Original Paper

Phonetica 2005;62:14–54
DOI: 10.1159/000087223

Received: October 3, 2003
Accepted: February 6, 2005

Adaptive Dispersion Theory and Phonological Vowel Reduction in Russian

Jaye Padgett[a] Marija Tabain[b,c]

[a]University of California, Santa Cruz, Calif., USA; [b]Macquarie University, Sydney; [c]La Trobe University, Melbourne, Australia

Abstract

Russian exhibits a rich pattern of phonological vowel reduction, by which some vowel contrasts are neutralized in unstressed syllables. Recent work in phonology suggests a mechanism by which *phonetic* vowel reduction – compression of the overall vowel space due to target undershoot – might lead to patterns like Russian. Presenting acoustic data from 9 speakers of Russian, we use Euclidean distance measures, measures of F1–F0 and F2–F1, and Bayesian classification to provide a basic picture of how the overall vowel space, as well as the distribution of vowels, change as stress is reduced. We are particularly interested in whether contraction of the vowel space in unstressed positions is primarily due to raising, and in whether contrasting pairs of vowels are evenly spaced within and across contexts. Our results provide qualified support for the first hypothesis, but largely do not support the hypothesis of equal spacing, in particular across contexts. Of additional interest, we find that some impressionistically described neutralizations are incomplete.

Introduction

The term ‘vowel reduction’ has two meanings, depending on whether used by a phonologist or a phonetician, as Fourakis [1991] notes. ‘Phonetic’ vowel reduction refers to undershoot of vowel targets, due either to coarticulation or a tendency to centralize, or both. It is a gradient, subphonemic process, dependent on (at least) speech rate and register, stress, and segmental context. The result is a shrinkage of the overall vowel space. ‘Phonological’ vowel reduction typically refers to the neutralization of vowel phoneme contrasts, often (but not always) resulting in a [ə]-like pronunciation. This occurs for example in English *expl**a**nation* and *emph**a**sis* (cf. *expl**ai**n* and *emph**a**tic*, respectively). It is a categorical substitution of sounds, and not gradient undershoot: it does not depend on speech rate or register, and [eɪ] is not an option in *expl**a**nation* no matter how careful the speech.

Though it seems plausible that there could be a connection between these two kinds of reduction, only recently has work emerged, in phonology, trying to forge a link

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/05/0621–0014
$22.00/0
Accessible online at:
www.karger.com/journals/pho

Jaye Padgett
Department of Linguistics Stevenson College
University of California
Santa Cruz, CA 95060 (USA)
Tel. +1 831 459 3157, Fax: +1 831 459 3334
E-Mail padgett@ucsc.edu

[Flemming, 1995/2002, in press; cf. Crosswhite, 2001, in press; Barnes, 2002]. Flemming's work suggests roughly the following view. Unstressed vowels are shorter than their stressed counterparts; decreased vowel duration is known to correlate with formant undershoot [Lindblom, 1963]. When this leads to undershoot of F1, the vowel space shrinks, the bottom of that space raising. In this compressed space vowel phonemes are more prone to be confused, with neutralization (loss of contrast) resulting. Flemming appeals to a particular implementation of the principle of sufficient contrast of Adaptive Dispersion Theory [Liljencrants and Lindblom, 1972; Lindblom, 1986, 1990] in order to do this, and to predict the extent to which phonetic reduction will lead to phonological reduction in a given language. Though these ideas are attractive and plausible, there have been no attempts yet to test them by detailed analysis of quantitative data from a language having phonological vowel reduction.[1]

This paper examines phonological vowel reduction in Russian with this aim in mind. A preliminary goal is to provide enough data on the full and reduced vowels of Russian to test the impressionistic descriptions of Russian vowel reduction, including claims about neutralization. Though acoustic properties of Russian vowels have been reported [Jones, 1959; Fant, 1960; Lobanov, 1971; Purcell, 1979; Bolla, 1981; Kouznetsov, 2001], there are no quantitative studies of Russian phonological vowel reduction, employing controlled environments and a large number of speakers, in the published literature, so far as we know. We present data from 9 speakers (8 female and 1 male) in both palatalized and non-palatalized environments, covering stressed, prestressed, and unstressed (non-prestressed) vowels. (The latter two contexts are traditionally distinguished in descriptions of Russian vowel reduction.) Our second goal is to analyse the data in order to test some predictions of the above account of phonological vowel reduction: (1) vowels are shorter in unstressed syllables; (2) the overall vowel space shrinks in unstressed syllables; (3) shrinkage largely involves raising of the vowel space floor. Finally, we consider in detail how vowels are spaced with respect to each other both within and across contexts. Though we can draw only limited conclusions about the general phonetic bases of phonological reduction from a study of one language, our hope is to contribute to a larger research agenda that will eventually include other languages as well.

2. Background

2.1 Russian Vowel Reduction

Phonological vowel reduction in Contemporary Standard Russian has been well described [Avanesov, 1956, 1972; Halle, 1959; Lightner, 1965, 1972; Ward, 1975; Hamilton, 1980; Crosswhite, 2001, in press]. In stressed syllables, Russian contrasts the five phonemes /i,e,a,o,u/.[2] In unstressed syllables, this five-way contrast is neutralized to a two-way contrast after palatalized consonants, and to a three-way contrast

[1] Herrick [2003] is a study of phonological vowel reduction in Catalan, having much the same goals as ours. Our work and Herrick's developed simultaneously and have influenced one another.

[2] Throughout the paper we employ the standard phonological notation / / to indicate a vowel's underlying representation, in the phonological sense. This is a matter of convenience, and it could be taken instead to indicate what vowel an unstressed vowel alternates with when under stress.

elsewhere. The forms in (1) illustrate vowel reduction in one elsewhere context, when the preceding consonant is non-palatalized. The underlined vowels are stressed in (1)a. They are unstressed in the derived adjectives shown in (1)b. As can be seen, high vowels do not change, but /e/ is described as raising to [i], and /o,a/ are said to neutralize to [ɐ] or [ə]. According to descriptions, /a/ and /o/ reduce to [ə] in most unstressed syllables, but to [ɐ], a low-mid central vowel, in prestressed syllables – the syllable immediately preceding the stressed syllable.[3] Our transcriptions here reflect these assumptions. They depart from the norm in one respect: the vowel /i/ after non-palatalized consonants is usually transcribed as [ɨ] (or 'y' in Slavicist literature), but here we transcribe [i] with velarization on the preceding consonant, following Padgett [2001].

(1)	a.	'dˠim	'smoke'	b.	dˠimɐ'voj	(adj.)
		'sudnə	'ship'		sudɐ'voj	(adj.)
		'tsˠex	'(factory) shop'		tsˠixɐ'voj	(adj.)
		'got	'year'		gədɐ'voj	'annual'
		'praf	'law'		prəvɐ'voj	'legal'

High vowels do not change after palatalized consonants either, as can be seen in (2). In this context, though, every non-high vowel is described as neutralizing to [i].

(2)	a.	'vʲit	'species'	b.	vʲidɐ'voj	(adj.)
		'klʲutʃ	'key'		klʲutʃi'voj	(adj.)
		'dʲelə	'business'		dʲilɐ'voj	(adj.)
		'slʲos	'tears' (gen.pl.)		slʲizətɐ'tʃivˠij	'tear' (gas) (adj.)
		'rʲat	'row, file'		rʲidɐ'voj	'average (rank and file)'

2.2 The Phonetic Sources of Phonological Vowel Reduction

In a classic work, Lindblom [1963] argued that phonetic vowel reduction is target undershoot due to a decrease in duration. Though 'reduction' is commonly assumed to mean centralization, i.e., tending toward a realization that is schwa-like, Lindblom [1963] concluded that reduction was in essence assimilatory: formants are perturbed in the direction of the formants of neighbouring consonants and vowels. Though this can lead to centralization as a side effect, there is no imperative to centralize per se. In studies of phonetic vowel reduction in Midwestern American English and in Dutch (respectively), Fourakis [1991] and Van Bergem [1993] similarly found formant undershoot, rather than any tendency for vowels to centralize per se. Though not supporting centralization, these studies do support the view that phonetic vowel reduction involves a reduction in the overall vowel space employed. This shrinkage of the overall vowel space under decreased duration is a common feature of all phonetic vowel reduction phenomena. It is this, rather than schwa-like realizations, that justifies the term 'reduction'.

It is natural to wonder whether *phonological* vowel reduction as in Russian results from phonetic vowel reduction. The following seems a plausible line of thought [cf. Van Bergem, 1993]. Because the overall vowel space contracts under phonetic reduction, vowels are more crowded. A decrease in distance between vowels plausibly leads

[3] Also in word-initial onsetless syllables, and before vowels within a word, two contexts not considered in this paper. Note also that in the literature [ɐ] is often transcribed as [ʌ], though as Barnes [2002] notes, it is not IPA [ʌ].

to the confusion of vowel categories. If speakers give up on these distinctions, or language learners fail to acquire them, then phonological vowel reduction results. (There is a question whether this connection between phonetic and phonological reduction is synchronic or purely historical: crudely speaking, do speakers phonologically reduce because they know that contrast cannot be maintained in unstressed syllables, or is phonological reduction simply the consequence of a failure to learn or perceive contrasts in unstressed syllables? We deliberately put this question aside.)

There is at first blush an impediment to linking phonetic and phonological reduction in such a way. As we have noted, though phonetic reduction involves a contraction of the overall vowel space, it does not necessarily involve centralization of vowel quality; instead vowel targets can assimilate to those of surrounding segments. Yet according to impressionistic descriptions, phonological reduction in many languages leads precisely to something schwa-like. This is the typical impressionistic characterization of the phonologically reduced vowel in English words like *sofa* [see for example Chomsky and Halle, 1968; Ladefoged, 1993]. Similarly, Russian /a/ and /o/ are reduced to [ə] in some contexts, as seen above. It is therefore not immediately obvious how neutralization might have its roots in phonetic vowel reduction.

The view that unstressed vowels centralize under phonological vowel reduction oversimplifies matters in at least three respects, however. First, typological surveys indicate that phonological vowel reduction targets *non-high* vowels disproportionately [Crosswhite, 2001; Barnes, 2002]. This is true of Russian. Second, in phonological vowel reduction, though schwa might be produced, not all neutralized vowels neutralize to schwa. Russian /e/, in fact, reduces to [i] in unstressed syllables, as seen earlier. This is no quirk of Russian. The surveys of Crosswhite [2001] and Barnes [2002] show that phonologically reduced vowels often *raise*. Phonological reduction in Catalan exemplifies both of the above properties. The Standard Catalan stressed inventory of /i,e,ɛ,a,ɔ,o,u/ reduces to [i,ə,u] in unstressed syllables. The high vowels do not change, and while /e,ɛ,a/ reduce to [ə], /ɔ,o/ reduce to [u]. Similarly, in some eastern dialects of Bulgarian the six stressed vowel phonemes /i,e,ə,a,o,u/ neutralize to [i,ə,u] when unstressed; in this case the vowels /e,a,o/ all raise, to [i,ə,u], respectively. Pettersson and Wood [1987] and Wood and Pettersson [1988] note the difficulty of characterizing this pattern by appeal to traditional vowel height: assuming that [ə] is a mid vowel like [e] and [o], there is no simple way to express the class [e,a,o] targeted for raising. Based on a study of Bulgarian vowel production, these authors conclude that jaw height and tongue constriction degree must be factored apart, with [e,a,o] having a lowered jaw position and [i,ə,u] a raised one. The generalization is then that vowel reduction involves jaw raising. Finally, as Flemming [in press] notes, phonologically reduced vowels described as schwa in at least some cases can in fact be high, perhaps more appropriately transcribed as [ɨ]. For example, Kondo [1994] finds that British English [ə] has an F1 target between 270 and 320 Hz, suggesting a rather high realization. (However, results of Browman and Goldstein [1992] suggest a mid vowel target for American English schwa.) As Flemming [in press] points out, a high realization of schwa, especially when between consonants, follows from the assumption that it is a minimal effort vowel, employing the minimal jaw opening consistent with a vocalic articulation. All of the above observations about phonological reduction suggest that, as with phonetic reduction, it may not be a matter of centralization. Mid vowels often raise to high vowels, and schwa often results from the merger of non-high vowels only.

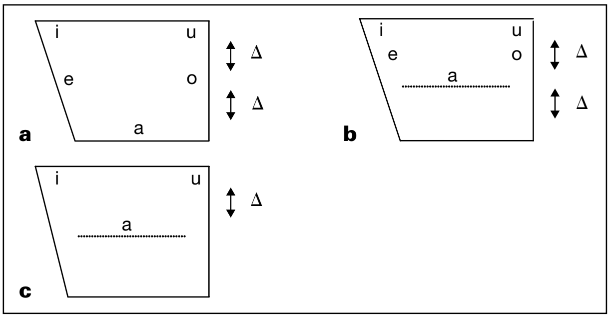


Fig. 1. Hypothesized effect of jaw and/or tongue body raising, due to decreased duration, on the five-vowel inventory of Russian, assuming a minimal perceptual distance threshold Δ. **a** Stressed vowels; **b** Unstressed vowels, with raised vowel height floor; **c** Neutralization.

Flemming [1995, in press] proposes an account of phonological vowel reduction that links the properties of phonetic and phonological reduction noted above.[4] Though his model is one of phonology, cast within Optimality Theory [Prince and Smolensky, 1993], it is based on the principles of Adaptive Dispersion Theory, and it assumes that vowels within a language are subject to a principle of sufficient perceptual distance [Lindblom, 1986, 1990]. Taking Russian as our example, the account assumes a distance threshold (which we indicate with the symbol Δ) holding among adjacent vowels, as in figure 1a. (Only distances in height are shown here.) If unstressed vowels are shorter than stressed vowels, then the formant undershoot theory of Lindblom [1963] predicts undershoot of F1 in unstressed vowels. As Flemming [in press] notes, Lindblom [1963] found that F1 of Swedish vowels decreased exponentially as vowel duration decreased. Flemming also notes the well-known correlation between vowel height and vowel duration in general, lower vowels being intrinsically longer than higher ones [Lehiste, 1970]. The plausible reason for this connection is the greater articulatory displacement of jaw and/or tongue body, and therefore extra time, required to achieve lower vowel targets. If this is so, then the prediction is that under shortening the vowel height floor is effectively raised, assuming speakers do not compensate by increasing articulatory effort [Lindblom, 1983; Moon and Lindblom, 1994].[5] This in turn entails more crowding among the vowel phonemes. This state of affairs is depicted in figure 1b. Under these circumstances, vowel quality would obviously change, but we use the same symbols as in figure 1a to clarify the underlying, or target, values of these vowels. The scenario in figure 1b is not stable, however, if the assumed minimal distance Δ is violated, as shown. The result will be neutralization, as in figure 1c.

An important feature of this account is the focus it places on reduced vowel *inventories* such as [i,ɐ,u], rather than on particular reduced vowels like [ɐ] or [ə]. This is important for two reasons. First, phonological vowel reduction is proposed to result from phonetic vowel reduction specifically when the latter leads to violation of a language's minimal perceptual distance threshold. Appeal to such a threshold obviously requires reference to the overall inventory as opposed to isolated vowels. Second, under phonological vowel reduction, vowels often reduce to *higher* vowels, even to [i] and [u], as we have seen. To see how this might be, consider figure 1 again. Hypothesized stage 1b assumes

[4] This need not be the only source of phonological vowel reduction, which might sometimes result from entirely different mechanisms [see Crosswhite, 2001, in press].

[5] Decreased duration in itself need not imply undershoot; it does not for comparatively shortened vowels before voiceless consonants in English, for example, according to Summers [1987]. Our concern here is only with decreased duration due to lack of stress.

that, prior to neutralization, vowels become more crowded under vowel height floor raising, approaching or reaching the minimal distance threshold Δ. Suppose that vowels must remain equidistant. This could be either because speakers try to counteract the neutralizing potential of undershoot, or because listeners store or give more weight only to examplars that are maximally unambiguous [on the latter see Guy, 1996, and Wedel, in progress]. Then raising of /a/ necessarily leads to raising of /e/ and /o/, too, since otherwise /e/ and /o/ would not lie midway between the high and low vowels. It follows that the chance of confusing /e/ with /i/ would be equal to that of confusing /e/ with /a/, as far as F1 is concerned. (However, /e/ might resemble /i/ more than /a/ in its F2 value.) The account here has nothing to say about which would occur, and this is possibly either a matter of chance or of small, language-specific deviations from the assumed equidistance. Indeed, the simple scenario depicted in figure 1b does not explain why Russian /e/ raises to [i] while /o/ (along with /a/) becomes [ə]. Nevertheless, the point is that this account assumes that raising of /e/ to [i] can be motivated in part by the requirement that /e/ remain distinct from /a/, even as /a/ raises, an appeal to the overall inventory again.

One question raised by this account of phonological vowel reduction involves the status it gives to F1 over F2. Does the fact that phonological vowel reduction across languages affects especially non-high vowels, and involves raising, imply that undershoot of F1 causes more perceptual confusion than undershoot of F2? If so, why should that be? In this context it is interesting to note that, while Fourakis [1991] found no overall tendency toward centralization in Midwestern American English unstressed vowels, his figure 2 suggests that all vowels except for [i] and [ɪ] *raised* somewhat on average, and that this was the primary reason for a shrunken vowel space under phonetic reduction. According to his tables VI and VII, all vowels but [i] had lower average F1 values under phonetic reduction for women. (The data for men are less clear.) On the other hand, Lindblom [1963] specifically notes that undershoot in his data is most conspicuous for F2. There are two facts which qualify this conclusion. (For the sake of this discussion we computed undershoot values using Lindblom's equations, assuming a vowel duration of 80 ms.) First, when undershoot is interpreted in ERB (see section 3), the effect of F1 increases overall with respect to that of F2. This is significant, since we are interested precisely in the perceptual consequences of undershoot. In fact, the effect of F1 now exceeds that of F2 for three of the eight vowels Lindblom studied, /a,æ,e/, including both low vowels. (Lindblom observes only the lax Swedish vowels.) Nevertheless, F2 undershoot remains larger for the rest of the vowels. Second, in Lindblom's data F1 undershoot has a more consistent effect on vowel quality across contexts than does F2 undershoot. In particular, while non-high vowels uniformly raise due to undershoot (high vowels do not substantially change), the effect of F2 on a vowel depends on the consonantal context, front vowels being affected least consistently. (Back vowels generally increase in F2.) Thus F1 undershoot might be more consistent in its effects, and at least for low vowels, greater in degree, than F2 undershoot.

These observations nevertheless leave much room for a role for F2 undershoot in phonological vowel reduction. Apart from Lindblom's results and the well-known existence of F2 undershoot in general, the fact that some phonologically reduced inventories are [i,ɐ,u] rather than [i,ɛ,ɔ,u], lacking an F2 contrast between non-high vowels, itself suggests a role for F2 undershoot. It is true that this fact might follow in principle from vowel raising alone: if /e,o/ neutralized with /i,u/ (respectively), then there would be no possibility of such a contrast. However, Russian /o/ neutralizes with /a/, not /u/. F2 undershoot could affect non-high vowels disproportionately because non-high vowels occupy an intrinsically shrunken F2 space in comparison to high vowels. Finally, there is reason to believe that differences in F2 are in general harder to perceive than those in F1 (see section 3.2.4). If this is so, it would be especially puzzling if neutralization resulted from F1 undershoot only. Though undershoot of F1 might play a large role in phonological vowel reduction, we should not ignore the possible role of F2.

Though Flemming's [1995, in press] account for phonological vowel reduction is attractive, there have so far been no attempts to test these ideas in detail by analyzing data from a language having phonological vowel reduction. This is one of our goals here. The explanation above, and indeed the theory inspiring it, Adaptive Dispersion Theory, raise several questions in this regard. First, are stressed vowels actually longer than unstressed vowels in Russian? More interestingly, do prestressed vowels occupy an intermediate position in terms of duration? They are predicted to, since /a/ in this position is claimed not to reduce to the extent it does in other unstressed syllables (see the discussion above). Second, does the overall vowel space in fact shrink in unstressed syllables? Third, is this shrinkage attributable primarily to raising of the vowel height floor? (Once again, prestressed syllables should occupy an intermediate position.)

Given the importance of the ideas of dispersion, and perceptual confusability, in all dispersion-based accounts, we also systematically investigate the spacing among vowels both within and across the relevant phonological contexts in Russian – stress level and consonantal palatalization. Can the stressed inventory of Russian be characterized as having equally spaced vowels? Can the unstressed inventory? Further, are the distances observed among stressed (or palatalized) vowels comparable to those among unstressed (or non-palatalized) vowels? If not, how do they differ? For example, the Russian transcriptions suggest the possibility that reduced [i,u] in prestressed position differ more from [ɐ] than stressed [i,e] do from [e,o], or [e,o] do from [a]. This is because [ɐ] is a low-mid vowel. Presumably, the facts of vowel reduction involve a complex interplay of perceptual and articulatory constraints. Answers to questions about spacing will help provide a basis for future quantitative modeling of vowel dispersion and vowel reduction.

An unexpected finding of this study is the existence of incomplete neutralization in the Russian data. For some of the neutralizations described impressionistically, the relevant vowels remain acoustically distinct. We discuss some implications of this finding as well.

3. The Experiment

3.1 Method

3.1.1 Speakers

Nine speakers of Russian were recorded for this study – 8 female and 1 male. Speakers were aged between 19 and 64, and had spent between 1 and 44 years in Australia. All of the speakers were recruited from the Russian Department at Macquarie University in Sydney, where they were either teaching staff or students of translation. In addition, some of the speakers taught Russian at their local community school on Saturday mornings (this is a typical activity in ethnic communities in Australia). All speakers except speaker MK, who holds a doctorate in syntactic theory and has interests in cognitive linguistics, were naïve as to the purpose of the experiment.

Table 1 gives a list of the speakers with their ages, the number of years spent in Australia at the time of recording, and where they grew up learning Russian. It will be seen that about half the speakers are not from Russia: 3 are from China, 1 is from Ukraine (Kiev), and another had spent time in Ukraine, Uzbekistan and Moscow. The 2 students who had spent time in Ukraine said they were

Table 1. List of speakers (note that all speakers except AC are female)

Speaker	Age	Years in Australia	Grew up in
AC (male)	19	8	Moscow
DR	19	1	Ukraine, Uzbekistan, Moscow
JD	53	5	St. Petersburg
MK	45	10	Moscow
NR	63	44	China
TM	40	25	China
TO	30	5	St. Petersburg
VS	23	10	Kiev (Ukraine)
ZL	64	40	China

fluent in both Ukrainian and in Russian (related East Slavic languages). Those born in China reflect a significant subgroup of Russian speakers in Australia, consisting of White Russians born and raised in China, who attended Russian-speaking schools in China and who did not go back to Russia until after the fall of Communism. Two of these speakers reported that when they went back to Russia in the 1990s, native Russians did not realize that they were not from Russia. One of the China-born speakers, NR, had completed a doctoral thesis which examined language maintenance by the Russian communities in China vs. Australia. She reported that the greater similarity between Australian English culture and Russian culture led to greater language loss, whereas the greater dissimilarity between Russian and Chinese culture led to the White Russians forming an enclave where language maintenance was highly valued. To further give us reasonable confidence that the data we acquired reflected not only native speech but the standard dialect described above, speaker MK (see above), who helped with recruiting, screened potential subjects for standard pronunciation. We finally note that, with the exception of speaker TO, at the time of recordings all of the speakers lived in households where Russian was spoken daily with family members.

3.1.2 Stimuli

Stimuli consisted of 30 words (plus 4 filler words used for another experiment) placed in a carrier phrase ['maʂ̡ə skɐ'zalə _____] 'Masha said _____'. The stimuli were real words devised by the first author and speaker MK. All of the words contained at least two syllables and all words were familiar to the 9 speakers. The list of stimuli is given in table 2.

Each word contained one of the five target vowels, usually in the first syllable (with exceptions noted below). This first syllable was either Stressed, Prestressed or Unstressed. In addition, each target vowel followed either a palatalized or a non-palatalized consonant. (For ease of reference, these vowels will be referred to as 'palatalized' and 'non-palatalized', respectively. Palatalized vowels will be denoted by /ji/, /je/, etc.) The consonant following the target vowel was non-palatalized in all cases, and the following vowel was always [a], [ɐ] or [ə]. There were 30 words in all (5 vowels $\times$ 3 stress levels $\times$ 2 consonantal contexts). Each speaker produced 15 differently randomized repetitions of a list, giving about 450 tokens per speaker (the number varies slightly from speaker to speaker according to mispronunciations, hesitations, repetitions etc.).

For the non-palatalized vowels, the consonant preceding the vowel was a labial (one of /p b v/) and the following consonant was an alveolar stop (one of /t d/). The same conditions were true for the palatalized vowels /ji je ja/, but for the vowels /jo ju/, the preceding consonant was a lateral, due to phonotactic gaps in the lexicon (the following consonant was however still an alveolar stop). Also due to lexical gaps, 6 of the stimuli consisted of prepositional phrases, and 3 of these contained an extra segment in word-initial position. (The phrase meaning 'in one's stride' contains an /s/ before the lateral, and the phrases meaning 'about the story' and 'about the stage' contain an unstressed /o/ before the bilabial.) In phonological accounts, a preposition is assumed to join with a following open class word to form one phonological word, indistinguishable in most phonological respects from a single open class word [see Halle, 1959, for example].

Table 2. List of stimuli

Target vowel	Consonant	Stress	Stimulus	Phonemic	Gloss
i	Non-palatalized	S	'pˠ**i**tkə	p**i**tka	‘torture’
		P	pˠ**i**'tatsə	p**i**tatʲsʲa	‘to try’
		U	bˠ**i**tɐ'voj	b**i**tovoj	‘involving way of life’
	Palatalized	S	'vʲ**i**dnə	vʲ**i**dno	‘evidently’
		P	pʲ**i**'tatʲ	pʲ**i**tatʲ	‘to feed’
		U	vʲ**i**tɐ'mʲin	vʲ**i**tamʲin	‘vitamin’
e	Non-palatalized	S	'vˠ**e**təm	v **e**tom	‘in that’
		P	ɐbˠ**i**'tapʲi	ob **e**tapʲe	‘about the stage’
		U	ɐbˠ**i**tɐ'ʒe	ob **e**taʒe	‘about the story’ (of a building)
	Palatalized	S	'vʲ**e**dətʲ	vʲ**e**datʲ	‘to manage’
		P	bʲ**i**'da	bʲ**e**da	‘misfortune’
		U	bʲ**i**tɐ'nʲirəvətʲ	bʲ**e**tonʲirovatʲ	‘to concrete’
a	Non-palatalized	S	'p**a**dələ	p**a**dalo	‘fell’ (neut.)
		P	v**ɐ**'tagə	v**a**taga	‘throng; gang’
		U	vətɐ'manʲe	v atamanʲe	‘in the Cossack chief’
	Palatalized	S	'pʲ**a**təjə	pʲ**a**taja	‘fifth’ (fem.)
		P	pʲ**i**'ta	pʲ**a**ta	‘heel’
		U	pʲ**i**tɐ'tʃʲok	pʲ**a**tatʃʲok	‘five-kopeck coin’
o	Non-palatalized	S	'v**o**tkə	v**o**dka	‘vodka’
		P	v**ɐ**'da	v**o**da	‘water’
		U	p**ə**tɐ'lok	p**o**tolok	‘ceiling’
	Palatalized	S	'slʲ**o**tə	s lʲ**o**ta	‘in one’s stride’
		P	lʲ**i**'talə	lʲ**o**talo	‘flew’ (neut.)
		U	lʲ**i**dɐ'xot	lʲ**o**doxod	‘ice drift’
u	Non-palatalized	S	'p**u**tənʲitsə	p**u**tanʲitsa	‘confusion’
		P	b**u**'tan	b**u**tan	‘butane’
		U	v**u**dɐ'lʲenʲii	v **u**dalʲenʲii	‘in the moving off’
	Palatalized	S	'lʲ**u**təjə	lʲ**u**taja	‘fierce’ (fem.)
		P	lʲ**u**t'skajə	lʲ**u**dskaja	‘involving people’ (fem.)
		U	lʲ**u**dɐ'jet	lʲ**u**dojed	‘cannibal’

In this and all subsequent tables, S = Stressed, P = Prestressed, and U = Unstressed.

It should be noted that non-palatalized consonants in Russian can be velarized [Fant, 1960; Öhman, 1966; Purcell, 1979; Evans-Romaine, 1998; Padgett, 2001; Kochetov, 2002]. Therefore even non-palatalized consonants can carry inherent vocalic specifications that exert a coarticulatory influence on adjacent vowels. However, the phonetic studies cited suggest that velarization is weaker than palatalization, and not invariably present. It is consistently present before front vowels, however, most likely as a means of keeping such consonants distinct from palatalized ones in the same context. Our transcriptions assume velarization only before front vowels.

3.1.3 Recordings and Labelling

All data were recorded in a sound-treated room at the Speech, Hearing and Language Research Centre at Macquarie University under the supervision of a recording technician and the second author. Speakers were paid for their time. Data were recorded onto DAT at a sampling rate of 20 kHz, and transferred to SUN workstations where tokens were segmented and labelled using the EMU speech analysis system [Harrington et al., 1993; Cassidy and Harrington, 2001] by a paid phonetically trained labeller. Formants and fundamental frequency were tracked automatically, using LPC with a default of 12 coefficients, where the frame-shift is 5 ms. Mistracked formants were hand-corrected.

3.1.4 Analysis

Most statistical analyses were carried out by the second author using the EMU system, the R statistical package [R Development Core Team, 2003], and for ANOVAs, SPSS. Preliminary observations of the data suggested that the most appropriate sampling point for formant measurement was the 75% mark of total vowel duration for the Stressed contexts, and the 50% mark of total vowel duration for the Prestressed and Unstressed contexts, where total vowel duration does not include the stop burst from any preceding consonant. (The difference between voiced and voiceless preceding bilabial stops should not be important, since in Russian the voiceless stops are not aspirated.) Observation of the spectrographic data suggested that the influence of the secondary palatalization gesture (or of the secondary velarization gesture for non-palatalized /i/ and /e/) had a much more noticeable influence on the early part of the vowel in Stressed syllables than in non-Stressed ones. (Effects of the following consonant were generally smaller for all vowels; palatalization and velarization in Russian generally affect following vowels much more than preceding ones, and in our data the following consonant was non-palatalized, and in a context calling for little or no velarization.) It will be seen below that the duration of Stressed vowels is much greater than that of non-Stressed vowels, usually well over 100 ms compared to 40–80 ms. Given this length, measurement of the formants at 75% of vowel duration for Stressed vowels did not show much influence of the following consonant. By contrast, however, a measurement at 75% for the much shorter Prestressed and Unstressed vowel contexts was much more likely to include part of the formant transition into the following consonant. For this reason, we chose to sample these data at the 50% mark.

All formant and F0 data were smoothed using Tukey's median filter with a window width of 3 samples; i.e. the middle value of 3 successive samples was set as the median value of those 3 samples, with this process being repeated until convergence. Given the importance here of perceptual (rather than acoustical) vowel spacing, we converted all data to equivalent rectangular bandwidths (ERB) using the following formula [Moore and Glasberg, 1996]:

(3) $21.4 \times \log10\,(f \times 0.00437 + 1)$

where f is the frequency value in Hertz. Based on experiments where a given tone is masked by a simultaneous but different tone or by a narrow band of noise, it has been shown that the ear integrates two tones produced simultaneously within 1 ERB.

In the early stages of our analyses, we used the F2′ algorithm given by Ménard et al. [2002], in order to imitate the spectral integration of the higher formants carried out by the human ear. (This algorithm was found to be particularly useful in distinguishing the front rounded vs. unrounded vowels in French.) However, we found that our measurements of F3 and F4 were unreliable for the female speakers (8 of the 9 speakers being female), resulting in large amounts of variation around the mean vowel values. In addition, our impression was that the relative spacing of the vowel means did not differ greatly according to whether we used F2 or F2'. We therefore restricted the analysis to F1 and F2.

The following sections give results involving duration, changes in the overall vowel space, and spacing between vowels within and across contexts. To make it easier to follow, we provide more detail about methodology in the relevant sections.

3.2 Results

3.2.1 Duration

Table A1 in the 'Appendix' gives the duration values (mean and standard deviation) for all vowels and all speakers, together with results from a one-way ANOVA for each speaker. With the exception of 1 speaker (AC) whose durations are much shorter, vowels in the Stressed context have a duration of around 100–160 ms, while the Unstressed and Prestressed vowels have a mean duration of between 40 and 80 ms. The distinction between Stressed and Prestressed vowels, as well as Stressed and Unstressed vowels, is strongly maintained by all speakers in both the palatalized and the non-palatalized contexts, according to the ANOVA results. However, while most speakers (7 out of 9) show a significant difference between the Prestressed and Unstressed vowels in the non-palatalized context, the opposite is true in the palatalized context, where 7 out of 9 speakers do not show a significant difference between the Prestressed and Unstressed vowels.

3.2.2 Overall Vowel Space

Figure 2 gives plots of all six vowel spaces for all speakers (3 stress contexts × 2 consonantal contexts, palatalized and non-palatalized). Figure 3 presents all 8 of the female speakers pooled. Table 3 presents measures of the size, and location of the edges, of each vowel space, averaging across the female speakers. These averages are based on the values in table A2 in the 'Appendix', which presents results for each speaker. To obtain the latter results, for each speaker we took as the maximum F1 value the mean F1 for /a/ plus one standard deviation, and as the minimum F1 value the mean F1 for either /i/ or /u/ minus one standard deviation. The lower of the latter two was chosen, giving the larger estimate of the F1 range. Maximum F2 is the mean F2 for /i/ plus one standard deviation, and minimum F2 the mean F2 for /u/ minus one standard deviation. The point of including one standard deviation at the extremities was to capture a fair proportion of the data without including outliers.

We consider first the position of the non-palatalized vowel space extremities. Since 1 ERB represents a kind of critical band, it seems reasonable to treat as perceptually insignificant differences that do not approach this value. With this in mind, the vowel plots in figures 2 and 3 and the data in table 3a reveal that speakers raise the F1 minimum value by about 1 ERB between Stressed and Unstressed position on average, with Prestressed position seeming to occupy an intermediate position. Raising of minimum F1 is not completely consistent, though, a notable exception being speaker ZL, whose F1 minimum lowers from Stressed to non-Stressed contexts. (In addition, speaker VS's /u/ seems to behave oddly.[6]) Speakers likewise lower the F1 maximum values moving from Stressed toward Unstressed, in this case by about 1 ERB at each step. There are no exceptions to this lowering of maximum F1. F1 maxima are affected more than F1 minima overall, judging by these averages (and see figure 3). Turning to F2, speakers lower the maximum F2 value by about 1.5 ERB between Stressed and Prestressed position; the change from Prestressed to Unstressed is well below 1 ERB. The values for F2 minima raise by about 1 ERB overall between Stressed and Unstressed positions. This effect is inconsistent, however; speakers JD, TO, VS, and ZL do not seem to raise F2 minima in one or both of the non-Stressed contexts. Overall then, it would appear that the non-palatalized vowel space is centralized as stress becomes weaker, with lowering of maximum F1, and of maximum F2 to a lesser degree, being the largest of the effects.

Given the above, it is not surprising that the size of the non-palatalized vowel space (as measured by 'difference' in table 3a) reduces in F1 for all speakers, moving from Stressed to non-Stressed contexts. This decrease occurs progressively, by roughly 1.5 ERB at each step on average, though TO and ZL have virtually identical Stressed and Prestressed F1 ranges. The F2 range likewise diminishes, by about 2 ERB on average from Stressed to Prestressed, and by about 1 ERB from Prestressed to Unstressed.

[6] Speakers VS and JD had rather high-pitched voices, with mean F0 values of about 275 and 250 Hz, respectively. Speaker VS's high vowel data in particular seem to behave differently from that of the other speakers. We can only assume that due to the high pitch of her voice, there were problems tracking F1 in the high vowel contexts for this speaker (although the data were carefully checked by the second author).

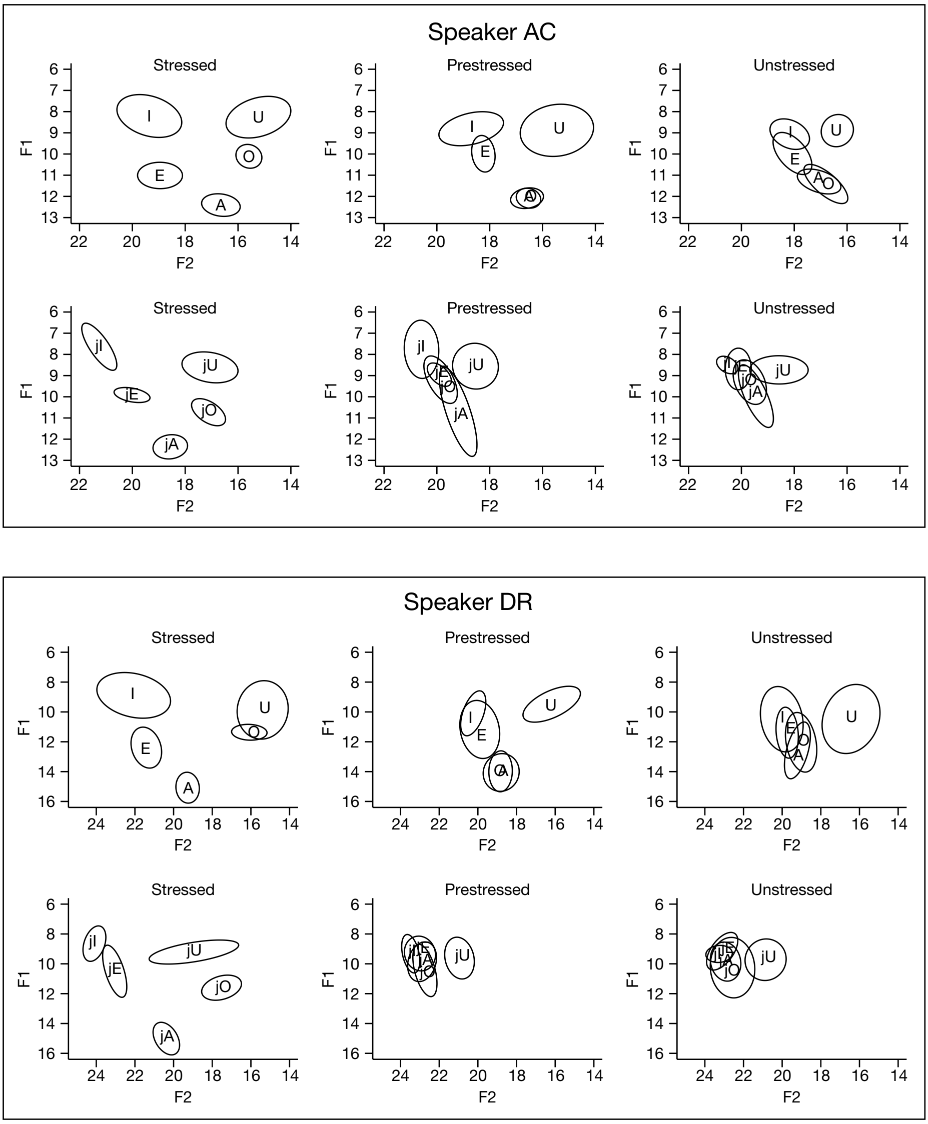


Fig. 2. Plots of non-palatalized (upper row) and palatalized (bottom row) vowel data for all 9 speakers in all stress contexts. Vowels following palatalized consonants are indicated in the figures with a preceding 'j'. All scales are in ERB. Ellipses represent 2.45 SD around the mean.

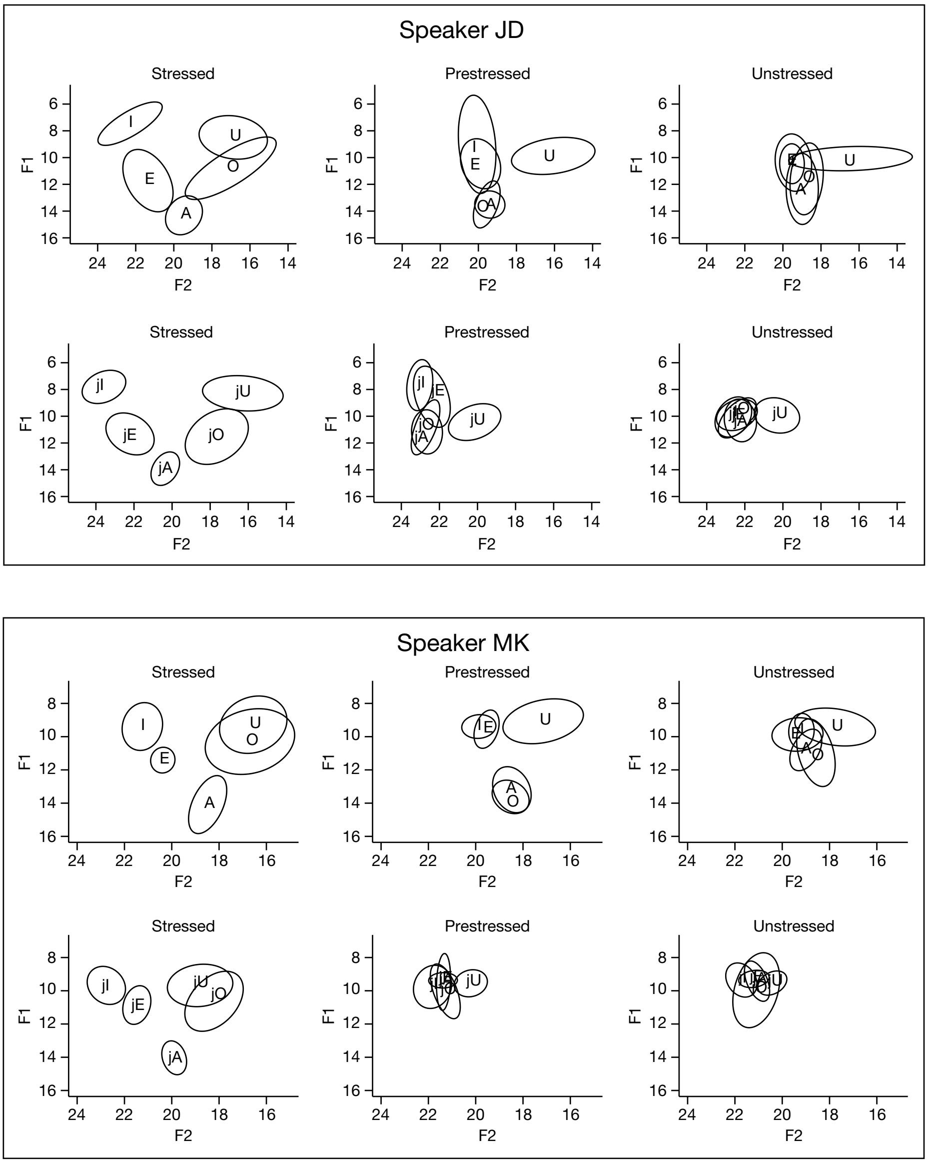


Fig. 2 *(continued).* Plots of non-palatalized (upper row) and palatalized (bottom row) vowel data for all 9 speakers in all stress contexts. Vowels following palatalized consonants are indicated in the figures with a preceding 'j'. All scales are in ERB. Ellipses represent 2.45 SD around the mean.

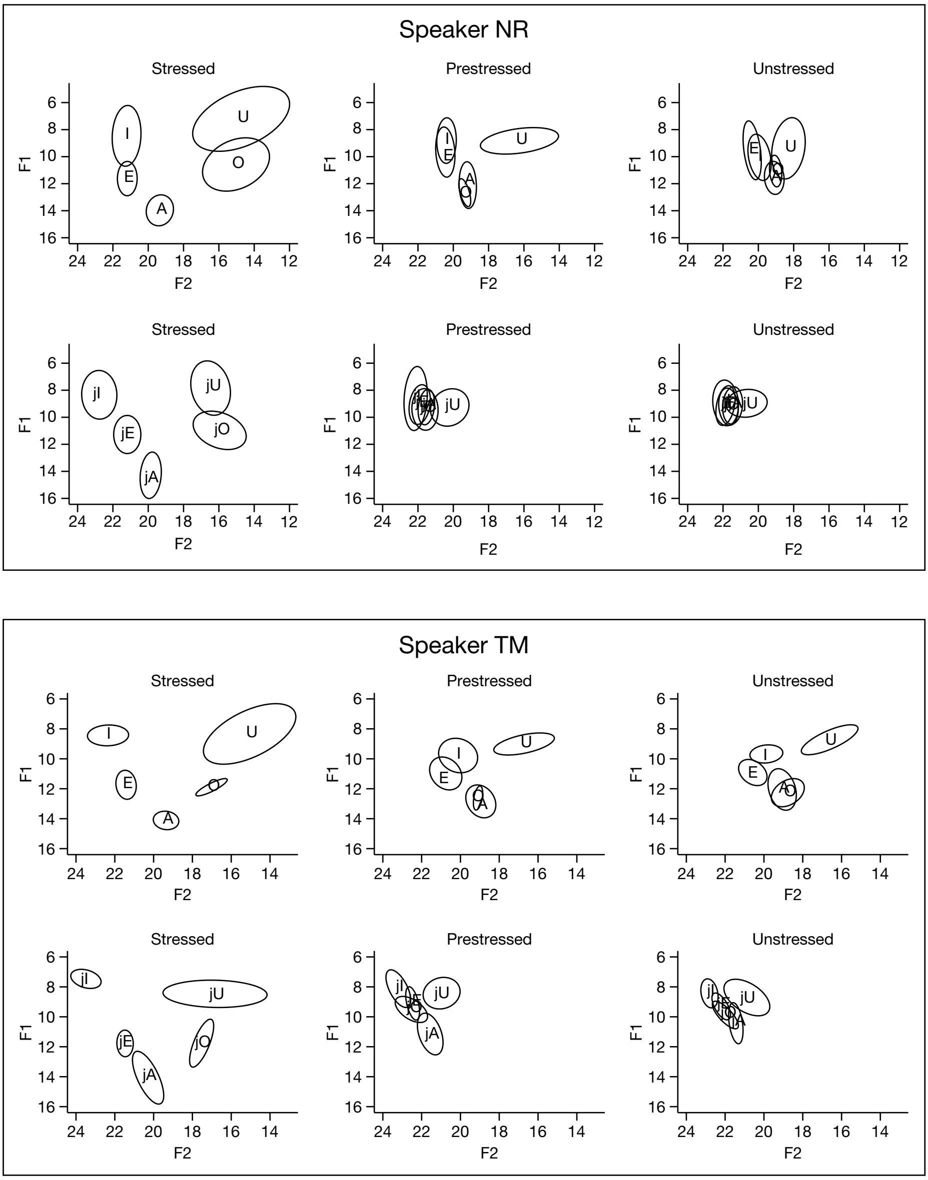
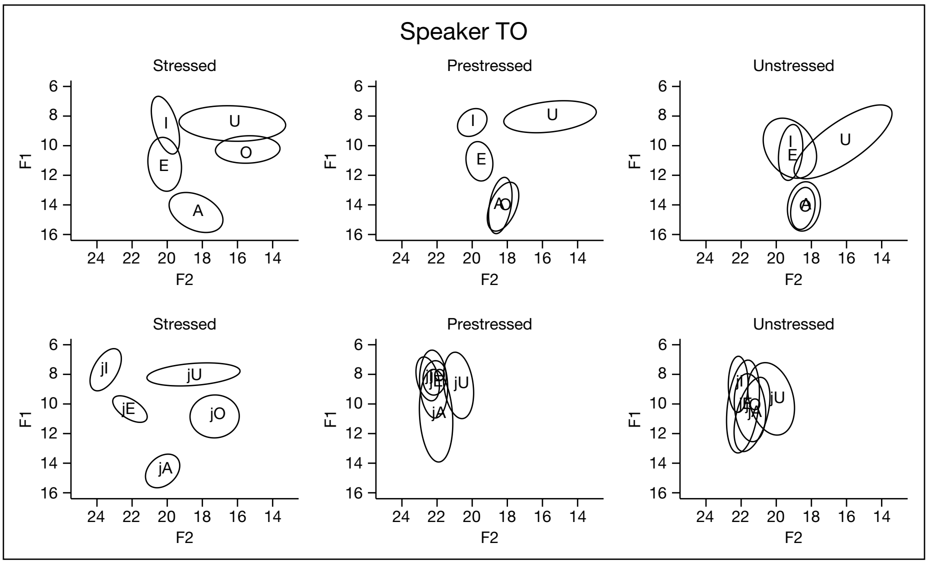


Fig. 2 *(continued).* Plots of non-palatalized (upper row) and palatalized (bottom row) vowel data for all 9 speakers in all stress contexts. Vowels following palatalized consonants are indicated in the figures with a preceding 'j'. All scales are in ERB. Ellipses represent 2.45 SD around the mean.



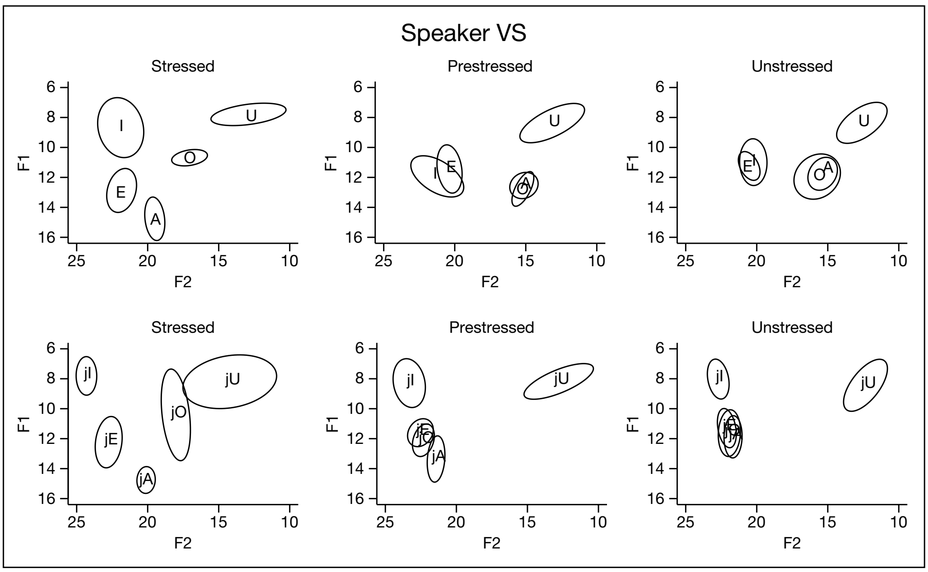


Fig. 2 *(continued).* Plots of non-palatalized (upper row) and palatalized (bottom row) vowel data for all 9 speakers in all stress contexts. Vowels following palatalized consonants are indicated in the figures with a preceding 'j'. All scales are in ERB. Ellipses represent 2.45 SD around the mean.

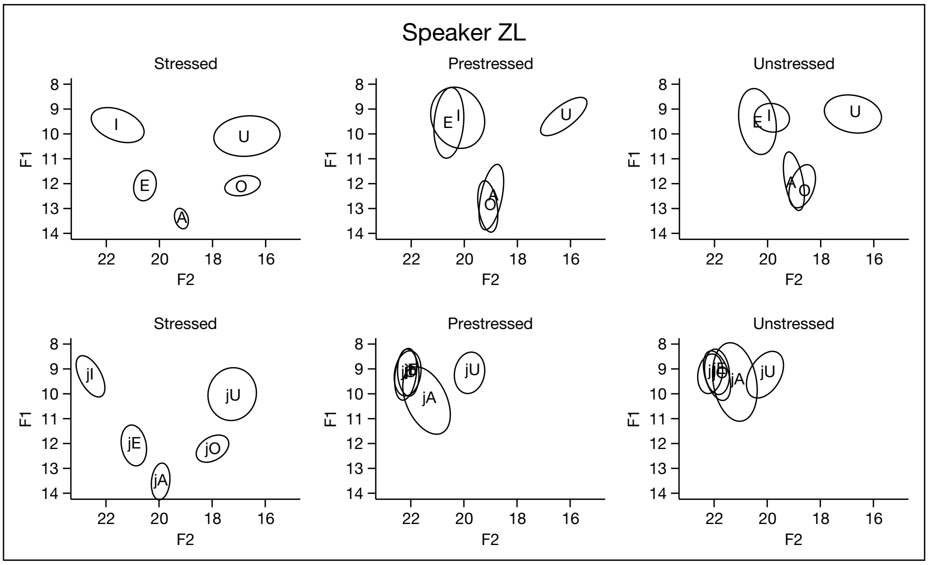


Fig. 2 *(continued).* Plots of non-palatalized (upper row) and palatalized (bottom row) vowel data for all 9 speakers in all stress contexts. Vowels following palatalized consonants are indicated in the figures with a preceding ‘j’. All scales are in ERB. Ellipses represent 2.45 standard deviations around the mean.

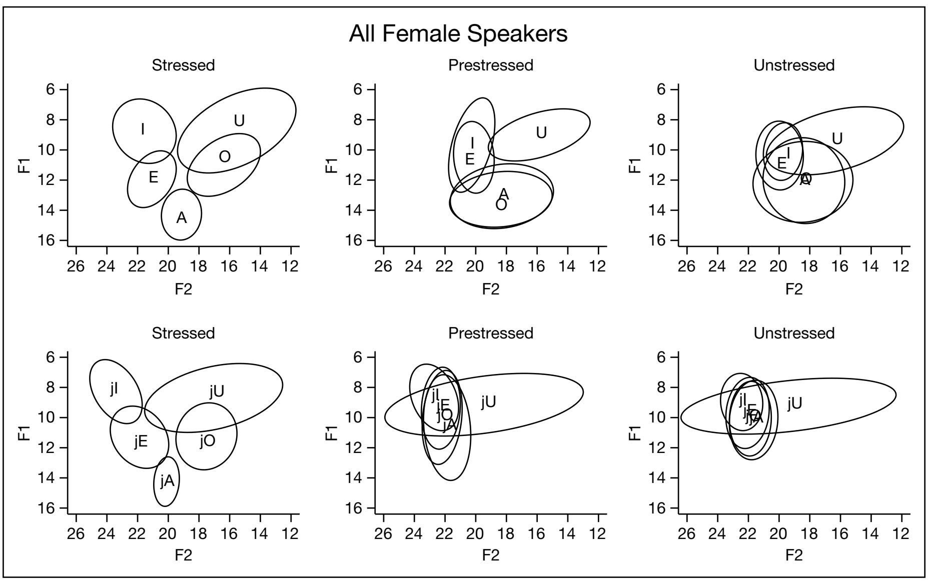


Fig. 3. Plots of non-palatalized and palatalized vowel data in all stress contexts, for all 8 female speakers combined. Data presentation as in figure 2.

Table 3. F1 and F2 vowel space measures for each vowel context, averaged across all female speakers

a Data for non-palatalized tokens

Context	F1			F2		
	min.	max.	diff.	min.	max.	diff.
S	7.75	14.75	6.99	14.72	22.09	7.37
P	8.28	13.65	5.37	15.17	20.67	5.50
U	8.81	12.77	3.97	15.67	20.10	4.43

b Data for palatalized tokens

Context	F1			F2		
	min.	max.	diff.	min.	max.	diff.
S	7.76	14.74	6.98	16.32	23.71	7.39
P	8.02	11.33	3.31	19.04	22.99	3.95
U	8.47	10.76	2.29	19.05	22.69	3.65

All values are in ERB. See text for details of calculation.

This F2 reduction generally occurs for all speakers, though TO actually shows an increase from Stressed to Prestressed, and some speakers, e.g., DR and TM, seem to reduce very little from Prestressed to Unstressed.

Turning now to the palatalized data (figures 2, 3, table 3b), there is again a tendency to raise minimum F1 overall from Stressed to non-Stressed contexts. This effect is small and not uniform, however, some speakers preserving F1 or lowering it. In contrast, the overall lowering effect on F1 maxima is by roughly 3.5 ERB from Stressed to Prestressed. It is much weaker from Prestressed to Unstressed, and the change is slight or non-existent here for DR, MK, NR, and TO. F2 minima raise on average by about 3 ERB between Stressed and Prestressed contexts, though speakers vary, and 1 (VS) even lowers this value. There is no apparent change overall from Prestressed to Unstressed. Finally, F2 maxima lower by roughly 1 ERB moving from Stressed to Unstressed. Overall, then, the effect is again one of centralization, though the lowering effect on F1 maxima, and raising of F2 minima, between Stressed and Prestressed contexts, are the predominant effects.

Once again this leads to a contraction of the overall vowel space. The average F1 range for female speakers shrinks by roughly 4 ERB from Stressed to Prestressed, and by about 1 ERB from Prestressed to Unstressed. Though there is a great deal of speaker variation in the degree of F1 contraction between contexts, all speakers evince increasing contraction moving from Stressed toward Unstressed. (The one exception appears to be MK, whose Prestressed and Unstressed F1 ranges are roughly the same.) For F2, the average range contracts by roughly 3 ERB from Stressed to Prestressed, with all speakers but VS showing contraction. No clear difference emerges between Prestressed and Unstressed syllables.

The averages given above, and the figures, suggest the following major differences between the palatalized and non-palatalized vowel spaces. First, a greater lowering of maximum F1 occurs for palatalized vowels. In consequence, the overall F1 range is notably smaller for these vowels in non-Stressed contexts. (Given the impressionistic descriptions, this should not be surprising: non-palatalized vowels reduce to [i,u,ɐ], palatalized vowels to [i,u].) Second, F2 maxima are generally higher across contexts for palatalized vowels, but it is the F2 minima that distinguish palatalized and non-palatalaized vowels most notably, those of palatalized vowels being much higher. This seems largely because [u] is fronted in the palatalized context, and [i] backed in the velarized context, in non-Stressed syllables. The result is again a more contracted space for palatalized vowels.

3.2.3 Incomplete Neutralization

Inspection of figures 2 and 3 suggests that the phonetic facts are in some respects out of line with the phonological description of neutralization as given in the introductory section. On the one hand, there appears to be complete overlap between /o/ and /a/ in the non-Stressed contexts for non-palatalized vowels. This holds for all speakers. On the other hand, while non-palatalized /i/ and /e/ seem close together in the vowel space in non-Stressed contexts, and show quite a bit of overlap, they still retain elements of their standard relationship. For example, /i/ has a lower F1 in many cases. Something similar seems true for palatalized /ji,je,jo,ja/, which are described as neutralized. Again these vowels are often quite close and overlapped. Yet abstracting away from a great deal of interspeaker variation, they seem to retain traces of their characteristic distinctions. This can be seen clearly in figure 3.

This apparent distinction between the complete neutralization of /a/ and /o/ on the one hand, and the other incomplete neutralizations, is supported by an analysis of variance. When the data were put through a multivariate ANOVA with fixed factors speaker × vowel × stress, and dependent variables F1 and F2, all factors and all interactions of factors were significant at $p < 0.001$. (Palatalized and non-palatalized data were run separately.) Post-hoc tests on the vowels showed that all vowel contrasts as well were significant at that level. However, when the Stressed context was removed from the analysis, leaving only Prestressed and Unstressed tokens in the dataset, post-hoc tests on the non-palatalized vowels showed that /a/ and /o/ were the only vowels which were not significantly different from each other, neither in F1 nor F2, even though there was an overall effect of stress on both of these dependent variables. All other non-palatalized vowels, and all palatalized vowels, were different from each other in both F1 and F2 at the $p < 0.001$ level. (The one exception is non-palatalized /i/ and /e/, which were not significantly different in F2.)[7]

These statistical results support a systematic difference in production for all pairs but non-palatalized /a/ and /o/ in non-Stressed positions. However, they tell us nothing

[7] We also ran a repeated-measures ANOVA on all of the data using the median F1 and F2 values, with factors consonant × vowel × stress. The main effect for vowel type was significant, and for post-hoc analyses of this factor we set alpha to 0.008 (0.05/6, since there were six vowel comparisons in each context, as before). All stressed pairs were significantly different. Among the non-Stressed non-Palatalized pairs, all were significantly different except for /a/ – /o/ and /i/ – /e/, exactly those that are impressionistically neutralized. However, results for the non-Stressed Palatalized vowel pairs were puzzling. In this context, recall, only /ju/ remains impressionistically distinct. Consistent with this, /ja/ – /jo/ and /je/ – /jo/ were not significantly different; however, /ji/ – /je/ and /je/ – /ja/ were. More troubling, /ju/ was not distinguished from its comparison vowels /ji/ and /jo/. Failure to distinguish /ju/ may have occurred due to a great deal of interspeaker variation in the pronunciation of /ju/ (fig 3), but it is hard to understand why this impressionistically distinct vowel was statistically 'neutralized' while some impressionistically neutralized pairs were significantly different. Perhaps the difficulty of sampling formant values for palatalized vowels across speakers and stress contexts played a role. Whatever the cause, since this test failed to distinguish even some impressionistically distinct vowels, it was too weak for our purposes.

about the behaviour of individual speakers. Equally important, they do not tell us about the likely perceptual distinctiveness of pairs whose productions do differ, but in many cases by only a small amount. In this section we apply Euclidean distance measures, measures of F1–F0 and F2–F1, and a Bayesian classification to the Russian data in order to flesh out our results. While these cannot provide direct evidence about perception, they do give some basis for conclusions about the nature and precise patterning of incomplete neutralizations in Russian.

How to measure the acoustic distance between vowels, in a way likely to be perceptually relevant, is a complex question about which much still remains unknown. Some works compare vowels based only on spectral prominences, in particular vowel formant and possibly fundamental frequencies [Liljencrants and Lindblom, 1972; Boë et al., 1994; Schwartz et al., 1997; Ménard et al., 2002]. Others, such as Lindblom [1986, 1990] and Bladon and Lindblom [1981], advocate comparing entire vowel spectra (suitably transformed in order to take into account acoustic-to-auditory transformations). The latter works argue that we cannot know a priori what spectral properties are relevant, and they cite evidence that properties other than frequency prominences matter (e.g., intensity).[8] A synthesis of these positions is found in Diehl et al. [2003]. The latter work, as well as Boë et al. [1994] and Schwartz et al. [1997], point out that the relative contribution made by spectral properties other than prominences often appears to be relatively slight. Our analyses rely on (auditorily transformed) spectral prominences.[9]

We first consider Euclidean distances, following Liljencrants and Lindblom [1972], Lindblom [1986], and others. We measure distances only between vowels which are adjacent in the vowel space along either F1 or F2: hence, the distance measures presented below are for [i~e], [e~a], [a~o], [o~u], [i~u] and [e~o] in both the palatalized and non-palatalized vowel spaces.

Refer once again to figures 2 and 3. (Refer also to the Euclidean distance measures presented in table A3 in the ‘Appendix’, if desired. The highlighted numbers in that table indicate distances below 1 ERB, a critical band.) We continue to take a distance of 1 ERB or less as a rough indication that two vowels may be perceptually indistinguishable. Beginning again with non-palatalized vowels, the distance between /a/ and /o/ in Prestressed and Unstressed contexts falls below this threshold for every speaker. The vowels /i/ and /e/ fall within 1 ERB in both non-Stressed contexts for 4 speakers, but only in the Unstressed context for 4 other speakers, and not at all for 1 speaker. However, this distance remains within 2 ERB for 8 of the 9 speakers. There are no other pairs of vowels that systematically approximate a distance of 1 ERB or less.

Among the palatalized vowels, /ju/ differs from both /ji/ and /jo/ systematically across contexts. It is among /ji,je,jo,ja/, the vowels described as neutralizing, that we indeed find a tendency for distances to fall within 1 ERB in non-Stressed contexts. However, this effect is much less systematic. Depending on the pair of vowels measured, we find only between 3 and 7 of the speakers collapsing the pair to within 1 ERB in both non-Stressed contexts; some collapse a given pair only in the Unstressed context, and some not at all. (Speaker TO, oddly, collapses /ji/ and /je/ to within 1 ERB only in the Prestressed context.) Only 2 speakers, MK and NR, collapse all of these

[8] Bladon and Lindblom [1981] also note the potential difficulties of applying spectral peak measures to sounds having zeros, such as nasalized vowels.
[9] However, in section 3.2.4 we do consider in general terms the effect of giving more weight to F1 distinctions over F2 distinctions.

pairs to within 1 ERB in both non-Stressed contexts. In addition, we cannot be certain how a given speaker will treat one pair of vowels based on how she treats another. For example, speaker ZL collapses the pairs /ji/ and /je/, and /je/ and /jo/, to within 1 ERB in both non-Stressed contexts, but collapses /je/ and /ja/, and /ja/ and /jo/, only in the Unstressed context. Yet speaker JD collapses /ja/ and /jo/ in both contexts, and other pairs only in the Unstressed context. Still, there is a greater likelihood overall for /ji/ and /je/, and for /je/ and /jo/, to be collapsed.[10]

Consider now a different way of assessing neutralization. It has been proposed that vowels can be well classified by height using F1–F0, and by backness using F2–F1, the auditorily transformed difference between F1 and F0 or F2 and F1, respectively [Syrdal and Gopal, 1986; Ménard, et al., 2002]. (Note that both of these employ Bark rather than ERB.) These measures are used here to generate independent hypotheses regarding how the vowels in the non-Stressed contexts may be perceived by Russian listeners. We provide mean F1–F0 and F2–F1 values in ERB for each vowel in each stress context in table A4 in the 'Appendix'.

As a baseline, consider first the Stressed vowel data. Speaker means for F1–F0 lie roughly between 0.5 ERB and 4.5 ERB for the high vowels (with the exception of speaker ZL's /u/ and /ju/), between 4 and 8 ERB for the mid vowels, and between 7.5 and 10 ERB for the low vowel. This is true for both palatalized and non-palatalized vowels. In finding a relatively clear distinction among these vowel heights, these results are in line with those presented in Ménard et al. [2002]. The results for F2–F1 are somewhat surprising, suggesting the following ordering from back to front in the perceived vowel space: /a/ (4–6 ERB), /o/ (4–8 ERB), /u/ (5–10 ERB), /e/ (8–13 ERB) and /i/ (11–17 ERB), again regardless of consonantal context.

If we take a difference of 1 ERB or less as suggesting indistinguishability, then two vowels whose values for F1–F0 lie within 1 ERB of each other would be categorized together in height, and two with F2–F1 values within 1 ERB of each other would be categorized together in backness. Considering first the non-palatalized vowels, F1–F0 and F2–F1 for /a/ and /o/ are both within this threshold in both Prestressed and Unstressed contexts for every speaker. (This can be seen in table A5, if desired, where differences of less than 1 ERB are shaded.) F1–F0 and F2–F1 values for /i/ and /e/ are both within 1 ERB for 3 speakers in both contexts; for 3 speakers in only the Unstressed context; and for 3 speakers in neither context. There is therefore broad agreement with the Euclidean distance results. This agreement extends to the behaviour of particular speakers.

The palatalized vowel data also broadly parallel the Euclidean distance results. For pairs involving /ji,je,ja,jo/, we find between 3 and 6 of the speakers collapsing to within 1 ERB in both non-Stressed contexts, depending on the vowel pair involved; between 1 and 4 collapse only in the Unstressed context, and between 1 and 5 not at all. (Once again TO collapses /ji/ and /je/ to within 1 ERB only in the Prestressed context.) The behaviour of particular speakers is largely similar across the two sets of results. Here, unlike with Euclidean distance data, we can consider the separate contributions of height and backness in neutralization. We note that cases where F1–F0 but not F2–F1 fall within 1 ERB outnumber the reverse (3 out of 3 for /ji,je/, 3 out of 4 for /je,ja/, and 1 out of 1 for /ja,jo/).

[10] Speaker VS maintains an unusually large contrast between /ji/ and /je/ in all three stress contexts. We cannot discount the possibility that F0 was mistracked as F1 for this speaker's high vowels, as already mentioned.

Finally, we present confusion matrices based on a Bayesian classification of the data. While these confusion matrices are not based on perceptual data, we believe that they will provide an indication of which vowels are likely to be confused with each other given their distribution in the vowel space and their variability across speakers. The Bayesian classification was carried out on the 8 female speakers' ERB-transformed F1 and F2 data using the Round-Robin training and testing method. In this method, the vowel data for each speaker are classified following training on the remaining speakers' data (in this case, each speaker's data are classified following training on the other 7 speakers' data). The results from the 8 training and testing sessions are then pooled, and the resulting classifications are presented as percentage values in a standard confusion matrix.

Table 4 presents the confusion matrices resulting from Bayesian classification of the data. Confusion matrices are presented separately for the palatalized and the non-palatalized data, and for each stress context. The total percentage of correctly classified tokens is given at the top of each table.

It can be seen that the stressed vowel data are extremely well classified in both the palatalized and the non-palatalized contexts, with overall correct classification rates of over 90%. This is despite the fact that no attempt was made to normalize the data across speakers (except for excluding the male speaker). Non-palatalized /i e a/ have correct classifications of nearly 100%, as does /ja/. /ji/ and /je/ have correct classifications of around 90%; it is not surprising that the front vowels may fail to be distinguished in the palatalized context. There is also somewhat less success classifying /o/ and /u/, as well as /jo/ and /ju/. In the first case, this may be due to the [uo]-like pronunciation of /o/ by some speakers (see below), whereas in the second case, it may be due to the palatalization gesture pulling the tongue body upward for /jo/.

Turning to the Pre- and Unstressed data, the overall percentage correct is much lower, as is to be expected given that we have not collapsed the data according to descriptions of neutralization; that is, we present results according to the underlying phonemes. In the non-palatalized context, there is a much poorer distinction between /i/ and /e/; however, they are both correctly classified better than chance in both the Pre- and Unstressed contexts (although only marginally so for /e/ in the Unstressed context). The vowels /a/ and /o/ are even more poorly distinguished from each other. Though /o/ is correctly classified at better than chance, /a/ is more often classified as /o/ than /a/. This suggests neutralization. We note that both vowels also have a good chance (around 15%) of being classified as /u/. This was not true for all speakers' classifications; it may be that a more careful normalization of the data would reduce the number of /u/ classifications. /u/ in turn is correctly classified almost 100% of the time in the Prestressed context, and around 80% of the time in the Unstressed context, suggesting that it maintains its phonological identity well across stress contexts.

The Pre- and Unstressed palatalized data are much less consistent than the non-palatalized data. /ju/, which phonologically should be distinct from all the other vowels in these contexts, is often classified as /ja/: it is correctly classified 70% of the time in the Prestressed context, but only 50% of the time in the Unstressed context. Regarding the remaining vowels, there is a good deal of misclassification among them, as is to be expected. However, the classification rates suggest that vowels continue to maintain some degree of their underlying distinctions. The vowel /ji/ is more likely to be classified as /ji/ or /je/ than as either /jo/ or /ja/. The vowel /je/ is classified as /ji/, /je/, or /ja/ most often, with /jo/ a lesser possibility. We find /ja/ classified most often as either /je/,

Table 4. Confusion matrices based on Bayesian classification of 8 female speakers' data using the round-robin training and testing method

a Non-palatalized tokens

	I	E	A	O	U
Stressed (91.3%)					
I	99.1	0.0	0.0	0.0	0.9
E	1.6	97.6	0.8	0.0	0.0
A	0.0	0.8	97.7	1.6	0.0
O	0.0	0.0	0.0	85.0	15.0
U	0.0	0.0	0.0	22.6	77.4
Prestressed (60.3%)					
I	60.5	38.7	0.0	0.8	0.0
E	41.9	55.6	1.7	0.9	0.0
A	0.0	6.0	29.1	50.4	14.5
O	0.0	2.5	24.2	60.0	13.3
U	0.0	0.0	1.8	0.0	98.2
Unstressed (55.0%)					
I	61.1	31.0	0.0	5.3	2.7
E	37.3	51.7	1.7	7.6	1.7
A	6.5	12.9	21.0	44.4	15.3
O	9.3	5.9	9.3	59.3	16.1
U	6.5	1.6	5.7	3.3	82.9

b Palatalized tokens

	jI	jE	jA	jO	jU
Stressed (90.8%)					
jI	91.9	8.1	0.0	0.0	0.0
jE	10.2	89.0	0.0	0.0	0.8
jA	0.0	1.6	98.4	0.0	0.0
jO	0.0	0.0	0.0	88.1	11.9
jU	0.0	0.0	0.0	13.2	86.8
Prestressed (44.7%)					
jI	53.5	25.4	4.4	9.6	7.0
jE	31.7	35.8	17.5	8.3	6.7
jA	4.4	20.4	46.0	23.9	5.3
jO	9.2	40.3	27.7	18.5	4.2
jU	0.0	2.5	21.2	5.9	70.3
Unstressed (39.2%)					
jI	60.7	23.1	1.7	7.7	6.8
jE	34.5	22.7	21.0	16.8	5.0
jA	10.2	19.5	48.3	15.3	6.8
jO	20.7	28.4	31.9	13.8	5.2
jU	0.9	4.3	44.4	0.0	50.4

Data were classified using F1 and F2. Confusion matrices are presented separately for each stress context, and confusions are expressed as a percentage. The label for each row gives the true identity of the data to be classified, and the label for each column gives the classification result; hence, rows sum to 100% whereas columns do not. At the top of each table is given the total percentage of correctly classified tokens.

/ja/, or /jo/. Finally, /jo/ is often classified as any of /ji/, /je/, /ja/, or /jo/, especially in the Unstressed context, and it is notably poorly classified as /jo/ itself.

3.2.4 Pairwise Vowel Spacing

Here we present results bearing on vowel spacing more generally. We ran univariate ANOVA on the distance values presented in table A3 in the 'Appendix', with Speaker and Stress as fixed factors; separate analyses were conducted for the non-palatalized and palatalized data. Alpha was set at 0.025 (rather than 0.05), given that non-palatalized and palatalized data were kept separate, and posthoc analyses (for Stressed vs. Prestressed, and Prestressed vs. Unstressed) were adjusted to 0.0125, according to the Bonferroni method. Note that speaker VS was removed for these analyses, due to the extreme values of her distance means.

For the non-palatalized data presented in table A3a, the mean distance in the Stressed vowel context was 3.74 ERB; 2.73 ERB in the Prestressed context, and 1.83 ERB in the Unstressed context. The effect of Stress was significant at $p < 0.001$ [$F(2, 144) = 20.23$]; there was no effect of Speaker, and no interaction between Speaker and Stress. Post-hoc tests showed that Stressed was significantly different from Prestressed (difference in means is 1.01 ERB, with a 95% confidence interval of 0.29 to 1.75), and Prestressed was significantly different from Unstressed (difference in means is 0.90 ERB, with a 95% confidence interval of 0.16 to 1.62).

For the palatalized data presented in table A3b, the mean distance in the Stressed vowel context was 3.76 ERB; 1.36 ERB in the Prestressed context, and 0.96 ERB in the Unstressed context. The effect of Stress was significant at $p < 0.001$ [$F(2, 144) = 117.27$], and the effect of Speaker was just significant at $p = 0.024$ [$F(7, 144) = 2.40$]. However, there was no interaction between Speaker and Stress. Post-hoc tests showed that Stressed was significantly different from Prestressed (difference in means is 2.40 ERB, with a 95% confidence interval of 1.90 to 2.85); however, Prestressed was *not* significantly different from Unstressed (difference in means is 0.40 ERB, with a 95% confidence interval of −0.06 to 0.90).

From these results we can conclude, first, that the overall spacing among Stressed vowels is roughly the same for palatalized and non-palatalized vowels. Second, stress has a significant effect on overall vowel spacing; this includes a distinction between Prestressed and Unstressed for non-palatalized vowels, but not for palatalized vowels. Finally, palatalized vowels seem to undergo the greater reduction in spacing.

Given the existence of (in)complete neutralization in non-Stressed contexts, it is no surprise that the overall distance among underlying vowel pairs shrinks outside of stress. However, since our results above, and impressionistic descriptions, suggest that in many instances the incompletely neutralized vowels cannot be distinguished, it is worth considering distances only among vowels that are impressionistically distinct in a systematic way. The main question considered here is therefore, assuming that incompletely neutralized vowels are indeed indistinguishable and so count as the same, how constant is vowel spacing within and across contexts?

Our analyses reveal that the Euclidean distance between vowels in the Stressed vowel space (where no neutralizations occur) is anything between 1 and 7 ERB, with most of the distances ranging between 2 and 5 ERB. There is a tendency for the Euclidean distance to be between 2 and 4 ERB in the Prestressed context for pairs of vowels that do not impressionistically neutralize, and between 1 and 3 ERB in the Unstressed context. There is therefore a good deal of variability in distances.

However, when we consider average values across speakers, the Russian vowels are more evenly spaced, as can be seen in table 5a. Since listeners successfully categorize vowels as spoken by many different speakers, it seems reasonable to speculate that they have some implicit knowledge of these average values, and that they therefore have more than an abstract validity.

Considering first the stressed vowels (which are averages of the data in table A3 in the 'Appendix'), for those differing primarily in F1, we could say that average distances are consistently around 3 ERB if it weren't for the position of /o/, which seems too close to /u/ and too far from /a/. We suspect that this is due to a tendency for the vowel /o/ to be slightly diphthongized to /uo/ under stress. This tendency is noted in particular among younger female speakers, as borne out by our data where speakers DR, MK and TO, as well as AC who is male but younger, have Euclidean distances of less than 2 ERB between /o/ and /u/. Our auditory impression and spectral observations of these speakers' stressed /o/ utterances suggest a strong diphthongization of this vowel in the stimulus word 'vodka'. For vowels differing in F2 the distances are much larger. Note the small effect that palatalization seems to have on these values.

Works that seek to predict vowel inventories based on principles of dispersion, such as Lindblom [1986] and Schwartz et al. [1997], observe that in order for predictions to match attested patterns well, distances in backness and/or roundness must be weighted less heavily than those in height. According to Schwartz et al.[1997], the vowel space must be distorted until the full F2 range is 0.5–0.75 the length of the F1 range, the distances between vowels affected accordingly. Table 5b shows Euclidean distances between Russian vowels assuming a weighting for F2 of 0.625 × F1, adapting a strategy employed by Herrick [2003].[11] Given the purpose of the weighting, it is not surprising that distances in F2 are now much more like those in F1, in the neighbourhood of 3 ERB. There is evidence that F1 does indeed contribute more to vowel perception than do higher formants [Nooteboom, 1968; Lindblom, 1975; Benkí, 2003; Diehl et al., 2003, and references therein]. Known perceptual biases that might explain this underrating of the F2 dimension include the relative intensity and resistance to masking of lower-frequency components. However, the value of 0.5–0.75 itself is not deduced from such considerations, since the latter are too poorly understood; rather, Schwartz et al. [1997] posit this value in order for their dispersion model to output the correct vowel inventories. Table 5b should be read with this caution in mind.

Consider now the non-Stressed vowels in table 5. To get these averages, raw values for impressionistically neutralized vowels were combined within speakers before taking speaker averages, and the overall averages reported here are based on those speaker averages. For the Prestressed, non-palatalized data, the impressionistically neutralized vowels are /i,e/ vs. /a,o/, /a,o/ vs. /u/, and /i,e/ vs. /u/. As can be seen, these

[11] Working from the discussion in Schwartz et al. [1997, pp. 275–276], we first found the full F1 range in ERB for each speaker, by subtracting the lowest F1 value in table A2 from the highest for that speaker. We did the same for F2. (The lowest and highest values were not necessarily found in the same context. For instance, the lowest and highest F2 for a speaker were typically found in the non-palatalized and palatalized contexts, respectively.) We then computed F2/F1 for each speaker and took the average: 1.253. In order to bring this ratio to 0.625, it must be multiplied by 0.499 (0.499 × 1.253 = 0.625). We therefore multiplied the actual F2 distance values by 0.499, and computed Euclidean distances based on these adjusted F2 distances.

Table 5. Euclidean distance measures between vowel pairs: averages across speakers

a Simple Euclidean distances

Non-palatalized consonant	Context						
		/i/ vs. /e/	/e/ vs. /a/	/a/ vs. /o/	/o/ vs. /u/	/i/ vs. /u/	/e/ vs. /o/
	S	3.22	3.04	4.13	2.59	5.85	4.62
		/i,e/ vs. /a,o/		/a,o/ vs. /u/		/i,e/ vs. /u/	
	P	3.87		4.77		4.31	
	U	2.48		3.30		3.38	

Palatalized consonant	Context						
		/ji/ vs. /je/	/je/ vs. /ja/	/ja/ vs. /jo/	/jo/ vs. /ju/	/ji/ vs. /ju/	/je/ vs. /jo/
	S	3.24	3.40	3.85	2.87	6.03	4.28
		/ji,je,ja,o/ vs. /ju/					
	P	2.74					
	U	2.46					

b Euclidean distances adjusted by weighting 72 (see text)

Non-palatalized consonant	Context						
		/i/ vs./e/	/e/ vs. /a/	/a/ vs. /o/	/o/ vs. /u/	/i/ vs. /u/	/e/ vs./o/
	S	3.17	2.56	3.53	2.38	3.44	2.81
		/i,e/ vs. /a,o/		/a,o/ vs. /u/		/i,e/ vs. /u/	
	P	3.38		4.26		2.38	
	U	2.06		2.81		1.85	

Palatalized consonant	Context						
		/ji/ vs. /je/	/je/ vs. /ja/	/ja/ vs. /jo/	/jo/ vs. /ju/	/ji/ vs. /ju/	/je/ vs. /jo/
	S	3.00	3.08	3.24	2.63	3.53	2.55
		/ji,je,ja,o/ vs. /ju/					
	P	1.54					
	U	1.31					

All values are in ERB. The stressed vowel data are averages of the data in table A3 in the 'Appendix'. For the non-stressed data, raw data for impressionistically neutralized vowels were combined within speakers before taking speaker averages, and the overall averages reported here are based on those speaker averages.

distances are overall only slightly larger than those seen among the Stressed vowels. However, more can be said when vowel pairs are separated according to whether they differ primarily in height or backness. The distances between /i,e/ and /a,o/, or /a,o/ and /u/, differing primarily (but not entirely) in height, are roughly 1 ERB larger than the distances among Stressed vowels differing primarily in height. (This is true with or without weighted F2.) In other words, Russian Prestressed [i] and [ɐ], or [ɐ] and [u], really do seem somewhat farther apart than Stressed [i] and [e], or [e] and [a], as the transcriptions imply they are. On the other hand, the other non-Stressed vowels are closer together than are Stressed vowels. Non-Stressed /ji,je,ja,jo/ versus /ju/ are particularly close.

To summarize the results here: First, there is a good deal of interspeaker variability in vowel distances. Second, differences in F2 seem overall much larger than those in F1 unless we employ a weighted measure. Third, assuming weighted measures, distances between Stressed vowels seem consistently around 3 ERB. But Prestressed [ɐ] is further than this from [i,u]. On the other hand, non-Stressed vowels otherwise seem closer together than Stressed, with palatalized [ji,ju] being notably close. Finally, we can also conclude that impressionistically contrasting Russian vowels maintain an absolute minimal distance of at least 2.5 ERB across contexts in unweighted distances, or roughly 1.5 ERB in weighted distances, averaging across speakers.

4. Discussion

4.1 Incomplete Neutralization

Our phonetic investigation supports the impressionistic descriptions of Russian phonological vowel reduction only in some respects. The claims involving underlying /a,o/ after non-palatalized consonants are largely borne out. By all the criteria set out in section 3 – comparison of means, Euclidean distances, formant differences, and Bayesian classification – these vowels neutralize in non-Stressed syllables.

The facts are different for non-palatalized /i,e/ and for palatalized /ji,je,ja,jo/. Means across speakers for these vowels are all significantly different. According to the other three measures, some speakers likely maintain a perceptible distinction between /i/ and /e/ in Prestressed (but not in Unstressed) position, and some in both non-Stressed positions, contrary to the described neutralization. For palatalized vowels, the expected full-scale perceptual neutralization between /i,e,a,o/ seems borne out by few speakers (MK and NR; also JD for Unstressed position). For most, the data suggest only a strong tendency for these vowels to raise and front toward /ji/, and lingering effects of their underlying differences are often evident. There are many speaker differences for these incompletely neutralized vowels.

This state of affairs resembles the well-known finding of near merger of vowels documented in Labov [1994], and incomplete neutralization of voicing in word-final obstruents in languages such as German and Catalan [see for example Port and Crawford, 1989, Charles-Luce, 1993, and references therein]. It is doubtful that our findings can be attributed solely to orthography. Though Russian spelling does distinguish between /i/ and /e/, as well as among /ji,je,jo,ja/[12], it equally distinguishes between /a/ and /o/. Since the latter neutralize completely in our data, it is not the case that speakers are allowing spelling distinctions to override neutralization in any general sense. A reviewer notes that orthography might nevertheless matter. In particular, it is known that neutralization between /a/ and /o/ is historically older than the other neutralizations. The vowels /je,ja,jo/ were impressionistically distinct from /ji/ in Prestressed position

[12] Except for /je/ vs. /jo/, which are never distinguished in non-Stressed positions (and rarely even when stressed). As noted in the last section, neutralization between these vowels seems more relatively successful.

within the last century in Standard Russian. And today there is a great deal of dialectal variation in the existence of, and the precise patterning of, neutralizations among /ji,je,ja,jo/. Incomplete neutralization among these vowels (and possibly between /i/ and /e/) plausibly reflects a linguistic change still in progress. Perhaps the more recent neutralizations are in some sense less entrenched in the language and so less resistant to orthography. However, this scenario still presupposes a distinction between the entrenched neutralization of /a,o/ and a less entrenched tendency to neutralize the other vowels, the latter arguably what is implied by incomplete neutralization.

It is also possible that failure to neutralize completely is related to the dialectal background of our subjects. As noted earlier, the subjects were vetted by a linguistically trained native Russian speaker for both native speaker status and standard dialect. However, we cannot rule out the possibility that they are part of a community in which pre-revolutionary pronunciation norms (including failure to neutralize in some cases) may have had an effect. (See discussion in section 3.1.1.) Having said this, we see no correlation in our data between speaker background, e.g., having been raised in China, and a tendency to neutralize. Further, the question of precisely which dialect(s) incompletely neutralize does not bear on the theoretical interest of our findings.

A possible drawback of our methodology should be acknowledged. Each phonetic context is represented by only one stimulus item in our study. The finding that /e/ and /i/ do not neutralize completely, for example, rests on stimuli containing a word of French origin. (The occurrence of /e/ after non-palatalized consonants is very restricted in native Russian, because this vowel historically caused palatalization of a preceding consonant.) These words have long been in everyday use in Russian, however, and are standardly described as undergoing vowel reduction. (The other stimuli evincing incomplete neutralization include common historically native words.) Our study employed a large number of speakers, having different backgrounds, and many repetitions per stimulus. Another approach worth considering would employ perhaps fewer repetitions per word but more than one stimulus item per phonetic context.

One interpretation of incomplete neutralization is that reduction of (for example) non-palatalized /e/, and palatalized /je,jo,ja/, is not in fact phonologized, but remains a phonetic effect, the degree of reduction subject to factors such as speech rate or style. Since phonological models are designed to effect only categorical changes, such as a substitution of [i] for /e/, they are poorly suited to handle incomplete neutralization. However, the fact that reduction is so often perceptually (if not acoustically) neutralizing, as suggested by our results and by impressionistic descriptions, raises deeper questions for these phonological models. Phonologists have long regarded the maintenance versus loss of phonemic contrast – normally judged impressionistically – as being at the core of their domain of explanation. The existence of incomplete neutralization does not sit easily with this view. [See related discussion in Manaster-Ramer, 1996, and Port, 1996.] If incomplete neutralization is real, and phonology is to model it, it may be necessary (at the least) to distinguish articulatory and perceptual representations, with substitution of categories a fact about perception and not production.

4.2 Duration

Our finding that stressed vowels are much longer than other vowels is consistent with the hypothesis that phonological vowel reduction (in particular, neutralization) has its roots in phonetic undershoot, to which shorter, unstressed vowels are especially prone. On the other hand, no support is found for a general durational distinction between prestressed syllables and other unstressed syllables. Recall that in impressionistic descriptions this positional distinction is motivated mainly by the facts of /a/ and /o/ after non-palatalized consonants: these vowels reduce to [ɐ] in prestressed syllables and to [ə] in other unstressed syllables. A plausible hypothesis, suggested by Crosswhite [2001, in press] and Barnes [2002], is that prestressed syllables are inherently longer than other unstressed syllables, so that reduction is comparatively inhibited in the former context. Our results in fact support such a durational distinction after non-palatalized consonants, as we have seen. However, this result may be due to an inherent durational difference between [ɐ] and [ə] themselves, since no distinction is found after palatalized consonants. This result cannot therefore be taken as support for any general durational distinction between prestressed syllables and other unstressed syllables, a distinction that would *motivate* the difference between [ɐ] and [ə]. Nor can it be taken as strong evidence *against* this hypothesis, however. It is possible, for example, that failure to find a durational distinction after palatalized consonants is a floor effect: only high vowels occur there, and these are already comparatively short. Furthermore, though our duration results do not support a distinction between prestressed and other unstressed syllables, the two contexts are often distinguished in other ways: unstressed syllables tend to have more compressed vowel spaces and (by the criteria we employed) a greater likelihood to perceptually neutralize.

Barnes [2002] presents evidence (based on 1 speaker) that pre- and unstressed syllables have the same target [ɐ], with any further raising toward [ə] in unstressed syllables due to phonetic undershoot. Our own data do not support this hypothesis, as shown in figure 4, which plots the duration of Prestressed and Unstressed /a/ against F1–F0 for all 9 speakers. Figure 4 confirms that our Prestressed and Unstressed /a/ differ rather reliably in duration and tend to differ in height (see below for more discussion of the latter). But for most speakers, [ə] does not approach [ɐ] under increasing duration. However, unlike Barnes [2002], we did not manipulate speech rate and style or attempt to control duration in any other way.

4.3 Overall Vowel Space

Our results bear out the hypothesis that the vowel space should contract in non-Stressed syllables, again consistent with an undershoot explanation for phonological vowel reduction. A more complex question concerns the nature of this contraction. As we saw in section 2.2, some accounts suggest that compression should be more serious in the vowel height dimension than in the vowel backness dimension for languages exhibiting phonological vowel reduction. This does not seem true of Russian. The non-palatalized vowel space contracts roughly 3 ERB on average in both F1 and F2. In the palatalized vowel space there is a difference, roughly 5 ERB versus 4 ERB for F1 and F2 respectively.[13] Yet this seems too small to support a major distinction between

[13] One might view the palatalized context as irrelevant, since the impressionistic inventory there is [i,u] when not stressed. Assuming this, it is no surprise that F2 contracts more, and we could not use this to *explain* neutralization in height. However, since neutralization is not in fact complete in this context, it seems reasonable to compare F1 and F2 here too.

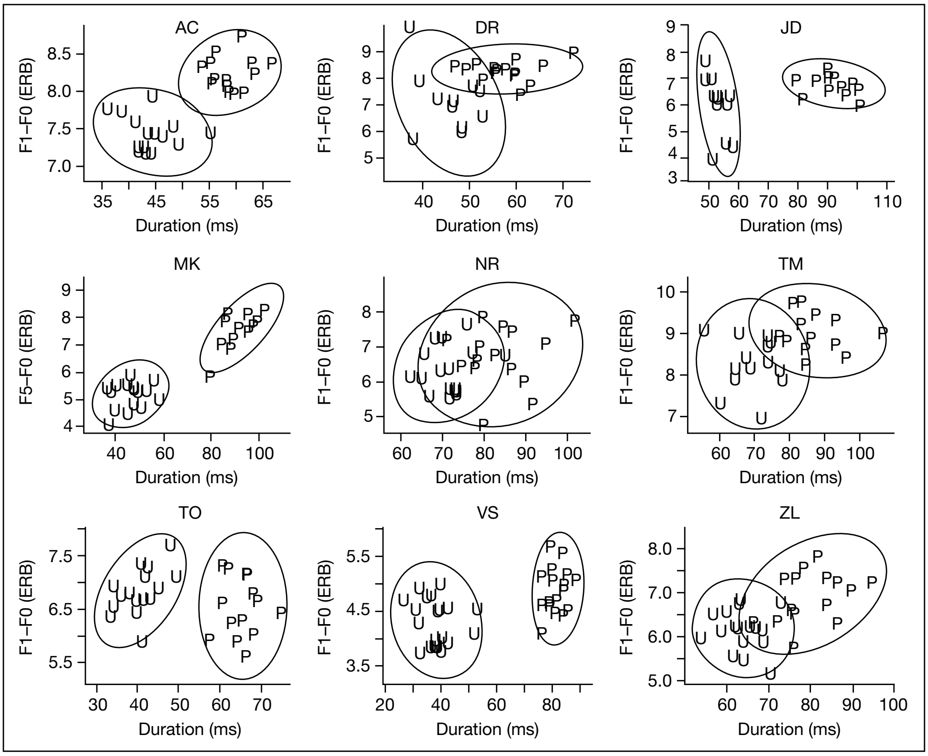


Fig. 4. Plots of duration X F1–F0 (in ERB) for the vowels /a/ and /o/, in Prestressed (P) and Unstressed (U) contexts, for all 9 speakers.

contraction in F1 versus F2. In other words, if vowel space contraction is due to undershoot, we have reason to think that undershoot in Russian has affected F1 and F2 equally.

This conclusion might be moderated given the evidence discussed in section 3.2.4 that differences in F1 are perceptually more consequential than those in F2, even when equal in ERB. It may be that in Russian the F1 contraction counts more than appearances suggest. Assuming an F2 weighting of 0.625 $\times$ F1, following Schwartz et al. [1997], for example, contraction of F1 is notably larger than that of F2 even in Russian. Even granting this comparison, however, the effect on F2 remains substantial in Russian.

An interesting question concerns the extent to which F2 undershoot in Russian is due to the existence of a palatalization contrast in the language. In a study of Catalan vowel reduction similar to ours, Herrick [2003] finds a great deal of compression of the vowel space in terms of F1, and *none* for F2, a finding more in line with the account of vowel reduction presented in section 2.2. Herrick's [2003] stimuli were nonsense words in which the relevant vowel was flanked by labial stops. Catalan lacks a palatalization contrast. In addition, labials are generally assumed to make few articulatory demands that conflict with the demands of adjacent vowels; for example, they are relatively unresistant to vowel-to-vowel coarticulation [on Catalan specifically see

Recasens et al., 1997]. In our study, the target vowel was preceded by (in most cases) a labial stop, which was either palatalized or not, and followed by a non-palatalized dental. There are questions about the degree of velarization (or velopharyngealization) in Russian non-palatalized consonants, and about the consonants most affected. But there is some agreement that labials are among the Russian sounds most prone to velarization, and that velarization is strong for all non-palatalized consonants before front vowels (see the references cited in 3.1.2). Our target vowels are therefore likely to have occurred in the environment of either a palatalized or a velarized sound, or both. These Russian consonants, unlike Catalan labials, are constrained to achieve targets for the tongue body (and possibly lip rounding). They are therefore similar to Catalan sounds like [ɲ] and [ɫ], which Recasens et al. [1997] characterize as having a high 'degree of articulatory constraint' (DAC) for the tongue dorsum. According to the DAC model of coarticulation, a high DAC value means that a sound will be more resistant to coarticulation, and more likely to cause coarticulation of neighbouring sounds. Indeed, it is well known that Russian palatalization exerts large effects on neighbouring vowels, and their resistance to vowel-to-vowel coarticulation was first noted by Öhman [1966]. Finally, the pattern of F2 contraction in our Russian data seen in section 3.2.2 is consistent with these considerations: it seems largely due to fronting of [ju] (i.e., [u] after palatalized consonants) and backing of [i] (i.e., [i] after non-palatalized consonants) in non-Stressed syllables. In sum, we have good reason to attribute F2 contraction in Russian vowel reduction to the effect of coarticulation with neighbouring consonants which bear an inherent palatalization contrast. It is a question for further research whether phonological vowel reduction ever involves large contractions of the F2 space in languages lacking such a contrast.

As for F1 contraction itself, the account of phonological vowel reduction laid out in section 2.2 predicts that it should be due specifically to raising of the vowel space floor. The results for palatalized vowels certainly support this view. (This result is not trivial, since neutralization to [i,u] is incomplete.) For non-palatalized vowels we found a more complex picture: though raising of the floor is the greater effect, lowering of the ceiling does occur (an average of 1.9 versus 1 ERB, respectively.)

Studies have shown that female speakers tend to disperse vowels more than men, in some cultures at least. [For an overview and discussion see Diehl et al., 1996.] Given this fact, one might wonder whether our finding that the vowel space contracts under reduction is related to the fact that 8 of our 9 speakers are female. Though this is a question worth testing in future work, we note that our 1 male speaker, AC, exhibits an overall vowel space contraction comparable to that of the females.

Overall, our results lend support to the hypothesis that phonological vowel reduction (neutralization) results from a compression of the vowel space due to target undershoot in unstressed syllables. Though undershoot might affect predominantly vowel height distinctions in most cases of phonological vowel reduction, in Russian it does affect backness distinctions as well, most likely because consonants bear palatalization and velarization specifications which directly compete with specifications of vowels.

4.4 Pairwise Vowel Spacing

The existence of incomplete neutralization and the question of vowel spacing bear on each other in an important way. If we were to count incompletely neutralized vowels as different for the purposes of discussing spacing, then we would have to conclude that spacing is grossly uneven, and that there is in fact no generally applicable spacing requirement at all, as our vowel plots make clear. This conclusion might hold of speaker productions, but it does not address the perceptual side. It seems clear that incompletely neutralized vowels are in fact perceptually neutralized, at least in many instances. It is therefore interesting to ask about the spacing requirements among impressionistically distinct vowels.

Beginning with Stressed vowels, we can conclude that these are roughly evenly spaced in Russian, with two important caveats. First, this is only true if we apply a weighted Euclidean measure in which differences in F1 count more than those in F2. In unweighted terms, backness distinctions are disproportionately large. Second, even spacing holds better of average values across speakers than of values within speakers. This is not surprising, but the fact that vowels are in fact roughly evenly spaced does not itself follow from averaging. (If vowels are unevenly spaced in the same way for all speakers, then average values are also. See the discussion of Prestressed [i,u,ɐ] below.) Since listeners must recognize the vowel productions of a variety of speakers, we might speculate that they have implicit knowledge of these averages.

Among non-Stressed contexts, only the non-palatalized ones contrast more than two vowels (assuming the described neutralizations), having surface [i,u,ɐ/ə]. In Prestressed position, average distances among these three vowels are roughly similar only in unweighted terms. Given the weighted measures, we have from table 5b 3.38 ERB for [i,ɐ], 4.26 ERB for [ɐ,u], and 2.38 ERB for [i,u]. Since the weighted Euclidean measures are needed to achieve even spacing among Stressed vowels, we can conclude that vowel spacing is not consistent within contexts, even using averages across speakers.

Finally, we found that distances varied also as a function of context. First, the claim that non-palatalized /a,o/ are realized as something like [ɐ] in Prestressed position and [ə] in Unstressed position received some support. The vowel [ɐ] in particular seems lower than expected given the F1 spacing among Stressed vowels. This seems puzzling given the view of vowel reduction outlined at the outset. Further raising of [ɐ] should be favoured on articulatory grounds, given that account, since reaching a higher position in shorter vowels requires less effort. Following Crosswhite [2001, in press], we might suppose that the shorter duration of non-Stressed vowels must in fact be taken into account in a complete theory of perceptual distance. Assuming that vowels separated by some spectral distance are harder to distinguish as they become shorter in duration, perhaps the perceptual distance between Prestressed [i] and [ɐ] is more comparable to that between Stressed [i] and [e] than it appears to be. In other words, ‘perceptual distance’ takes into consideration not only formant values (and relative intensity of formants), but duration of cues, a possibility that only perceptual studies can address.

Second, we found that distances are otherwise consistently shorter in non-Stressed positions compared to Stressed positions. This was particularly true of palatalized /ji/ and /ju/. This fact has at least two interpretations. It is possible that something like a 3-ERB difference is maintained among vowels in all contexts, when phonetic *targets* are intended; apparent violations of this 3-ERB minimal distance would then be attributed to phonetic undershoot. Under this scenario, /ji/ and /ju/ might move apart under hyperarticulated speech. The alternative is that there is significant variation in spacing among vowel targets themselves. Further testing would be needed in order to address this question.

5. Conclusion

On the empirical side, our results suggest that traditional descriptions of Russian vowel reduction are correct in some respects, but need modification in others. Most significantly, our results suggest that non-Stressed vowels do not always neutralize completely where they are said to do so, with the notable exception of /a/ and /o/ after non-palatalized consonants. However, though phonetic differences seem to be maintained among such vowels, the results here do suggest that these distinctions would be perceptually minimal. Distances in ERB, differences in backness and height based on F1–F0 and F2–F1, and Bayesian classification all suggest this conclusion.

On the more theoretical side, we find generally good support for an approach to vowel reduction like that of Flemming [1995, in press]. First, non-Stressed vowels are significantly shorter than Stressed vowels (a fact also expected given Crosswhite's [2001, in press], typology of phonological vowel reduction). Second, the vowel space contracts in non-Stressed syllables, with the largest component of contraction due to raising of the vowel space floor. On the other hand, significant lowering of the vowel place ceiling also occurs for non-palatalized vowels. In addition, significant undershoot in F2 is also found. Since we might attribute the latter fact to the inherent palatalization or velarization of consonants in Russian, it does not seem inconsistent with the proposed account of vowel reduction.

Finally, we provided an analysis of vowel spacing in Russian, showing how this varies both within and across contexts. This provides an empirical base for models of dispersion and/or vowel reduction, and we hope, points up important questions for future research, including the relative importance of F1, F2, other spectral properties, and vowel duration, on perception of vowel categories.

Acknowledgements

We would like to thank Richard Beare, Chris Callaghan, Marika Kalyuga and Natalie Molloy for help with various stages of the data collection and analysis. We thank Adam Albright, Jonathan Barnes, Katherine Crosswhite, Randy Diehl, Ed Flemming, Dylan Herrick, Junko Ito, Lucie Ménard and Armin Mester for helpful comments. We would also like to thank the Russian Department at Macquarie University for its support and our speakers for their time. This research was supported by a Macquarie University Research Grant and an Australian Research Council postdoctoral fellowship to the second author. The same author's affiliation changed to the University of Western Sydney during the revision stages of this paper.

Appendix

Table A1. Mean and standard deviation (SD) for vowel duration (in milliseconds), together with results from a one-way ANOVA for each speaker

a Results for non-palatalized tokens

Speaker	Context	Duration		n	ANOVA			Post-hoc	
		mean	SD		d.f.	F	p	S~P	P~U
AC	S	68.50	16.82	78	2, 221	102.90	***	***	***
	P	45.98	16.05	74					
	U	36.29	7.28	72					
DR	S	93.73	24.74	79	2, 213	193.01	***	***	n.s.
	P	45.09	16.75	70					
	U	41.40	7.26	67					
JD	S	124.44	24.51	78	2, 209	245.73	***	***	***
	P	73.30	24.65	66					
	U	47.47	12.38	68					
MK	S	101.65	26.69	68	2, 213	105.48	***	***	***
	P	72.39	29.74	72					
	U	44.70	9.65	76					
NR	S	132.65	37.18	76	2, 229	195.16	***	***	n.s.
	P	66.94	19.82	77					
	U	60.23	11.67	79					
TM	S	138.38	23.32	64	2, 195	300.10	***	***	**
	P	74.81	19.97	66					
	U	64.25	10.82	68					
TO	S	83.84	30.10	74	2, 216	83.29	***	***	***
	P	56.53	19.86	71					
	U	38.95	7.74	74					
VS	S	139.29	30.45	84	2, 247	444.98	***	***	***
	P	65.23	21.44	84					
	U	40.43	9.44	82					
ZL	S	121.07	23.27	84	2, 246	239.32	***	***	*
	P	73.72	16.81	83					
	U	66.10	9.72	82					

b Results for palatalized tokens

Speaker	Context	Duration		n	ANOVA			Post-hoc	
		mean	SD.		d.f.	F	p	S~P	P~U
AC	S	84.31	14.98	76	2, 220	308.28	***	***	n.s.
	P	40.66	10.11	74					
	U	40.22	11.83	73					
DR	S	109.14	20.09	71	2, 209	410.54	***	***	**
	P	41.78	10.74	71					
	U	47.11	14.36	70					
JD	S	147.84	21.26	76	2, 209	659.13	***	***	n.s.
	P	64.97	13.95	68					
	U	55.59	13.80	68					
MK	S	124.39	21.47	70	2, 203	408.23	***	***	n.s.
	P	52.15	16.18	67					
	U	46.32	15.40	69					

Table A1. (continued)

b Results for palatalized tokens

Speaker	Context	Duration		n	ANOVA			Post-hoc	
		mean	SD.		d.f.	F	p	S~P	P~U
NR	S	164.93	26.26	79	2, 228	666.16	***	***	n.s.
	P	64.32	15.45	77					
	U	63.11	16.15	75					
TM	S	160.13	25.04	66	2, 195	590.14	***	***	n.s.
	P	71.07	11.68	66					
	U	65.04	13.68	66					
TO	S	101.48	20.32	72	2, 215	356.03	***	***	n.s.
	P	44.51	9.82	72					
	U	42.38	13.29	74					
VS	S	154.84	25.83	85	2, 245	912.47	***	***	***
	P	65.02	12.13	80					
	U	45.18	10.68	83					
ZL	S	142.74	17.43	80	2, 242	603.75	***	***	n.s.
	P	72.73	13.02	83					
	U	71.00	14.19	82					

Bonferroni-adjusted post-hoc tests are also given for the pairs 'Stressed vs. Prestressed' and 'Prestressed vs. Unstressed'. Statistical significance ***$p < 0.001$; **$p < 0.01$; *$p < 0.025$.

Table A2. F1 and F2 vowel space measures for each speaker and each vowel context

a Data for non-palatalized tokens

Speaker	Context	F1			F2		
		min.	max.	diff.	min.	max.	diff.
AC	S	7.84	12.67	4.83	14.77	19.87	5.10
	P	8.38	12.28	3.90	14.90	19.20	4.30
	U	8.58	11.53	2.95	16.12	18.47	2.35
DR	S	8.26	15.49	7.23	14.85	22.84	7.99
	P	8.99	14.57	5.58	15.59	20.52	4.93
	U	9.53	13.35	3.82	15.85	20.50	4.65
JD	S	6.80	14.90	8.10	16.14	22.98	6.84
	P	7.51	13.92	6.41	15.15	20.44	5.29
	U	9.70	13.42	3.72	15.16	19.82	4.66
MK	S	8.57	14.80	6.23	15.97	21.59	5.62
	P	8.60	13.83	5.23	16.49	20.19	3.70
	U	9.03	11.38	2.35	17.02	19.36	2.34
NR	S	6.22	14.43	8.21	13.62	21.55	7.93
	P	8.12	12.81	4.69	15.38	20.64	5.25
	U	8.45	12.06	3.61	17.99	20.28	2.29
TM	S	7.51	14.36	6.85	14.08	22.78	8.70
	P	8.71	13.29	4.58	16.11	20.55	4.44
	U	8.32	12.61	4.29	16.02	20.22	4.20
TO	S	7.82	15.06	7.24	15.05	20.43	5.38
	P	7.60	14.81	7.21	14.47	20.32	5.85
	U	8.72	14.78	6.06	15.07	19.85	4.78

Table A2. (continued)

a Data for non-palatalized tokens

Speaker	Context	F1			F2		
		min.	max.	diff.	min.	max.	diff.
VS	S	7.49	15.35	7.86	11.82	22.56	10.74
	P	7.84	12.90	5.06	12.28	22.06	9.78
	U	7.80	12.20	4.40	11.89	20.65	8.76
ZL	S	9.36	13.57	4.21	16.19	21.97	5.78
	P	8.85	13.07	4.22	15.91	20.67	4.76
	U	8.89	12.38	3.49	16.36	20.11	3.75

b Data for palatalized tokens

Speaker	Context	F1			F2		
		min.	max.	diff.	min.	max.	diff.
AC	S	7.19	12.59	5.40	16.63	21.52	4.89
	P	7.20	11.57	4.37	18.19	20.85	2.66
	U	8.33	10.57	2.24	18.13	20.70	2.57
DR	S	8.13	15.42	7.29	18.00	24.33	6.33
	P	8.76	10.36	1.60	20.65	23.68	3.03
	U	9.06	10.34	1.28	20.45	23.57	3.12
JD	S	7.25	14.37	7.12	15.41	24.03	8.62
	P	6.85	11.79	4.94	19.57	23.29	3.72
	U	9.35	10.91	1.56	19.86	23.03	3.17
MK	S	9.12	14.41	5.29	18.18	23.04	4.86
	P	9.19	10.14	0.95	19.95	22.18	2.23
	U	8.90	9.88	0.98	20.11	21.94	1.83
NR	S	7.02	15.00	7.98	16.01	23.17	7.16
	P	7.64	9.78	2.14	19.79	22.43	2.64
	U	8.24	9.68	1.44	20.26	22.22	1.96
TM	S	7.18	14.76	7.58	15.74	23.80	8.06
	P	7.58	11.69	4.11	20.61	23.49	2.88
	U	8.01	10.96	2.95	20.40	23.00	2.60
TO	S	7.11	15.00	7.89	17.42	23.86	6.44
	P	7.70	12.07	4.37	20.40	22.78	2.38
	U	7.90	11.73	3.83	19.76	22.40	2.64
VS	S	7.28	15.12	7.84	12.87	24.60	11.73
	P	7.62	13.98	6.36	11.76	23.74	11.98
	U	7.48	12.45	4.97	11.73	23.03	11.30
ZL	S	8.95	13.81	4.86	16.89	22.81	5.92
	P	8.80	10.81	2.01	19.58	22.35	2.77
	U	8.83	10.15	1.32	19.82	22.36	2.54

All values are in ERB. See text for details of calculation.

Table A3. Euclidean distance measures between vowel pairs in the F1–F2 plane

a Data for non-palatalized tokens

Speaker	Context	i/e	e/a	a/o	o/u	i/u	e/o
AC	S	2.83	2.69	2.53	1.88	4.10	3.47
	P	1.26	2.66	0.05	3.37	3.24	2.70
	U	0.91	1.66	0.28	2.55	1.80	1.91
DR	S	3.56	3.43	4.89	1.64	6.74	5.42
	P	1.13	3.08	0.08	5.21	4.10	3.00
	U	0.94	1.29	0.57	3.01	3.58	0.96
JD	S	4.35	3.20	4.28	2.31	5.51	4.44
	P	1.46	3.08	0.16	5.03	4.08	3.06
	U	0.11	2.02	0.78	2.75	3.09	1.39
MK	S	2.19	3.27	4.19	1.02	4.68	3.84
	P	0.36	3.79	0.43	4.75	2.73	4.19
	U	0.26	0.99	0.45	1.74	1.44	1.34
NR	S	3.16	2.98	5.48	3.38	6.60	6.24
	P	0.85	2.75	0.58	4.92	4.12	3.19
	U	0.66	2.39	0.51	1.83	1.76	2.03
TM	S	3.42	3.14	3.22	4.06	7.29	4.40
	P	1.35	2.59	0.26	4.32	3.48	2.33
	U	1.42	1.88	0.36	4.13	3.39	2.25
TO	S	2.64	3.71	5.17	1.96	3.81	4.83
	P	2.65	3.21	0.17	6.63	4.44	3.34
	U	0.29	3.72	0.14	5.04	3.05	3.84
VS	S	4.19	3.00	4.75	5.05	9.03	5.25
	P	0.99	5.35	0.23	4.85	8.83	5.33
	U	0.38	5.19	0.42	4.77	8.08	4.83
ZL	S	2.64	1.90	2.65	2.00	4.88	3.67
	P	0.39	3.38	0.40	4.61	3.99	3.66
	U	0.56	2.76	0.37	3.47	3.05	3.09

b Data for palatalized tokens

Speaker	Context	i/e	e/a	a/o	o/u	i/u	e/o
AC	S	2.59	2.84	2.18	2.11	4.30	3.00
	P	1.37	1.89	1.65	1.46	2.19	0.24
	U	0.45	1.44	0.70	1.24	1.97	0.75
DR	S	2.12	5.24	4.44	2.78	5.17	5.63
	P	0.47	0.56	1.59	1.97	2.54	1.16
	U	0.24	0.71	0.54	1.79	2.43	1.22
JD	S	3.77	3.18	3.61	3.51	7.30	4.55
	P	1.07	2.63	0.35	2.60	3.99	2.30
	U	0.22	0.24	0.45	2.10	2.29	0.22
MK	S	1.74	3.54	3.79	1.11	3.96	3.27
	P	0.59	0.09	0.56	1.18	1.63	0.66
	U	0.53	0.18	0.54	0.80	1.32	0.68
NR	S	3.33	3.29	5.12	3.16	6.32	5.24
	P	0.50	0.42	0.19	1.41	2.04	0.46
	U	0.19	0.40	0.25	1.06	1.19	0.18
TM	S	4.75	2.57	3.59	3.36	6.72	3.94
	P	1.23	2.23	1.90	1.90	2.25	0.40
	U	1.06	1.34	0.76	1.61	1.94	0.61

Table A3. (continued)

b Data for palatalized tokens

Speaker	Context	i/e	e/a	a/o	o/u	i/u	e/o
TO	S	2.99	4.57	4.72	3.08	5.00	4.84
	P	0.48	2.14	2.55	1.45	1.81	0.42
	U	1.52	0.74	0.47	1.34	2.11	0.40
VS	S	4.69	3.64	4.82	4.39	10.09	5.06
	P	3.40	2.07	1.53	10.30	10.52	0.54
	U	3.32	0.72	0.35	10.10	10.37	0.37
ZL	S	3.22	1.76	2.35	2.31	5.38	2.96
	P	0.03	1.31	1.31	2.40	2.33	0.11
	U	0.27	0.78	0.71	1.80	2.06	0.15

All values are in ERB. Values falling at or below 1 ERB are highlighted.

Table A4. Mean values of F1–F0 and F2–F1 for each vowel in each stress context for each speaker

a Data for non-palatalized tokens

		F1–F0					F2–F1				
		i	e	a	o	u	i	e	a	o	u
AC	S	4.61	7.33	8.86	6.55	4.22	11.11	7.92	4.22	5.47	6.95
	P	5.05	6.14	8.24	8.20	4.98	9.90	8.25	4.51	4.45	6.58
	U	5.24	6.16	7.46	7.62	4.82	9.10	8.09	5.76	5.38	7.47
DR	S	3.25	6.83	9.42	5.82	3.78	13.17	9.02	4.19	4.71	5.52
	P	5.12	5.30	8.30	8.25	3.48	10.16	8.74	4.75	4.86	6.71
	U	4.96	5.68	7.02	6.62	4.70	9.56	8.38	6.66	7.01	5.99
JD	S	0.59	5.35	7.65	4.25	1.78	14.83	9.63	5.15	6.21	8.38
	P	1.93	3.71	6.71	6.69	2.95	11.03	9.40	5.86	6.03	6.17
	U	3.77	3.99	5.99	5.13	3.59	9.09	9.10	6.67	7.19	6.40
MK	S	3.21	5.27	7.95	4.44	2.51	11.85	8.96	4.39	6.39	7.27
	P	3.70	3.90	7.59	7.96	3.35	10.43	9.94	5.23	4.85	8.02
	U	4.08	4.28	5.19	5.53	3.74	9.40	9.35	8.07	7.48	8.21
NR	S	3.08	6.61	9.28	5.81	1.72	12.76	9.56	5.37	4.44	7.53
	P	3.72	4.42	6.75	7.40	3.39	11.60	10.82	7.11	6.72	7.45
	U	5.05	4.31	6.42	5.98	4.21	9.95	10.89	7.60	8.03	8.98
TM	S	4.51	7.84	10.19	8.02	4.26	13.91	9.68	5.24	5.11	6.71
	P	6.02	7.19	9.00	8.86	5.25	10.36	9.80	6.13	6.50	7.73
	U	6.01	7.23	8.21	8.44	4.71	10.20	9.65	7.01	6.50	7.88
TO	S	1.39	4.94	7.67	4.05	0.47	11.48	8.88	3.84	5.16	7.78
	P	0.96	3.71	6.55	6.74	0.93	11.54	8.51	4.33	4.10	7.50
	U	2.98	3.38	6.83	6.99	2.75	9.06	8.73	4.33	4.26	6.47
VS	S	1.76	6.09	8.11	3.84	0.68	13.22	8.96	4.73	6.36	5.10
	P	5.40	3.79	4.89	5.17	0.77	9.37	9.00	2.66	2.51	4.84
	U	3.62	3.92	4.32	4.80	0.83	9.27	9.30	3.61	3.81	4.25
ZL	S	4.96	7.57	8.83	7.55	5.21	11.92	8.46	5.76	4.78	6.61
	P	3.62	3.84	6.96	7.32	3.54	10.90	11.03	6.45	6.20	6.96
	U	3.67	3.85	6.16	6.40	3.40	10.49	10.89	7.12	6.60	7.60

Table A4. (continued)

b Data for palatalized tokens

		F1–F0					F2–F1				
		i	e	a	o	u	i	e	a	o	u
AC	S	3.79	6.16	8.87	7.07	4.70	13.61	10.09	6.20	6.38	8.44
	P	4.03	5.01	6.86	5.28	4.59	12.82	10.95	8.48	10.66	9.98
	U	4.57	4.96	6.06	5.33	4.73	12.03	11.43	9.48	10.37	9.84
DR	S	2.83	4.83	9.49	5.99	3.24	15.48	12.58	5.40	5.95	9.77
	P	3.55	3.48	4.55	4.53	3.66	14.20	13.74	13.07	12.34	11.38
	U	3.51	3.43	4.18	4.57	3.92	13.99	14.01	13.14	12.37	11.19
JD	S	0.76	4.74	7.86	4.97	1.24	15.80	10.92	6.47	6.16	8.03
	P	1.59	1.35	4.57	3.77	3.54	15.38	13.88	11.51	11.93	9.74
	U	3.51	3.64	3.53	3.27	3.42	12.63	12.35	12.00	12.60	10.46
MK	S	3.34	4.75	8.09	4.61	3.37	13.06	10.59	5.83	7.55	9.09
	P	4.32	3.65	4.08	4.14	3.64	12.12	12.06	11.95	11.31	10.71
	U	3.68	3.63	3.98	4.35	3.77	12.35	11.89	11.62	11.11	10.86
NR	S	3.21	6.29	9.69	6.30	2.65	14.43	9.91	5.56	4.98	8.62
	P	3.23	3.92	4.06	3.73	3.61	13.55	12.87	12.31	12.21	10.96
	U	3.73	3.96	3.97	4.08	3.64	12.99	12.80	12.41	12.55	11.77
TM	S	3.61	7.89	10.15	7.81	4.30	16.04	9.72	6.24	5.79	8.39
	P	4.46	5.29	7.64	5.59	4.49	15.13	13.41	10.46	13.07	12.60
	U	4.79	5.37	6.62	6.07	4.89	14.40	12.92	11.05	12.13	12.19
TO	S	0.54	4.33	7.46	5.01	0.33	15.80	11.77	5.73	6.45	10.49
	P	1.03	1.27	3.81	0.55	1.40	14.20	13.51	11.23	13.87	12.00
	U	1.46	2.90	3.32	3.04	2.50	13.50	11.72	10.67	11.25	10.64
VS	S	0.91	5.55	8.03	3.69	1.24	16.64	10.50	5.34	7.60	6.02
	P	1.41	3.87	6.11	4.30	0.45	14.98	10.89	8.03	10.17	4.57
	U	0.55	3.90	4.76	4.15	0.92	14.70	10.82	9.80	10.31	3.92
ZL	S	4.46	7.51	9.17	7.67	5.16	13.31	8.90	6.44	5.80	7.25
	P	3.40	3.40	4.77	3.42	3.39	12.97	12.98	11.17	13.01	10.66
	U	3.36	3.39	3.96	3.60	3.42	12.99	12.81	11.73	12.65	10.89

All values are in ERB.

Table A5. Difference between mean F1–F0 values, and between mean F2–F1 values, for each pair of vowels, based on values in table A4

a Data for non-palatalized tokens

		i/e		e/a		a/o		o/u		i/u		e/o	
		F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1
AC	S	2.72	3.19	1.53	3.7	2.31	1.25	2.33	1.48	0.39	4.16	0.78	2.45
	P	1.09	1.65	2.1	3.74	0.04	0.06	3.22	2.13	0.07	3.32	2.06	3.8
	U	0.92	1.01	1.3	2.33	0.16	0.38	2.8	2.09	0.42	1.63	1.46	2.71
DR	S	3.58	4.15	2.59	4.83	3.6	0.52	2.04	0.81	0.53	7.65	1.01	4.31
	P	0.18	1.42	3	3.99	0.05	0.11	4.77	1.85	1.64	3.45	2.95	3.88
	U	0.72	1.18	1.34	1.72	0.4	0.35	1.92	1.02	0.26	3.57	0.94	1.37
JD	S	4.76	5.2	2.3	4.48	3.4	1.06	2.47	2.17	1.19	6.45	1.1	3.42
	P	1.78	1.63	3	3.54	0.02	0.17	3.74	0.14	1.02	4.86	2.98	3.37
	U	0.22	0.01	2	2.43	0.86	0.52	1.54	0.79	0.18	2.69	1.14	1.91
MK	S	2.06	2.89	2.68	4.57	3.51	2	1.93	0.88	0.7	4.58	0.83	2.57
	P	0.2	0.49	3.69	4.71	0.37	0.38	4.61	3.17	0.35	2.41	4.06	5.09
	U	0.2	0.05	0.91	1.28	0.34	0.59	1.79	0.73	0.34	1.19	1.25	1.87

Table A5. (continued)

a Data for non-palatalized tokens

		i/e		e/a		a/o		o/u		i/u		e/o	
		F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1
NR	S	3.53	3.2	2.67	4.19	3.47	0.93	4.09	3.09	1.36	5.23	0.8	5.12
	P	0.7	0.78	2.33	3.71	0.65	0.39	4.01	0.73	0.33	4.15	2.98	4.1
	U	0.74	0.94	2.11	3.29	0.44	0.43	1.77	0.95	0.84	0.97	1.67	2.86
TM	S	3.33	4.23	2.35	4.44	2.17	0.13	3.76	1.6	0.25	7.2	0.18	4.57
	P	1.17	0.56	1.81	3.67	0.14	0.37	3.61	1.23	0.77	2.63	1.67	3.3
	U	1.22	0.55	0.98	2.64	0.23	0.51	3.73	1.38	1.3	2.32	1.21	3.15
TO	S	3.55	2.6	2.73	5.04	3.62	1.32	3.58	2.62	0.92	3.7	0.89	3.72
	P	2.75	3.03	2.84	4.18	0.19	0.23	5.81	3.4	0.03	4.04	3.03	4.41
	U	0.4	0.33	3.45	4.4	0.16	0.07	4.24	2.21	0.23	2.59	3.61	4.47
VS	S	4.33	4.26	2.02	4.23	4.27	1.63	3.16	1.26	1.08	8.12	2.25	2.6
	P	1.61	0.37	1.1	6.34	0.28	0.15	4.4	2.33	4.63	4.53	1.38	6.49
	U	0.3	0.03	0.4	5.69	0.48	0.2	3.97	0.44	2.79	5.02	0.88	5.49
ZL	S	2.61	3.46	1.26	2.7	1.28	0.98	2.34	1.83	0.25	5.31	0.02	3.68
	P	0.22	0.13	3.12	4.58	0.36	0.25	3.78	0.76	0.08	3.94	3.48	4.83
	U	0.18	0.4	2.31	3.77	0.24	0.52	3	1	0.27	2.89	2.55	4.29

b Data for palatalized tokens

		i/e		e/a		a/o		o/u		i/u		e/o	
		F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1	F1–F0	F2–F1
AC	S	2.37	3.52	2.71	3.89	1.8	0.18	2.37	2.06	0.91	5.17	0.91	3.71
	P	0.98	1.87	1.85	2.47	1.58	2.18	0.69	0.68	0.56	2.84	0.27	0.29
	U	0.39	0.6	1.1	1.95	0.73	0.89	0.6	0.53	0.16	2.19	0.37	1.06
DR	S	2	2.9	4.66	7.18	3.5	0.55	2.75	3.82	0.41	5.71	1.16	6.63
	P	0.07	0.46	1.07	0.67	0.02	0.73	0.87	0.96	0.11	2.82	1.05	1.4
	U	0.08	0.02	0.75	0.87	0.39	0.77	0.65	1.18	0.41	2.8	1.14	1.64
JD	S	3.98	4.88	3.12	4.45	2.89	0.31	3.73	1.87	0.48	7.77	0.23	4.76
	P	0.24	1.5	3.22	2.37	0.8	0.42	0.23	2.19	1.95	5.64	2.42	1.95
	U	0.13	0.28	0.11	0.35	0.26	0.6	0.15	2.14	0.09	2.17	0.37	0.25
MK	S	1.41	2.47	3.34	4.76	3.48	1.72	1.24	1.54	0.03	3.97	0.14	3.04
	P	0.67	0.06	0.43	0.11	0.06	0.64	0.5	0.6	0.68	1.41	0.49	0.75
	U	0.05	0.46	0.35	0.27	0.37	0.51	0.58	0.25	0.09	1.49	0.72	0.78
NR	S	3.08	4.52	3.4	4.35	3.39	0.58	3.65	3.64	0.56	5.81	0.01	4.93
	P	0.69	0.68	0.14	0.56	0.33	0.1	0.12	1.25	0.38	2.59	0.19	0.66
	U	0.23	0.19	0.01	0.39	0.11	0.14	0.44	0.78	0.09	1.22	0.12	0.25
TM	S	4.28	6.32	2.26	3.48	2.34	0.45	3.51	2.6	0.69	7.65	0.08	3.93
	P	0.83	1.72	2.35	2.95	2.05	2.61	1.1	0.47	0.03	2.53	0.3	0.34
	U	0.58	1.48	1.25	1.87	0.55	1.08	1.18	0.06	0.1	2.21	0.7	0.79
TO	S	3.79	4.03	3.13	6.04	2.45	0.72	4.68	4.04	0.21	5.31	0.68	5.32
	P	0.24	0.69	2.54	2.28	3.26	2.64	0.85	1.87	0.37	2.2	0.72	0.36
	U	1.44	1.78	0.42	1.05	0.28	0.58	0.54	0.61	1.04	2.86	0.14	0.47
VS	S	4.64	6.14	2.48	5.16	4.34	2.26	2.45	1.58	0.33	10.62	1.86	2.9
	P	2.46	4.09	2.24	2.86	1.81	2.14	3.85	5.6	0.96	10.41	0.43	0.72
	U	3.35	3.88	0.86	1.02	0.61	0.51	3.23	6.39	0.37	10.78	0.25	0.51
ZL	S	3.05	4.41	1.66	2.46	1.5	0.64	2.51	1.45	0.7	6.06	0.16	3.1
	P	0	0.01	1.37	1.81	1.35	1.84	0.03	2.35	0.01	2.31	0.02	0.03
	U	0.03	0.18	0.57	1.08	0.36	0.92	0.18	1.76	0.06	2.1	0.21	0.16

All values are in ERB. Values falling at or below 1 ERB are highlighted.

References

Avanesov, R.I.: Fonetika sovremennogo russkogo literaturnogo iazyka (Izdatel'stvo moskovskogo universiteta, Moscow 1956).

Avanesov, R.I.: Russkoe literaturnoe proiznoshenie (Prosveshchenie, Moscow 1972).

Barnes, J.A.: Positional neutralization: a phonologization approach to typological predictions; doct. diss. University of California, Berkeley (2002).

Benkí, J.: Analysis of English nonsense syllable recognition in noise. Phonetica *60:* 129–157 (2003).

Bladon, R.A.W.; Lindblom, B.: Modeling the judgment of vowel quality differences. J. acoust. Soc. Am. *69:* 1414–1422 (1981).

Boë, L.-J.; Schwartz, J.-L.; Vallée, N.: The prediction of vowel systems: perceptual contrast and stability; in Keller, Fundamentals of speech synthesis and speech recognition, pp.185–213 (Wiley, Chichester 1994).

Bolla, K.: A conspectus of Russian speech sounds (Adakemiai Kiado, Budapest 1981).

Browman, C.; Goldstein, L.: 'Targetless' schwa: an articulatory analysis; in Docherty, Ladd, Papers in Laboratory Phonology. II: Gesture, Segment, Prosody, pp. 26–56 (Cambridge University Press, Cambridge 1992).

Cassidy, S.; Harrington, J.: Multi-level annotation in the EMU speech database management system. Speech Commun *33:* 61–77 (2001).

Charles-Luce, J.: The effects of semantic context on voicing neutralization. Phonetica *50:* 28–43 (1993).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Crosswhite, K.: Vowel reduction in Optimality Theory (Routledge, New York 2001).

Crosswhite, K.: Vowel reduction; in Hayes, Kirchner, Steriade, Phonetic bases of markedness (Cambridge University Press, Cambridge, in press).

Diehl, R.L.; Lindblom, B.; Creeger, C.P.: Increasing realism of auditory representations yields further insights into vowel phonetics. Proc. 15th Int. Cong. Phonet. Sci., 2003.

Diehl, R.L.; Lindblom, B.; Hoemeke, K.A.; Fahey, R.P.: On explaining certain male-female differences in the phonetic realization of vowel categories. J. Phonet. *24:* 187–208 (1996).

Evans-Romaine, D.K.: Palatalization and coarticulation in Russian; PhD diss., University of Michigan, Ann Arbor (1998).

Fant, G.: Acoustic theory of speech production (Mouton, The Hague 1960).

Flemming, E.: Contrast and perceptual distinctiveness; in Hayes, Kirchner, Steriade, Phonetic bases of markedness, (Cambridge University Press, Cambridge, in press).

Flemming, E.S.: Auditory representations in phonology; doct. diss. UCLA (1995; Routledge, New York 2002).

Fourakis, M.: Tempo, stress, and vowel reduction in American English. J. acoust. Soc. Am. *90:* 1816–1827 (1991).

Guy, G.R.: Form and function in linguistic variation; in Guy, Feagin, Schiffrin, Baugh, Towards a social science of language: papers in honor of William Labov, vol. 1: Variation and change in language and society, 221–252 (Benjamins, Amsterdam 1996).

Halle, M.: The sound pattern of Russian (Mouton, The Hague 1959).

Hamilton, W.S.: Introduction to Russian phonology and word structure (Slavica Publishers, Columbus 1980).

Harrington, J.; Cassidy, S.; Fletcher, J.;McVeigh, A.: The mu+ system for corpus based speech research. Computer Speech Lang. *7:* 305–331 (1993).

Herrick, D.: Variation, location, and inventory size: the case of Catalan vowel reduction; doct. diss. University of California, Santa Cruz (2003).

Jones, L.G.: Excursus on the contextual variants of the Russian vowels; in Halle, The sound pattern of Russian (Mouton, 's-Gravenhage 1959).

Kochetov, A.: Production, perception, and emergent phonotactic patterns (Routledge, New York 2002).

Kondo, Y.: Targetless schwa: is that how we get the impression of stress timing in English? In Proc. Edinburgh Linguistics Dep. Conf., pp. 63–76 (1994).

Kouznetsov, V.: Spectral dynamics and classification of Russian vowels. Proc. 11th Session of the Russian Acoustical Society, pp. 439–442 (2001).

Labov, W.: Principles of language change, vol. 1: Internal factors (Blackwell, Oxford 1994).

Ladefoged, P.: A course in phonetics (Harcourt Brace Jovanovich, Fort Worth 1993).

Lehiste, I.: Suprasegmentals (MIT Press, Cambridge 1970).

Lightner, T.M.: Segmental phonology of Modern Standard Russian; doct. diss., MIT (1965).

Lightner, T.M.: Problems in the theory of phonology, vol. 1: Russian phonology and Turkish phonology (Linguistic Research, Edmonton 1972).

Liljencrants, J.; Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language *48:* 839–862 (1972).

Lindblom, B.: Spectrographic study of vowel reduction. J. acoust. Soc. Am. *35:* 1773–1781 (1963).

Lindblom, B.: Experiments in sound structure. Revue Phonét. appl. (Université de l'État Mons, Belgique) *51:* 155–189 (1975).

Lindblom, B.: Economy of speech gestures; in MacNeilage, The production of speech, pp. 217–245 (Springer, New York 1983).

Lindblom, B.: Phonetic universals in vowel systems; in Ohala, Jaeger, Experimental phonology, pp. 13–44 (Academic Press, Orlando 1986).

Lindblom, B.: Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle, Marchal, Speech production and speech modelling, pp. 403–439 (Kluwer, Dordrecht 1990).

Lobanov, B.M.: Classification of Russian vowels spoken by different speakers. J. acoust. Soc. Am. *49:* 606–608 (1971).
Manaster-Ramer, A.: A letter from an incompletely neutral phonologist. J. Phonet. *24:* 477–489 (1996).
Ménard, L.; Schwartz, J.-L.; Boë, L.-J.; Kandel, S.; Vallée, N.: Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. J. acoust. Soc. Am. *111:* 1892–1905 (2002).
Moon, S.J.; Lindblom, B.: Interaction between duration, context, and speaking style in English stressed vowels. J. acoust. Soc. Am. *96:* 40–55 (1994).
Moore, B.; Glasberg, B.: A revision of Zwicker's loudness model. Acta acust. *82:* 335–345 (1996).
Nooteboom, S.G.: Perceptual confusions among Dutch vowels presented in noise. IPO annu. Prog. Rep, Inst. Perception Res., Eindhoven *3:* 68–71 (1968).
Öhman, S.E.G.: Coarticulation in VCV utturances: spectrographic measurements. J. acoust. Soc. Am. *39:* 151–168 (1966).
Padgett, J.: Contrast dispersion and Russian palatalization; in Hume, Johnson, The role of speech perception in phonology, pp. 187–218 (Academic Press, San Diego 2001).
Pettersson, T.; Wood, S.: Vowel reduction in Bulgarian and its implications for theories of vowel production – a review of the problem. Folia ling. *21:* 261–279 (1987).
Port, R.; Crawford, P.: Incomplete neutralization and pragmatics in German. J. Phonet. *17:* 257–282 (1989).
Port, R.F.: The discreteness of phonetic elements and formal linguistics: response to A. Manaster-Ramer. J. Phonet. *24:* 491–511 (1996).
Prince, A.; Smolensky, P.: Optimality theory: constraint interaction in generative grammar (unpublished, Rutgers University, New Brunswick University of Colorado, Boulder 1993).
Purcell, E.T.: Formant frequency patterns in Russian VCV utterances. J. acoust. Soc. Am. *66:* 1691–1702 (1979).
R Development Core Team: R: a language and environment for statistical computing (R Foundation for Statistical Computing, Vienna 2003).
Recasens, D.; Pallarès, M.D.; Fontdevila, J.: A model of lingual coarticulation based on articulatory constraints. J. acoust. Soc. Am. *102:* 544–561 (1997).
Schwartz, J.-L.; Boë, L.-J.; Vallée, N.; Abry, C.: The dispersion-focalization theory of vowel systems. J. Phonet. 25: 255–286 (1997).
Summers, W.V.: Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. J. acoust. Soc. Am. *82:* 847–863 (1987).
Syrdal, A.K.; Gopal, H.S.: A perceptual model of vowel recognition based on the auditory representation of American English vowels. J. acoust. Soc. Am. *79:* 1086–1100 (1986).
Van Bergem, D.R.: Acoustic vowel reduction as a function of sentence accent, word stress, and word class. Speech Commun. *12:* 1–23 (1993).
Ward, D.: Unaccented vowels in Russian. Russian Ling. *2:* 91–104 (1975).
Wedel, A.: Categorical behavior through self-organization: deriving constraint dominance; (PhD diss. University of California, Santa Cruz in progress).
Wood, S.A.J.; Pettersson, T.: Vowel reduction in Bulgarian – the phonetic data and model experiments. Folia ling. *22*: 239–262 (1988).