Listen, Repeat, Decide: Investigating Pronunciation Variation in Spoken Word Recognition among Russian Speakers

Vladislav Zubov, Elena Riekhakaynen
St Petersburg University
St Petersburg, Russia
Vladzubov21@gmaul.com, e.riehakajnen@spbu.ru

Abstract

Variability is one of the important features of natural speech and a challenge for spoken word recognition models and automatic speech recognition systems. We conducted two preliminary experiments aimed at finding out whether native Russian speakers regard differently certain types of pronunciation variation when the variants are equally possible according to orthoepic norms. In the first experiment, the participants had to repeat the words with three different types of pronunciation variability. In the second experiment, we focused on the assessment of words with variable and only one standard stress. Our results support the hypothesis that listeners pay the most attention to words with variable stress, less to the variability of soft and hard consonants, and even less to the presence / absence of /j/. Assessing the correct pronunciation of words with variable stress takes significantly more time than assessing words which have only one correct pronunciation variant. These preliminary results show that pronunciation variants can provide new evidence on how a listener accesses the mental lexicon during natural speech processing and chooses among the variants stored in it.

Keywords: pronunciation variants; spoken word recognition; Russian

1. Introduction

Spoken word recognition (SWR) studies quite often address the problem of variability in the speech signal, since variability (or variation) is one of the important features of natural speech (Brouwer, 2010) and also a challenge for both spoken word recognition models and automatic speech recognition systems (Luce, McLennan, 2005). The variability of a speech signal can include individual characteristics of the speaker (timbre, dialect, accent, etc.), emotional state (tempo, intonation), speech style (formal or informal, etc.), features of the communication environment (noise and interference), see (Pufahl, Samuel, 2014) for a review. Particular attention is paid to the pronunciation variation: duration and quality, sound changes, reduction, stress. Pinnow et al. (2017) provided an example of how spoken variants can be used to assess different approaches to how a listener accesses words: either there is a set of different variants in the lexicon, or information is available in the speech signal that allows a successful comparison between the surface form and the canonical form, the latter being stored in the mental lexicon.

In particular, the paper examines reduced words and analyzes their role in the activation of unreduced canonical forms. Reduction in general is most often in the scope of the studies on pronunciation variability (see (Tucker, Ernestus, 2016)). Another type of variability is discussed by Cutler and Jesse (2021), who suggest that the stress patterns should be represented in the mental lexicon of a particular language and play a role in the process of spoken word recognition. Stress can serve as an important marker in the process of lexical access, determining which lexical items are activated in the native speaker's mental lexicon. Thus, we assume that the use of words with variable pronunciation as material for research in the field of SWR will provide new data on the lexical access and on the organization of the mental lexicon.

Despite a significant number of studies of variable pronunciation and the mental lexicon across various languages, researchers often encounter a challenge as pronunciation variants may influenced by sociolinguistic parameters. These variants can belong to different dialects, age groups, or hold varying degrees of prestige. Such characteristics impose limitations on research, as illustrated by Warren and Hay (2006).

Based on the Russian language material, descriptive studies of variation are usually carried out within the framework of orthoepy and sociolinguistics. Many papers provide rich data on modern pronunciation norms and sociocultural factors of speakers that influence the choice between pronunciation options (Kalenchuk, Savinov, 2021). However, until now, perceptual studies of pronunciation variants have not been systematically carried out. At the same time, in our opinion, the Russian language is a promising source of data on the processing of variability during SWR, since unfixed stress and active lexical processes associated with borrowing words result in numerous items with different pronunciation variants. Particularly interesting are the cases when pronunciation variants are noted by researchers as equal, i.e. there is no evidence for significant factors influencing the choice of a certain variant (context, frequency, style of speech, social status of the speaker, etc.). Thus, two or more pronunciation variants of a word exist in parallel in everyday speech and are used without any restrictions, e.g. variation of stress patterns (núzhny or nuzhný 'are needed') or variation of the consonant before the following vowel /e/ (soft or hard) ([sʲérvʲis] or [sérvʲis] 'service'), and so on. We assume that such variants can be useful for studying the ways a listener accesses the mental lexicon during natural speech processing and chooses among the variants stored in it. As far as we can conclude from the literature, such equally possible variants are not frequent in other languages. Thus, Russian data can provide quite rare evidence

The Workshop on Cognitive Aspects of the Lexicon, pages 129–132

on how a listener processes variation not influenced by sociolinguistic factors.

In this paper, we describe two preliminary experiments that we conducted to answer the following questions:

(1) To what extent do listeners generally notice variability in the speaker's speech, and does this depend on the variability type (different stress, substitution of sounds, or changes in the number of sounds)?

2) Is it possible to equate access to words that have several pronunciation variants with access to words that have one pronunciation variant, but are pronounced correctly or incorrectly?

The answer to the first question is explicitly stated in a few Russian-language papers (Pozharitskaya, 2004; Kasatkin, 2011; Kalenchuk, Savinov, 2021), which show that listeners pay attention to the place of stress much more often than to the segmental structure of words, but these assumptions are not supported by any experimental data. In our paper, we report Experiment 1, which offers empirical support for this proposition.

As for the second question, it is necessary to carry out preliminary studies to describe the mechanism of the recognition of words with incorrect pronunciation, and then compare these results with data obtained on the material of words with variable pronunciation. In Section 3 of the paper, we describe a pilot Experiment 2, which will be the beginning of such work.

Both experiments were conducted in the accordance with the Declaration of Helsinki and the existing Russian and international regulations concerning ethics in research.

2. Experiment 1

2.1 Method

As the goal of the experiment was to find out whether listeners pay attention to how the words with equally possible pronunciation variants are realized, we decided to ask participants to repeat the phrases they heard. There are at least two types of repetition tasks, one being the shadowing and the other – the imitation task. In the former, the participants are not given any special instructions on how accurate their repetition should be, whereas in the latter they are "explicitly instructed to imitate the productions they were exposed to" (Dufour, Nguyen, 2013). Dufour and Nguyen (2013) have shown that the general mechanism revealed by these two experimental paradigms is probably the same and provides evidence on how the words are stored in the long-term memory. Thus, we instructed our participants that they should just repeat what they heard. We supposed to obtain the information on how accurately participants process different types of pronunciation variation.

2.2 Stimuli

We chose the material for the experiment from the Big Orthoepic Dictionary of the Russian Language (https://gramota.ru/biblioteka/slovari/bolshoj-orfoepicheskij-slovar-russkogo-yazyka). According to it, all the words we used in the stimuli can have two pronunciation variants and these variants do not depend on the age and other parameters of the speakers and are considered equally appropriate to be used by the native speakers of Russian. We compared three types of variation: 1) Stress: variation of stress patterns (e.g. núzhny or nuzhný 'are needed'); 2) CV: variation of the quality of a consonant before the following vowel /e/ (soft or hard) ([sʲérvʲis] or [sérvʲis] 'service'); 3) VJV: presence or absence of the consonant [j] between two vowels (proekt [proekt] or [project] 'project'). For each group, we chose 12 words. These were mainly nouns (26 out of 36), but also five adjectives, four verbs and one adverb. Nouns are the most frequent words in the Russian language, and it seems that the phonetic variation of the three types we chose for our study occurs in these words more often than in other parts of speech. We included in the experiment 12 fillers (the words without pronunciation variants) which were also mainly nouns.

We created two-word constructions with all the words, which were read by a male speaker and audio-recorded. For all the stimuli, we recorded both pronunciation variants; fillers were recorded only once, as they had only one possible pronunciation. Then, we arranged all words into two stimuli lists. Each list included 12 fillers and one of the two possible pronunciation variants for each of 36 stimuli. The duration of both stimuli lists was about 3.5 minutes.

2.3 Procedure

During the experiment, participants listened to one of the two audio recordings via headphones and were asked to repeat after the speaker exactly what they heard. They were given 3 seconds to respond to each stimulus. The experimenter documented whether the variant pronounced by the participant matched the one in the recording.

2.4 Participants

96 native speakers of Russian took part in the experiment (62 female; M_{age} = 19 y.o.). None of them reported any hearing problems. All participants provided an oral consent to take part in the experiment.

2.5 Results

The number of correct repetitions (CORR) after the speaker for each individual stimulus was analyzed (regardless of the pronunciation variants, since the number of their presentations was equal). The mean CORR (Max = 96) and standard deviation (SD) for each type are provided in Table 1.

Variation type	CORR Mean	SD
VJV	51.75	5.29
CV	66.08	8.21
Stress	85.75	6.65

Table 1: Average correct repetitions and standard deviations for each condition

The smallest number of correct repetitions was in the group 3 VJV (with the presence or absence of the intervocalic /j/) – 53.9%, and the largest – in the group of words with variable stress (89.3%). To test whether the differences were significant, a linear regression model was fit. The outcome variable was CORR, and the predictors were the type of variation, which had three levels: VJV, CV, Stress (see Table 2).

	Estimate	SE	t	p
(Intercept)	54.85	1.91	28.73	< .001
CV	15.95	2.70	5.91	< .001
Stress	38.10	2.70	14.11	< .001

Table 2: Summary of significant effects in the number of correct repetitions

Neither the frequency of word forms of the selected words, nor the part of speech had a significant effect on the number of correct repetitions after the speaker, and thus these parameters were not included in the model. It can be concluded that the number of correct repetitions strongly depends on the type of variability. Listeners pay the most attention to words with variable stress, which was noted in previous papers (Pozharitskaya, 2004; Kasatkin, 2011; Kalenchuk, Savinov, 2021). The change in the quality of the consonant sound before the vowel is less prominent for the native speakers of Russian, whereas the presence or absence of an intervocalic /j/, apparently, is not noticed in speech, since the number of correct repetitions behind the speaker is close to random. In the next experiment, we decided to test how lexical access to words with variable stress occurs.

3. Experiment 2

3.1 Method

Reaction time is a measure which is commonly used to study lexical access. Most often the reaction time is measured while participants perform a lexical decision task (LDT). As in our study we focus on pronunciation variation, we measured reaction time while participants were deciding whether the given word is correct or not. Thus, we used a modified version of the LDT.

3.2 Stimuli

We recorded 30 isolated words for the experiment. Their pronunciation was checked in the same orthoepic dictionary as in the first experiment. There were three groups of words: 1) those with variable stress (for each of them we recorded two stimuli with both variants); 2) with the only one standard stress and pronounced correctly by the speaker; 2) with the only one standard stress but pronounced incorrectly by the speaker.

3.3 Procedure

The experiment was conducted in PsychoPy. Each participant was presented with 30 isolated words in random order through headphones; one of the two possible stimuli for every word with variable stress was chosen randomly by the program. After listening to each stimulus, participants answered whether the word sounded correct or not by pushing one of two buttons on the computer. Reaction time (from the beginning of listening to making a decision) and the correctness of answers to questions were measured.

3.4 Participants

25 people took part in the experiment (20 female; Mage = 18 y.o.). None of them reported any hearing problems. All participants provided an oral consent to take part in the experiment.

3.5 Results

We analyzed the average reaction time (RT, ms) in each of the stimulus groups (747 reactions in total), as well as the answers of the participants (in which cases the stimulus was considered correct, the percentage of the total number of responses).

Pronunciation	RT (ms)	SD	Answers “correct”
Variable	2248.69	859.31	64.7%
Incorrect	2090.45	786.12	6.4%
Correct	1592.17	509.68	100%

Table 3: Average mean RT, standard deviations and the percentage of the answers “correct” for each condition

The Table 3 shows that words with variable stress are rated as correctly pronounced in 64.7% of all cases, while the words with the only one correct stress (group 2) are rated as correct by all participants and the incorrectly pronounced stimuli are most often considered incorrect.

Words with variable stress required the greatest amount of time for participants to react, but we should note that the standard deviation in this group is the largest.

To assess the statistical significance of the results obtained, we used linear regression with RT as the dependent variable. The group of stimuli, in which there were three levels (variable, irregular and correct) and the frequency of word forms according to the Russian National Corpus (https://ruscorpora.ru/en/) were used as predictors (Table 4). Length of stimuli in number of sounds and part of speech did not show a statistically significant effect and were not included in the model.

	Estimate	SE	t	p
(Intercept)	2219.75	47.97	46.27	< .001
Incorrect	-99.29	68.14	-1.46	0.146
Correct	-562.00	70.93	-7.92	< .001
Freq (log10)	-67.37	20.40	-3.30	0.001

Table 4. Summary of effects in RT

The word form frequency plays a role in the evaluation of stimuli, even though a modified LDT technique is used, so this factor needs to be considered in future studies.

The RT for the group of words with regular stress turned out to be statistically significantly lower compared to the other groups, but no statistical difference was achieved between the groups of stimuli with irregular and variable stress. It is assumed that the lack of a statistically significant difference may be due to the heterogeneity of the stimuli in these groups, since it is not easy to select words of the same length, part of speech and frequency for the Russian language, because we lack a database of words with pronunciation variants.

4. Discussion and Conclusion

In this study, we expand on the concept of variability in the Russian language from a perceptual perspective and try to assess the role of pronunciation variability in the process of SWR. Based on the results of two experiments, we can conclude that, firstly, listeners notice variability in speech in different ways, and when repeating after the speaker, in some cases they activate the same units that they heard, and in the others – those that are stored in their mental lexicon and not necessarily matching the heard variant.

We provided experimental evidence for the assumption of Russian orthoepy experts that a naive native speaker of Russian, when assessing pronunciation, pays more attention to word stress, less to the variability of soft and hard consonants, and even less to the presence / absence of /j/ (Pozharitskaya, 2004; Kasatkin, 2011; Kalenchuk, Savinov, 2021). Secondly, assessing the correct pronunciation of words with variability takes significantly more time than assessing words which have only one correct pronunciation variant. However, it is not clear how the process of accessing words with variable pronunciation occurs: whether it is similar to how words with incorrect stress are recognized or differs from it. We hypothesize that further exploration of the variability phenomenon in Russian from a perceptual perspective will yield insights into these questions.

The limitations of the current study include the following:

a) the level of conducted experiments is rather shallow, since the results do not allow us to draw conclusions about the access to the listener's mental lexicon. However, the results obtained show the promise of further research into the described language material;

b) the sets of stimuli for both experiments were unbalances because of the absence of a database containing pronunciation variants in Russian;

c) it is necessary to compare the results with similar data from other languages, in which we can find equally possible pronunciation variants;

d) the documentation of accurate repetitions after the speaker in the first experiment relying on the experimenter's hearing might have influenced the results (particularly in the VJV group).

In our further research on Russian, we intend to conduct a more careful selection of stimuli. This selection will allow to include various factors into the model (frequency, morphological features, morphemic composition, etc.). We also plan to develop designs for more complex experiments aimed at gathering data on the process of SWR.

5. Acknowledgements

We thank Elizaveta Zhukatinskaya and the students who took part in the linguistic program in the Educational Centre "Sirius" (Sochi, Russia) in October 2021 for the help with collecting and analyzing the data.

6. Bibliographical References

Brouwer, S. (2010) *Processing strongly reduced forms in casual speech*. PhD Thesis, Radboud University Nijmegen, Nijmegen.

Cutler, A., and Jesse, A. (2021). Word Stress in Speech Perception. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The Handbook of Speech Perception*, Wiley, pp. 239–265). https://doi.org/10.1002/9781119184096.ch9

Dufour, S., and Nguyen, N. (2013) How much imitation is there in a shadowing task? *Frontiers in Psychology:* 4. https://doi.org/10.3389/fpsyg.2013.00346

Kalenchuk, M. L., Savinov D. M. (Eds.). (2021). Pronunciation standards in usage and codification (in Russian). Moscow: Russian Language Institute.

Kasatkin, L. L. (2011). Orthoepeme as the Basic Unit of Orthoepy. *Voprosy Jazykoznanija*, 2: 31–38.

Luce, P. A., and McLennan, C. T. (2005). Spoken Word Recognition: The Challenge of Variation. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*, Blackwell Publishing Ltd. pp. 590–609. https://doi.org/10.1002/9780470757024.ch24

Pinnow, E., Connine, C. M., and Ranbom, L. J. (2017). Processing pronunciation variants: The role of probabilistic knowledge about lexical form and segmental co-occurrence. *Journal of Cognitive Psychology*, 29(4): 393–403. https://doi.org/10.1080/20445911.2017.1279619

Pozharitskaya, S. K. (2004). Orthoepy: Ideas and Practices. In G. E. Kedrova, Potapova V. V. (Eds.) *Language and Speech: Problems and Solutions*. Moscow: MAKS.

Pufahl, A., and Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, 70: 1–30. https://doi.org/10.1016/j.cogpsych.2014.01.001

Tucker, B., and Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon*, 11(3): 375-400. https://doi.org/10.1075/ml.11.3.03tuc

Warren, P., and Hay, J. (2006). Using sound change to explore the mental lexicon. In. C. M. Flinn-Fletcher & G. M. Haberman (Eds.), *Cognition and Language: Perspectives from New Zealand*, Australian Academic Press. pp. 105-125.