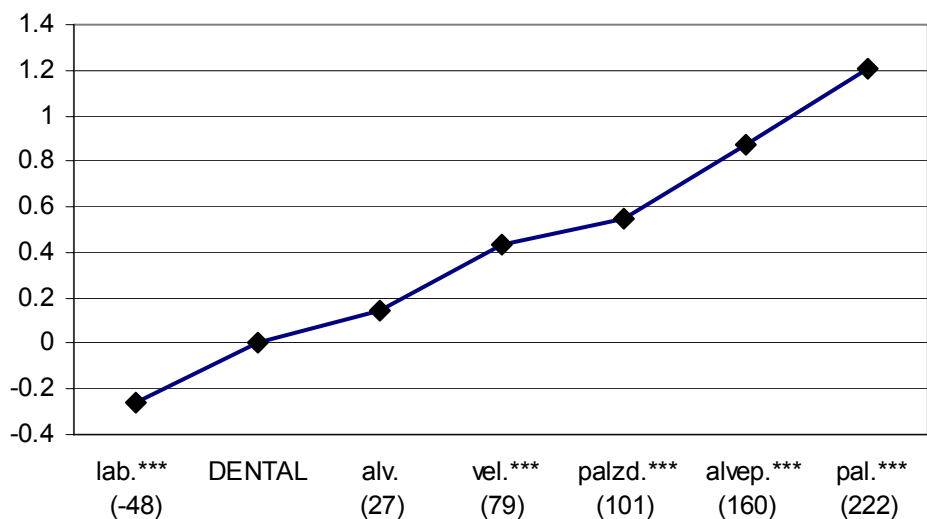


Figure 6.3.3.3.1 MLR-estimated effect of the individual levels of the variable following segment place on F2 of central vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 183 Hz as the average value of standard deviation for the F2 of central vowels.

As was the case with front vowels, the effect of following velars on central vowels differs from the effect of preceding velars in that they condition F2 values significantly higher than dentals. This effect raises the same difficulties with interpretation (cf. section 6.2.3.3). In contrast to front vowels, following labials condition significantly lower F2 values than following dentals. Central vowels followed by alveolars do not differ significantly from central vowels followed by dentals, unlike the situation observed for preceding alveolars.

With dentals as the reference category, the variability of central vowels due to differences in following consonant place can be estimated at 260 Hz (from -48 Hz for labials to +222

for palatals). For front vowels, the same variability could be estimated at 395 Hz. Thus the F2 variability related to following consonant place differences is greater in front vowels than in central vowels, which is the opposite of the situation observed for preceding consonant place, as discussed in the previous section. It should be noted, however, that this is not the case if palatals, which have a particularly strong effect on front vowels, are excluded. In this case, the values are 142 Hz for front vowels and 208 Hz for central vowels.

In our investigation of interaction effects involving following consonant place further on in this chapter, we will refer to labials as *backing* consonants, to alveolars and dentals as *neutral* consonants and to velars, palatals, alveopalatals and palatalized consonants as *fronting* consonants. As was explained in the preceding sections, these names are used for convenience rather than to reflect specific beliefs about the true impact of these consonant with regard to their idealized target F2 values.

6.3.3.4 Vowel duration

The regression coefficient for the variable vowel duration for central vowels was estimated at -0.002, significant at $p < 0.01$. With the average standard deviation of F2 values of central vowels estimated at 183 Hz, this value indicates that for each additional millisecond in vowel duration, the F2 value drops by 0.37 Hz.

As was argued in the introduction to this chapter, it is likely that when F2 is considered, some consonants have a lowering effect and others a raising effect. This is particularly likely to be the case with central vowels, since their F2 values are neither extremely high nor extremely low. This means that a single duration coefficient is unlikely to fully capture the impact of duration. It is likely to partially reflect the role of duration in the contexts in which the impact of consonants is the strongest. The coefficient for duration has a negative sign, indicating that in general central vowels become more back in the vowel space as their duration increases. This is not surprising given the fact that the strongest consonantal influence can be expected to be fronting induced by consonants with a palatal component. However, a better understanding of the role of duration requires the consideration of interactions between the place variables and vowel duration.

Both the interactions between preceding consonant place and duration and between following consonant place and duration significantly improve model fit (increases in R^2 by 0.011 and 0.248, respectively, $p < 0.001$ for both). The latter interaction also improves prediction (increase in “ R^2 ” by 0.0054). The former results in a very small decrease in prediction accuracy (drop in “ R^2 ” by 0.0018). We will consider them both in detail, noting that the interaction between preceding consonant place and duration has to be treated with caution.

Table 6.3.3.4.1 presents the regression coefficients for the variable duration calculated separately for six models (vowels preceded/followed by fronting/neutral/backing

consonants; for the definition of fronting, neutral and backing see sections 6.3.3.2 and 6.3.3.3). The models were calculated using all the predictors included in the validated model plus duration.

Condition	fronting	neutral	backing
preceding consonant	-0.014 (-2.5 Hz/ms)	-0.0018 (-0.31 Hz/ms)	0.00035 (0.064 Hz/ms)
following consonant	-0.008 (-1.45 Hz/ms)	-0.0002 (-0.037 Hz./ms)	0.0031 (0.57 Hz/ms)

Table 6.3.3.4.1 Regression coefficients for the variable duration in different consonantal contexts, F2 of central vowels. Fronting, backing and neutral consonants for the preceding and the following segment condition are defined in sections 6.3.3.2 and 6.3.3.3, respectively. The values are in standard deviations and can be interpreted as change in standardized F2 value per each additional millisecond of vowel duration. The values in brackets were estimated using 183 Hz as the average value of standard deviation of F2 for central vowels.

For each group of consonants, the sign of the regression coefficient is the same for preceding consonants and for following consonants. The effect of duration is clearly the greatest in the context of fronting consonants. Increased duration results in lowered F2 values of vowels; from the perspective of the undershoot model, this indicates that fronting consonants do in fact have a fronting influence on the formant frequencies of central vowels. Since the sign of regression coefficients is also negative for the neutral consonants, we can conclude that they in fact also have a fronting influence on the F2 of central vowels. (Of course, in this case it is a less strong fronting influence, as can be judged from the regression coefficients obtained for different places of articulation, reported in the preceding sections). When the backing consonants are considered, the situation is different. Both regression coefficients have a positive sign, indicating that as the vowel gets longer, its F2 formant frequency increases. This suggests that the backing

consonants do in fact have a backing influence on the target F2 frequencies of central vowels.

In other words, the duration data discussed here enable us to conclude that the target F2 frequencies of central vowels are lower than those observed in the context of fronting and neutral consonants, but higher than those observed in the context of backing consonants. Central vowels do appear to exhibit undershoot effects along the lines suggested by Lindblom (1963).

6.3.3.5 Postpausal phonological word

The F2 of central vowels was found to be raised in postpausal phonological words, by 0.17 sd (~ 34 Hz). The effect was significant at $p < 0.001$. It is unclear what the source of this effect is. In chapter 5, we observed that both front and mid vowels had higher F1 values in postpausal phonological words (cf. sections 5.3.3.12 and 5.4.3.12). For the first formant, it could be interpreted as less reduction than in other contexts, for unknown reasons. For the second formant, such a determination cannot be made without considering differences in consonantal contexts. The interaction between the postpausal variable and following segment place did not result in a significant improvement in model fit. The interaction between the postpausal variable and preceding consonant place significantly improved model fit (increase in R^2 by 0.009, $p < 0.001$), but it resulted in a

decrease in prediction accuracy (drop in “R²” by 0.007). We nonetheless informally investigated the latter interaction in detail.

We constructed separate models for vowels preceded by fronting, neutral and backing consonants, using all the predictors included in the validated model plus the postpausal variable. Central vowels preceded by fronting consonants in postpausal phonological words were estimated to have *lower* F2 values than such vowels in non-prepausal phonological words, by approximately 5 Hz. Central vowels preceded by backing consonants in postpausal phonological words were estimated to have *higher* F2 values than such vowels in non-prepausal phonological words, by approximately 14 Hz. Since these effects point in the opposite direction from the general effect of these consonants, they could indicate that there is less vowel reduction in postpausal phonological words. However, central vowels preceded by neutral consonants in postpausal phonological words were estimated to have *higher* F2 values than such vowels in non-postpausal phonological words, by approximately 83 Hz. Since in the preceding section we concluded that the impact of the neutral consonants vis-à-vis the target F2 frequency of central vowels is likely to be fronting, this effect cannot indicate that reduction is more limited in postpausal phonological words.

6.3.3.6 Word content/function

Vowels in other languages have been found to be more reduced in function words (e.g. van Bergem 1993 for Dutch). We have found evidence for this effect in our investigation of F1 of mid vowels (cf. section 5.3.3.10). For central vowels, F2 values were found to be higher in function words, by 0.13 sd (~ 25 Hz). To judge whether this can indeed be understood as illustrating the expected increased reduction in function words, we need to consider the way in which this variable interacts with the place variables.

In fact, both the interaction of the word content/function distinction with preceding consonant place and with following consonant place significantly increased model fit (increases in R^2 by 0.0022 and 0.0011, respectively, $p < 0.001$ for both) and prediction accuracy (increases in “ R^2 ” by 0.003 and 0.005, respectively). Table 6.3.3.6.1 shows the difference between content and function words in each specific consonantal context.

Condition	fronting	neutral	backing
preceding consonant	1.37 (250 Hz)	-0.017 (-3 Hz)	0.00035 (0.064 Hz)
following consonant	0.3 (55 Hz)	0.27 (50 Hz)	-0.2 (-35 Hz)

Table 6.3.3.6.1 Differences between F2 values of central vowels in function and content words in different consonantal contexts. Different classes of consonants are defined in sections 6.3.3.2 and 6.3.3.3.

In the case of following consonants, the results are in full agreement with the predictions that more reduction should be observed in function words. F2 values are higher in function words before the fronting consonants and the neutral consonants, and in section 6.3.3.4 we concluded that all of these consonants have a F2-raising effect vis-à-vis the

target frequency. Conversely, central vowels have lower F2 values in function words when followed by backing consonants, as expected. We also observe strong F2-raising in central vowels preceded by fronting consonants. However, for the preceding neutral and backing context, the directionality of the effect does not agree with our predictions. It has to be noted, however, that these unexpected effects are very small in size.

It seems possible to conclude that, by and large, we observe greater reduction of central vowels in the F2 domain in function words than in content words.

6.3.3.7 Other variables potentially interacting with place variables

Predictors reflecting differences in style, syllable position, word frequency and the syllable content/function distinction were not selected into the validated model. However, as was argued in the introduction to this chapter, their relative importance could be obscured by their differential effects in different consonantal contexts. For this reason, we have explored the potential interactions between these variables and the place variables. The results are summarized in table 6.3.3.7.1. As can be seen, only the interaction between following segment place and the syllable content/function distinction fulfilled our criteria, i.e. it improved significantly model fit and improved the predictive power of the model.

Predictor	preceding segment place	following segment place
style	0.006, n/s	0.008, p<0.1, -0.008
stress	0.006, p<0.1, -0.005	0.015, p<0.001, -0.012
syllable content/function	0.01, p<0.001, -0.005	0.01, p<0.001, 0.019
word frequency	0.021, p<0.001, -0.05	0.006, p<0.05, -0.012

Table 6.3.3.7.1 Exploration of interactions between the place variables (columns) and other selected variables (rows) for F2 of central vowels. Each cell shows in turn change in model fit (i.e. increase in R^2 of the training set) resulting from the addition of the interaction term, significance of the R^2 increase and change in the predictive power of the model (change in “ R^2 ” of the validation set). Changes in predictive power were not calculated for interaction that did not significantly improve model fit. Cells are in bold if a given interaction both significantly improves model fit and improves prediction. The variable stress is the simplified version of the variable syllable position, cf. section 5.3.3.7.

We calculated separate regression models for central vowels followed by the fronting, neutral and backing consonants, using all the variables in the validated model and the syllable content/function variables. The resulting regression coefficients revealed that vowels followed by the neutral consonants were estimated to have F2 values lower by approximately 24 Hz in function syllables, while vowels followed by the fronting consonants and the backing consonants were all estimated to have F2 values higher by approximately 21 Hz in function syllables. Thus with the exception of the fronting context, it is not possible to assert that more reduction is observed in function syllables.

This result is different from that obtained for the content/function distinction on the word level, as reported in section 6.3.3.6. Of course, there is a significant overlap between these two variables and since the word content/function variable is present in the validated model, the interaction between these variables could be obscuring the effect of this distinction on the syllable level. We therefore recalculated the regression coefficients for the syllable content/function status having dropped the word content/function distinction from the model. When this is done, vowels in function syllables in all

consonantal contexts are predicted to have higher F2 values, by approximately 41 Hz before fronting consonants, 5 Hz before neutral consonants and 6 Hz before backing consonants. Except for the last context, these results are more in line with the results obtained for the word content/function distinction in section 6.3.3.6, and thus with the undershoot model of vowel reduction. However, the observed effects are smaller in size. This difference must be caused by vowels that appear in function syllables, but in content words, i.e. in inflectional endings of content words. It seems that they behave more like vowels in stem syllables in content words, rather than like vowels in function syllables, with the possible exception of the fronting contexts.

6.3.3.8 Preceding segment manner

Figure 6.3.3.8.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

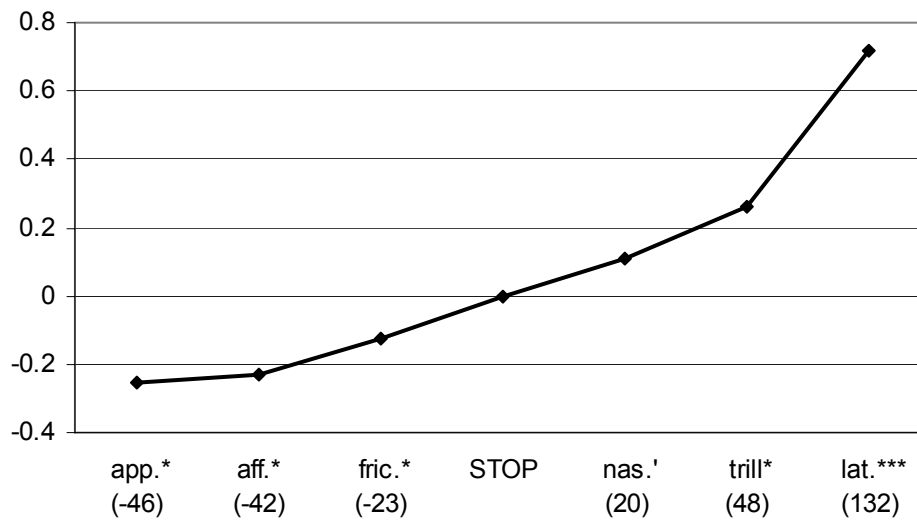


Figure 6.3.3.8.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F2 of central vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 183 Hz as the average value of standard deviation for the F2 of central vowels.

As was the case with front vowels, central vowels were found to have significantly lower F2 values when preceded by affricates and fricatives (vis-à-vis stops). F2 of central vowels was found to be significantly lower after approximants (vis-à-vis stops); for front vowels, lowering was also observed, but it was not significant. F2 of central vowels is significantly or nearly significantly higher after nasals, trills and laterals. For front vowels, F2-raising was also observed after trills and nasals. Thus the main difference between the effect of preceding segment manner on front and central vowels appears to lie in the fact that in the former case, there is small non-significant F2-lowering after laterals, while in the latter, there is strong and significant F2-raising in the same context. It should be noted in this context that in Polish the vowel *y* cannot follow laterals, thus

the difference is in fact one between front vowels on the one hand and *a* on the other. It is not clear what the source of this difference is nor, in fact, what the source of the manner differences is in the first place.

Adding to the validated model an interaction term between preceding segment place and manner nearly significantly improved model fit (increase in R^2 by 0.05, $p < 0.1$) but it did not improve prediction (decrease in “ R^2 ” by 0.001).

6.3.3.9 Following segment manner

Figure 6.2.3.9.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

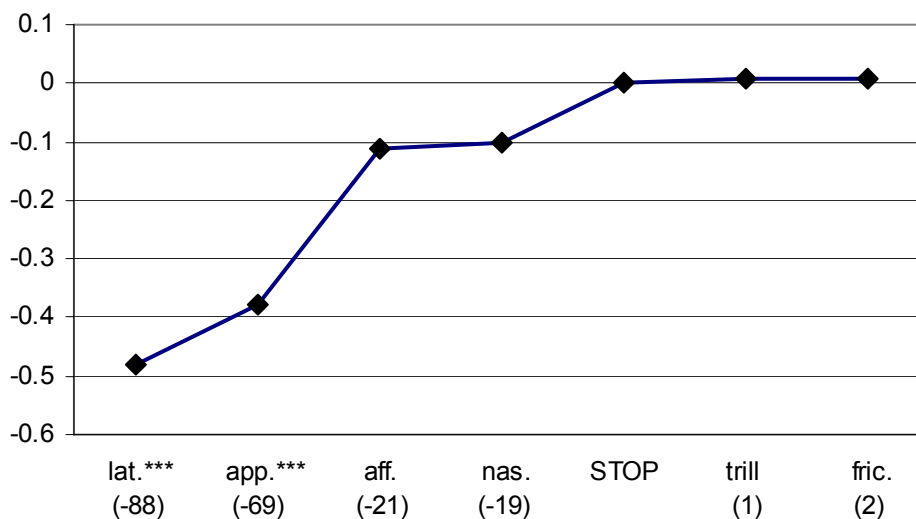


Figure 6.3.3.9.1 MLR-estimated effect of the individual levels of the variable following segment manner on F2 of central vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 183 Hz as the average value of standard deviation for the F2 of central vowels.

There are few differences in F2 values of central vowels conditioned by following segment manner vis-à-vis stops, except for the significant lowering effect of approximants and laterals. The lowering effect was also observed with preceding approximants (as well as with preceding and following approximants in the case of front vowels). The lowering effect of laterals is surprising, since the effect observed with preceding laterals was one of fronting. However, as was mentioned in the preceding section, the effect of preceding laterals on central vowels in fact illustrates their effect on the vowel *a* only, due to phonological restriction. This is not the case with the following segment effect, as the vowel *y* can be followed by laterals in Polish. This could thus be the source of the difference between the effect of preceding and following laterals.

However, this possibility was not confirmed by investigating the interaction between following segment manner and the vowel variable. This interaction significantly improved model fit (increase in R^2 by 0.007, $p < 0.01$) and prediction (increase in “ R^2 ” by 0.007). Figure 6.3.3.9.2 shows the following segment manner regression coefficients calculated separately for the vowels *y* and *a*. As can be seen, there is basically no difference in F2 frequency of the vowel *y* when it is followed by a lateral and when it is followed by a stop (the reference category). The F2 frequency of the vowel *a* is much lower when it is followed by a lateral (vis-à-vis stop), in contrast to the F2-raising effect observed for the preceding lateral. Other differences between *a* and *y* are smaller; in general, *y* appears to be less influenced by variation in following consonant manner, with the exception of approximants and trills.

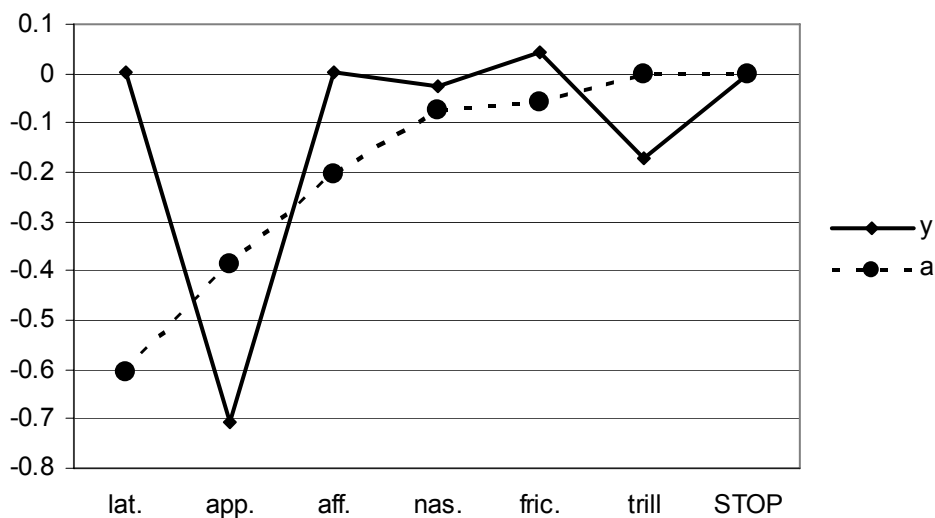


Figure 6.3.3.9.2 Interaction between vowel and following segment manner. The two lines link regression coefficients for following segment manner calculated separately for the two central vowels. All values are standard deviations.

In this case it is also unclear what the source of manner-driven variation in F2 is. The interaction between following segment manner and following segment place does not significantly improve model fit.

6.3.3.10 Vowel

The central vowel was estimated to have standardized F2 0.19 sd. (~35 Hz) higher if it was *y* rather than *a*, significant at $p < 0.001$. Since the data were standardized, this does not indicate that the second formant values of *y* were higher than those of *a*. What the vowel term indicates instead is that the two front vowels interacted with other variables in somewhat different ways.

Both the interaction between preceding segment place and vowel and between following segment place and vowel significantly increased model fit (increases in R^2 by 0.011 and 0.02, respectively, both significant at $p < 0.001$) and improved the predictive power of the model (increases in “ R^2 ” of the validation set by 0.011 and 0.019, respectively). The former interaction is shown in figure 6.3.3.10.1, the latter in figure 6.3.3.10.2.

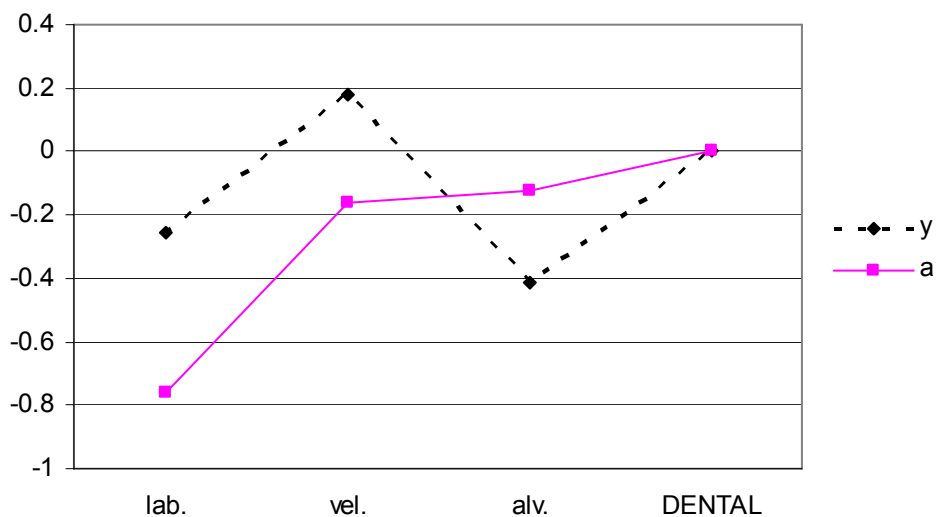


Figure 6.3.3.10.1 Interaction between vowel and preceding segment place, F2 of central vowels. The two lines link regression coefficients calculated separately for the two different front vowels. The vowel *y* cannot be preceded by consonants with palatal component and these places are not included. All values are standard deviations.

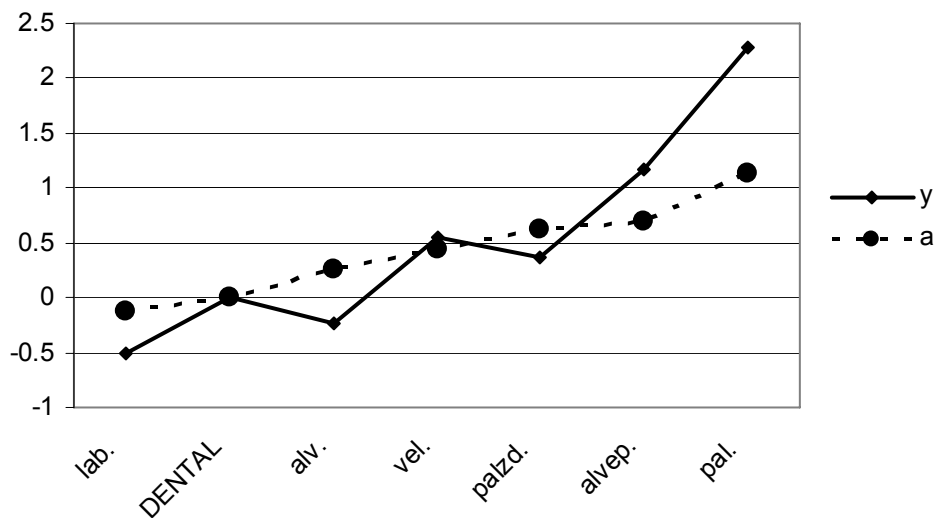


Figure 6.3.3.10.2 Interaction between vowel and following segment place, F2 of central vowels. The two lines link regression coefficients calculated separately for the two different vowels. All values are standard deviations.

Figure 6.3.3.10.1 shows that the two central vowels do not behave uniformly with regard to consonants at the two different places of articulation. In particular, there is much more labial-induced F2 lowering (vis-à-vis dentals) in the case of *a* than in the case of *y*. The opposite is the case for alveolars. Finally, an interesting difference is that the F2 of *y* is higher after velars than after dentals, but the reverse is the case for the vowel *a*.

In figure 6.3.3.10.2 the lines linking regression coefficients for *y* and *a* are largely parallel, indicating that the differences are not very big. The vowel *y* undergoes more lowering (vis-à-vis dentals) when followed by alveolars than the vowel *a*; in the case of labials, this is the reverse of the situation observed for preceding segment place (as described in the previous paragraph). The vowel *y* also undergoes more F2-raising before alveopalatals and palatals. This last difference may in fact largely account for the vowel term in the regression equation. The strongest consonantal effects are the strongly F2-raising effects of consonants with a palatal component. The vowel *y* appears to be more sensitive to this effect, and the positive regression coefficient represents the difference between the fronting estimated for central vowels in general and for *y* in particular.

6.3.4 Summary

The variation of central vowels in the F2 domain appears to be primarily a function of consonantal context, in particular place of articulation. The duration variable emerges as a significant predictor of vowel reduction as well. The investigation of its interactions

with the place variables reveals that its importance may in fact be underestimated due to the non-uniform directionality of consonant place effects. It appears that the mechanism that governs the reduction of central vowels is target undershoot in Lindblom's sense.

On the undershoot analysis, it can be inferred from our results that the target F2 frequencies of central vowels are higher than those observed after labials and alveolars, but lower than those observed after consonants with other places of articulation. On the other hand, they may be inferred to be higher than those observed before labials, but lower than those observed before consonants with other places of articulation.

There is no evidence that stress, speech style, word frequency and the syllable content/function play an important role in the reduction of central vowels. On the other hand, the results for the word content/function distinction suggested that central vowels are more reduced in function words.

The manner variables have a relatively strong impact on F2 frequency of front vowels as well. The results are complex, they differ for preceding and following consonants and they escape a full explanation. Likewise, the small effect of the postpausal variable could not be satisfactorily accounted for.

The vowel term is significant, indicating that there are some differences in the behavior of the two central vowels. In particular, it appears that the vowel *y* is more sensitive to following consonants with a palatal component than the vowel *a*.

6.4 Back vowels

6.4.1 Selecting the model

There were 1501 tokens of the back vowels *o* and *u* in the entire dataset (see section 3.1 for a detailed breakdown). Following the procedures outlined in chapters 2 and 4, they were randomly divided into a training set, comprising 75% of tokens (1126) and a validation set, comprising 25% of tokens (375). The training set was used to select an optimal combination of variables, using a bidirectional automatic search algorithm employing the Akaike Information Criterion. Of the twenty-two independent variables considered, thirteen were selected into the model. They were then used to predict normalized F2 scores for vowel tokens in the validation set and “R²” was calculated. Predictors were then dropped one-by-one in the order that reversed the order in which they were selected by the AIC procedure, and “R²” was recalculated. It was found to be maximized without dropping any further variables from the model. Accordingly, all the thirteen variables were used to construct the validated model of the training set.

Table 6.4.1.1 shows the independent variables considered, listed in their order of inclusion in the AIC-based model as well as “ R^2 ” values resulting from dropping the given variable and all lower-ranked variables from the model, with variables included in the validated model appearing above the thick line. Table 6.4.1.2 compares the statistical data of the validated model of the training set with the predictions for the validation set obtained using the same independent variables and the same regression coefficients. Figures 6.4.1.1 and 6.4.1.2 show the correlations between predicted and obtained values of the dependent variable in the training set and the validation set, respectively, providing a graphical illustration of the strength of the relationship.

Independent variable	AIC rank	“R²”
preceding segment place	1	<i>n/a</i>
following segment place	2	0.3359
vowel duration	3	0.4885
preceding segment manner	4	0.5394
preceding segment voicing	5	0.5539
speech style	6	0.5680
following segment manner	7	0.5745
postpausal phonological word	8	0.5832
word frequency	9	0.5878
following segment voicing	10	0.5921
clause-initial phonological word	11	0.5946
prepausal phonological word	12	0.5964
number of following consonants	13	0.5972
<i>full AIC model</i>	<i>n/a</i>	0.5979
speech rate	<i>not selected</i>	<i>n/a</i>
word content/function	<i>not selected</i>	<i>n/a</i>
phonological word size	<i>not selected</i>	<i>n/a</i>
syllable content/function	<i>not selected</i>	<i>n/a</i>
subject	<i>not selected</i>	<i>n/a</i>
comparative intensity	<i>not selected</i>	<i>n/a</i>
vowel	<i>not selected</i>	<i>n/a</i>
clause-final phonological word	<i>not selected</i>	<i>n/a</i>
syllable position	<i>not selected</i>	<i>n/a</i>

Table 6.4.1.1 Variable selection for F2 back vowels model. AIC rank refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). “R²” is the correlation between the predicted and the obtained scores in the validation set obtained by dropping the independent variable listed in the same line and all independent variables with a lower AIC rank. The maximum “R²” is shown in bold and the bold line separates independent variables included in the validated model (above the line) from the ones that were not included (below the line).

Statistic	R²	“R²”	RSE	“RSE”
Value	0.5979	0.5614	0.6303	0.6443
Difference	-0.037		0.14	

Table 6.4.1.2 F2 in back vowels. Comparison of the key statistics of the validated model of the training set and its predictions for the validation set. R² = squared multiple correlation coefficient of the validated model of the training set; “R²” = squared correlation coefficient between observed values of F1 in the validation set and values predicted by the validated model of the training set. RSE = residual standard error of the validated model of the training set; “RSE” = standard deviation of the differences between values observed in the validation set and values predicted for the validation set by the validated model of the training set. Difference indicates the difference between the statistics for the validation set and the statistics for the training set.

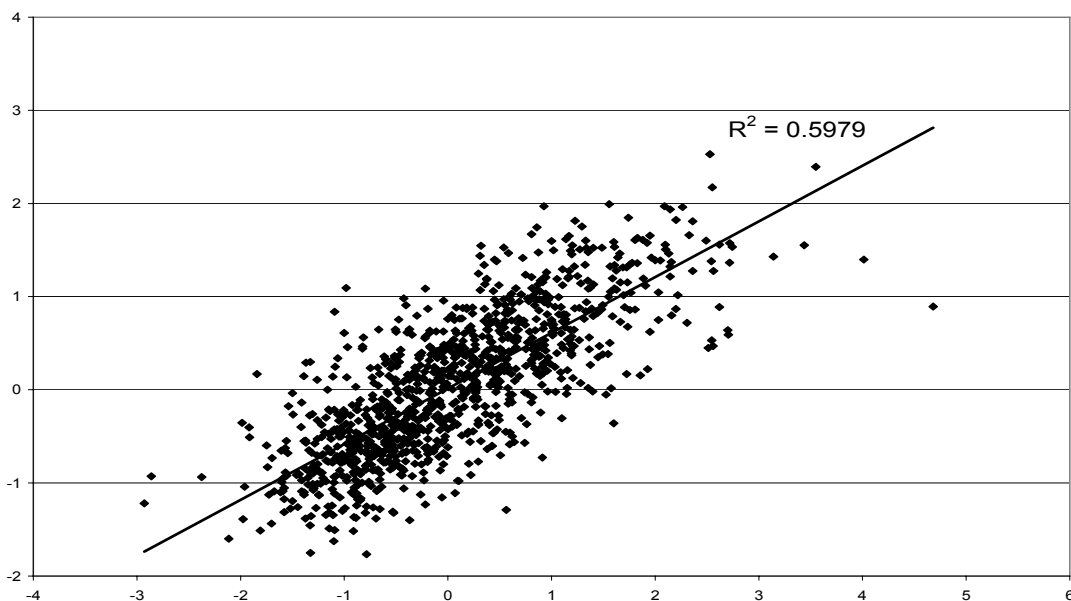


Figure 6.4.1.1 The validated model of F2 of back vowels. Correlation between F2 values observed in the training set (horizontal axis) and values predicted for the training set by the 13 predictors included in the validated model (vertical axis). All units are standard deviations.

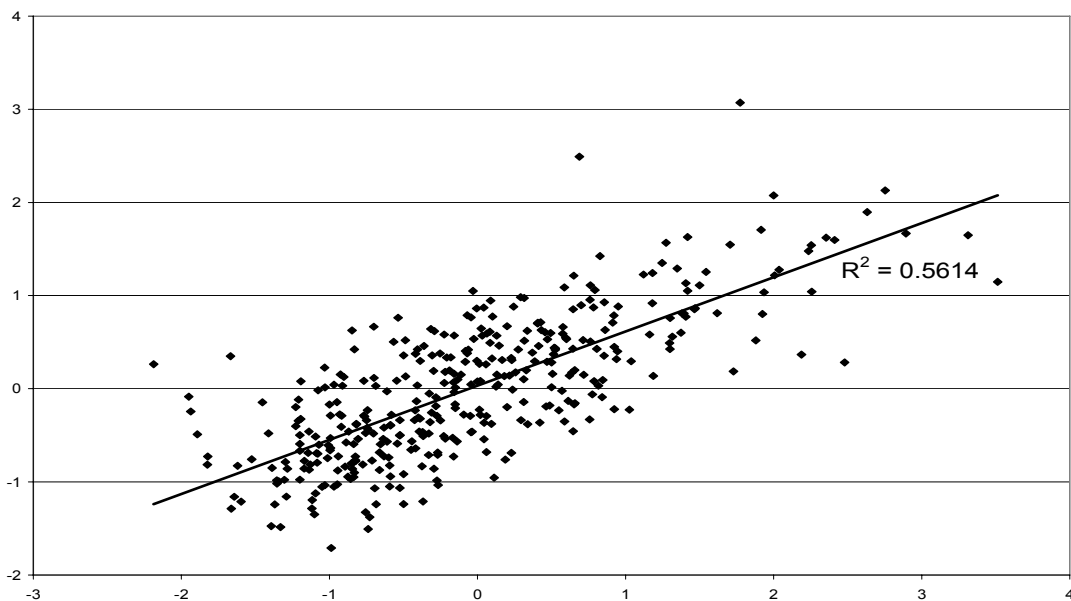


Figure 6.4.1.2 The validated model of F2 of back vowels. Correlation between F2 values observed in the validation set (horizontal axis) and values predicted for the validation set by the 13 predictors included in the validated model (vertical axis). All units are standard deviations.

The validated model is fairly accurate. R^2 values between 0.5614-0.5979 indicate a multiple correlation coefficient R between 0.75-0.77, i.e. a fairly strong linear relationship between the dependent variable and the predictors. As far as model validation is concerned, both the R^2 and the “ R^2 ” values and the RSE and “RSE” values are similar.

It should be noted that unlike the situation with front and central vowels, in this case the variable vowel was not selected into the AIC model.

6.4.2 Categorical regression

The thirteen predictors that were included in the validated model were used to construct a categorical regression (CR) model of the training set, so that the relative contribution of different variables could be meaningfully assessed by comparing regression coefficients computed on a uniform standardized scale. The R^2 of the categorical regression model was 0.613, indicating correlation coefficient $R=0.78$. The key statistics were thus very similar to the values obtained by MLR. Table 6.4.2.1 summarizes the results, providing also information about the usefulness of each variable calculated for the regular multiple regression validated model, as discussed in the preceding section. All values reported are absolute, i.e. ignoring the sign.

Variable	AIC	usefulness	b	0	partial	part
preceding segment place	1	0.2025	0.638	0.593	0.685	0.586
following segment place	2	0.1276	0.381	0.357	0.508	0.367
vowel duration	3	0.0396	0.242	0.169	0.338	0.224
preceding segment voicing	6	0.0107	0.152	0.224	0.207	0.132
preceding segment manner	4	0.0148	0.149	0.101	0.196	0.125
following segment manner	5	0.0125	0.134	0.053	0.195	0.124
speech style	7	0.0062	0.083	0.108	0.129	0.081
following segment voicing	10	0.0023	0.068	0.021	0.102	0.064
clause-initial PW	11	0.0017	0.051	0.054	0.072	0.045
postpausal PW	9	0.0026	0.047	0.065	0.067	0.042
word frequency	8	0.0029	0.043	0.127	0.060	0.038
number of following consonants	13	0.0007	0.040	0.164	0.055	0.034
prepausal PW	12	0.0008	0.035	0.102	0.054	0.034

Table 6.4.2.1 Comparisons of MLR and categorical regression results, validated F2 back vowel models of the training set. *AIC rank* refers to the order of inclusion of the variables in the initial model of the training set using the stepwise AIC algorithm (the AIC model). Usefulness refers to the drop in R^2 of the MLR validated model when a given variable is dropped. *b* is the categorical regression regression coefficient, *0* is the categorical regression zero-order correlation coefficient, *partial* is the categorical regression partial correlation coefficient and *part* is the categorical regression part (semipartial) coefficient. All values are absolute.

Figure 6.4.2.1 compares the usefulness of predictors, as estimated by MLR, with their squared semipartial coefficients, as estimated by categorical regression. The relative values of these two indicators are virtually identical, which justifies treating MLR and categorical regression as equivalent.

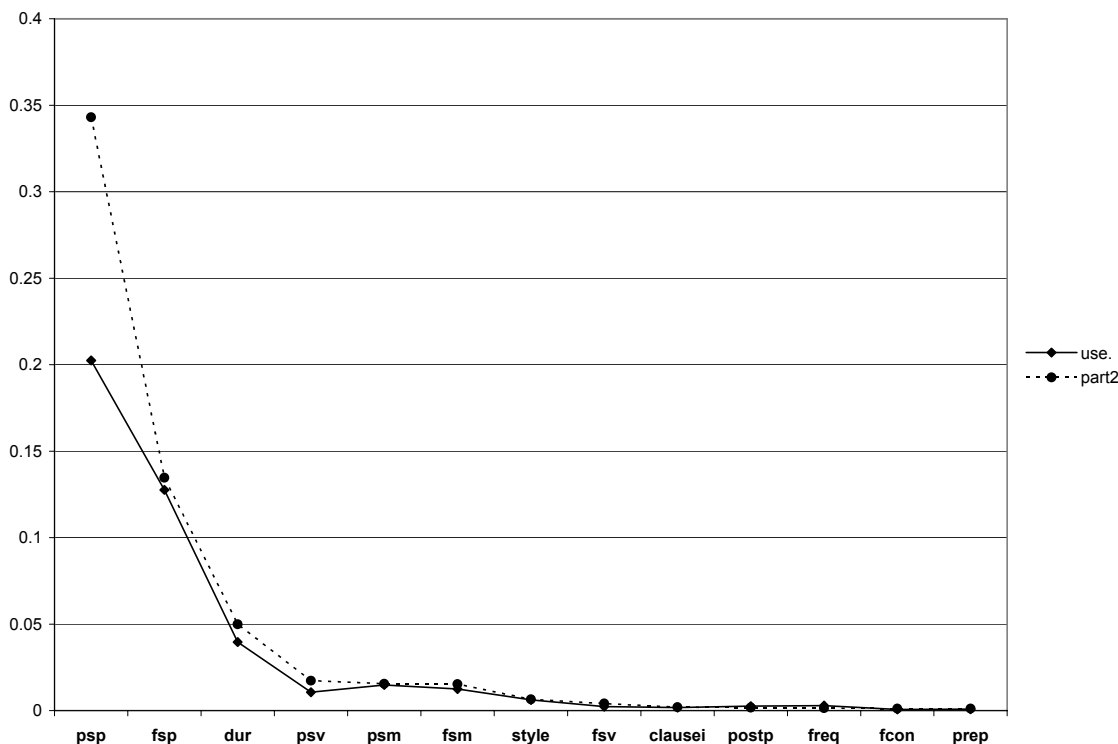


Figure 6.4.2.1 F2 of back vowels. Comparison of the MLR validated model of the training set (variable usefulness) with the categorical regression model of the training set (squared part correlation, “part2”) built with the same independent variables.

Figure 6.4.2.2 shows graphically the relative contribution of independent variables to F2 of central vowels estimated by the absolute values of the CR regression coefficients. As was the case with front and central vowels, the place variables emerge as the strongest predictors, with the preceding consonant playing a greater role than the following one. In the case of back vowels, duration is ranked very high, in the third place. The manner variables, and, somewhat surprisingly, preceding segment voicing also appear to be fairly strong, while other variables play a lesser role.

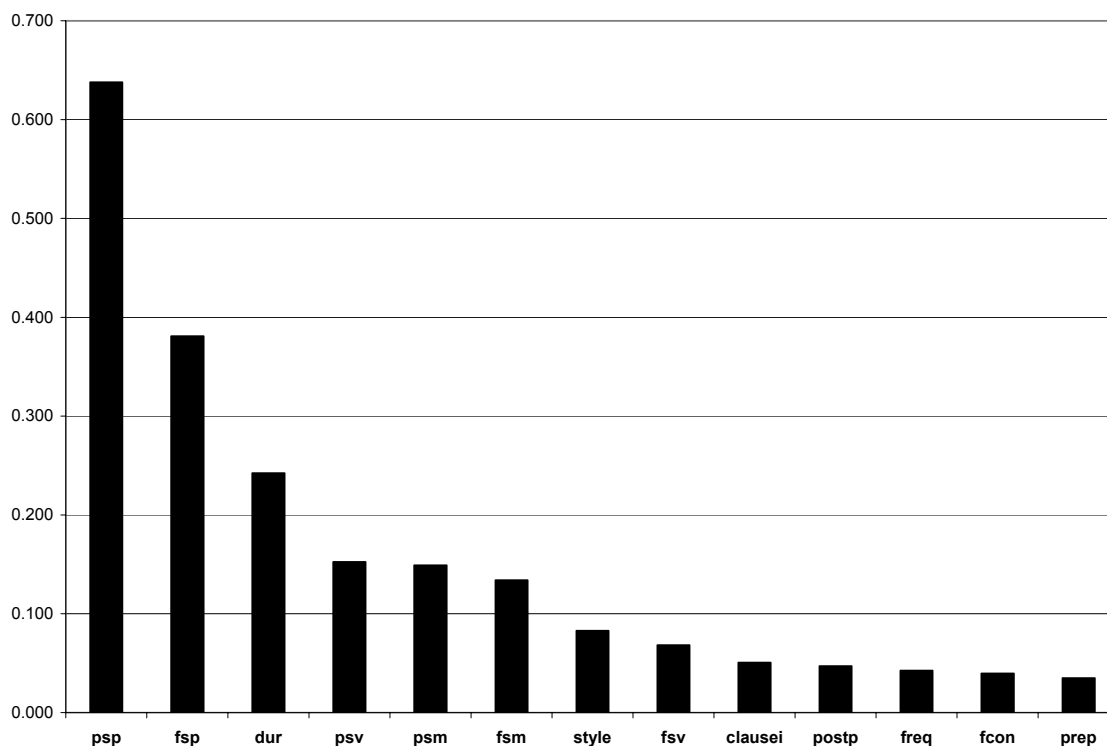


Figure 6.4.2.2 Relative impact of predictors on F2 of back vowels, estimated from the regression coefficients (b-coefficients) of the CR validated model of the training set. The abbreviations used in this table were defined in table 2.9.1 in chapter 2.

6.4.3 Individual predictors (MLR)

6.4.3.1 Intercept

The intercept of the MLR equation was 0.8. In other words, a back vowel was predicted to have an F2 higher by 0.8 standard deviations (~177 Hz) than its mean value when the contribution of all variables was equal to 0. As was explained in section 4.3.1, this is a

purely mathematical construct with no inherent meaning, since some of the variables cannot have the value of 0 (in this case, duration).

6.4.3.2 Preceding segment place

Figure 6.4.3.2.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

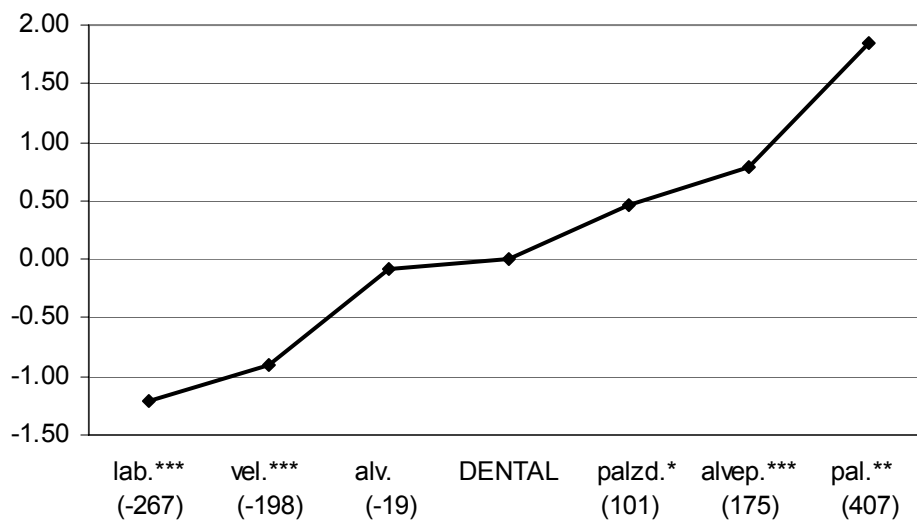


Figure 6.4.3.2.1 MLR-estimated effect of the individual levels of the variable preceding segment place on F2 of back vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 221 Hz as the average value of standard deviation for the F2 of back vowels.

For back vowels, preceding labial and velar consonants pattern together, conditioning significantly lower F2 values than preceding dentals. This is different from the situation observed both for front vowels (where no consonants condition F2 values significantly lower than dentals) and for back vowels (where F2 values significantly lower vis-à-vis dentals were observed for preceding labials and alveolars). In the case of back vowels, alveolars pattern with dentals. Consonants with a palatal component predictably condition significantly higher F2 values than do dentals.

It is worth noticing that in absolute terms, the amount of F2 variation that can be attributed to variation in preceding consonant place is very large, i.e. 674 Hz (from -267 for labials to 407 for palatals). This is much more than the variation observed for front and central vowels (283 Hz and 367 Hz, respectively, cf. section 6.3.3.2), despite the fact that the standard deviations of F2 values are comparable for front, central and back vowels (229 Hz, 183 Hz and 221 Hz, respectively).

In our investigation of interaction effects involving preceding consonant place further on in this chapter, we will refer to labials and velars as *backing* consonants, to alveolars and dentals as *neutral* consonants and to palatals, alveopalatals and palatalized consonants as *fronting* consonants. In this case also these are merely names of convenience, reflecting the *relative* properties of the consonants. That is to say, we are not claiming that consonants in the fronting group have a fronting influence vis-à-vis the idealized target F2 frequency of the vowel, those in the backing group have a backing influence, and

those in the neutral group have little or no influence. In fact, given the fact that back vowels can be expected to have inherently low F2 values, it is likely that the low values observed in the context of labial consonants are close to the target frequencies of the back vowels. It is therefore likely that consonants of all or almost all places of articulation result in F2-raising vis-à-vis the target formant frequency of the vowel.

6.4.3.3 Following segment place

Figure 6.4.3.3.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment place. The reference category is shown in capitals (in this case, dentals). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

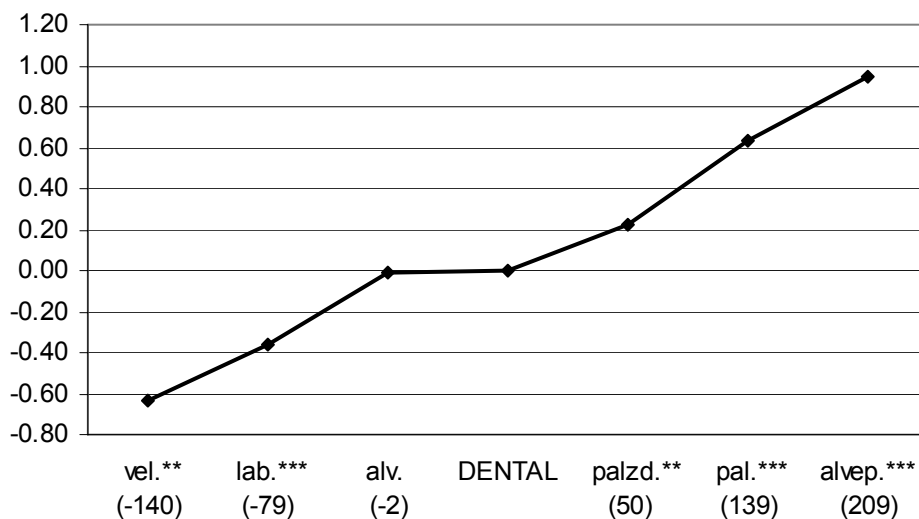


Figure 6.4.3.3.1 MLR-estimated effect of the individual levels of the variable following segment place on F2 of back vowels, in standard deviations, with *dental* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 221 Hz as the average value of standard deviation for the F2 of back vowels.

In relative terms, the way in which following consonants with different places of articulation influence back vowels is basically identical to the way preceding consonants do. The same groups of consonants condition significantly lower F2 values than do dentals (velars and labials) and significantly higher F2 values (all the consonants with a palatal component). The only differences lie in the fact that lower F2 values are observed in the context of velars rather than labials and higher F2 values in the context of alveopalatals rather than palatals, but within the current statistical paradigm it cannot be verified whether these differences are significant.

It should be noted that this consistency of the effect of preceding and following consonants on the second formant is unique to the back vowels. Most importantly, for both front and central vowels following velars differ from preceding velars in that they condition significantly higher F2 values than dentals.

With dentals as the reference category, the variability of central vowels due to differences in following consonant place can be estimated at 349 Hz (from -140 Hz for velars to +209 for alveopalatals). In section 6.3.3.3, we estimated the same variability for front vowels at 395 Hz, and for central vowels at 260 Hz. Thus the following-consonant-induced variability of the second formant does not increase as the vowel becomes more back, which is what was observed for the preceding-consonant-induced variability (cf. the preceding section).

In our investigation of interaction effects involving following consonant place further on in this chapter, we will refer to labials and velars as *backing* consonants, to alveolars and dentals as *neutral* consonants and to palatals, alveopalatals and palatalized consonants as *fronting* consonants. This is the same grouping as the one used for preceding consonants; as was explained in the preceding sections, these labels do not reflect claims about the effect of these consonants vis-à-vis the target frequencies of the back vowels, but rather their relative influence on them.

6.4.3.4 Vowel duration

The regression coefficient for the variable vowel duration for central vowels was estimated at -0.0081, significant at $p < 0.001$. With the average standard deviation of F2 values of central vowels estimated at 183 Hz, this value indicates that for each additional millisecond in vowel duration, the F2 value drops by 1.79 Hz.

As was argued in the introduction to this chapter, it is likely that when F2 is considered, some consonants may have a lowering effect and others may have a raising effect. For back vowels, this complication is less crucial. Given the fact that back vowels are likely to have inherently low F2 values, a great majority of consonantal effects is likely to be fronting (F2-raising). Nonetheless, it is still important to consider interactions between the place variables and duration, as the precise role of duration may be different for different classes of consonants.

In fact, both the interactions between preceding consonant place and duration and between following consonant place and duration significantly improve model fit (increases in R^2 by 0.001, $p < 0.01$ and 0.0116, $p < 0.001$ respectively) and the predictive power of the model (increases in “ R^2 ” by 0.003 and 0.014, respectively). Table 6.4.3.4.1 presents the regression coefficients for the variable duration calculated separately for six models (vowels preceded/followed by fronting/neutral/backing consonants; for the

definition of fronting, neutral and backing see sections 6.4.3.2 and 6.4.3.3). The models were calculated using all the predictors included in the validated model plus duration.

Condition	fronting	neutral	backing
preceding consonant	0.0048 (1.06 Hz)	-0.0077 (-1.7 Hz)	-0.008 (-1.8 Hz)
following consonant	-0.015 (-3.32 Hz)	-0.0086 (-1.9 Hz)	-0.0043 (-0.95 Hz)

Table 6.4.3.4.1 Regression coefficients for the variable duration in different consonantal contexts, F2 of back vowels. fronting, backing and neutral consonants for the preceding and the following segment condition are defined in sections 6.4.3.2 and 6.4.3.3, respectively. The values are in standard deviations and can be interpreted as change in standardized F2 value per each additional millisecond of vowel duration. The values in brackets were estimated using 221 Hz as the average value of standard deviation of F2 for back vowels.

When solely the effect of the following consonants is considered, the data in table 6.4.3.4.1 are fully consistent with the undershoot model, on the assumption that the target F2 frequencies of the back vowels are so low that even the backing consonants (i.e. labials and velars) increase them somewhat. As the vowel gets longer, the F2-raising effect of the following consonant wanes; hence the negative signs of the regression coefficients. This analysis is also applicable to the preceding consonant effect when the backing and neutral consonants are considered, but the data for the fronting consonants contradict it. It should be noted, however, that the last coefficient may not be wholly reliable. Consonants with a palatal component are rarely followed by back vowels in Polish; as a result, the effect of preceding fronting consonants was estimated on the basis of merely 42 data points (vis-à-vis 499 for the neutral consonants and 586 for the backing consonants). In the case of the following consonants, the distribution is much more even (229 vs. 459 vs. 439). In fact, when the individual submodels are considered, the regression coefficient for duration of vowels preceded by fronting consonants is the only

one that is not statistically significant ($p=0.6$). It seems appropriate to say, therefore, that the back vowel data support the existence of the undershoot mechanism in the F2 domain.

6.4.3.5 Word frequency

The MLR validated model estimated the regression coefficient for the continuous variable word frequency at 0.0000013. This indicates an increase of approximately 29 Hz per increase in word frequency of 100,000. The frequency of words in the training set containing back vowels varied between 1-265,857, which means that in the most frequent word in the dataset, F2 frequency was approximately 74 Hz higher than in the least frequent word. Given the inherently low F2 values of back vowels, this can be interpreted as indicating increased reduction in more frequent words. This is not unexpected in the light of previous findings (e.g. Jurafsky et al. 2001).

This clear picture is somewhat complicated when the interaction between frequency and preceding segment place is considered. This interaction significantly increased model fit (increase in R^2 by 0.0043, $p<0.05$) and its predictive power (increase in “ R^2 ” by 0.007). Regression coefficients calculated separately for the three different types of consonants were positive in the case of the fronting and the neutral consonants, but negative in the case of the backing consonants. The last result suggests that back vowels are more back after labials and velars in more frequent words. This is unexpected given the fact that, as

shown in the preceding section, back vowels become more back as their duration increases regardless of the consonantal context.

The interaction between following segment place and word frequency significantly increased model fit (increase in R^2 by 0.0093, $p < 0.05$), but it decreased the predictive power of the model (drop in “ R^2 ” by 0.04).

6.4.3.6 Speech style

Figure 6.4.3.6.1 presents the MLR-estimated regression coefficients for the levels of the variable style. The reference category is shown in capitals (in this case, normal reading style). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

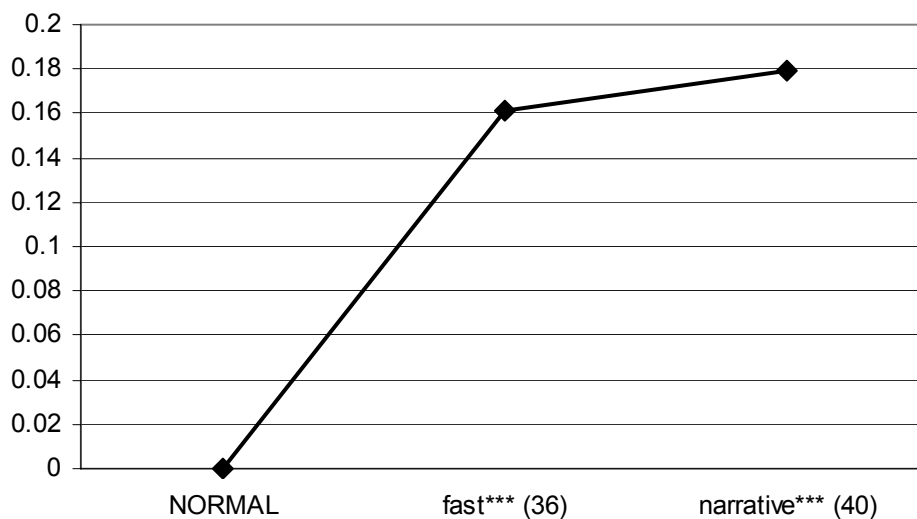


Figure 6.4.3.6.1 MLR-estimated effect of the individual levels of the variable style on F2 of back vowels, in standard deviations, with *normal reading* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$ and '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 221 Hz as the average value of standard deviation for the F2 of back vowels.

Back vowels have higher F2 values in the narrative style and the fast reading style than in the normal reading style. This means that they are more front in the two-dimensional vowel space, which can be interpreted as increased reduction in the less “formal” speech styles. This is consistent with previous findings on other languages (e.g. Koopmans-van Beinum 1980) and our findings on the role of style in F1 variability of mid and low vowels reported in the preceding chapter.

The interaction between preceding segment place and style significantly improved model fit (increase in R^2 by 0.01, $p < 0.01$), but it decreased the predictive power of the model (drop in “ R^2 ” by 0.01). The interaction between following segment place and style

significantly improved model fit (increase in R^2 by 0.012, $p < 0.001$) and it neither increased nor decreased the predictive power of the model. Regression coefficients calculated separately for the three different consonant types are shown in figure 6.4.3.6.2. As can be seen, vowels preceding the backing consonants stand out from the others: there is generally less reduction, and the fast style does not pattern with the narrative style.

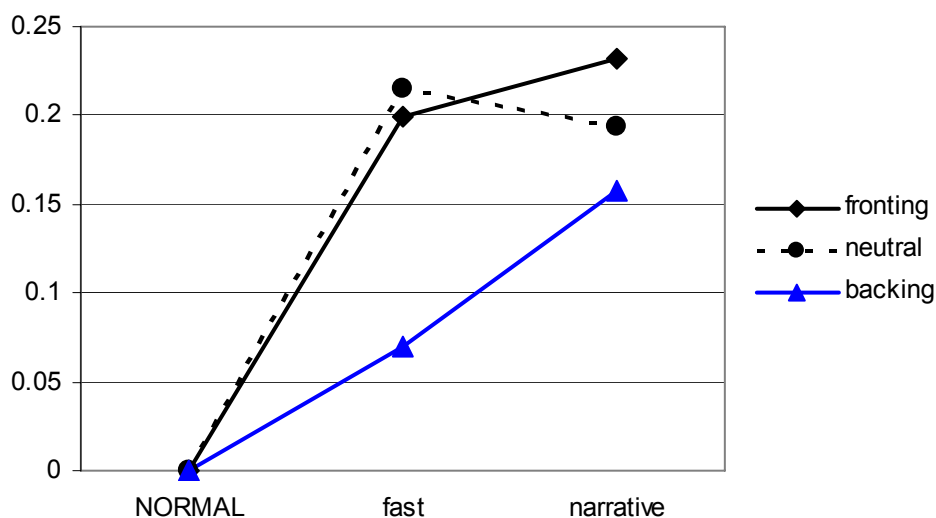


Figure 6.4.3.6.2 Interaction between style and following segment place. The three lines link speech style regression coefficients calculated separately for back vowels preceding consonants of the three different place classes. All values are standard deviations.

6.4.3.7 Preceding segment voicing

Back vowels were found to have F2 value 0.4 sd. (~66 Hz) lower after voiceless consonants. This effect was significant at $p < 0.001$. It can be interpreted as decreased reduction in this context. As was noted in section 4.3.8, there is the possibility that the semi-automatic segmentation algorithm we used underestimated the duration of vowels

after voiceless stops. For this reason, the regression coefficient estimated for this variable may be a correction of this artifact, at least in part. We have previously suggested this explanation for the more limited reduction observed after voiceless consonants in the F1 domain for mid and low vowels, cf. section 5.3.3.8 and 5.4.3.9. In addition, as was suggested in section 5.3.3.8, voiceless segments tend to have greater than average glottal opening, which may induce acoustic effects mimicking those of nasalization, due to the coupling between the vocal tract and the subglottal cavity (John Ohala, personal communication, Ohala and Ohala 1993). As will be discussed in section 6.4.3.14, preceding nasals do indeed have an F2-lowering effect on back vowels.

6.4.3.8 Following segment voicing

Back vowels were found to have F2 values 0.11 sd. lower (~ 25 Hz) before voiceless consonants. This effect was significant at $p < 0.05$. It can be interpreted as more limited reduction before voiceless consonants. A possible reason is the vowel's assimilation of the large glottal opening of voiceless consonants and the ensuing acoustic coupling between the supraglottal cavity and the subglottal cavity (cf. 5.3.3.8 and the preceding section).

6.4.3.9 Prepausal phonological word

Back vowels were found to have lower F2 values in prepausal phonological words, by 0.08 sd. (approximately 18 Hz). This could indicate more limited reduction in this context, which may be due to the fact that domain-final material has more limited information value (cf. section 3.18). The interactions between this variable and the place variables failed to simultaneously significantly increase model fit and the predictive power of the model.

6.4.3.10 Postpausal phonological word

Back vowels were found to have higher F2 values in postpausal phonological words, by 0.16 sd. (~ 36 Hz). This effect was significant at $p < 0.01$, and it can be interpreted as increased reduction in this context. To the extent that pauses coincide with syntactic boundaries, it is possible that back vowels are less reduced in domain-initial contexts, where the lexical material is somewhat more likely to convey known information. It should be noted, however, that in the F1 domain *less* reduction was found for mid and low vowels.

The interaction between this variable and preceding segment place failed to significantly increase model fit. The interaction with following segment place did significantly

increase model fit (increase in R^2 by 0.05, $p < 0.01$), but it decreased the predictive power of the model (drop in “ R^2 ” by 0.012).

6.4.3.11 Clause-initial phonological word

Back vowels were found to have higher F2 values in clause-initial phonological words, by 0.11 sd. (~23 Hz). This effect was significant at $p < 0.01$ and it can be interpreted as indicating increased reduction in this context. As was argued in section 3.19, it could reflect greater predictability of lexical material in domain-initial contexts. It should be noted, however, that a similar effect was not observed for any other types of vowels. It was not found to significantly interact with the place variables (failure to significantly increase model fit).

6.4.3.12 Number of following consonants

The MLR validated model estimated the regression coefficient for the ordinal variable number of following consonants at -0.05. Although this variable passed the AIC search and validation, the regression coefficient was not significantly different from 0 and has to be treated with caution. Between 0-3 consonants followed back vowels in the same word before another vowel. This indicates that the contribution of this variable varied between subtracting nothing ($-0.05 * 0$) from the intercept and subtracting 0.15 sd. ($-0.05 * 3$) from it. In other words, vowels followed by no consonants had F1 0.15 sd. (approximately 33

Hz) higher than vowels followed by three consonants. Thus the F1 value was lowered by approximately 11 Hz with each additional following consonant. To the extent that the number of following consonants reflects syllable structure, this indicates more reduction in open syllables, which is rather unexpected. (Note that for the low vowel, less reduction was observed in more open syllables, cf. section 5.4.3.14).

The interaction between this variable and preceding segment place significantly increased model fit (increase in R^2 by 0.07, $p < 0.01$), and the predictive power of the model (increase in “ R^2 ” by 0.07). The regression coefficients calculated separately for the three consonant classes indicate that back vowels have higher F2 values in more closed syllables after the fronting consonants (increase of approximately 30 Hz for each additional consonant) and lower F2 values in more closed syllables after the backing consonants (drop of approximately 7 Hz for each additional consonant), with no change after the neutral consonants. These results are thus inconsistent and do not shed further light on the role played by this predictor.

The interaction between the number of following consonants and following segment place did significantly increase model fit (increase in R^2 by 0.013, $p < 0.001$), but it decreased the predictive power of the model (drop in “ R^2 ” by 0.015).

6.4.3.13 Other variables potentially interacting with the place variables

Predictors reflecting differences in syllable position and the content/function status of word or syllable were not selected into the AIC model. We have, however, considered their potential interactions with the place variables. The results are summarized in table 6.4.3.13.1. As can be seen, only the interactions between the word and syllable content/function distinction and preceding segment place fulfilled our criteria, i.e. increased significantly model fit and improved (or did not change) the predictive power of the model.

Predictor	preceding segment place	following segment place
stress	0.0021, n/s	0.016, p<0.001, -0.006
syllable content/function	0.0081, p<0.01, 0	0.01, p<0.001, -0.013
word content/function	0.0056, p<0.01, 0.002	0.007, p<0.01, -0.004

Table 6.4.3.13.1 Exploration of interactions between the place variables (columns) and other selected variables (rows) for F2 of back vowels. Each cell shows in turn change in model fit (i.e. increase in R^2 of the training set) resulting from the addition of the interaction term, significance of the R^2 increase and change in the predictive power of the model (change in “ R^2 ” of the validation set). Changes in predictive power were not calculated for interaction that did not significantly improve model fit. Cells are in bold if a given interaction both significantly improves model fit and improves prediction. The variable stress is the simplified version of the variable syllable position, cf. section 5.3.3.7.

We calculated separate regression models for back vowels preceded by the neutral and the backing consonants, using all the variables in the validated model and the word content/function variable on the one hand, and the syllable content/function variable on the other. (There were no instances of back vowels preceded by fronting consonants in function syllables and function words). In the case of the syllable-based predictor, we found that back vowels have higher F2 values in function syllables both after the neutral

consonants (by approximately 14 Hz) and after the backing consonants (by approximately 22 Hz), indicating that they are more reduced in this context. This is in line with findings concerning the content/function distinction in other languages (cf. for example van Bergem 1993). In the case of the word-based predictor, we found that back vowels had higher F2 values in function words after the backing consonants (by approximately 57 Hz), but lower after the neutral consonants (by approximately 17 Hz). That last finding is unexpected.

6.4.3.14 Preceding segment manner

Figure 6.4.3.14.1 presents the MLR-estimated regression coefficients for the levels of the variable preceding segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

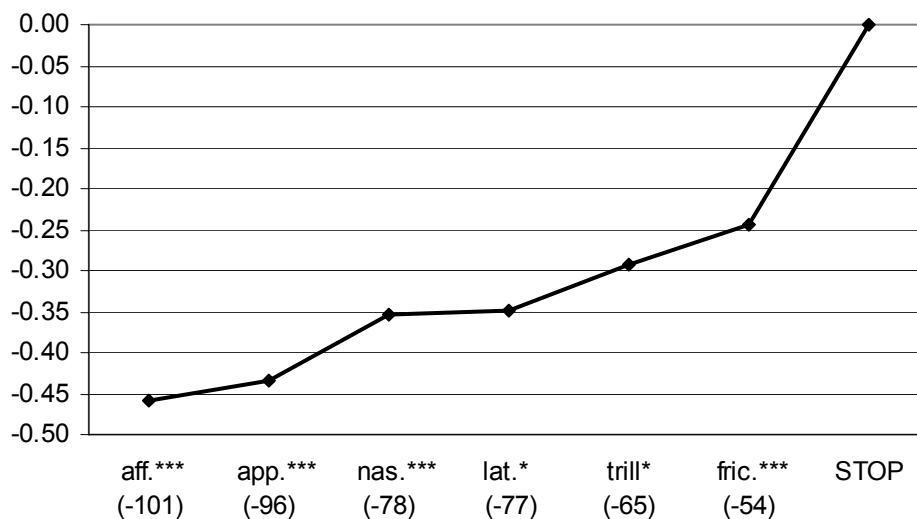


Figure 6.4.3.14.1 MLR-estimated effect of the individual levels of the variable preceding segment manner on F2 of back vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 221 Hz as the average value of standard deviation for the F2 of back vowels.

Back vowels were found to have significantly lower F2 values after consonants of all manners, vis-à-vis stops. This consistency sets this result apart from the more complex findings for front and central vowels. It is not clear what the source of this manner-related variation is.

Adding to the validated model an interaction term between preceding segment place and manner significantly improved model fit (increase in R^2 by 0.07, $p < 0.01$) and it improved prediction (decrease in “ R^2 ” by 0.002). It is impossible to present graphically the manner regression coefficients calculated separately for fronting, neutral and backing consonants, because not all categories were represented in each place group. (In fact, only preceding

stops with all three different place of articulation types could be found). However, the regression coefficients for each dummy variable in each submodel were negative, indicating that the F2-lowering vis-à-vis stops occurs regardless of the place of articulation of the preceding consonant.

6.4.3.15 Following segment manner

Figure 6.4.3.15.1 presents the MLR-estimated regression coefficients for the levels of the variable following segment manner. The reference category is shown in capitals (in this case, stops). Levels that are significantly different from the reference level are marked, and their estimated impact in Hertz is also shown.

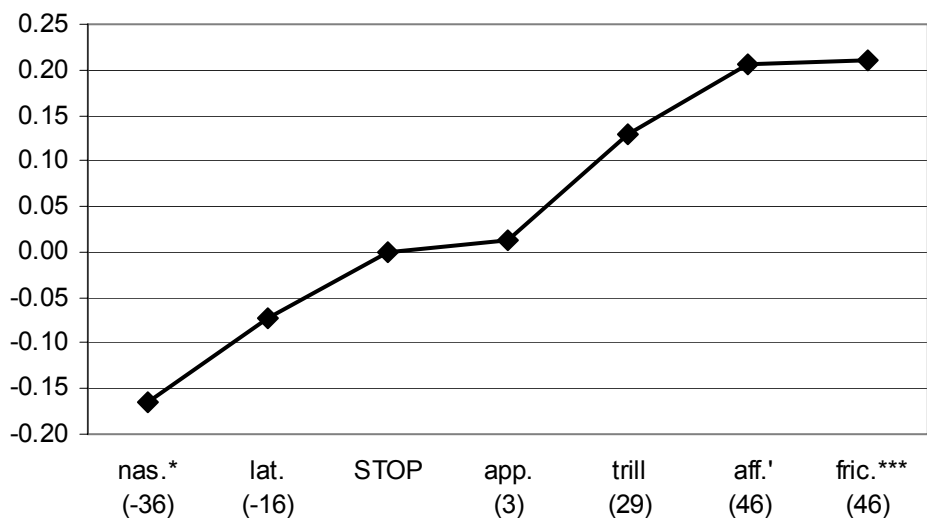


Figure 6.4.3.15.1 MLR-estimated effect of the individual levels of the variable following segment manner on F2 of back vowels, in standard deviations, with *stop* as the reference category (shown in capitals). If applicable, the fact that the given level is statistically significantly different from the reference level is shown ('***' indicates $p < 0.001$, '**' indicates $p < 0.01$, '*' indicates $p < 0.05$ and the apostrophe indicates $p < 0.1$ (almost significant)). The numbers in brackets indicate the approximate effect in Hertz, assuming 221 Hz as the average value of standard deviation for the F2 of back vowels.

Variation related to following consonant manner differs from that attributable to preceding consonant manner. Stops do not condition the highest F2 values, fricatives and affricates do. The only effect of following consonant manner that is consistent with the effect of the preceding consonant with the same manner of articulation is the F2-lowering found in the context of nasals. It is not clear what causes the manner-related differences, nor why there is an asymmetry in the impact of preceding and following consonants.

Adding to the validated model an interaction term between preceding segment place and manner significantly improved model fit (increase in R^2 by 0.011, $p < 0.01$) and it improved prediction (decrease in “ R^2 ” by 0.011). In this case also it is impossible to present graphically the manner regression coefficients calculated separately for fronting, neutral and backing consonants, because not all categories were represented in each place group. (Only following stops, nasals and fricatives with all three different place of articulation types could be found).

6.4.4 Summary

It appears that the variation of the back vowels in the F2 domain can be explained by the undershoot model of vowel reduction. The three key variables – preceding and following consonant place and duration – emerge as the strongest predictors of reduction. The analysis of their interactions is by and large consistent with the undershoot model. On the

undershoot account, it can be inferred that the target F2 frequencies of back vowels are lower than those observed in any consonantal context.

Although the results were not always absolutely clear-cut, it does appear that the back vowels are less reduced in the normal reading speech style, in content syllables and in less frequent words. On the other hand, the results for the content/function distinction in words are unexpected and no evidence for an effect of stress has been found.

Back vowels are also less reduced in prepausal phonological words, non-postpausal phonological words and clause-initial phonological words. There is a possibility that this variation reflects the differing information value of lexical material appearing in different positions. Differences in the scope of reduction have been found for different levels of the voicing variables, and for different syllable shapes (as represented by the variable number of following consonants).

As is the case with front and central vowels, the impact of different manner classes on F2 is complex and escapes a clear account. On the other hand, back vowels are unique in that the vowel term has not been selected into the model, which indicates that their behavior in the F2 domain is uniform regardless of their height.

6.5 General discussion

In this section we will attempt to synthesize the findings reported in this chapter. We will first consider the variables related to Lindblom's (1963) undershoot model of vowel reduction. Next, we will turn to other phonetic and phonological variables and finally to nonphonetic variables.

6.5.1 Consonant place and vowel duration

In the second formant models, vowel duration appears to play a smaller role than in the first formant models. In both cases, it was not selected as an important predictor of vowel reduction for one group of vowels (front and high, respectively). The situation is different for other vowel classes. In the F1 domain, it is ranked third (for mid vowels) and first (for the low vowel). In the F2 domain, it is ranked seventh (for the central vowels) and third (for the back vowels). The results for F1 and F2, however, are not directly comparable in this respect. As was argued above, for the non-high vowels, consonants of various places of articulation have a uniformly F1-lowering effect. This means that within the undershoot model, the role of duration is largely independent of the place of articulation of the consonant: increased duration results in all cases in an increase in F1 frequency. The situation is much more complex in the F2 domain. Because different classes of consonants can have either formant-raising or formant-lowering influences, increased duration may result either in formant lowering or formant raising. This means that a

single regression coefficient for duration is inadequate, and the variable-selection procedures we used are likely to underestimate the true contribution of vowel duration. To understand its role more fully, interaction effects need to be considered.

In fact, when these are taken into consideration, we find that there are good reasons to assume that a Lindblom-style undershoot mechanism is at work. This is particularly clear for central vowels. For back vowels, only the data for preceding fronting consonants were inconsistent with the undershoot model; however, we have argued that this finding is likely to be unreliable, as back vowels in Polish are rarely preceded by the fronting consonants. For the front vowels, only the data for following consonants were consistent with the undershoot model.

At the end of chapter 5, we compared our findings to Lindblom's (1963) original undershoot model, concluding that there is a strong similarity between them. In particular, Lindblom's model predicts no undershoot for high vowels, and then increased contribution of duration (rate of change) the lower the vowel is in the vowel space, i.e. the further apart its target F1 frequency and the consonant F1 locus are. In the case of F2, it is not possible to estimate the durational effect Lindblom's model predicts for Polish vowels, because the formulas require the knowledge of formant frequencies at the onset of the vowel. In fact, as was discussed in section 1.1, in Lindblom's model the size of the duration effect is directly related to the distance between the formant frequency at the vowel onset and the target formant frequency of the vowel. This means that for F2,

Lindblom's model also predicts that the contribution of duration should increase the further the vowel's target frequency is from consonantal locus, to the extent that differences in F2 at vowel onset reflect differences in consonantal loci.

Lindblom only incorporated F2 frequency at vowel onset into his formulas; however, he only considered symmetrical consonantal context. It is likely that in the case of asymmetrical contexts, the distance between the vowel target and the following consonant would play a role as well. In fact, for our findings this appears to clearly be the case for following fronting consonants. (For all vowel classes, these include all the consonants with a palatal component; for front and central vowels the class also includes velar consonants). The regression coefficient for the variable duration for front vowels followed by fronting consonants was -0.0018, for central vowels -0.008 and for back vowels -0.015. These coefficients indicate that front vowels followed by fronting consonants "lose" approximately 0.4 Hz in F2 formant frequency for each additional millisecond of duration, central vowels "lose" approximately 1.45 Hz and back vowels "lose" approximately 3.3 Hz. In other words, the increase in duration that is necessary to counteract the fronting influence of the following fronting consonants is greater the more back the vowel is.

Our data are thus very consistent with the predictions of Lindblom's (1963) undershoot model in the case of the consonants that have the greatest relative impact on the F2 frequencies of the Polish vowels, vis-à-vis other following consonants. This is not the

case, however, for preceding fronting consonants and difficult to assess for other classes of consonants due to differences in the directionality of consonantal effects.

Comparing the consonantal effects observed by Lindblom with those found in our study is more complicated. It has to be limited to the consonants with places of articulation that were used by Lindblom (labials, dentals and velars). An additional complexity is caused by the fact that there is no one-to-one correspondence between the vowels of Polish and those of Swedish. We have decided to use those of the Swedish vowels analyzed by Lindblom that have fairly close analogs in Polish: the front vowels [ɪ] and [ɛ], the central vowel [a] and the back vowels [ɔ] and [ʊ]. We assumed that those can be taken to stand in a rough correspondence with the Polish vowels *i*, *e*, *a*, *o* and *u*, respectively.

To mimic the vowel durations observed in Lindblom's study, we obtained the mean duration values of the vowels appearing in the training sets. These were 48 ms. for *i*, 58 ms. for *e*, 63 ms. for *a*, 56 ms. for *o* and 49 ms. for back. Using these values, we obtained the formant frequency values predicted by Lindblom's equations for the "corresponding" Swedish vowels in the three symmetrical consonantal contexts (labial, dental and velar). We then calculated the differences between values predicted for the labial context and the dental context, and for the velar context and the dental context. Next, we averaged these values for the front, central and back Swedish vowels.

For the Polish vowels, we used the sums of the standardized regression coefficients for preceding and following labials, dentals and alveolars, for the three classes of vowels. These values estimate the duration-independent differences between vowels surrounded by labials and vowels surrounded by dentals, and between vowels surrounded by velars and vowels surrounded by dentals. As such, they are comparable to the values predicted for Swedish by Lindblom's equations, as described in the preceding paragraph. We then converted the standardized regression coefficients to Hertz values, using the average standard deviations of F2 values for the male speakers only (since Lindblom's speaker was a male).

The results of this comparison are presented in figure 6.5.1.1. The absolute sizes of the consonantal effects differ greatly between Polish and Swedish, but in a regular way: Lindblom's equations predict a greater impact of labials and velars (vis-à-vis dentals) for all classes of vowels. This is indicated by the fact that the Swedish data points are always further away from 0 (which represents dentals, the reference category) than the corresponding data points for Polish. The relative effects of labials and velars vis-à-vis dentals are very similar. In both cases, vowels surrounded by labials are predicted to have lower F2 values than vowels surrounded by dentals, and the differences between them increase the more back the vowels are. In both cases, front and central vowels are predicted to have higher F2 values when surrounded by velars than when surrounded by dentals, with small differences between front and central vowels. Finally, in both cases back vowels are predicted to have lower F2 values when surrounded by velars than when

surrounded by dentals. Considering the differences in methodology and language, as well as the fact that both studies are based on small samples of speakers, these similarities are striking.

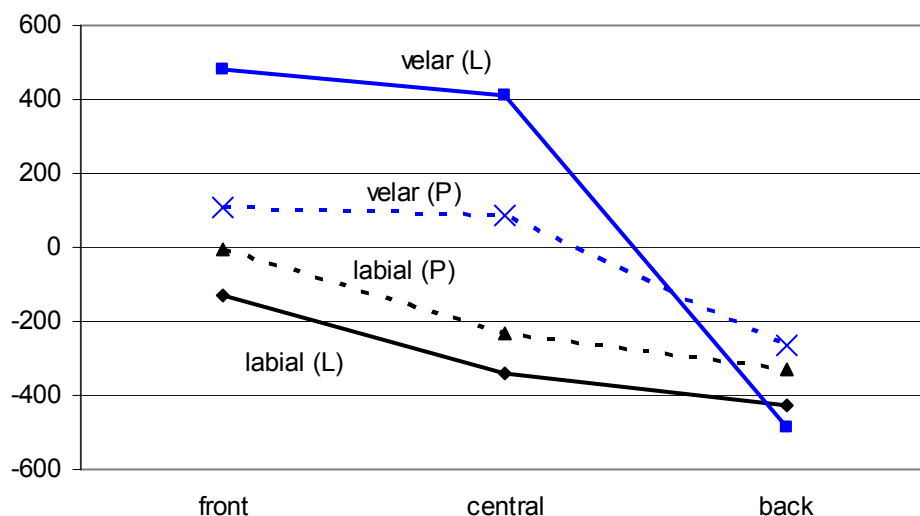


Figure 6.5.1.1 Differences between vowels in symmetrical labial and velar context and in symmetrical dental context predicted for Swedish vowels by Lindblom's (1963) equations (L) and between "corresponding" Polish vowels spoken by the male speakers in the current study (P). Refer to the text for details.

6.5.2 Other phonetic and phonological variables

The **manner** of following and preceding consonants was found to be a fairly strong predictor of vowel reduction. Unlike the situation observed for F1, the effects are complicated, differ strongly for different vowel classes and escape a straightforward explanation.

Speech rate was found to affect the F2 of front vowels, but the source of this effect is unclear. **Syllable position** was not found to determine the degree of reduction of vowels in the F2 domain, although the simplified variable stress did enter into interactions with the place variables for front vowels, suggesting that it does play a (small) role in the degree of reduction. This is, however, very different from the situation observed in the F1 domain, where syllable position was found to be a strong predictor of formant frequency in mid and low vowels.

The **voicing** variables and **number of following consonants** were found to play a small role in the F2 variability of back vowels. Although there is a potential explanation in the former case (cf. the relevant sections), it is striking that none of these effects were found for vowels with other degrees of frontness.

The **vowel** variable was selected into the models for front and central vowels, indicating that there are differences in the behavior of different vowels with similar degree of frontness. There was no evidence for differences in the F2 variability of *o* and *u*.

6.5.3 Other relevant variables

As far as the **content/function** distinction is concerned, the word-based variable was selected into only one model (central vowels) and the syllable based one into none. However, the examination of interactions indicated that in some cases, the expected

increased reduction is found in the functional domains. In other cases, the evidence was unclear or contradictory. Back vowels were also found to be more reduced in more **frequent** words, and front and back vowels were found to be less reduced in the normal reading **speech style**.

The **postpausal** variable was found to play some role in the variability of F2 in central and back vowels, and the **prepausal** variable and the **clause-initial** variable were found to influence back vowels. To a certain extent, these effects were found to be compatible with an explanation related to syntactic structure. However, since higher formant values were found in domain-initial contexts (postpausal phonological words and clause-initial phonological words) and lower F2 values in domain-final contexts (prepausal phonological words), these effects may be also analyzed as reflecting pitch-related differences in the vertical position of the larynx, associated with the phenomenon of pitch declination (cf. section 5.5.3).

6.5.4 Variables not selected into any model

As was mentioned above, the syllable position variable and syllable content/function distinction variable were not selected into any models, but there was some evidence for their role when interactions with the place variables were examined. In contrast to the F1 domain, the second formant of vowels was not found to correlate with **intensity** of the vowel. This is not surprising, for while F1 is likely to be strongly correlated with the

degree of vertical opening of the vocal tract, which in turn is likely to be correlated with intensity, there is no such straightforward relationship for F2. **Phonological word size**, **clause-final phonological word** and the **subject** variable also failed to be selected into any models. The non-significance of the subject variable indicates that the formant-frequency standardization technique used was successful, and that individual speakers did not employ different vowel reduction strategies.

6.5.5 Conclusions

As was the case with F1, Lindblom's (1963) model of target undershoot appears to be applicable to Polish vowels in the F2 domain. While explicit comparisons with Lindblom's model are more difficult due to the varying directionality of consonantal effects, there are indications that, as in Lindblom's model, the role of duration increases as the distance between the vowel target and the loci of the neighboring consonants increases. For the consonant places that could be compared, our observations match fairly well those of Lindblom.

Some effects of such factors as stress, speech style and the content/function distinction were found, but they appear to be far less robust in the F2 domain than in the F1 domain (in the case of non-high vowels).

7 General discussion

7.1 Vowel target undershoot in Polish

Lindblom (1963) introduced the target undershoot model of vowel reduction as an explanation for the differences in formant frequencies of stressed and unstressed vowels that are observed in many languages. His reasoning was based on the observation that vowel reduction is characteristic of languages in which “stress is known to be positively correlated with vowel duration” (517). He also invoked Polish as an example of a language in which little reduction is observed in unstressed syllables. This claim was based on the findings of Jassem (1959), who compared stressed and unstressed vowels in a small number of carefully articulated words. Although Lindblom (1963) does not state this explicitly, it appears that in the context of his model the alleged absence of vowel reduction from Polish must be a consequence of the limited correlation between stress and vowel duration.

In fact, it does appear that the duration of unstressed vowels in Polish differs from the duration of stressed vowels less dramatically than in some other languages. Jassem (1962) estimated the duration ratio of stressed to unstressed nonfinal vowels at 1.17:1 on average. In our data, the grand mean of duration of vowels that are expected to be stressed based on their position in the word is 65 ms.; for vowels that are not expected to be stressed, the grand mean of duration is 53 ms. This yields a ratio of 1.22:1. For

English, Crystal and House (1988b) estimated the same ratio at 2:1 for monophthongs and 1.5:1 for diphthongs.

Lexical stress, however, is not the sole determinant of vowel duration. In this work, we have found that in addition to its position within the word, the duration of Polish vowels is strongly influenced by many other factors, including most obviously speech rate, but also speech style, the manner and voicing of the neighboring consonant, syllable structure and the position of the vowel in the clause. As a result, the durational variation of Polish vowels is quite considerable. Unsurprisingly, we have found strong evidence for vowel target undershoot in our data.

In fact, the presence of this phenomenon in Polish has not gone unnoticed. Stieber 1966, Steffen-Batogowa 1975 and Sawicka 1995 have all pointed out that, impressionistically, Polish vowels tend to have fronted and raised variants in the vicinity of “soft” consonants (i.e. palatalized consonants, palatals and alveopalatals). These authors, however, have not thought about this variations as a type of vowel reduction, because they did not approach it from the standpoint of the target undershoot model. In this work, we have found that the “soft” consonants have a very strong impact on Polish vowels in both the F1 domain and the F2 domain (although the impact of many other consonants is also non-negligible) and that this impact wanes as the duration of the vowel increases. We take it to be a clear demonstration of the applicability of the target undershoot model to the variability of

Polish vowels and we believe that it is appropriate to describe this variability as “vowel reduction”.

Acoustic patterns reflect articulatory configurations. It can therefore be assumed that a greater distance between the acoustic targets of the vowel and the neighboring consonant indicates a greater difference between the articulatory configurations required for the vowel and the consonant in question. It is entirely intuitive that in such a situation we should observe greater coarticulation, i.e. that the speaker is less likely to achieve the target articulatory configuration of the vowel (or the consonant or both) than when the vowel and the consonant are more compatible articulatorily. The palatal component of the Polish “soft” consonants require greater displacement of the tongue front and the tongue body than any other consonants (Wierzchowska 1967). It is therefore unsurprising that the greatest amount of target undershoot is observed in their context.

It should be noted at this point that we are assuming a rather broad definition of what the articulatory target of the vowel is. We think it is likely that a fully canonical, absolutely context-free realization of a vowel – i.e. one which is unlikely if not impossible to be found in real speech data – involves specific articulatory requirements for all parts of the vocal tract, not just the tongue body. We do not take the fact that the target formant frequency is not reached in a given vowel token to indicate that a specific distance between the highest part of the tongue body and the palate is not achieved. That may or may not be the case in any given instance. However, to the extent that there is a

difference between the observed formant frequencies and the target formant frequencies, and that difference can be ascribed to the influence of the neighboring consonants, some aspect of the ideal articulatory configuration of the vowel has not been achieved because the articulatory gesture required for the consonant has not ended in time or has began “too early”. For example, while the tongue body may be in the required position, the tongue tip may be too elevated or the lips spread too wide. Of course, such issues cannot be fully resolved through an acoustic investigation alone.

The role of the acoustic and articulatory distance between the vowel and the consonant in target undershoot does not end here, but it is closely related to the contribution of duration. In this respect, we have found a very good match between our findings and Lindblom’s (1963) predictions. For the first formant, Lindblom predicts no undershoot in high vowels and then increasingly greater contribution of duration as the vowel becomes more open. Since all the consonants of Polish can be assumed to have a lowering influence on the first formant, the prediction is essentially that the role of duration increases as the distance between the vowel and the consonant increases. For the second formant, this prediction is explicitly written into Lindblom’s formulas: the duration term is an exponent of a power term in which part of the base is the distance between the formant frequency at vowel onset and the target frequency of the vowel.

At this point, it should be clarified what is meant by the contribution of duration. We are referring here to consonant-independent contribution of duration. In our multiple

regression results, this is directly represented by the regression coefficients for the variable duration. As was noted in chapter 2, regression coefficients represent the *unique* contribution of a given variable. For a continuous variable like vowel duration, regression coefficients essentially represent the rate of change in Hertz per millisecond. To say that this rate of change is greater for one vowel than another is to say that, given the same consonantal context, the same change in duration will result in a greater change in formant frequency for the former than for the latter. To say that this rate of change is positively correlated with the distance between the target for the vowel and the target for the consonant is to say that the greater the acoustic incompatibility between the vowel and the consonant is, the faster the acoustic changes must be. This in turn can be understood to indicate that the greater the distance between the articulatory configuration for the vowel and for the consonant, the faster the articulatory gestures must be.

It is important to note at this point the major limitation of multiple linear regression. As the name indicates, this statistical tool only models linear relationships; it cannot capture more complex relationships. The fact that we have found a greater rate of change of F1 for, say, the vowel *a* than for the vowel *e* may indicate that in the former case, the articulators move with greater constant speed than in the latter. However, it could also indicate that there is constant acceleration as the articulators move from the consonantal configuration to the vocalic one (or vice-versa). Such a phenomenon could not be captured with multiple linear regression.

Sawicka (1995) captures the variability of Polish vowels in the context of “soft” consonants by postulating a number of allophonic rules that substitute more close and more front variants of vowels for their more open and more back counterparts. The fact that such variation can be successfully modeled using the target undershoot model does not preclude the possibility that more categorical changes also occur. Within the approach taken here, this simply could not be verified. However, we did observe that the vowel *e* was more sensitive to consonantal influences in the F1 domain than the vowel *o*, and that it appeared to be more influenced in the F2 domain than the vowel *i*. This might indicate that in the vicinity of the “soft” consonants it changes in a way that exceeds coarticulatory requirements.

Our results agree with Sawicka’s impressionistic descriptions in another respect. She suggests that the fronting and raising processes are more common between “soft” consonants and after them than before them. In our data, target undershoot due to the influence of the “soft” consonants is more pervasive when the source of the effect is the preceding consonant than when it is the following consonant. This is also supported by the fact that Nowak (2006) showed that the perception of Polish alveopalatals depends to a greater extent on the following vowels than on the preceding vowels. It is unclear what the source of this difference is, but it does not seem to be limited to the “soft” consonants. For example, the impact of preceding labials on the two formants is greater than that of the following labials for all vowel categories, with the exception of front vowels, where the difference is in the opposite direction, but only 6 Hz. Lindblom’s (1963) undershoot

model does not make any predictions in this respect, because he only considered symmetrical consonantal context. It is however imaginable that there are different coarticulatory strategies at the CV boundary and at the VC boundary. Velars, for instance, are well known to have different coarticulatory patterns with preceding and following vowels in many languages (Stevens 1998). Such a difference could for example be a function of different dynamics of the consonantal release gesture and the consonantal closing gesture (Alan Timberlake, p.c.).

The mechanism of phonetic undershoot in its general form is easier to capture in the F1 domain than in the F2 domain. This is of course related to the fact that for non-high vowels, consonants of most places of articulation have a uniformly F1-lowering influence, but their influence on F2 may differ greatly depending on the target formant frequency of a given vowel. This is not a great problem when the “soft” consonants are considered, as they increase the second formant of F2 of all vowels (except *i*). The effect of other consonants is more difficult to assess. Nonetheless, we have found evidence that strongly indicates that the non-fronting consonants have an F2-lowering influence on front vowels (cf. section 6.2.3.4) while the backing consonants have an F2-raising influence on the back vowels (cf. section 6.4.3.3). (These consonantal classes were defined separately for different vowel classes and for the following and the preceding consonants, but in each case they include labials). The result of such conflicting influences is ostensible centralization. While Sawicka (1995) is probably right in asserting that vowel reduction understood as centralization to schwa does not frequently

occur in Polish, it can be argued that its seeds are present when the “non-soft” consonants are considered. Since the effects of the “non-soft” consonants are far smaller than those of the “soft” ones, it is possible that the lack of this type of reduction is constrained by the admittedly limited durational variation of Polish vowels (*vis-à-vis* many other languages).

7.2 Stress

As was mentioned before, the main goal of Lindblom’s (1963) undershoot model was to account for the fact that vowels in some languages are more reduced in unstressed syllables. He attributed this to the fact that unstressed vowels tend to be shorter. However, it has since been shown that stress has effects on vowel reduction that go beyond what can be attributed to the changes in duration it causes (Harris 1978, Nord 1986). In this work, we did not consider stress directly, i.e. we did not record our impressionistic judgments about the degree of stress. We did consider, however, the position of the vowel in the word. Since lexical stress in Polish is predictable, we were able to compare the behavior of vowels that are expected on phonological grounds to carry lexical stress and those that are not.

We found strong evidence that non-high vowels are more reduced in the non-stressable positions in the F1 domain, independently of duration. There was also some evidence that this is the case in the F2 domain, albeit it was less clear. In addition, we found that

intensity of the vowel was a good predictor of the degree of reduction of the vowel in the F1 domain; it is possible that this variable captured additional accentuation effects (e.g. phrasal accent).

It is not entirely clear what the mechanism of the duration-independent impact of stress on the degree of vowel reduction is. One possibility is that stressed syllables involve faster articulator movements; an explanation along these lines was suggested by Moon and Lindblom (1994) for style-related differences in undershoot. Another possibility, which may not contradict the first one, is that in unstressed syllables there is a greater overlap between the consecutive articulatory gestures, or “truncation” (Edwards et al. 1991).

Barnes (2002) and Flemming (2006) relate Lindblom’s (1963) target undershoot model to phonological vowel reduction in unstressed syllables observed in many languages. They both point out that the model correctly predicts reduction processes to be more common in the height domain. Flemming notes that phonological vowel reduction in the F2 domain is commonly found only when very short vowels of all types are reduced to schwa (as is the case in English). This is a natural consequence of the fact that most consonants have divergent effects on the second formant of vowels, depending on their degree of frontness. However, since palatal and palatalized consonants have a uniformly fronting influence, we should be able to find a pattern in which unstressed vowels are phonologically reduced to a front vowel in such a case, regardless of whether the

phonetic seeds of such a phonological process were to lie only in the shortening of unstressed vowels or an additional effect of destressing were involved. In fact, this seems to be the case in Russian, a language closely related to Polish. In Russian, vowels are often reduced to schwa; however, after palatalized consonants, non-high vowels are reduced to [ɪ] (Comrie 1990).¹

7.3 Non-phonetic and non-phonological variables

Previous studies on other languages have found that there are differences in the degree of vowel reduction between speech styles (Moon and Lindblom 1994, Koopmans-van Beinum 1980) and between content words and function words (van Bergem 1993). We have found evidence for the presence of this phenomenon in Polish, most consistently in the F1 domain for non-high vowels. More precisely, we have found that vowels tend to be less reduced in content words or syllables, and in the normal reading style, arguably the most “formal” of the speech styles we observed. It appears likely that these processes are governed by the same mechanism as stress-related differences, but we are not in a position to confirm this intuition.

On the other hand, there was little evidence that word frequency plays a major role in determining the scope of vowel reduction. The variable was selected into one model only, for back vowels, and its impact was minor. Reduction processes are known to be more

¹ Barnes (2002) presents an analysis of vowel reduction in Russian, but he does not discuss this specific case in detail.

common in more frequent words, because word frequency is a (crude) measure of predictability of lexical material and greater predictability allows greater reduction (cf. Jurafsky et al. 2001). However, Wright and Kilanski (2006) have recently argued that frequency may play a major role in coarticulatory processes (as opposed to categorical processes such as deletion) only as a factor affecting the difficulty of lexical items. This is a much more complex notion which we were not able to explore in this work.

Lexical items may also be more or less predictable depending on their position in the sentence. In this work, we have assumed that in an SVO language like Polish, more predictable words are more likely to occur phrase-initially and less predictable words phrase finally. Our data indicate that non-high vowels may indeed be more reduced in clause-final phonological words, and back vowels may be less reduced in clause-initial phonological words. All of these effects, however, were very small.

7.4 The effect of manner

The only variables whose impact was comparable in scope to the impact of the undershoot-related variables (i.e. place and duration) were preceding and following segment manner. These are more difficult to interpret, as there are few clear predictions about the impact of differences in manner on the degree of vowel reduction.

In the F1 domain, the greatest effect was that of nasal consonants. For each vowel type, both preceding and following nasals conditioned significantly higher F1 values (vis-à-vis dentals). F1-raising is an expected effect of contextual nasalization in the case of high and mid vowels, but for the low vowel previous studies make conflicting predictions (Wright 1986, Stevens 1998). Only Wright's findings are consistent with ours. In the F2 domain, nasals conditioned higher F2 values for the front vowels, but lower F2 values for the back vowels. This indicates that the vowel space is expanded in this domain under nasalization, which is unexpected. A further complication is that interactions between segment place and manner were observed (cf. section 6.2.3.7 and 6.2.3.8).

Liquids also had a strong effect on the formants (vis-à-vis stops), particularly in the F1 domain. Figure 7.4.1 presents regression coefficients for preceding and following laterals for all vowel classes, and figure 7.4.2 presents regression coefficients for preceding and following trills for all vowel classes.

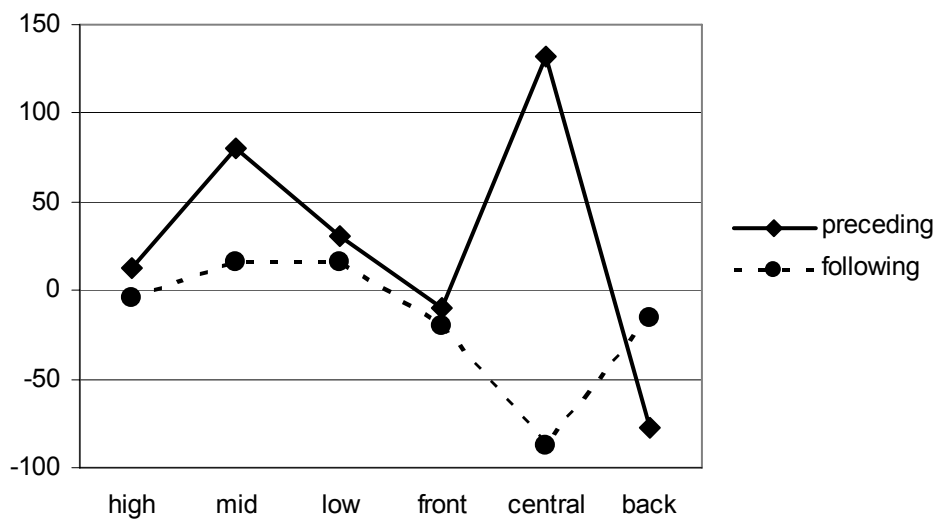


Figure 7.4.1 Regression coefficients for laterals for all six vowel categories (Hertz estimates). In each case, they estimate difference in F1 or F2 value vis-à-vis stops (i.e. the value for stops is 0 in each category).

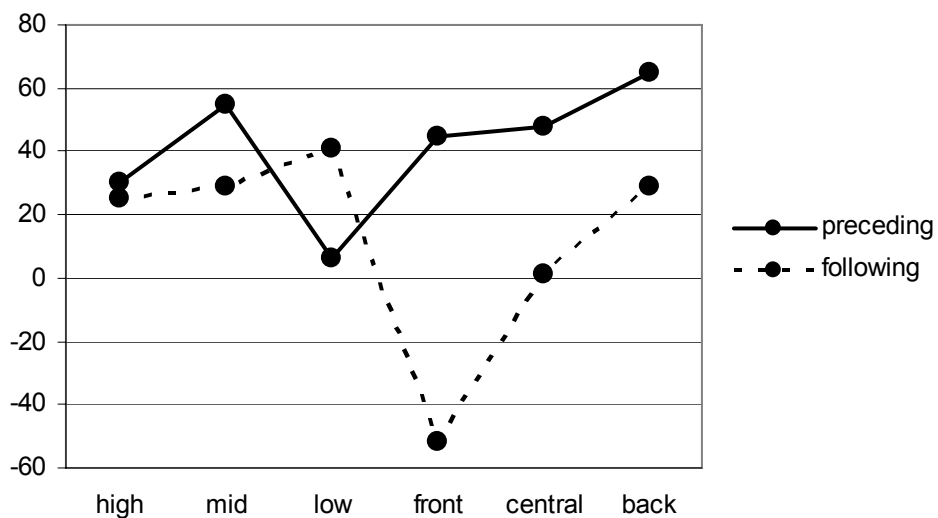


Figure 7.4.2 Regression coefficients for trills for all six vowel categories (Hertz estimates). In each case, they estimate difference in F1 or F2 value vis-à-vis stops (i.e. the value for stops is 0 in each category).

With the exception of the following laterals in the case of high vowels, higher F1 values are observed in the vicinity of liquids than in the vicinity of stops. This pattern is not

surprising, as it can be observed in other languages. Some English dialects neutralize vowel height distinctions before laterals in favor of the more open variant (Labov et al. 2006). Recasens (1991) and Recasens and Pallares (1999) have shown that trills require active lowering of certain parts of the tongue body. The situation observed for F2 is more puzzling. In general, it appears that following liquids condition lower F2 values than preceding liquids (but not necessarily lower vis-à-vis stops). While liquid consonants are known to have asymmetrical patterns of coarticulation with vowels (Stevens 1998), a precise account of these differences is beyond the scope of this work.

It should also be noticed that in the case of liquids, there is an additional complication. We have classified them as alveolar together with the “retroflex” fricatives and affricates. It is likely that these two groups do actually have somewhat different places of articulation; the regression coefficients for trills and liquids may therefore be a “correction” of our simplified classification of the place of articulation. In particular, the articulation of the “retroflex” consonants is likely to be more retracted (post-alveolar rather than alveolar, cf. Ladefoged and Maddieson 1996). This suggests that liquids should condition higher F2 values; consequently, the “corrective” regression coefficients for them should be positive. As can be seen from figures 7.4.1 and 7.4.2, this is the case only in some instances.

There are also difficulties with the interpretation of the impact of fricatives (vis-à-vis stops). Figure 7.4.3 presents regression coefficients for preceding and following fricatives for all vowel classes

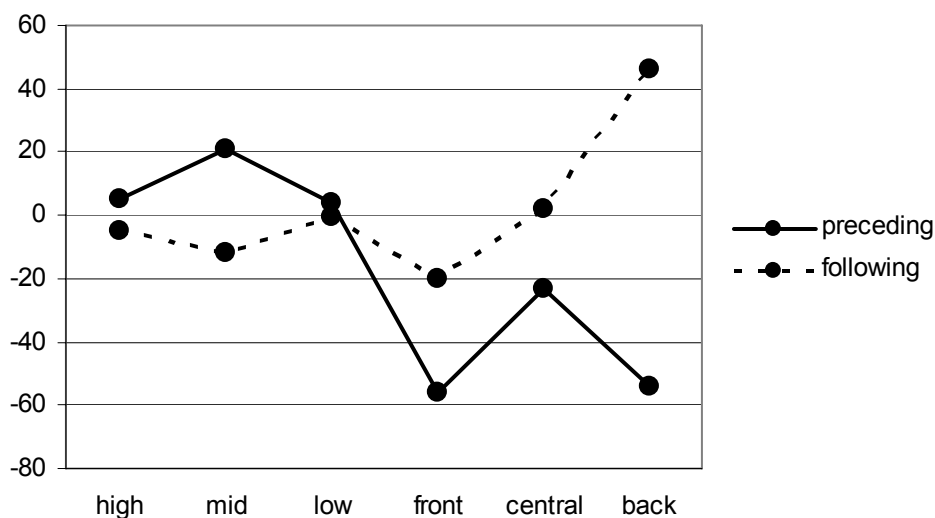


Figure 7.4.3 Regression coefficients for fricatives for all six vowel categories (Hertz estimates). In each case, they estimate difference in F1 or F2 value vis-à-vis stops (i.e. the value for stops is 0 in each category).

As can be seen, preceding fricatives condition higher formant values (vis-à-vis stops) in the F1 domain, which can be interpreted as more limited reduction. In the F2 domain, they condition lower F2 values (vis-à-vis stops). Given that the strongest consonantal effects in the F2 domain are F1-raising, the effect of fricatives can also be interpreted as more limited reduction. The effects of following fricatives are generally smaller and in the opposite direction. It is unclear why fricatives should differ from stops in terms of the consonantal impact on vowels. There are several possibilities which cannot be explored in detail here. First, once again the manner term may in fact be “correcting” errors

stemming from clustering together consonants with similar, but non-identical places of articulation. In the context of fricatives, the relevant place category we used is labials, which encompasses bilabial stops and nasals and labiodental fricatives. Second, there could be systematic differences in the estimated vowel duration, stemming from the use of the semi-automatic segmentation algorithm. Thus, since it seems that more limited reduction is observed after fricatives than after stops, it is conceivable that the true duration of post-fricative vowels was underestimated. Finally, there could be a true difference in the coarticulatory requirements imposed by fricatives and stops. Whichever of these answers is correct, a challenge lies in explaining the asymmetry observed between following and preceding fricatives.

As could be expected, in no case do the following affricates condition significantly different formant values than stops. Some differences, however, were observed, and for back vowels this difference was quite large and nearly significant (+46 Hz., $p < 0.1$). Preceding affricates tend to pattern with preceding fricatives, as expected.

The observed effects of approximants do not present a clear pattern. There is evidence that in the F2 domain their effect cannot be adequately considered without taking into account interactions with place of articulation (cf. section 6.2.3.8). In the F1 domain, their effects vis-à-vis stops are generally small; only for preceding approximants in the case of mid and low vowels significant formant-raising effect is observed.

7.5 Other phonetic and phonological variables

Every other single other phonetic and phonological variable was selected into at least one model. The sources of these effects are largely unclear. One exception is preceding segment voicing; it is likely that in those cases in which more limited reduction was found in this context, it reflects at least partially the underestimation of vowel duration after voiceless stops (cf. sections 4.3.8 and 5.3.3.8). The low vowel and back vowels were also found to be less reduced before voiceless consonants, for reasons that are unclear. Less reduction was found in more open syllables in the case of the low vowel in the F1 domain, but more in the case of the back vowels in the F2 domain. The low vowel was also found to be less reduced in shorter phonological words. Speech rate was found to be relevant for high and front vowels, and at least in the former case the effect was unexpected, in as much as speech rate is a determinant of vowel duration and duration itself was not relevant. Finally, no clear pattern of limited or enhanced reduction emerged from the examination of vowels in phonological words adjacent to pauses. However, the findings appear compatible with an explanation related to changes in pitch associated with pitch declination, as suggested by John Ohala (personal communication, cf. sections 5.5.3 and 6.5.3).

7.6 Conclusions

In this study, we created a number of statistical models that were able to account for a considerable part of the variance observed in Polish vowels in natural speech. Many of the effects that were discovered and validated on new data can be explained in a fairly straightforward way by Lindblom's (1963) target undershoot model of vowel reduction, despite the fact that the very paper that introduced this model invoked Polish as an example of a language in which this phenomenon is not found. Several other findings replicated previous findings for other languages (the effects of stress, style and the content/function distinction, among others). However, the exact source of several other effects discovered remains mysterious at present.

In addition, a significant proportion of the observed variance was not accounted for by our models. This can be a result of many factors. Most importantly, to the extent possible we have avoided predictors that require subjective judgments on the part of the investigator, which was motivated both by our interest in achieving the greatest possible degree of objectivity, as well as the consideration of efficiency, which is not unimportant in a study of this size. For the description of the consonantal context, we relied on phonological rather than phonetic criteria. Of course, just like vowels, consonants are not always realized in a canonical way. For example, phonological trills are often realized as taps or approximants; doubtlessly, this variation is in some way correlated with the formant variation observed in vowels. Other variables were chosen in such a way so as to

yield themselves to automatic or semi-automatic measurement (e.g. duration, speech rate and intensity). This left out other variables of potentially great importance, for which there are currently no automatized measurement methods. A key one is of course accentuation and the intonational contour of the utterance; others include for example the precise determination of the predictability of lexical material and other discourse-related factors.

However, automatic and semi-automatic procedures are not error-free. While we attempted to eliminate such errors through double-checking, it is not unlikely that the remaining ones are responsible for some part of the unexplained variance. Additional errors of similar nature may stem from the use of a Linear Predictive Coding algorithm to measure formant frequencies. It is likely that in some contexts the procedure was less reliable than in others (e.g. under nasalization).

In addition, we did not consider all the possible predictors of formant variability. While vowel-to-vowel coarticulation is known to occur in many languages, Choi and Keating (1991) have shown that its scope is limited in Polish, even when only the initial and final 5 ms. of the vowel are considered. We therefore did not consider it at all likely that any effect will be discovered at the steady-state part of the vowel. Wright (2003) and Scarborough (2004) have shown that coarticulatory processes are sensitive to the difficulty of lexical items. Difficulty is a complex metric that combines word frequency and neighborhood density. In the absence of data on neighborhood densities in the Polish

lexicon, it was simply not feasible to consider difficulty in the current work. Recent developments in exemplar-based phonology suggest that details of phonetic implementation may differ between individual lexical items (Johnson 1997, Pierrehumbert 2002). Unfortunately, there is no obvious way of considering this variation in multiple linear regression. Finally, since our interest was primarily in the undershoot model, which relies on the notion of a single target per vowel, we have ignored altogether dynamic aspects of vowel acoustics and articulation. We are of course aware that there is significant evidence that these are far from irrelevant (e.g. DiBenedetto 1989, Huang 1991).

There is no doubt that a more complete picture would emerge if we considered some or all of such variables. It is also likely that this would be the case if we used more fine-grained distinctions in some of our predictors, for example the place variables. More importantly, we constructed our models using feature-based classes of vowels. In most cases, the vowel variable was somewhat relevant, suggesting that vowels with the same degree of phonological height or frontness do not necessarily behave in the same way. To a certain extent, we explored such differences through interaction terms. There can be no doubt that a greater proportion of variance in observed formant frequencies would have been accounted for had many more interactions been considered. It is likely, however, that this would happen at the expense of preserving the clarity of the general picture, in as much as it can be judged to be clear. Our general strategy in this work has been to attempt to describe the forest, without ignoring the differences between the species of

trees, but without describing the idiosyncrasies of individual trees. Where to draw the line had to be a somewhat arbitrary decision.

References

Barnes, Jonathan (2002) *Positional Neutralization: A Phonologization Approach to Typological Patterns*. Doctoral dissertation, University of California, Berkeley.

Bernstein-Ratner, Nan (1985) "Dissociations between vowel durations and formant frequency characteristics". *Journal of Speech and Hearing Research* 28, 255-264.

Bybee, Joan (2000) "The phonology of the lexicon: evidence from lexical diffusion". In Michael Barlow and Suzanne Kemner (eds.) *Usage-based Models of Language*, 65-85. Stanford, CA:CSLI.

Carré, René and Samir Chennoukh (1995) "Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures". *Journal of Phonetics* 23.1-2, 231-241.

Carré, René and Pierre L. Divenyi (2000) "Modeling and perception of 'gesture reduction'". *Phonetica* 57, 152-169

Chiba, Tsutomu and Masato Kajiyama (1941) *The vowel, its nature and structure*. Tokyo: Tokyo Kaiseikan.

Cho, Taehong (2003) “Lexical stress, phrasal accent and prosodic boundaries in the realization of domain-initial stops in Dutch”. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 2657-2660. Barcelona, Spain.

Cho, Taehong and Patricia Keating (2001) “Articulatory and acoustic studies of domain-initial strengthening in Korean”. *Journal of Phonetics* 29, 155-190.

Choi, John D. and Patricia Keating (1991) “Vowel-to-vowel coarticulation in three Slavic languages”, *UCLA Working Papers in Phonetics* 78, 78-86.

Cohen, Jacob, Patricia Cohen, Stephen G. West and Leona S. Aiken (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Third Edition.* Mahwah, N.J. and London: Lawrence Erlbaum Associates.

Comrie, Bernard (1990) “Russian”. In Bernard Comrie (ed.) *The World's Major Languages*, 329-347. New York and Oxford: OUP.

Crystal, Thomas H. and Arthur S. House (1988a) “Segmental duration in connected speech signals: current results”. *Journal of the Acoustical Society of America* 83.4, 1553-1573.

Crystal, Thomas H. and Arthur S. House (1988b) "Segmental duration in connected speech signals: syllabic stress". *Journal of the Acoustical Society of America* 83.4, 1574-1585.

Dejna, Karol (1993) *Dialekty polskie*. Wrocław: Zakład Narodowy im. Ossolińskich.

Delattre, Pierre (1981) "Vowel reduction in four languages". In Pierre Delattre, *Studies in comparative phonetics*". Heidelberg: Julius Gross Verlag.

DiBenedetto, Maria-Gabriella (1989) "Vowel representation: some observations on temporal and spectral properties of the first formant frequency". *Journal of the Acoustical Society of America* 86.1, 55-66.

Dukiewicz, Leokadia (1995) "Fonetyka". In Henryk Wróbel (ed.) *Gramatyka współczesnego języka polskiego. Fonetyka i fonologia*, 7-103. Warszawa: Wydawnictwo Instytutu Języka Polskiego PAN.

Edwards, Jan, Mary E. Beckman and Janet Fletcher (1991) "The articulatory kinematics of final lengthening". *Journal of the Acoustical Society of America* 89.1, 369-382.

Fant, Gunnar, Anita Kruckenberg and Lennart Nord (1991) "Durational correlates of stress in Swedish, French and English". *Journal of Phonetics* 19.3-4, 351-365.

Flemming, Edward (2006) "A phonetically-based model of phonological vowel reduction". Ms., available at <http://web.mit.edu/~flemming/www/paper/vowelred.pdf>.

Fourakis, Marios (1991) "Tempo, stress and vowel reduction in American English". *Journal of the Acoustical Society of America* 90.4, 1816-1827.

Gahl, Suzanne (2006) "Why do frequent words shorten? Homophones in spontaneous speech." Talk presented at the Annual Meeting of the Linguistic Society of America, Albuquerque, NM.

Garson, G. David (2006) *Statnotes: Topics in Multivariate Analysis*. Available online at <http://www2.chass.ncsu.edu/garson/PA765/statnote.htm>.

Gay, Thomas (1978) "Effect of speaking rate on vowel formant movements", *Journal of the Acoustical Society of America*, 63.1, 223-230.

Gravetter, Frederick J. and Larry B. Wallnau (2000) *Statistics for the Behavioral Sciences*. Stamford, CT: Wadsworth.

Grzegorzczkova, Renata (1996) *Wykłady z polskiej składni*. Warszawa: PWN.

Grzegorzczkowska, Renata, Roman Laskowski and Henryk Wróbel (1984) *Morfologia*. Warszawa: PWN.

Harrington, Jonathan, Janet Fletcher and Corinne Roberts (1995) "Coarticulation and the accented/unaccented distinction: evidence from jaw movement data". *Journal of Phonetics* 23.3, 305-322.

Harris, Katherine S. (1978) "Vowel duration change and its underlying physiological mechanisms". *Language and Speech*, 21.4, 354-361.

Houde, R. A. (1968) "A study of tongue body motion during selected speech sounds". *SCLR Monograph No. 2*. Santa Barbara: Speech Communications Research Laboratory.

Huang, Caroline B. (1991) *An acoustic and perceptual study of vowel formant trajectories in American English*. RLE Technical Report 583.

Jassem, Wiktor (1959) "The Phonology of Polish Stress". *Word* 15, 252-269.

Jassem, Wiktor (1962) "Akcent języka polskiego". *Prace Językoznawcze* 31, 1-116.

Johnson, Keith (1997) "Speech perception without speaker normalization: An exemplar model". In Johnson & Mullennix (eds.) *Talker Variability in Speech Processing*, 145-165. San Diego: Academic Press.

Johnson, Keith (2003) *Acoustic and Auditory Phonetics*. 2nd Edition. Oxford: Blackwell.

Johnson, Keith (2006) *Quantitative Methods in Linguistics*. Manuscript available at <http://corpus.linguistics.berkeley.edu/~kjohnson/quantitative>.

Joos, Martin (1948) "Acoustic phonetics". In *Language* 24, pp. 1-136.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. (2001) "Probabilistic Relations between Words: Evidence from Reduction in Lexical Production". In Joan Bybee and Paul Hopper (eds.). *Frequency and the emergence of linguistic structure*, 229-254.. Amsterdam: John Benjamins.

Keating, Patricia (2004) "Linguistic phonetics in the UCLA Phonetics Lab". Presentation at *From Sound to Sense*, MIT. Available at <http://www.rle.mit.edu/soundtosense/conference/pdfs/invitedspeakers/Keating%20PAPER.pdf>.

Koopmans-van Beinum, F.J. (1980) *Vowel contrast reduction*. Amsterdam: Academische Pers B.V.

Kuehn, David P. and Kenneth L. Moll (1976) "A cineradiographic study of VC and CV articulatory velocities". *Journal of Phonetics* 4.4, 303-320.

Labov, William, Sherry Ash, and Charles Boberg (2006) *Atlas of North American English: phonetics, phonology, and sound change*. New York: Mouton de Gruyter.

Ladefoged, Peter, Iris Kameny and William Brackenridge (1976) "Acoustic effects of style of speech". *Journal of the Acoustical Society of America* 59.1, 228-231.

Ladefoged, Peter and Ian Maddieson (1996) *The Sounds of the World's Languages*. Oxford and Malden: Blackwell.

Lavoie, Lisa Marie (2000) *Phonological Patterns and Phonetic Manifestations of Consonant Weakening*. Doctoral dissertation, Cornell University.

Lehiste, Ilse (1977) "Isochrony reconsidered". *Journal of Phonetics* 5.3, 253-263.

Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge: MIT Press.

Liljencrants, Johan and Björn Lindblom (1972) "Numerical simulations of vowel quality systems: The role of perceptual contrast". *Language* 48:839-862.

Lindblom, B Björn (1963) "Spectrographic study of vowel reduction". *Journal of the Acoustical Society of America* 35.11, 517-525.

Maddieson, Ian (1984) *Patterns of Sounds*. Cambridge: CUP.

Maddieson, Ian (1985) "Phonetic cues to syllabification". In Victoria A. Fromkin (ed.) *Phonetic linguistics: essays in honor of Peter Ladefoged*, 203-221. New York: Academic Press.

Moon, Seung-Jae and Bjorn Lindblom (1995) "Interaction between duration, context, and speaking style in English stressed vowels". *Journal of the Acoustical Society of America* 96.1, 40-55.

Navarro Tomás, T. (1916) "Cantidad de vocales accentuadas". *Revista de Filología Española*, III, 387-408.

Navarro Tomás, T. (1918) *Manual de pronunciación española*. Madrid.

Nooteboom, S.G. and I. Slis (1970) "A note on the degree of opening and the duration of vowels in normal and "pipe" speech". *IPO Ann. Rep.* 5:55-58.

Nooteboom, Sieb (1997) "The prosody of speech: melody and rhythm". In William J. Hardcastle and John Laver (eds.) *The Handbook of Phonetic Sciences*, 640-673. Oxford, UK and Malden, MA: Blackwell.

Nord, Lennart (1986) "Acoustic studies of vowel reduction in Swedish". In *Speech Transmissions Laboratory Quarterly Progress and Status Report*. Stockholm: KTH.

Nowak, Paweł M. (2006) "The role of vowel transitions and frication noise in the perception of Polish sibilants". *Journal of Phonetics*, in press.

Ohala, John J. (1972) "How is pitch lowered?" Manuscript, available electronically at <http://trill.berkeley.edu/users/ohala/papers/how-whole.pdf>

Ohala, John J. (1975) "Phonetic explanations for nasal sound patterns". In Charles A. Ferguson, Larry M. Hyman, & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 289-316. Stanford: Language Universals Project.

Ohala, John J. (1981) "Articulatory constraints on the cognitive representation of speech". In Terry Myers, John Laver and John Anderson (eds.) *The Cognitive Representation of Speech*, 111-122. Amsterdam, New York and Oxford: North-Holland Publishing Company.

Ohala, John J. (2001) *Phonology in your ear*. Ms., University of California, Berkeley.

Ohala, John J. and Manjari Ohala (1993) "The phonetics of nasal phonology: theorems and data". In Marie K. Huffman and Rena A. Krakow (eds.) *Nasals, nasalization, and the velum*, 225-249. [Phonetics and Phonology Series, Vol. 5] San Diego, CA: Academic Press.

Picheny, Michael A., Nathaniel I. Durlach and Louis D. Braida (1985) "Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech". *Journal of Speech and Hearing Research* 29, 434-446.

Pierrehumbert, Janet (2002) "Word-specific phonetics". In Carlos Gussenhoven and Natasha Warner (eds.) *Laboratory Phonology* 7, 101-139. The Hague: Mouton de Gruyter.

Ramus, Franck, Marina Nespors and Jacques Mehler (2003) "The psychological reality of rhythm classes: perceptual studies". *Proceedings of the 15th International Congress of the Phonetic Sciences*, 337-342.

Recasens, Daniel (1991) "Production and coarticulatory characteristics of alveolar taps and trills". *Journal of Phonetics* 19, 267-280.

Recasens, Daniel and Maria D. Pallarès (1999) “A study of /r/ and /rr/ in the light of the 'DAC' coarticulation model”. *Journal of Phonetics* 27, 143-170.

Rubach, Jerzy (1974) “Syllabic consonants in Polish”. *Journal of Phonetics* 2.2, 109-116.

Rubach, Jerzy (1984) *Cyclic and lexical phonology: the structure of Polish*. Dordrecht: Foris.

Sawicka, Irena (1995) “Fonologia”. In Henryk Wróbel (ed.) *Gramatyka współczesnego języka polskiego. Fonetyka i fonologia*, 105-194. Warszawa: Wydawnictwo Instytutu Języka Polskiego PAN.

Scarborough, Rebecca (2004) “Degree of Coarticulation and Lexical Confusability”. *Proceedings of the 29th Meeting of the Berkeley Linguistics Society*, 367-378.

Schouten, M.E.H. and L.C.W. Pols (1979a) “Vowel segments in consonantal contexts: a spectral study of coarticulation – part I”. *Journal of Phonetics* 7, 1-23.

Schwab, A. James (2006) “Multiple regression – split sample validation”. Available at www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/MultipleRegression_SplitSampleValidation.ppt.

Steffen-Batogowa, Maria (1975) *Automatyzacja transkrypcji fonetycznej tekstów polskich*. Warszawa: PWN.

Steffen-Batogowa, Maria (2000) *Struktura akcentowa języka polskiego*. Warszawa & Poznań: Wydawnictwo Naukowe PWN.

Stevens, Kenneth N. and Arthur S. House (1963) "Perturbation of vowel articulations by consonantal context: an acoustical study". *Journal of Speech and Hearing Research* 6, 111-128.

Stevens, Kenneth N. (1997) "Articulatory-Acoustic-Auditory Relationships". In William J. Hardcastle and John Laver (eds.) *The Handbook of Phonetic Sciences*, 462-506. Oxford, UK and Malden, MA: Blackwell.

Stevens, Kenneth N. (1998) *Acoustic Phonetics*. Cambridge, MA and London, UK: The MIT Press.

Stieber, Zdzisław (1966) *Historyczna i współczesna fonologia języka polskiego*. Warszawa, PWN.

Stone, Gerald (1990) "Polish". In Bernard Comrie (ed.) *The World's Major Languages*, 349-366. New York and Oxford: OUP.

Summers, W. Van, David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow and Michael A. Stokes (1988) "Effects of noise on speech production: acoustic and perceptual analyses". *Journal of the Acoustical Society of America* 84.3, 917-928.

Swaffield, J, J. N. Shearme and J. N. Holmes (1961) "Some measurements on the vowel sounds of conversational speech". *Journal of the Acoustical Society of America* 33.11, 1683.

Tuller, Betty, Katherine S. Harris and J.A. Scott Kelso (1982) "Stress and rate: differential transformations of articulation". *Journal of the Acoustical Society of America* 71.6, 1534-1543.

van Bergem, Dick R. (1993) "Acoustic vowel reduction as a function of sentence accent, word stress and word class". *Speech Communication* 12, 1-23.

VanSon, R.J.J.H. and Louis C.W. Pols (1990) "Formant frequencies of Dutch vowels in a text, read at normal and fast". *Journal of the Acoustical Society of America* 88.4, 1683-1693.

VanSon, R.J.J.H. and Louis C.W. Pols (1992) "Formant movements of Dutch vowels in a text, read at normal and fast rate". *Journal of the Acoustical Society of America* 92.1, 121-127.

Wierzchowska, Bożena (1967) *Opis fonetyczny języka polskiego*. Warszawa: PWN.

Wightman, C., S. Shattuck-Huffnagel, M. Ostendorf and P. Price (1992) "Segmental durations in the vicinity of prosodic phrase boundaries". *Journal of the Acoustical Society of America*, 91, 1707-17.

Wright, James T. (1986) "The Behavior of nasalized vowels in the perceptual vowel space". In John J. Ohala and Jeri J. Jaeger (eds.) *Experimental Phonology*. Orlando, FL: Academic Press.

Wright, Richard (2003) "Factors of lexical competition in vowel articulation", in John Local, Richard Ogden and Rosalind Temple (eds.) *Phonetic Interpretation. Papers in Laboratory Phonology VI*, 75-87. Cambridge: Cambridge University Press.

Wright, Richard and Kelley Kilanski (2006) "The effects of lexical factors on consonant variation". Talk presented at the Annual Meeting of the Linguistic Society of America, Albuquerque, NM.

Zanussi, Krzysztof (1978) "Barwy ochronne". In *Scenariusze filmowe*. Warszawa: Iskry.

Appendix A

This appendix presents examples of speech segmentation in this work. The amplitude-based procedure employed is explained in section 2.3. Figures A1-A12 present short speech samples from each speaker in each different speech style, chosen so as to show a representative sample of consonants. Each waveform and corresponding spectrogram is accompanied by three tiers. The bottom tier shows the minor amplitude slope maxima, the middle tier the major amplitude slope maxima and the top tier the actual segment boundaries chosen. In most cases, these boundaries correspond to the major amplitude slope maxima. Where necessary, they correspond instead to the minor amplitude slope maxima only.

Consonants with the following combinations of place and manner can be found in the figures indicated:

voiceless stops: A2, A3, A4, A6 (2), A7

voiced stops: A5 (2), A6, A7, A9, A10

voiceless affricates: A3 (2), A5

voiced affricate: A2

voiceless fricatives: A3, A8, A12

voiced fricatives: A4 (2), A5, A9, A10, A11, A12

trills: A4, A6

laterals: A2, A8, A12

nasals: A1, A7 (3), A8, A11 (2)

approximants: A6, A7, A8

The illustrations were prepared in Praat 4.3.

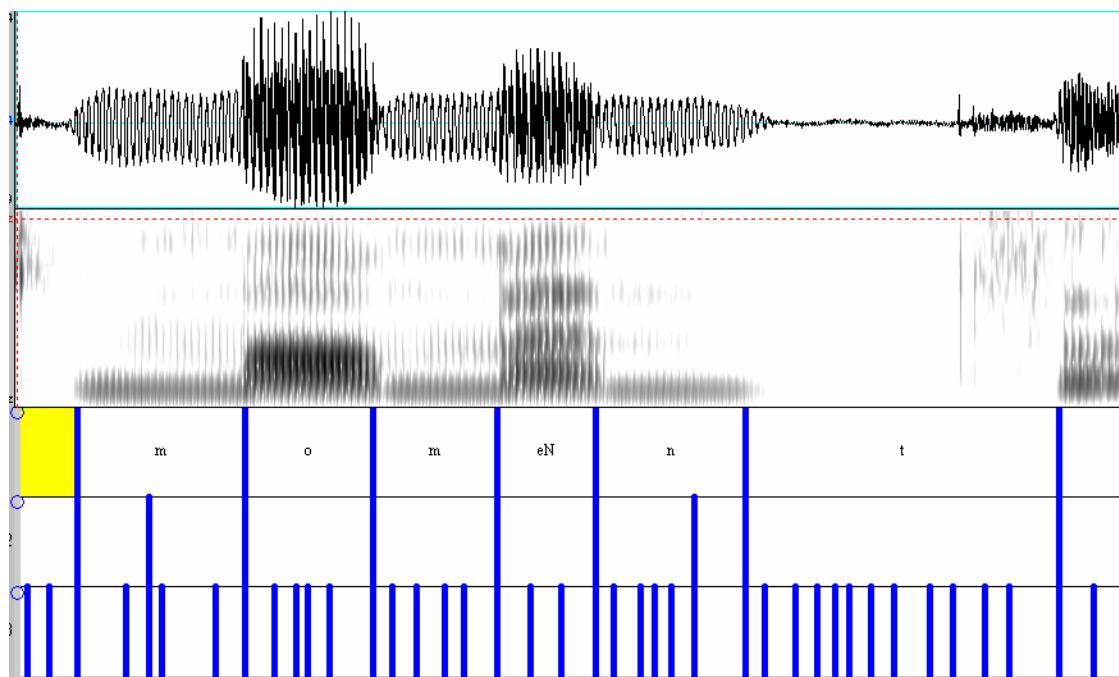


Figure A1 The word *moment* “moment” pronounced by speaker w1 in the narrative style. The phonological representation of this utterance is /mɔmənt/.

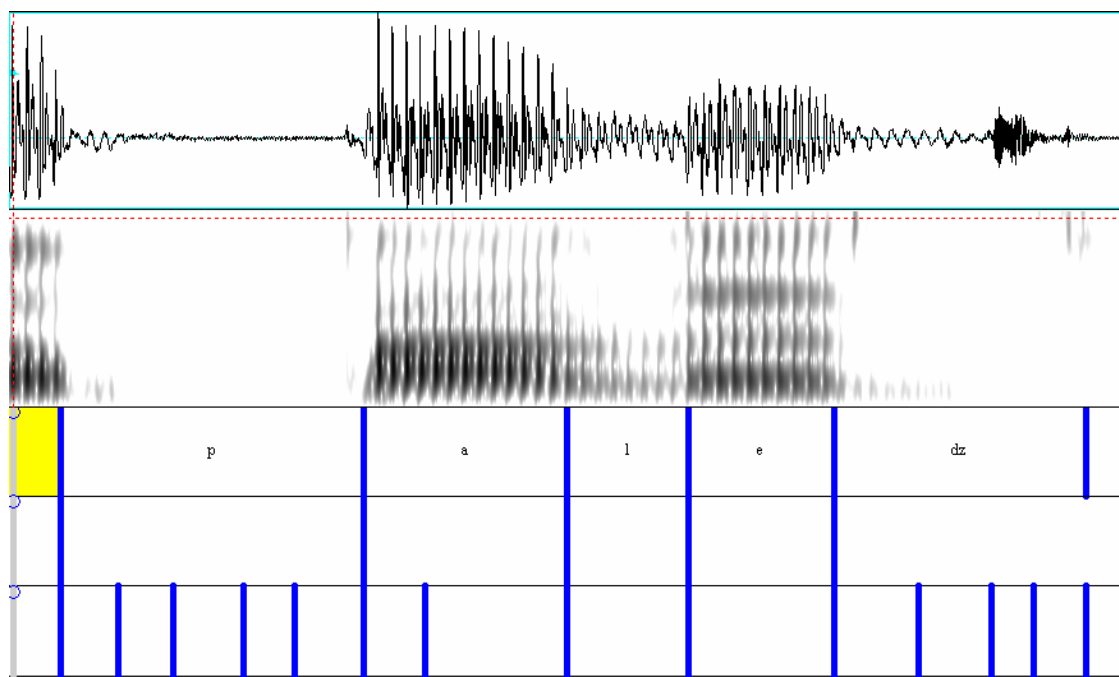


Figure A2 The word *palec* “finger” pronounced by speaker w1 in the slow reading style. The phonological representation of this utterance is /palets/.

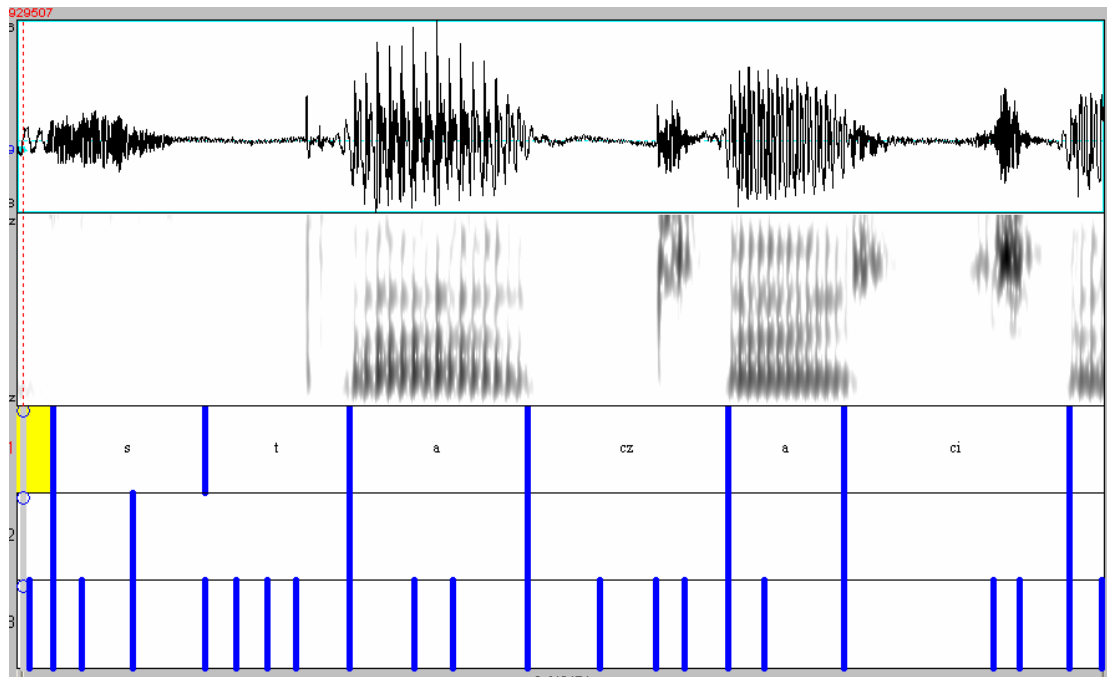


Figure A3 The word *staczać* "to roll down" pronounced by speaker w1 in the fast reading style. The phonological representation of this utterance is /statʂatɕ/.

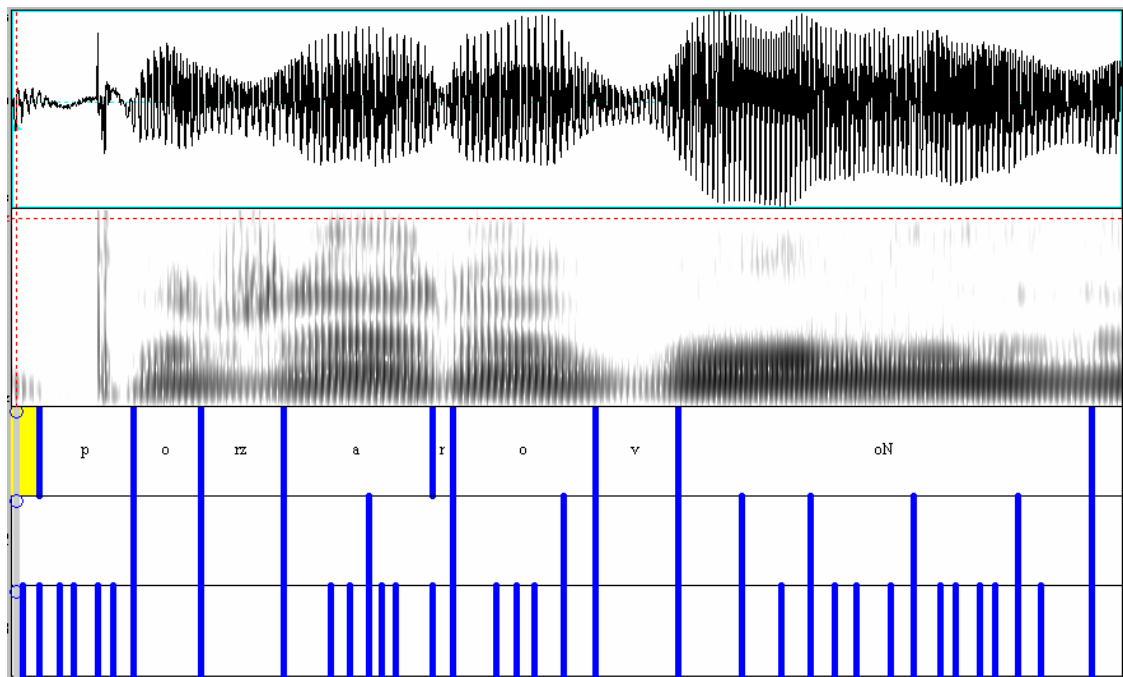


Figure A4 The word *pożarową* "fire" (adj.) pronounced by speaker w2 in the narrative style. The phonological representation of this utterance is /pɔʐarɔvɔ̃/. The extremely long final vowel is due to the speaker's hesitation.

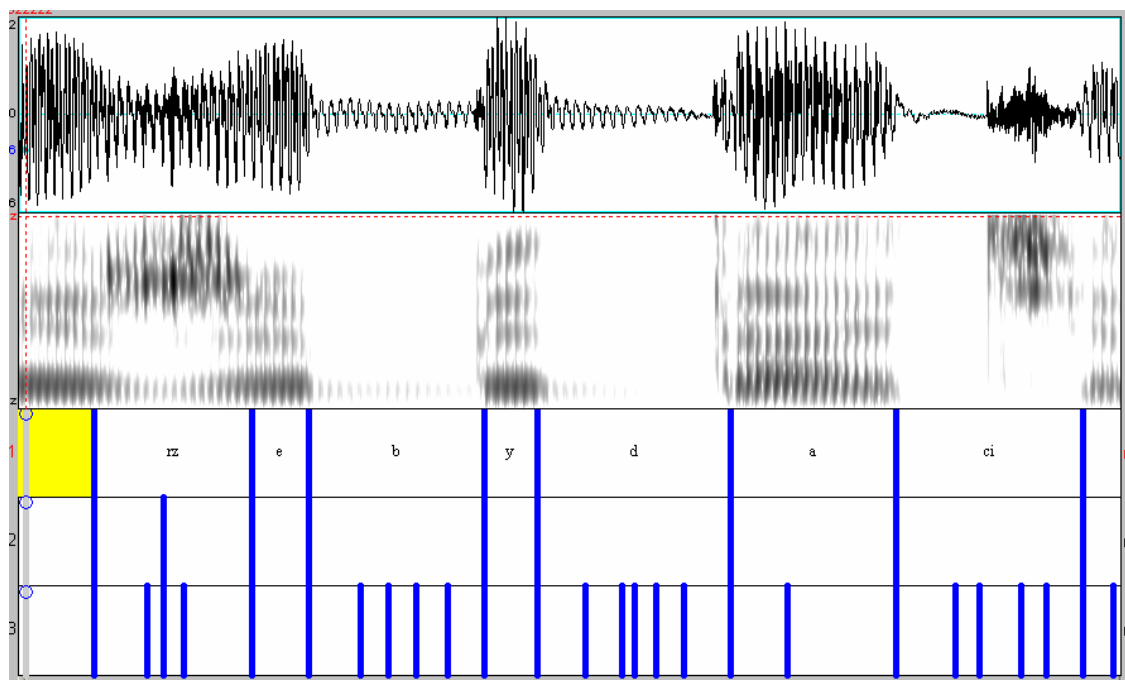


Figure A5 The words *żeby dać* “in order to give” pronounced by speaker w2 in the normal reading style. The phonological representation of this utterance is /zɛbɪdatɕ/.

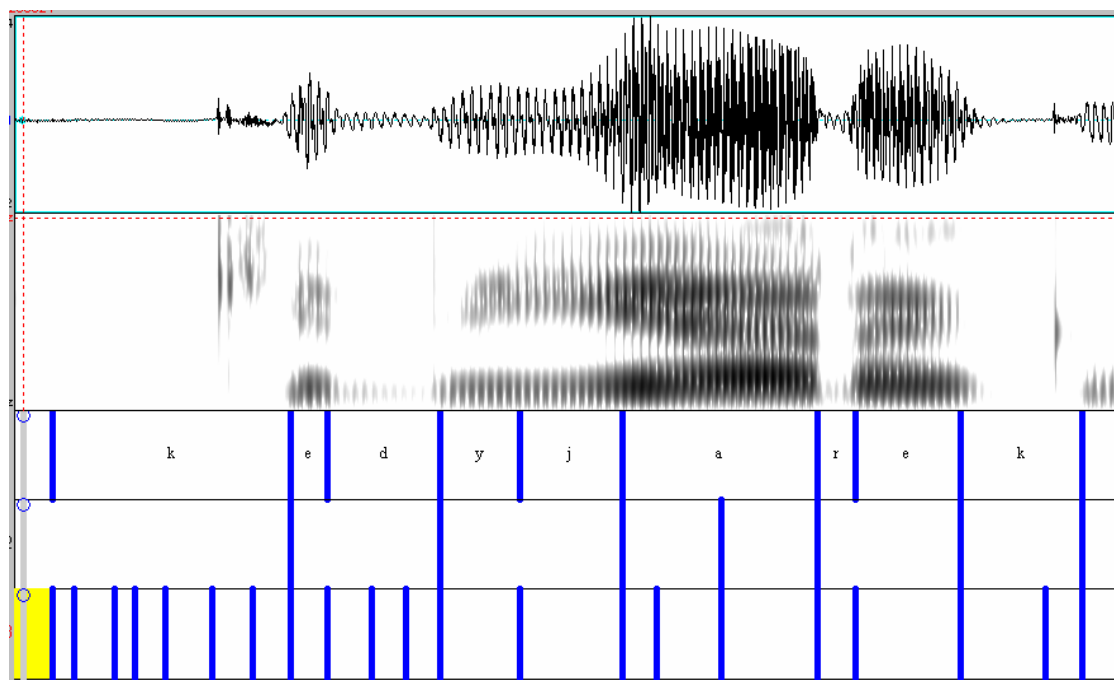


Figure A6 The words *kiedy Jarek* “when Jarek” pronounced by speaker w2 in the fast reading style. The phonological representation of this utterance is /cedɨjarek/.

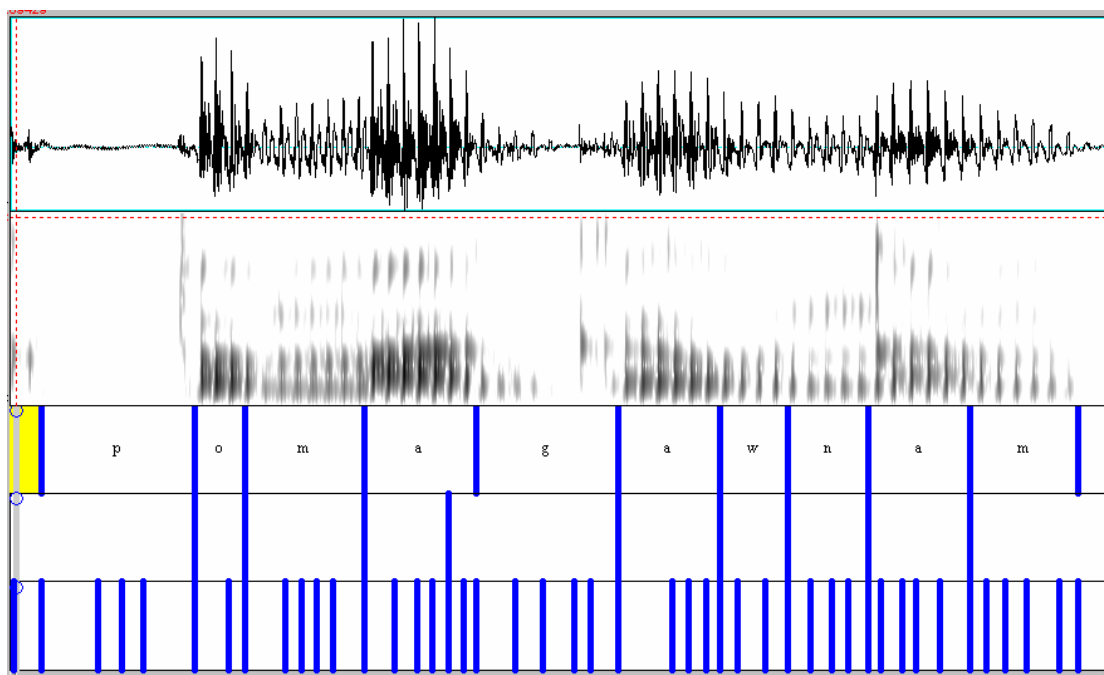


Figure A7 The words *pomagal nam* "he was helping us" pronounced by speaker m1 in the narrative style. The phonological representation of this utterance is /pɔmagawnam/.

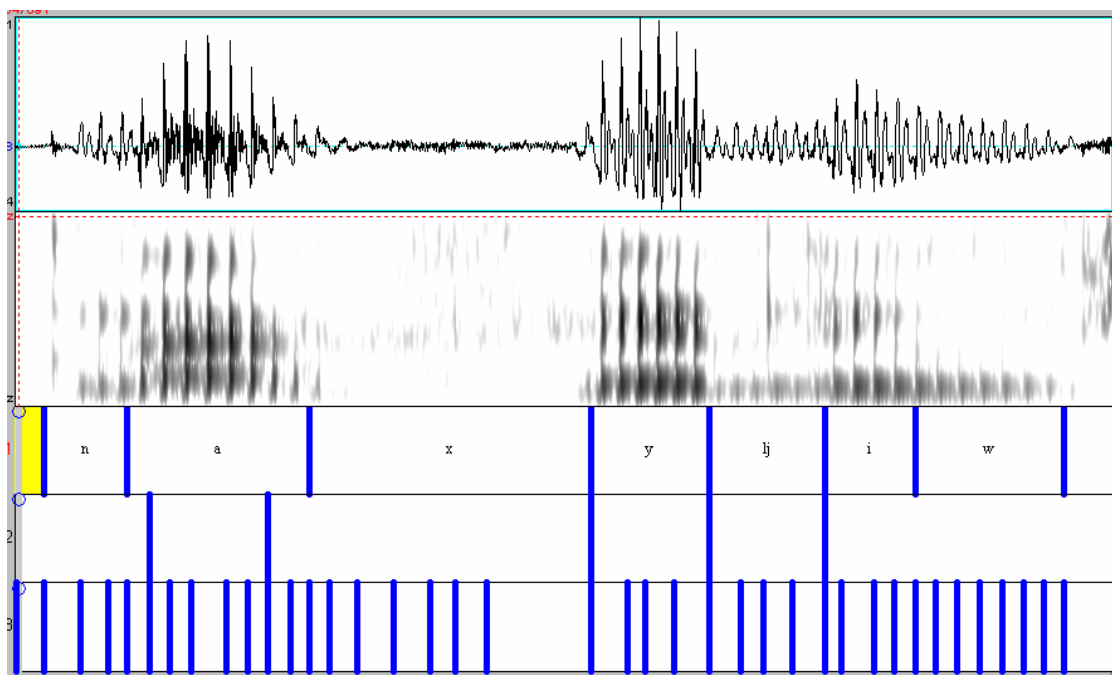


Figure A8 The word *naxylit* "he leaned down" pronounced by speaker m2 in the normal reading style. The phonological representation of this utterance is /naxil^hiw/.

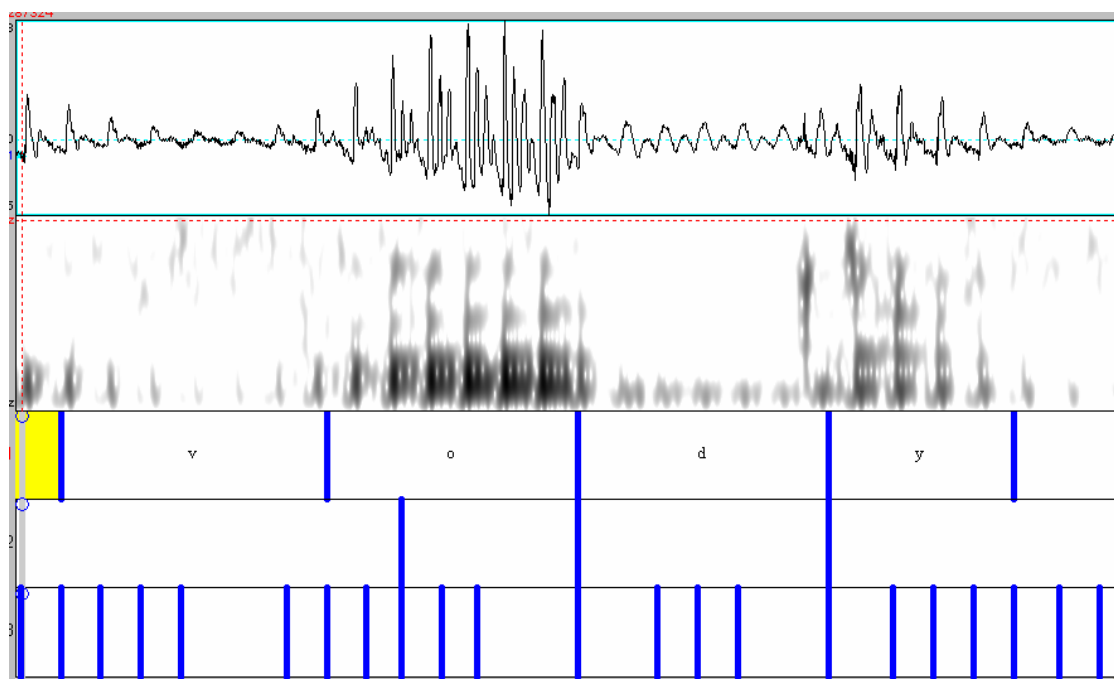


Figure A9 The word *wody* “water” (gen.) pronounced by speaker m1 in the fast reading style. The phonological representation of this utterance is /vɔdi/.

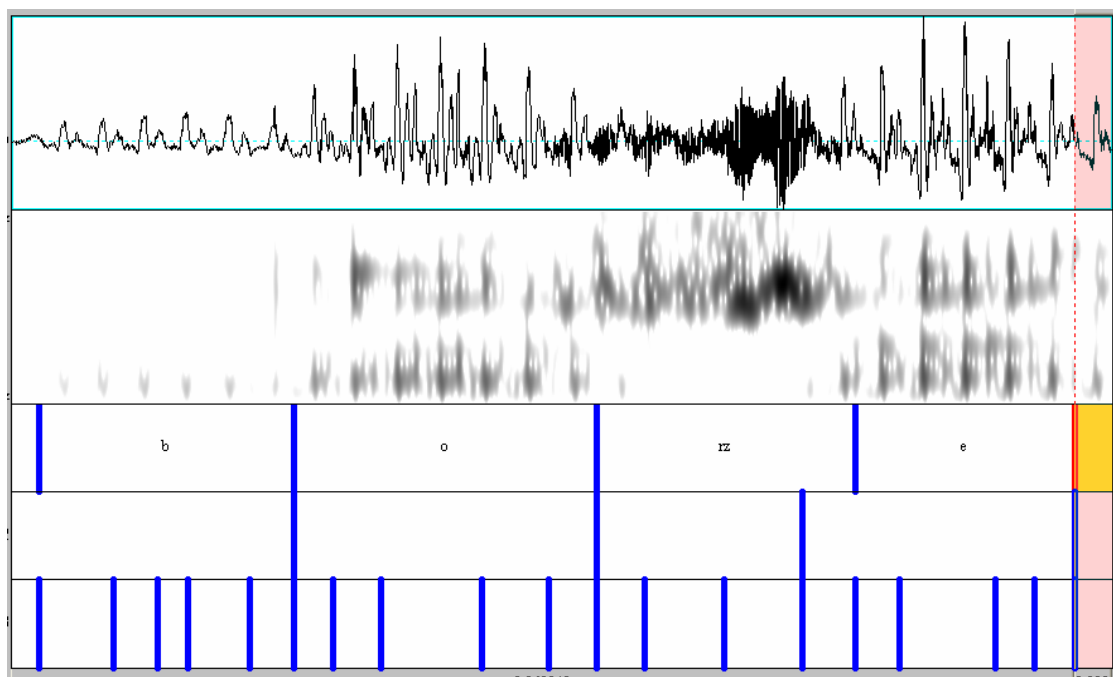


Figure A10 The word *boże* “god’s” spoken by speaker m2 in the narrative style. The phonological representation of this utterance is /bɔʒɛ/.

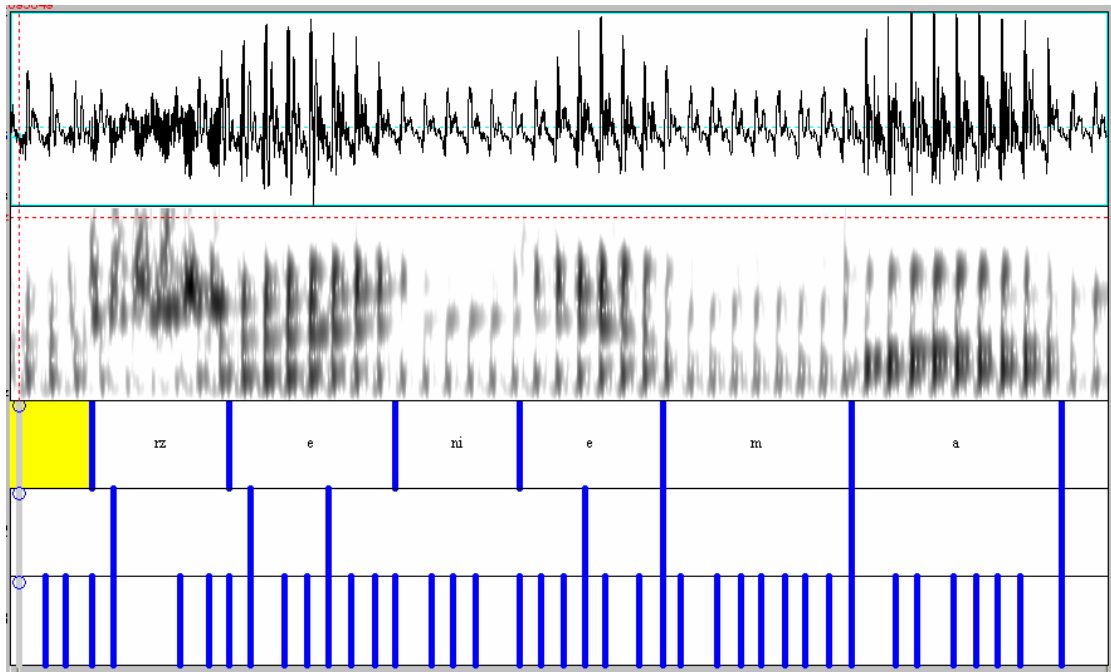


Figure A11 The words *ze nie ma* “that there is not” spoken by speaker m2 in the normal reading style. The phonological representation of this utterance is /zɛɲema/.

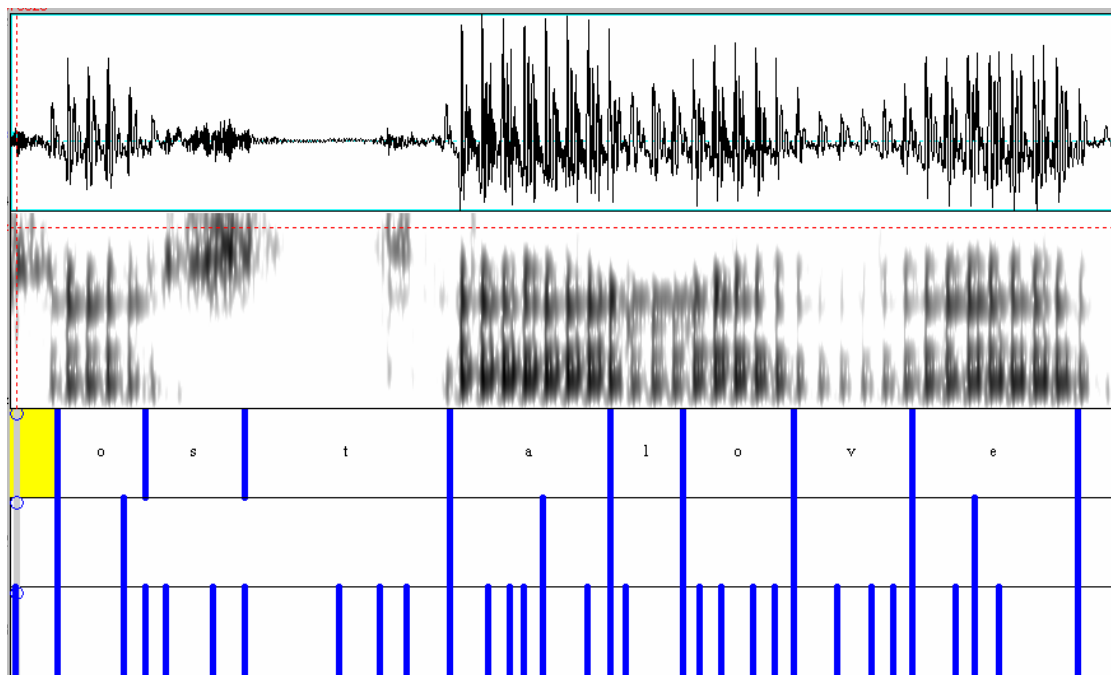


Figure A12 The words *o stalowe* “against steel (adj).” spoken by speaker m2 in the fast reading style. The phonological representation of this utterance is /ɔstalɔwɛ/.

Appendix B

This appendix presents examples of the automatic steady-state selection procedure used in this work, described in section 2.4. Figures B1-B12 show two tokens of each vowel, representing a variety of consonantal contexts. One example was taken from each speaker's contribution to each style. The bottom tier shows segmentation and transcription, while the top tier indicates the steady-state points selected. LPC-based estimates of the formants are also shown. The illustrations were prepared in Praat 4.3.

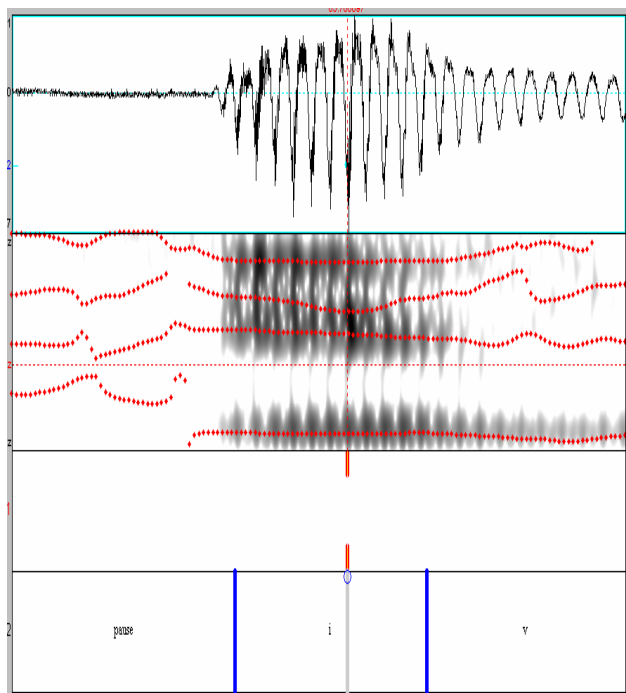


Figure B1 The vowel *i* between a pause and a voiced labiodental fricative pronounced by speaker w1 in the narrative style.

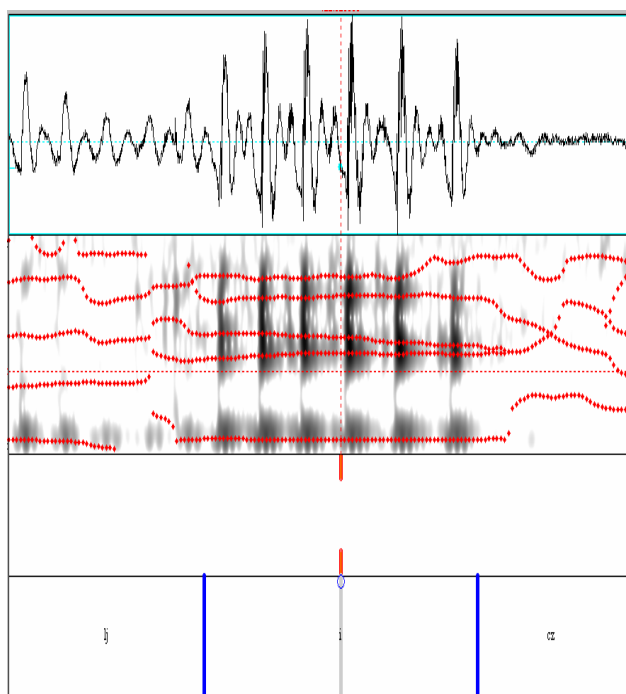


Figure B2 The vowel *i* between a palatalized alveolar lateral and a voiceless alveolar (retroflex) affricate pronounced by speaker m1 in the normal reading style.

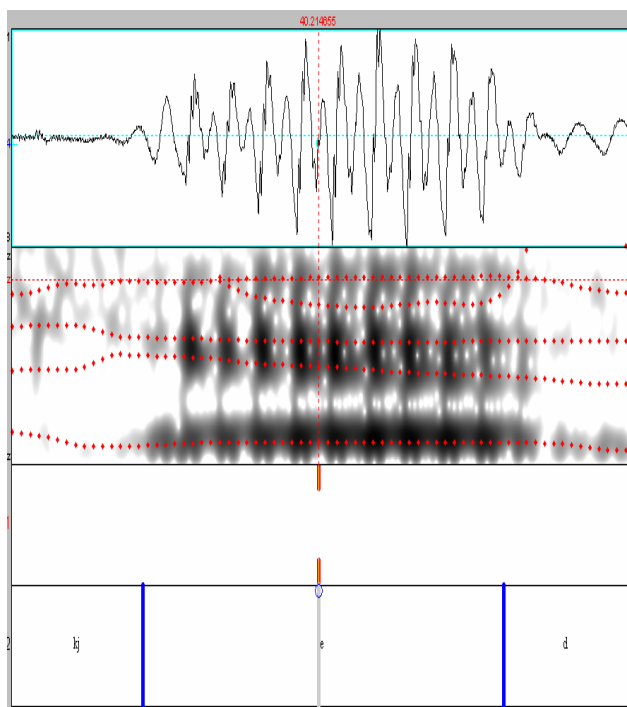


Figure B3 The vowel *e* between a voiceless palatal stop and a voiced dental stop pronounced by speaker w1 in the normal reading style.

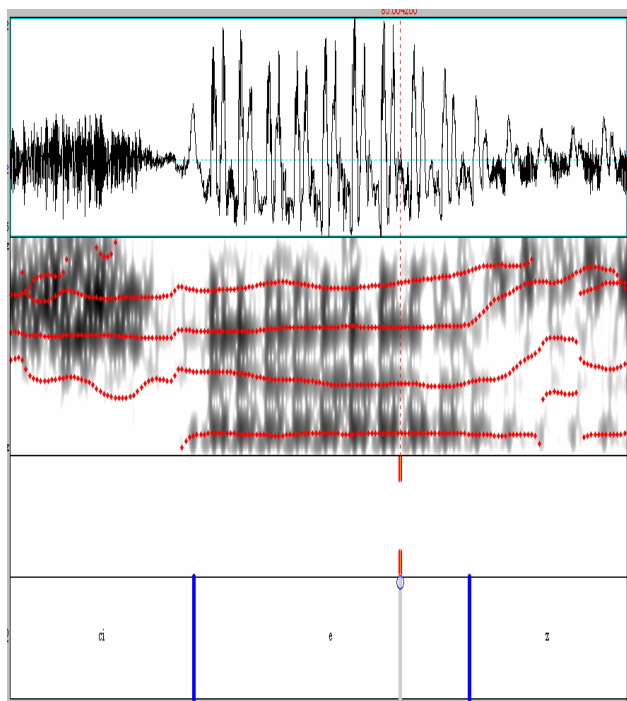


Figure B4 The vowel *e* between a voiceless alveopalatal affricate and a voiced dental fricative pronounced by speaker m2 in the narrative style.

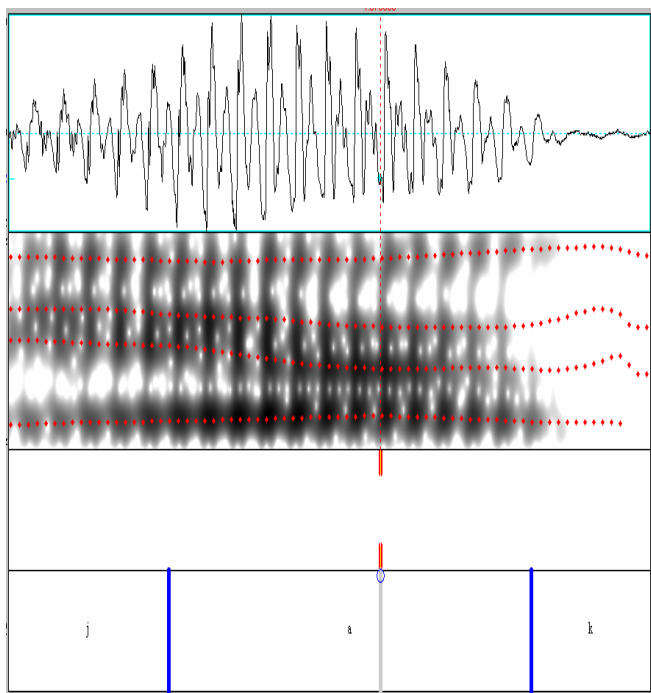


Figure B5 The vowel *a* between a palatal glide and a voiceless velar stop pronounced by speaker w1 in the fast reading style.

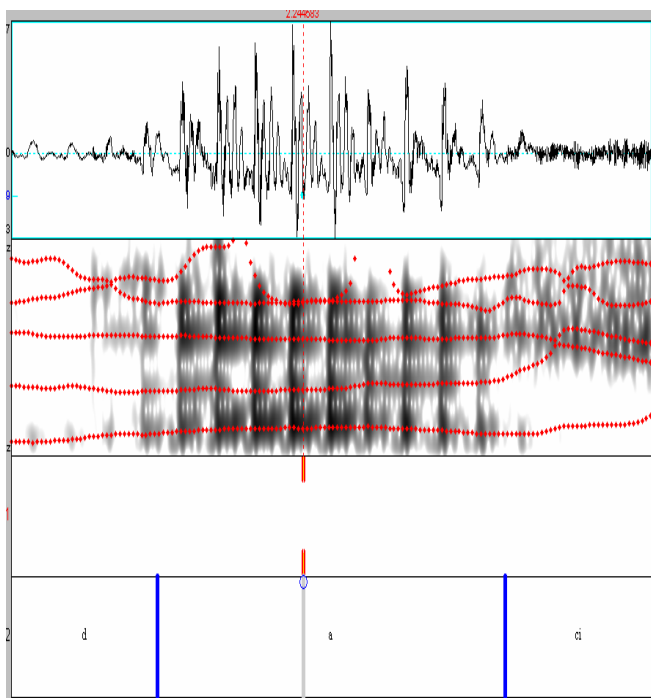


Figure B6 The vowel *a* between a voiced dental stop and a voiceless alveopalatal affricate pronounced by speaker m2 in the normal reading style.

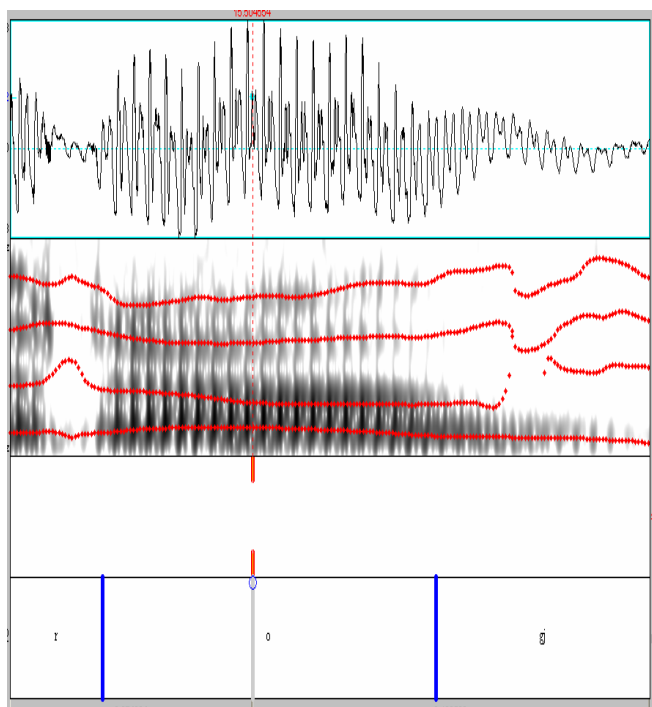


Figure B7 The vowel *o* between an alveolar trill and a voiced palatal stop pronounced by speaker w2 in the narrative style.

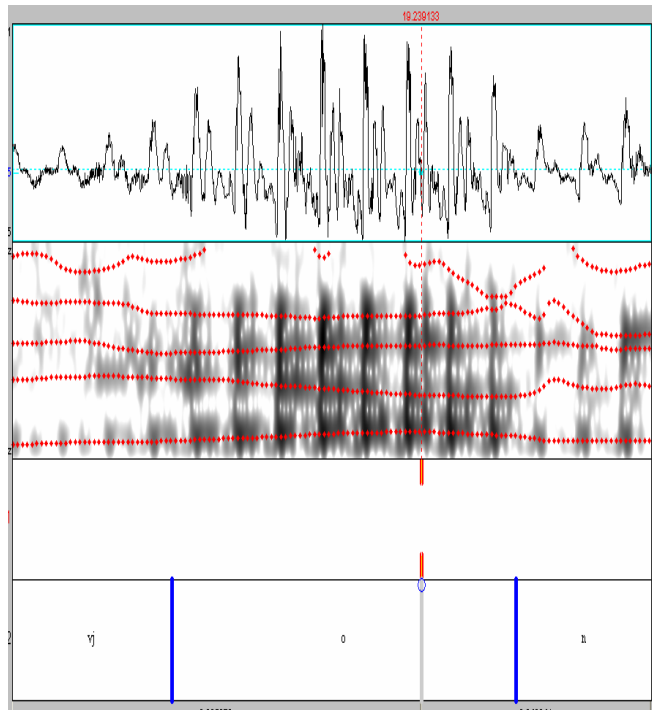


Figure B8 The vowel *o* between a voiced palatalized labiodental fricative and a dental nasal pronounced by speaker m2 in the fast reading style.

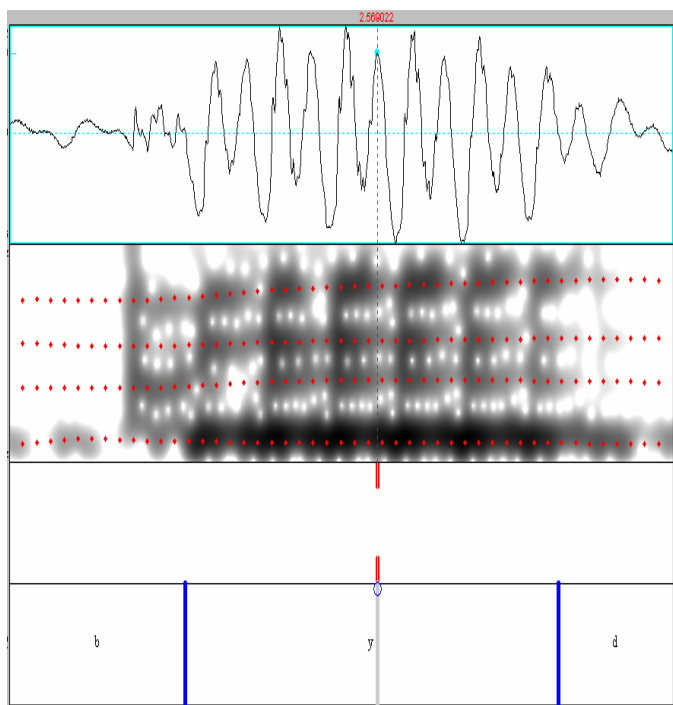


Figure B9 The vowel *y* between a voiced bilabial stop and a voiced dental stop pronounced by speaker w2 in the normal reading style.

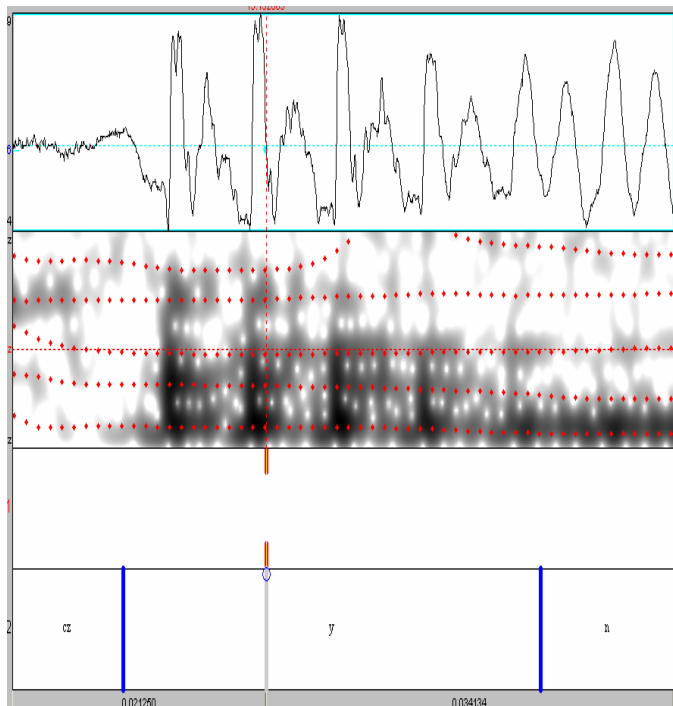


Figure B10 The vowel *y* between a voiceless alveolar (retroflex) affricate and a dental nasal pronounced by speaker m2 in the fast reading style.

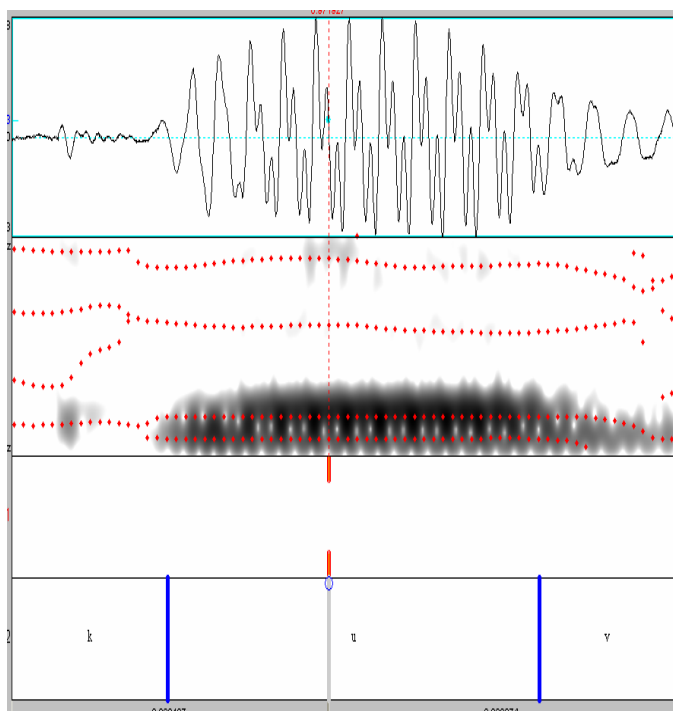


Figure B11 The vowel *u* between a voiceless velar stop and a voiced bilabial stop pronounced by speaker w2 in the fast reading style.

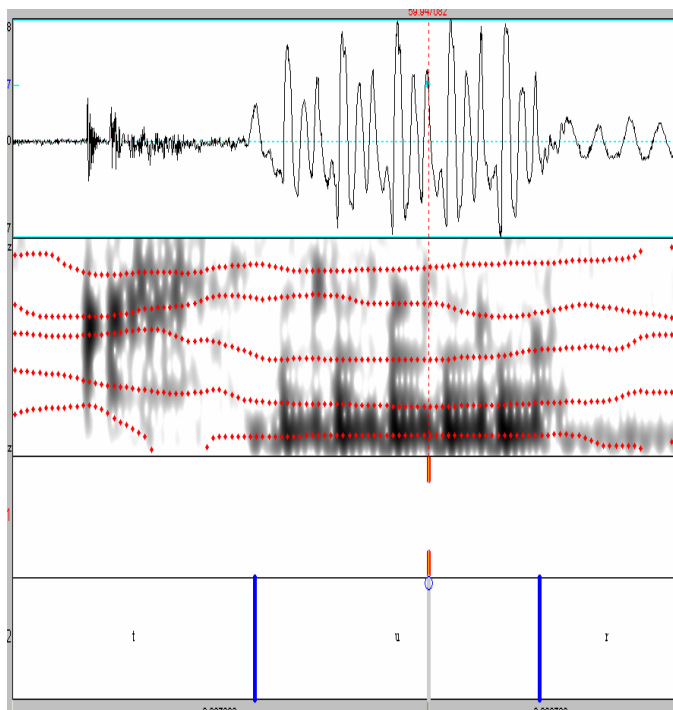


Figure B12 The vowel *u* between a voiceless dental stop and an alveolar trill pronounced by speaker m2 in the narrative style.