Neutralization of Word-Final Voicing in Russian*

Miriam Shrager

Abstract: This paper has two aims. The first is to describe a pilot instrumental study of the incomplete neutralization of Russian final dental stops /t/ and /d/. This study refutes the results of a previous instrumental study of word-final voicing neutralization, which suggested that /t/ and /d/ are completely neutralized word-finally. The study examines several phonetic quantities that might be correlated with incomplete neutralization and serve as cues for the correct classification of voiced and voiceless obstruents. The second aim is to bring forward an extensive summary and discussion of previous studies and theories on incomplete neutralization.

1. Introduction

Many languages have the phenomenon of voicing neutralization in consonants. In certain environments the voicing distinction is "suppressed" or neutralized, so that only a voiced or a voiceless sound is pronounced in that environment, but never both. One of the most familiar examples of voicing neutralization is the realization of underlying voiced obstruents as voiceless in word-final position. For example, in German /gib/ and /gip/ are pronounced [gip] according to the standard phonological description (Jakobson, Fant, and Halle 1952: 8).

Yet, ample evidence from the past 20 years suggests that various languages in which neutralization is supposed to occur actually show incomplete neutralization, i.e., the segment does not lose its voicing specification completely but realizes its underlying specification in production to an extent that is identifiable to listeners at a rate better than chance. More concretely, there are a number of acoustic cues for voicing that the listener uses to determine the segment intended by the speaker, such as duration of voicing into stop closure (a primary cue)

* I would like to thank Michael Thompson, Christina Bethin, and the two anonymous referees, for their helpful comments. The results of this experiment were first presented at the Annual Pan-American Meeting of the Acoustical Society of America (Shrager 2002).

Journal of Slavic Linguistics 20(1): 71–99, 2012.

and duration of burst release (a primary cue for aspiration that is often a secondary cue for voicing). In complete neutralization these cues are entirely absent, meaning that there is no difference in the acoustic measurements between the realizations of corresponding voiced and voiceless segments, in which case it is expected that listeners will identify the underlying voicing of the segment at rates comparable to those of chance. In incomplete voicing neutralization, however, at least one voicing cue differs significantly in the realizations of the underlying voiced and voiceless segments (the two populations on which statistical tests are run), which shows up experimentally as a statistically significant difference between the two populations; in cases of incomplete neutralization this will be used by listeners to identify the voicing of underlying segments at rates statistically better than chance. Experimental phonetic studies have shown that incomplete neutralization is found in Catalan, Polish, German, and Dutch where additional low-level features (that is, secondary acoustic cues) appear to be involved in incomplete neutralization.

In a study by Dinnsen and Charles-Luce (1984) on final devoicing in Catalan, five speakers read a list of carrier sentences with embedded minimal-pair words. The study found that final devoicing differs by speaker, and that the environment has an effect on incomplete neutralization. Thus, vowels shorten before a word-final obstruent if the next word begins with a consonant, but for some speakers they shorten significantly less if the word-final obstruent is underlyingly voiced (Dinnsen and Charles-Luce 1984: 56).

In a study on Polish neutralization (Slowiaczek and Dinnsen 1985), five speakers read frame sentences containing words from 15 minimal pairs differing in the voicing of the final obstruent. Overall the study found that underlying voicing affected the durations of several phonetic measures. For example, vowel duration before the underlying voiced obstruent was about 10 ms longer than before the voiceless one, a significant difference that held across all speakers. Similarly, the closure duration of the underlying voiced obstruents was found to be shorter than that of the underlying voiceless obstruents, though this was significant only in certain environments for some speakers. Additionally, for labial stops a statistically significant difference of about 13 ms was found in voicing after the acoustic offset.

Tieszen (1997) examined dialectal differences in word-final devoicing in Polish. Nine minimal word pairs were put in carrier sentences in different environments and some of them were embedded in a reading passage. Five speakers from three different dialect areas read the passage first and then the sentences. Tieszen looked at three variables: vowel duration, consonant closure duration, and glottal pulsing into the closure. The results showed that in two of the three examined dialects, word-final devoicing is not complete in Polish, and that the voicing cues associated with the incomplete neutralization of voicing differ by dialect. In the third dialect the duration differences were not statistically significant for the three variables; thus, it is possible that other voicing cues not examined in this study are involved.

The studies by Port and O'Dell (1985) and Port and Crawford (1989) deal with final voicing neutralization in German, for example, in such words as *Bund* 'group' and *bunt* 'colorful'. In their studies, the authors tried to determine whether neutralization is really complete and, in the case of incomplete neutralization, tried to find the variables that contribute to the difference. In a production experiment (Port and O'Dell 1985), 10 German subjects read a list of words that included 20 target words spread randomly among decoy words. Four variables were measured: the duration of the vowel preceding the final consonant, the final consonant closure, voicing into the closure, and the release burst of the final consonant. The results demonstrate that for three of the four variables the difference in duration was significant; vowel duration before underlying voiced stops was significantly longer by about 15 ms, voicing into the closure was longer by 5 ms, and release burst duration was shorter by 15 ms. In a perception experiment of the same study in which 10 German subjects had to identify the correct words, 59% of the words were correctly identified. A multiple regression analysis of the four acoustic variables on the number of voiced responses was performed. It showed that longer burst release correlated negatively with perception of [+voice], while a longer vowel nucleus correlated positively with the number of voiced responses.

In a later study, Port and Crawford (1989) tried to address issues that emerged from criticisms expressed by Fourakis and Iverson (1984). One criticism was that the results of Port and O'Dell (1985) were artificial, i.e., when subjects read a word list, they are influenced by orthography, causing a secondary process of "deneutralization" to occur. Moreover, it was possible that subjects tried to help the non-native tester by reading the list so as to disambiguate the minimal pairs. To address these issues, Port and Crawford (1989) used three minimal pairs read by five speakers under five different pragmatic conditions. First, the target words were embedded in sentences that the subjects read as a list. Second, the sentences were read to the subjects by the German-speaking assistant and recited from memory. Third, the subjects contrasted the test words with each other in explanatory phrases. Fourth, the subjects dictated sentences to the German-speaking assistant that emphasized the minimal pair words semantically. Fifth, the subjects read a list of words that contained the six target words of the three minimal pairs in random order among twelve distractors. In all five pragmatic conditions, five variables were measured: vowel duration, stop duration, burst duration, nasal duration (in *Bund/bunt*), and glottal pulsing into the stop closure. These variables were then subject to standard statistical analyses to determine which of them significantly differed with underlying voicing.

Analysis of variance, in which the distinctiveness of a single dependent variable at a time is examined, demonstrated that when the data were pooled across speakers, conditions, and word pairs, the duration of the burst release was significantly different between the two groups of voiced/voiceless stops in coda position; the other variables were not significantly affected by underlying voicing. However, discriminant analysis, which combines all variables to best distinguish between the two groups in the data (voiced/devoiced), succeeded in distinguishing the voiced from the voiceless tokens 64% of the time across all five pragmatic conditions used in the experiment (Port and Crawford 1989). Discriminant analysis was also tested separately on each condition. The least contrast was found when tokens were embedded in disguised sentences with no significant difference in identification rate between reading the sentences and reciting them (55% vs. 56%, respectively). The greatest contrast was found when subjects dictated the contrastive sentences (78%) or read from a word list (62%).

This experiment demonstrated that voicing neutralization in German is incomplete to various degrees depending on speech styles and that speakers are able to control the degree of neutralization (for example, in the contrastive sentence tasks). When discriminant analysis was used across all the data for each of the three word pairs separately, it identified the underlying voiced stop in 67% of the cases. A subsequent perception experiment in the same study showed that listeners had 69.2% correct identification of voiceless/voiced pairs. Since discriminant analysis showed a similar identification rate, the assumption was made that discriminant analysis can be regarded as comparable to human speech perception.

Warner, Jongman, Sereno, and Kemps (2004) examined final devoicing in Dutch. In their study they conducted production and perception experiments and investigated speaker-specific production differences. In addition, they looked at the influence of orthography on durational differences associated with Dutch incomplete neutralization. The production experiment was a large study as the regards the number of subjects and of items; 15 subjects read a list containing 20 minimal pair words with final voiced/voiceless coronal stops (/d/, /t/) following phonemically short and long vowels (10 minimal pairs each), and 16 minimal pair words in intervocalic position where neutralization does not occur. The study showed that only vowel and burst durations were significantly associated with underlying voicing. Vowel duration (averaged for long and short vowels), although much smaller than in Polish and German, was still significantly longer (by 3.5 ms) before underlying /d/ than /t/; burst duration was significantly longer (9 ms) after /t/ than /d/ in words with a long vowel before the target, but not following short vowels. The perception experiment with 30 listeners showed that listeners are able to distinguish words with final /d/ and /t/ at significantly better than chance accuracy for speakers who produced longer vowels before /d/ than /t/. Manipulation of vowel duration confirmed that listeners use it as a cue to underlying final voicing. Interestingly, the same was found to be true of closure duration when it was the only manipulated parameter, suggesting that in perception listeners can use just one cue to identify underlying voicing.

Ernestus and Baayen (2003, 2006, 2007) also tested neutralization effects in Dutch (specifically focusing on how speakers interpret the voicing of neutralized final segments). In the 2003 study, participants were presented with nonce-verbs with a final neutralized obstruent and asked to create past-tense forms from them by adding the past tense suffix (*-de*/*-te*), which makes the voicing of the underlying obstruent resurface. They found that voicing of a neutralized obstruent is correlated with its place and manner of articulation and with the characteristics of the preceding segment.[1] Their main conclusion was that speakers of Dutch rely on their knowledge of the underlying representations of similar words, which they use as exemplars.

In a subsequent study Ernestus and Baayen (2006) investigated whether listeners interpret the stem-final obstruent using subphonemic cues. In a production experiment, Dutch speakers read a list of pseudo-verb stems that were spelled either with voiced or voiceless final obstruent. The results showed a significant difference in the duration of release noises (burst plus the following period of aspiration), with the segments corresponding to the graphemes *b* and *d* shorter than those corresponding to *p* and *t* by a difference 23 ms. A subsequent experiment tested whether Dutch listeners add the voiced variant of the past tense suffix *-de* to pseudo-verbs when the final obstruent is not completely devoiced more often than to those in which it is completely voiceless. The results confirmed that Dutch listeners use subphonemic cues in the speech signal, here for incomplete devoicing, when choosing the voicing specification of the final obstruents in nonce-words. In addition, the results showed that listeners base their choices on exemplars of similar existing words (also Ernestus and Baayen 2003). In their subsequent study on Dutch final neutralization, Ernestus and Baayen (2007) found that vowels preceding voiced fricatives were on average 16 ms longer than vowels preceding voiceless fricatives.

In all the studies above, the durational differences between underlying voiced and voiceless obstruents are the same as those found in the language in non-neutralizing environments. Namely, vowels are longer before underlying voiced obstruents, and voicing into the closure, obstruent closure, and burst duration is shorter for the voiced obstruent.

2. Russian Neutralization

In the phonological system of the Russian language, almost all obstruents participate in an oppositional relationship of voiceless-voiced. According to the established view, this opposition is neutralized in certain weak positions (Shcherba 1911), such as (a) word-finally, (b) be-

---

[1] Twenty-four percent of the nonce-words with voiceless obstruent were interpreted as voiced, with velar fricatives having the highest percentage.

fore another obstruent other than /v/ with either voicing specification, and (c) across word boundaries, i.e., assimilation to the initial voicing specification of the initial obstruent in the next word. For instance, both /d/ and /t/ become [t] word finally, hence *kod* 'code' and *kot* 'cat' are both realized as [kot]. Thus, the underlying voicing specification of the final obstruent of a root or stem is only apparent when it occurs before a vowel, a sonorant, or *v*. Consider the following examples in (1), in which the obstruent in question is italicized:

(1)

Not neutralized	Neutralized
a. vo*d*a 'water' [d]	a. vo*d* 'of waters (gen pl)' [t]
b. vo*d*nyj 'water' (adj)[dn]	b. *s*dal '(I) gave away' [zd]
c. *d*ver 'door' [dv]	c. o*v*ca 'a sheep' [fc]
d. *T*ver' (city name) [tv]	

In spite of the fact that Russian word-final devoicing is commonly cited as a classic example of neutralization, there has been very little work addressing the question of whether the neutralization is complete or incomplete. In fact, only one instrumental study has been fully dedicated to incomplete neutralization in Russian, that of Pye (1986) in which she conducted a production experiment on the realization of word-final voicing of Russian obstruents: bilabial, coronal, and velar stops, and dental (*s, z*) and alveolar (*š, ž*) fricatives.[2] Five subjects recorded a randomized sentence list including 18 minimal target pairs and 17 dummy pairs. The words were read in two different sentence frames to introduce variation. For each place of articulation of the target obstruent, half the pairs were in frame 1 and the other half in frame 2; the dummy words were placed in frame 3. For each frame two sentences were used, the first (labeled (i) below) to introduce the target word, and a similar second sentence (labeled (ii) below) in which sentence stress would fall on the word *tože* 'also' rather than the target. Pye's (1986: 3–4) sentence frames 1–3 are summarized in (2) below:

(2) a. i. Učitel' skazal "______________".
The teacher (male) said "__________".

[2] The fricatives *s* and *z* are dental in Russian, and *š* and *ž* are alveolar.

(2) a. ii. Učitel'nica tože skazala "__________________".
The teacher (female) also said "__________".

b. i. Perevodčik povtoril "______________".
The translator (male) repeated "___________".

ii. Perevodčica tože povtorila "____________".
The translator (female) also repeated "___________".

c. i. Učenik napisal "______________".
The student (male) wrote "__________".

ii. Učenica tože napisala "___________".
The student (female) also wrote "__________".

As can be seen from (2), the target words were placed in sentence-final position to avoid voicing assimilation across word boundaries.

Four durations were measured: (a) vowel duration, (b) the duration of the vowel and liquid sequence /ol/ in the words *stolp*/*stolb*, (c) obstruent duration, and (d) the duration of voicing into the obstruent. Data pooled across the five subjects showed differences in vowel duration in both sentence contexts between underlying voiced and voiceless obstruents that varied according to the place of articulation. The durational differences were on the order of 5–20 ms, with the exception of vowels before labial stops in sentence context (i) and velar stops in sentence context (ii). The least difference in vowel duration was found preceding the non-palatalized coronals *t*/*d*.

A difference in consonant duration was also found, but it varied greatly by manner of articulation across speakers. In the two sentence contexts, the duration of the underlying voiced obstruent was found to be shorter than that of its voiceless counterpart by a difference on the order of 6–30 ms (with exceptions). Again, the least difference was seen in coronals, and the duration of /d'/ was not shorter than that of /t'/. The durational differences for the pair *stolp*/*stolb* containing the vowel + liquid /ol/ combination were similar to the differences found in the other pairs of words: in the two sentence contexts, vowel duration was longer before underlying /b/ than before /p/, and the obstruent duration was shorter for /b/ than for /p/. Voicing into the obstruent showed a large difference, with longer glottal pulsing before the underlying voiced obstruents than the voiceless ones. The overall differences that Pye found (excluding the /ol/ sequence) for the two sentence contexts ((2a) and (2b)) are summarized in Table 1 below.

Table 1. Durational Differences Associated with Voiced and Voiceless Final Obstruents (Pye 1986)

		VD (%)	CD (%)	ViC (ms)*
b/p	a.	35.8	–15.2	22.6
	b.	8	–14.9	15.6
d/t	a.	9	–1.8	29.1
	b.	0.7	–3.9	21.0
d’/t’	a.	13.6	–13	33.1
	b.	7	6.7	29.2
g/k	a.	16.8	–10	16.8
	b.	34	–40	28.0
z/s	a.	9.6	–12.2	12.25
	b.	9.8	–23	12.0
ž/š	a.	21.1	–5.1	2.3
	b.	11.5	–13.5	1.0

VD=Vowel Duration, CD= Consonant Duration, ViC=Voicing into Consonant
*Pye (1986) did not provide the percentages for ViC.

When the data for all obstruents were pooled for each subject, the results varied greatly.

Although no statistical tests were performed, Pye (1986) concluded that in Russian underlying voicing specifications can be distinguished by vowel duration except for the non-palatalized dental stops (coronals). Obstruent duration, however, was not found to distinguish voiced and unvoiced, as the differences varied greatly according to subject and manner of articulation. Thus, Pye demonstrated that final voicing neutralization in Russian is incomplete for certain obstruents, but not for others.

3. Current Study

As mentioned above, Pye (1986) found that voicing neutralization is incomplete for some obstruents in Russian, but not for the dental stops *t*/*d*. Accordingly, the goal of the current study is to examine the voicing neutralization of coronal stops using the methodologies and statistical analyses introduced in the experiments by Port and O'Dell (1985) and Port and Crawford (1989). Besides reexamining the variables that Pye used, an additional variable was added: the burst durations of final obstruents. Furthermore, while both Pye's study and this one are based solely on production, the current study employs discriminant analysis (a statistical test described in more detail below), which in the experiments of Port and O'Dell (1985) and Port and Crawford (1989) showed results similar to those of their perception experiment and which might therefore be considered a weak model of speech perception. Since the goal of the current study was to fill a gap in Pye's study, only the stops *t/d* were examined.

3.1. Methods

Three subjects were given a list of words containing randomly-distributed minimal pairs of voiced–voiceless coronal stops /t/ /d/ in final position. The subjects, while performing the reading task and afterwards, were kept ignorant of the purpose of the experiment.

3.1.1. Corpus

In Pye's study, coronal stops as opposed to the other consonants were found to completely neutralize word-finally in the durations measured. Therefore we wanted to put the target words in a context that would be most likely to induce a difference. According to the findings of Port and Crawford (1989), the greatest contrast due to underlying voicing in German was found when subjects dictated contrastive sentences (78%), but when subjects read a word list, the contrast was found to be 62%. We decided to apply the second context, the word list, since it induced sufficient contrast but was less artificial.

The target words were presented in a word list written by a native speaker in cursive Russian orthography. The eight target pairs (Table 2) were scattered among sixteen distractors, which together were presented as a list of random words. For the recordings, the word list was read three times in a row by each subject. The total analyzed corpus consisted of 144 tokens (16 words repeated 3 times by 3 speakers).

Table 2. Target Word Pairs

Voiced		Voiceless	
kod	'code'	kot	'cat'
l'ed	'ice'	l'ot	'flight'
l'ud	'people'	l'ut	'fierce'
mod	'fashions (gen)'	mot	'squanderer'
obéd	'dinner'	obét	'vow'
pod	'under'	pot	'sweat'
rod	'type'	rot	'mouth'
vod	'waters (gen)'	vot	'here'

3.1.2. Subjects and Procedure

At the time of the experiment (autumn 2001) it was not easy to find recently arrived native speakers of standard Russian of similar age and gender in a small American university town. Consequently, the experiment had to be conducted with a relatively small number of subjects, three female graduate students at Indiana University in Bloomington (IUB) who had recently come to the United States as adults to attend graduate school. Although two of the subjects were graduate students in linguistics, their specializations were outside of phonetics, and none of them had any background in acoustic phonetics. All three subjects were speakers of standard Russian and their stay in the United States at the time had been minimal. Two of the subjects were from Moscow and one was from St. Petersburg. The experiment was conducted entirely in Russian by a native Russian speaker.

Subjects were recorded individually in the recording laboratory of the IUB linguistics department onto DAT (Digital Audio Tape). Subjects were given the list of words as described above, which they read three times each. Words were randomly distributed in the lists, except that members of minimal pairs were not allowed to occur adjacent to each other in order to eliminate consciously-introduced contrasts. Subjects familiarized themselves with the word list prior to recording and the inflected forms were explained (e.g., *mod*: gen pl of *moda* 'fashion') as were words (*l'ot, l'ud, l'ut*) that are known to any native speaker, but rarely used. Subjects were asked to read the list at a normal reading speed with individual words separated by intervals of no less than two seconds and without any special emphasis on the pronunciation of the words, in order to eliminate hypercorrect slow "reading pronunciations," on the one hand, and voicing assimilation across word boundaries in a fast reading, on the other.

The subjects read the list without having been told it contained minimal pairs, and as they read it to a native speaker of Russian who had prepared the list (the author), there was no need for them to emphasize the correct words or spelling. The subjects did not know each other and were asked not to discuss the experiment with others.

3.1.3. Measurements

The author took the measurements by hand in the phonetics laboratory at IUB, using SoundScope implemented on a Mac workstation. For each target word four measurements were taken: (a) the duration of the vowel preceding the target stop, (b) the duration of the final stop, (c) duration of voicing into the stop closure, and (d) the burst duration of the final stop.

Both spectrograms and waveforms had to be used for every measurement in the temporal analysis. The vowel duration was measured from the onset in the regular waveform of vowel characteristics until the end of the regular waveform, which in the wideband spectrogram corresponded to the end of F2. Note that temporal measurements in words with initial liquids required more attention. Thus, in the pair /rod/–/rot/ it was hard to find the r/o boundary, as the rounding starts at the onset of /r/. The first two formants, F1 and F2, remained the same through the whole r/o combination, but a transition can be seen in F3 and F4 at a certain point after the initial liquid. In this case, the first glottal pulse on the waveform after the beginning of change in F3 and F4 as seen in the spectrogram was taken as the vowel onset. The measurements of the two pairs of the palatalized laterals, /l'ot/–/l'od/ and /l'ut/–/l'ud/, relied both on the waveform and the spectrogram in a similar way. The transition from the lateral to the vowel manifested itself in two ways: a sudden discontinuity in the envelope of the waveform and a corresponding pinch in F1 on the spectrogram.

The voicing into closure was indicated by glottal pulses after the end of the vowel that appeared in both the waveform and the wideband spectrogram. Measurements of consonant duration were taken from the point where these glottal pulses ceased to be visible to the release of the stop. In the final analysis these two measurements were combined to calculate the total duration of the final stop, which was added as an additional variable, (e) Total Closure.

The burst duration was measured from the start of the sudden burst of turbulent noise after stop release, as seen on the waveform, to the end of aspiration noise visible on the spectrogram.

3.1.4. Analysis

Results were analyzed statistically using the SPSS (Statistical Package for the Social Sciences) tests for analysis of variance (ANOVA) and repeated measures (RM) logistic regression (fit using generalized estimating equations (GEE)). Logistic regression was used because our outcome variable is binary (1 = voiced, 0 = voiceless).[3] Repeated measures are included to account for the 16 tokens and 3 replicas within each subject, which reduces interspeaker variability and allows us to have a general model. In addition, we used two classification tests, discriminant analysis and the logistic regression test. Discriminant analysis was used as a means of measuring the degree of contrast by combining several variables, as in the experiment by Port and Crawford (1989). Discriminant analysis is used when the cases being observed fall into discrete categories, such as voiced and voiceless, and the characteristics measured in the experiment have continuous values, such as durations. For two distinct categories, one linear combination of the measured variables—the discriminant function—is constructed that best predicts to which category an individual case belongs. In effect, a threshold value for the discriminant function is calculated, and if the value of the discriminant function for a given case is greater or less than the threshold, its predicted category membership falls in one or the other group (Klecka 1980; Port and Crawford 1989). In the social sciences, discriminant analysis is widely used to predict the class membership of future observations, but in perception exper-

[3] Linear statistical models give an expected probability that is calculated from a linear combination of independent variables, which means that in principle and often in practice, the model, unfortunately, will give expected probabilities greater than 1 or less than 0, which are meaningless. Logistic regression avoids this problem by using odds (the log odds of which can range in value between $-\infty$ and $+\infty$) rather than probabilities, which also makes it well-suited to problems involving a binary dependent variable.

iments it is most often used to study the relative significance of each variable and the overall adequacy of the model.

While, as previously mentioned, the results of discriminant analysis in previous studies of incomplete neutralization proved to be similar to those of a human perception test (Port and Crawford 1989), logistic regression better accounts for variability among speakers in a binary-choice test. Thus, both types of test were used in this study and their results were subsequently compared.

4. Results

The subjects showed great variability in measurements, both in degree and character of voicing into the closure: subject 1 showed a number of pulses consistently much lower than those of subjects 2 and 3. Sometimes the voicing extended all through the closure almost until the release. However, the overall variable mean pooled across all the three speakers (Table 3) shows that the mean durations of the four variables, the preceding vowel (VD), the target consonant closure (CD), the voicing into the closure (VC), and the burst (BD) all point to incomplete neutralization. In Table 3 the mean vowel duration for the voiced group is slightly longer than for the voiceless; closure duration is longer for the voiceless; voicing into the closure is longer for the voiced; and burst duration is longer for the voiceless. Although the differences between the voiced and voiceless obstruents are small, they are consistent across all five measures and match the expected characteristics of the opposed members, +voice/–voice, when they are not neutralized.

Table 3. Group Statistics. Variable Means According to Final Voicing Feature

Varbl	FV	Mean(ms)	Differences	Std.Dev
VD	d	172	3	47
	t	169		51
CD	d	94	–2	42
	t	96		41
VC	d	37	1	18
	t	36		21
BD	d	58	–22	28
	t	80		24

To test the significance of the differences for each variable across the three speakers, accounting for speaker variability,[4] we used the Repeated Measures Logistic Regression test. The results (in Table 4) show that the most significant variable was burst duration (BD). Also significant was consonant duration (CD), and somewhat significant was voicing into the consonant (VC). Vowel duration proved non-significant. Tables 5 and 6 give more details on the Regression test of Table 4; Table 5 shows the parameters estimates, and Table 6 shows goodness-of-fit.

Table 4. Repeated Measures Logistic Regression: Test of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	14.648	1	.000
VD	0.479	1	.489
CD	5.395	1	.020
VC	3.381	1	.066
BD	25.008	1	.000

Dependent Variable: voice; Model: (Intercept), VD, CD, VC, BD

[4] Thanks to an anonymous reviewer for making this suggestion.

Table 5. Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval	
			Lower	Upper
(Intercept)	9.068	2.3695	4.424	13.712
VD	–8.259	11.9306	–31.643	15.125
CD	–26.076	11.2269	–48.081	–4.072
VC	–12.360	6.7218	–25.534	0.815
BD	–69.015	13.8008	–96.064	–41.966
(Scale)	1			

Dependent Variable: Voice Model: (Intercept), VD, CD, VC, BD

Table 6. Goodness of Fit[b]

	Value
Quasi Likelihood under Independence Model Criterion (QIC)[a]	163.223
Corrected Quasi Likelihood under Independence Model Criterion (QICC)[a]	166.030

Dependent Variable: Voice Model: (Intercept), VD, CD, VC, BD

a. Computed using the full log quasi-likelihood function.

b. Information criteria are in small-is-better form.

Following the analysis of Port and Crawford (1989), we wanted to use a classification analysis. Initially we used discriminant analysis as in Port and Crawford (1989) for reasons previously stated. The data of the current study were pooled across all variables to find the best linear combination of the variables to predict into which of the two groups a given segment would fall. This analysis, summarized in Table 7, demonstrates that a total of about 74% of tokens overall were correctly classified as having voiced or voiceless final consonants; 69% of the underlying voiceless tokens were correctly classified, while voiced tokens had a higher rate of correct identification, 78%.

Table 7. Discriminant Analysis Classification Results

	FV	Predicted Group Membership 0	1	Total
Original	0	50	22	72
Count	1	16	56	72
Percent (%)	0	**69**	31	100.0
	1	22	**78**	100.0

On average 73.6% of original grouped cases (voiced and voiceless) were correctly classified. (Voiced = 1 Voiceless = 0)

Taking into account the variable analysis in Table 3, we expect that the most salient factor in the classification of the voiced tokens was shorter burst duration. And indeed, the structure coefficients in Table 8 show that the classification of voicing was based most strongly on the burst duration. (The greater the absolute value of an entry in the structure matrix, the more salient the corresponding factor.)

Table 8. Structure Matrix

	Function **1**
Burst Duration	.750
Vowel Duration	–.062
Voice into Closure	–.055
Closure Duration	.049

Pooled within-groups correlation between discriminating variables and standardized canonical discriminant functions.

In addition, we decided to use a different classification test—logistic regression—taking into account repeated measures for the differences of each speaker (Table 9). In this analysis a linear combination of the four variables was used to calculate a predicted probability of the word being voiced, based on an initial transformation from probability to odds. If the predicted probability was higher than 0.5 then it was considered as voiced (1) and if less than 0.5 then it was unvoiced (0). This classification was compared to the underlying voicing of the 144 words, showing that among the 72 words with underlying *t*, 53 words (74%) were classified correctly, and of the 72 words with underlying *d*, 55 words (77%) were classified correctly.

It is noteworthy that the overall average classification results of the logistic regression test (75%) are similar to those of discriminant analysis (74%), even taking into account the fact that the Logistic Regression test better accounts for variability among speakers.

Table 9. Logistic Regression Classification Results

voice * classify Crosstabulation

			classify 0.00	1.00	Total
voice	0	Count	53	19	72
		% within voice	73.6%	26.4%	100.0%
	1	Count	17	55	72
		% within voice	23.6%	76.4%	100.0%
Total		Count	70	74	144
		% within voice	48.6%	51.4%	100.0%

0 = underlying voiceless /t/, 1 = underlying voiced /d/
Overall, 75% of the data were correctly classified

The classification results are illustrated with the box plot graphs below. Figure 1 illustrates the difference of the predicted value of the mean response found by the classification test of the underlying voiced (1) and voiceless (0). The label "voic" in Figures 1–3 stands for "voicing. "

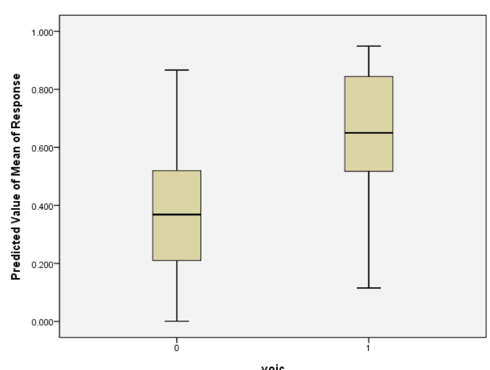


Figure 1. Predicted Value of Mean Response

In Figures 2 and 3 box plots were made in which only one variable at a time is treated as contributing to the classification results. Figure 2, which shows the voicing classification based on the burst duration (BD), illustrates well that BD contributed the most to the classification results in Tables 5 and 7. Figure 3, on the other hand, illustrates that VD contributed very little, if at all.

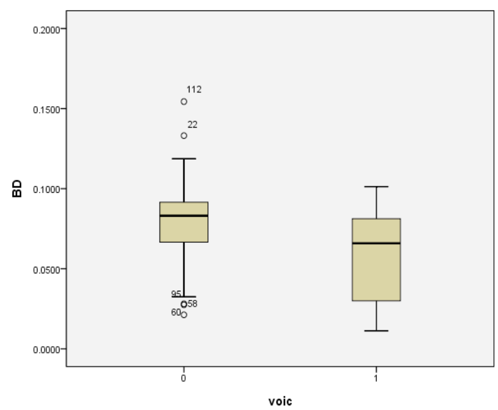


Figure 2. Classification of Voicing by Burst Duration (BD)

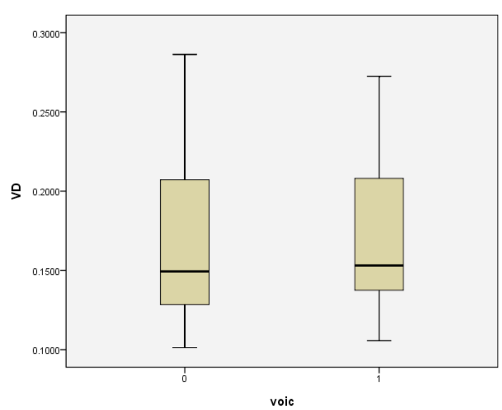
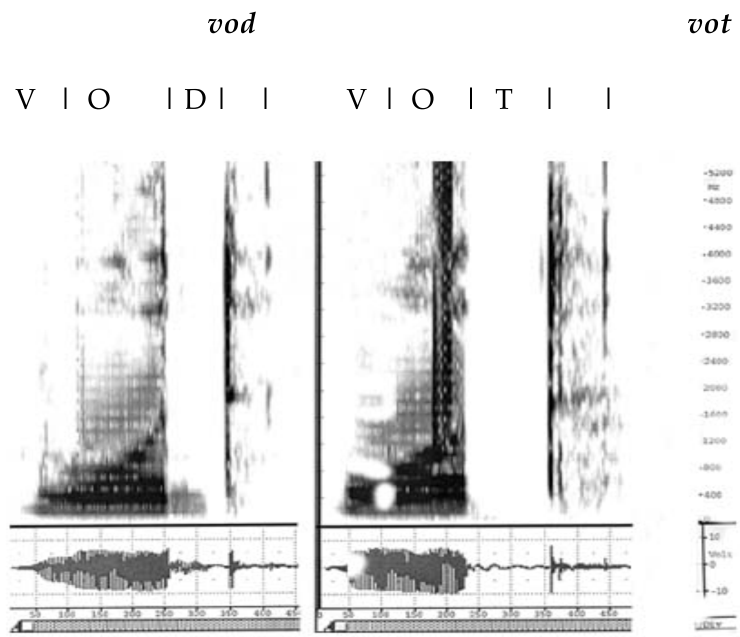


Figure 3. Classification of Voicing Based on Vowel Duration (VD)

Figure 4 most vividly illustrates incomplete neutralization as seen on the spectrogram and wave forms of two words, *vod* 'water' (pl gen) and *vot* 'here'. Both the spectrogram and the wave form show the differences between the two words: the closure duration is longer in *vot*, the vowel duration is longer in *vod*, and there is voicing into the closure of *vod*. The difference in burst duration, statistically the most significant variable, is especially apparent in the spectrogram, which strikingly exemplifies the greater burst duration and intensity for *t* compared to *d*.

Figure 4. Representative Spectrograms and Wave Forms of the Target Pair *vod* - *vot*



5. Summary and Discussion

If we compare this study with that of Pye (1986), we see that contrary to her results, the dental stops /t/ and /d/ show incomplete neutralization. As for the relative contributions of specific variables, this study accords with Pye's in finding that the duration of preceding vowels acts as the least salient cue for the underlying voicing of /t/ and /d/. This is contrary to what was found for Dutch coronals /t/ and /d/ in the experiment of Warner, Jongman, Sereno, and Kemps (2004). However, vowel duration is recognized as being significantly affected by the voicing of adjacent consonants in the Germanic languages but not in most Slavic languages, and thus this is to be expected.[5] That is, this study confirms that for Russian /t/ and /d/ it is not the vowel but the consonant that contains the cues for voicing in incomplete neutralization; primarily burst duration (a factor Pye did not include), then consonant duration and voicing into the consonant closure. Overall, it is clear that Russian can be included with German, Dutch, Catalan, and Polish as showing incomplete neutralization of word-final obstruents.

In this study as well as others (Port and Crawford 1989; Warner, Jongman, Sereno, and Kemps 2004; Ernestus and Baayen 2006), the burst duration of the final consonant emerges as a strong cue for incomplete neutralization. This happens because a partial devoicing gesture in time results in weaker or partial voicing, so that in word-final position longer burst duration acts as a cue for voiceless stops. Physically, this could be explained by the airflow mechanism. The glottis is more open with voiceless stops than with voiced stops. This difference, which is essentially one of the articulatory features encoded by the phonological feature of [voice], is presumably maintained by the stops even in neutralizing positions, in which the effect of a following voiceless segment or silence is to cause the vocal folds not to vibrate. Thus, there is less resistance to the air flow during the production of voiceless stops, and more pulmonic force is applied to the primary constriction. This affects the intensity and duration of the burst at

[5] Additional works besides those discussed in this paper are: for English, Jones 1950, House and Fairbanks 1953, Peterson and Lehiste 1960, Sharf 1962, Chen 1970, Cochrane 1970, Klatt 1973, 1976, Hogan and Rozsypal 1980, Luce and Charles-Luce 1985, Kozhevnikov and Chistovich 1966; for German, Chen 1970; for Norwegian, Chen 1970; for Dutch, Slis and Cohen 1969; for Czech and Polish, Keating 1985, Podlipsky 2008.

release. Since longer burst duration is the result of a more open position it may serve as a cue for voiceless obstruents, regardless of the lack of vocal fold vibration, both in production and in perception processes.

Although this study was based on a production experiment, it employed discriminant analysis, which proved to be similar to human perception in previous studies (Port and Crawford 1989). In our study discriminant analysis and the logistic regression test had similar percentages of correct classification (74%–75%). Thus, our analysis shows that acoustic cues present in the signal are sufficient to separate the categories statistically. A subsequent perception study would need to test the role played by temporal factors acting as cues for correct sound categorization. Additionally, it would be interesting to test whether a perception test would have similar percentage of classification as the statistical tests.

After the current paper had been submitted for publication, a new paper describing Russian incomplete naturalization came out in print (Dmitrieva and Jongman 2010).[6] Although the authors primarily examine the effect of second language proficiency in English on neutralization in Russian, their study also includes an experiment similar to that described in this paper. However, their experiment concerns obstruents with all the places of articulation for two groups of participants: those who had been significantly exposed to English (resident in the United States from one to ten years) and participants with no significant exposure to English (resident and recorded in Russia). The same durations were measured as in the current experiment: vowel, closure, voicing into the closure, and release. The analysis included a two-way ANOVA test with Underlying Voicing and Knowledge of Second Language as independent variables, but did not include any classification tests as in the current study. The results showed that knowledge of English affects Russian final neutralization. Thus, the averages of durational differences for Russians who lived in the United States were: vowels 6 ms, closure –16 ms, voicing into the closure 4 ms, and stop release 17 ms. All these differences were statistically significant. On the other hand, durational measurements of the Russian speakers with no English influence resembled those of the current study, except for closure duration (see Table 3): vowels 2 ms, closure –16 ms, voicing into

[6] Thanks to Christina Bethin for pointing this out to me.

closure 1 ms, release portion –16 ms. Only release duration and (to a lesser extent) closure duration were statistically significant. Overall, their study found a significant effect of knowledge of second language on voicing for all variables except for closure duration. In addition, they conducted a separate one-way ANOVA test for each group of Russian speakers whose results showed that for the group of Russian speakers without knowledge of English, the significant main effect was only for release duration. This result is very similar to those obtained here (see Tables 3, 4, 5, and 8). Closure duration in our experiment proved less significant than in Dmitrieva and Jongman 2010 due to the restricted scope of our experiment, the Russian dentals *d*/*t*, while they (like Pye 1976) examined all Russian obstruents. We reiterate that the obstruents *d*/*t* have smaller incomplete neutralization effects than do other Russian obstruents, which explains the differences between our results and those of Dmitrieva and Jongman (2010).

5.1. Further Discussion

There are other general and theoretical questions about incomplete neutralization that we would like to address here. The first question, whether incomplete neutralization happens and how it is manifested, has already been answered by ample evidence from all the studies described above. All of them show that temporal cues are involved in incomplete neutralization, such as the durations of the preceding vowel, closure, voicing into the closure, and release burst. It is possible that other non-temporal cues (particularly intensity of release burst, and perhaps for affricates and fricatives some measure of intensity of the noise of the segment) will also be found to be relevant by later research.

The second question is, what influences incomplete neutralization? Is it the orthography, the underlying representation, or something else? The answers to this question are not as straightforward as to the first. There are strong indications from several studies, for example, those of Port and Crawford (1989), Fourakis and Iverson (1984), and Ernestus and Baayen (2006), that orthography does influence incomplete final devoicing. Especially interesting in this regard is the study of the influence of orthography on underlying representations by Warner, Jongman, Sereno, and Kemps (2004: 270–73). In this study, pairs of words were examined with the same string of phonemes, for example, /kledən/, but with alternate spellings, for example, Dutch *kleden* 'to dress' and *kleedden* 'dressed'. The consonants in these words did not occur in neutralizing environments as they were intervocalic. Fifteen Dutch speakers read a list of words containing 20 orthographic minimal pairs in random order. Vowel and consonant durations were measured. The results show that there was a significant orthographic effect at a sub-phonemic level on consonant duration (of 3.4 ms) and vowel duration (only in the environment of /t/). These results suggest that orthography rather than underlying representation influence the difference of duration in speaker's performance.

Finally, Warner, Jongman, Sereno, and Kemps (2004) suggest that not only orthography but other non-phonemic factors, such as morphology, word frequency, and lexical neighborhood difficulty, may influence natural speech. Ernestus and Baayen (2003, 2006, 2007) suggest that a word's inferred morphological paradigm influences the voicing specification of the stem-final obstruent and conclude that Dutch speakers rely on the underlying representations of similar words as exemplars when dealing with nonce-words. Yu (2006, 2007) goes even further. He views tonal merger in Cantonese as a type of phenomenon similar to incomplete neutralization. His experiments support phonetic analogy in a case of tonal morphology in Cantonese and suggest that the surface realization of a rising tone is affected by the tonal specification of its paradigmatic neighbor. To explain such phenomena as tonal merger in Cantonese and incomplete neutralization, he proposes an exemplar-based model of phonological representations.

It is likely that both orthography and underlying representation have some influence on incomplete neutralization in Russian, but they do not tell the full story. We would like to suggest the following explanation: there are two opposite factors or phonetic processes that are working against each other. On the one hand, there is the spelling, which differentiates the voiced from the voiceless, *kod* vs. *kot*, and more fundamentally the grammatical knowledge that there is an underlying /d/ in *kod* (it reemerges in the gen sg, *kóda*). On the other hand, there is grammatical knowledge of the phonological rule that in final position /d/ → [t]. How does the speaker cope with the competing demands of the knowledge of the underlying forms and of the phonological rules in production and perception? A possible explanation is that the process of final devoicing is non-discrete and operates in continuous time. A native speaker who knows the spelling and the grammar recognizes the underlying final voiced obstruent, for example, in a written word. The speaker also needs to apply the devoicing rule in accordance with the standard pronunciation rules of word-final devoicing. The rule of final devoicing is applied simultaneously with or just after the native speaker's knowledge that the underlying obstruent is voiced comes into play. It is possible that this is the reason that the native speaker does not reach his target sound to the full, and thus, for example, that the underlying voiced /d/ tends to have shorter burst duration than the underlying voiceless /t/ would have, despite the cues for voicing being largely neutralized.[7]

A third question is what theory can better account for the incomplete neutralization effect. Can it be accounted for within the framework of a phonological theory with discrete units, such as binary oppositions and distinctive features? In the case of "incomplete" neutralization low-level phonetic features seem to be involved. One might suggest, for example, from the voicing difference between final /d/, /t/ in Russian, that there is an additional distinctive feature, +/– BD (burst duration). This would be wrong, however, since the results of this experiment demonstrate that significant overlap exists between the two sets of phones corresponding to underlying voiced and voiceless stops; after all, this is precisely what is meant by incomplete neutralization. Out of 72 tokens with an underlying voiced stop, 56 (78%) were classified as voiced according to discriminant analysis (see Table 4) while 16 tokens were classified as voiceless. Hence, these 16 tokens participate in the phonological neutralization of the voicing feature. This variance does not fit within the theoretical approach that a language can be fully described with discrete units only, using the values (+/–).[8] Even though Optimality Theory (OT) (Prince and Smolensky 1993) can account for variance with different hierarchies of constraints, unless constraints can be formulated in terms of phonetic parameters, it is diffi-

[7] Less schematically, we might view speech production as due to a number of distinct processes that can partially counteract each other. Thus, a word is spoken first by muscles in the vocal tract being triggered to produce the gestures appropriate for a given phoneme, then other processes are triggered by reaching the boundaries of higher-level linguistic units, processes that can act to reduce or eliminate distinctions encoded by the segment-level processes.

[8] On the acoustic level when there is no neutralization much higher identification rates (98%–100%) are the norm.

cult to see how OT would account for cases of incomplete neutralization.

Finally, note that the study of Warner, Jongman, Sereno, and Kemps (2004) showed that closure duration was found to be a cue for the perception of final devoicing when this was the only feature that was manipulated, in spite of the fact that the previous production and perception experiments did not show a significant difference for closure duration. This may point to the strong tendency of listeners to differentiate homonyms, applying even the smallest and otherwise insignificant cue to communicate a difference. In short, we agree with Manaster Ramer (1996) that only collaboration between practitioners of such distinct sciences as acoustic phonetics, auditory phonetics, and phonology can yield an adequate theoretical explanation for such phenomena as incomplete neutralization. Moreover, an adequate account of gradient, non-discrete phenomena like incomplete neutralization should take into consideration temporal and environmental variables. Gafos (2006), for example, proposes to reconcile phenomena like incomplete neutralization with discrete phonology by using nonlinear dynamical models that relate discrete aspects of the grammar and continuous, environmental variables. This might be a promising direction for future investigation.

In any case, a linguist's priority is to describe the phenomenon of incomplete neutralization as precisely as possible. Therefore, more experiments are needed in languages in which any type of neutralization occurs. A follow-up study should be conducted for Russian, including a perception experiment that compares all the phonetic variables across all Russian obstruents. Furthermore, in order to find out exactly which factors most influence speaker's performance—orthography, underlying representations, or phonetic intraparadigmatic analogy—an experiment could be conducted with children at the age prior to acquiring spelling rules

References

Chen, Matthew. (1970) "Vowel length variation as a function of the voicing of the consonant environment". *Phonetica* 22(3): 129–59.

Cochrane, G. R. (1970) “Some vowel durations in Australian English”. *Phonetica* 22(4): 240–50.

Dinnsen, Daniel A. and Jan Charles-Luce. (1984) “Phonological neutralization, phonetic implementation, and individual differences”. *Journal of phonetics* 12: 49–60.

Dmitrieva, Olga and Allard Jongman. (2010) “Phonological neutralization by native and non-native speakers: The case of Russian final devoicing”. *Journal of phonetics* 38: 438–92.

Ernestus, Mirjam and Harald R. Baayen. (2003) “Predicting the unpredictable: Interpreting neutralized segments in Dutch”. *Language* 79: 5–38.

———. (2006) “The functionality of incomplete neutralization in Dutch: The case of past-tense formation”. Louis M. Goldstein, Douglas H. Whalen, and Catherine Best, eds. *Laboratory phonology 8: Varieties of phonological competence*. Berlin: Mouton de Gruyter, 27–49.

———. (2007) “Intraparadigmatic effects on the perception of voice”. Jeroen van de Weijer and E. J. van der Torre, eds. *Voicing in Dutch: (De)voicing-phonology, phonetics, and psycholinguistics*. Amsterdam: Benjamins, 153–73.

Fourakis, Marios and Gregory K. Iverson. (1984) “On the ‘incomplete neutralization’ of German final obstruents”. *Phonetica* 41(3): 140–49.

Gafos, Adamantios. (2006) “Dynamics in grammar: Comment on Ladd and Ernestus and Baayen”. Louis M. Goldstein, D. H. Whalen, Catherine Best, eds. *Laboratory phonology 8: Varieties of phonological competence*. Berlin: Mouton de Gruyter, 51–78.

Hogan, T. John and Anton J. Rozsypal. (1980) “Evaluation of vowel duration as a cue for The voicing distinction in the following word final consonant”. *Journal of the Acoustical Society of America* 67: 1764–71.

House, Arthur S. and Gordon Fairbanks. (1953) “The influence of consonant environment upon the secondary acoustical characteristics of vowels”. *Journal of the Acoustical Society of America* 25(1): 105–13.

Jakobson, Roman, Gunnar Fant, and Morris Halle. (1952) *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.

Jones, Daniel. (1950) *The phoneme: Its nature and use.* Cambridge: Heffer.

Keating, Pat A. (1985) "Universal phonetics and the organization of grammars". Victoria A. Fromkin, ed. *Phonetic linguistics, Essays in honor of Peter Ladefoged*. London: Academic Press, 115–32.

Klatt, Dennis. H. (1973) "Interaction between two factors that influence vowel duration". *Journal of the Acoustical Society of America* 54(4): 1102–1104.

———. (1976) "Linguistic use of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5): 1208–21.

Klecka, William. (1980) *Discriminant Analysis.* Beverly Hills, CA: Sage. [*Sage University Paper series on Quantitative Applications in the Social Sciences*, 19.]

Kozhevnikov, Valery A. and Ludmila. A. Chistovich. (1966) *Speech: Articulation and perception* (English translation). Washington, DC: Joint Publications Research Service.

Luce, Paul A. and Jan Charles-Luce. (1985) "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production". *Journal of the Acoustical Society of America* 78(6): 1949–57.

Manaster Ramer, Alexis. (1996) "A letter from an incompletely neutral phonologist". *Journal of phonetics* 24: 477–89.

Peterson, Gordon and Ilse Lehiste. (1960) "Duration of syllable nuclei in English". *Journal of the Acoustical Society of America* 32(6): 693–703.

Podlipsky, Vaclav J. (2008) "Vowel duration as a cue for coda voicing and the perception of second language vowel quantity". *Acoustics 08.* Paris: Sociéte Française d'Acoustique (SFA), 155–59.

Port, Robert and Michael O'Dell. (1985) "Neutralization of syllable-final voicing in German". *Journal of phonetics* 13(4): 455–71.

Port, Robert and Penny Crawford. (1989) "Incomplete neutralization and pragmatics in German". *Journal of phonetics* 17: 257–82.

Prince, Allan and Paul Smolensky. (1993/2004) *Optimality theory: Constraint interaction in generative grammar.* Technical Report 2, Center for Cognitive Science, Rutgers University and University of Colorado at Boulder. [Revised version published by Blackwell, 2004.]

Pye, Shizuka. (1986) "Word-final devoicing of obstruents in Russian". *Cambridge papers in phonetics and experimental linguistics*. Cambridge University: Department of Linguistics, 5: 1–10.

Shcherba, Lev Vladimirovich. (1911) *Court exposé de la pronunciation russe*. Teubner.

Sharf, Donald. (1962) "Duration of post-stress intervocalic stops and preceding vowels". *Language and speech* 5: 26–30.

Shrager, Miriam. (2002) "Neutralization of word-final voicing in Russian". Poster presented at the Annual Pan-American Meeting of the Acoustical Society of America, Cancun, December, 2002.

Slis, Iman H. and Antonie Cohen. (1969) "On the complex regulating the voiced-voiceless distinction". *Language and speech* 12(3): 137–55.

Slowiaczek, Louisa and Daniel Dinnsen. (1985) "On the neutralizing status of Polish word-final devoicing". *Journal of phonetics* 13(3): 325–41.

Tieszen, Bozena. (1997) *Final stop devoicing in Polish: An acoustic and historical account for incomplete neutralization.* PhD dissertation, University of Wisconsin, Madison.

Trubetzkoy, Nikolai Sergeyevich. (1969) *Principles of phonology*. Berkeley: University of California Press. [Originally published as *Grundzüge der Phonologie*. C. Baltaxe, trans. Goettingen: Vandenhoeck and Ruprecht, 1958.]

Warner, Natasha, Allard Jongman, Joan Sereno, and Rachel Kemps. (2004) "Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch". *Journal of phonetics* 32: 251–76.

Yu, Alan C. L. (2006) "In defense of phonetic analogy: Evidence from tonal morphology". Paper presented at the Annual Meeting of the Linguistics Society of America, Albuquerque, New Mexico, January, 2006.

———. (2007) "Understanding near mergers: The case of morphological tone in Cantonese". *Phonology* 24: 187–214.

Slavic Language and Literatures
Indiana University
Ballantine Hall 502
Bloomington, IN 47405
mshrage@indiana.edu

Received: March 2009
Revised: September 2011