Правительство Российской Федерации

Федеральное государственное автономное образовательное

учреждение высшего образования

Национальный исследовательский университет

«Высшая школа экономики»

###### Факультет гуманитарных наук

###### Образовательная программа

###### «Фундаментальная и компьютерная лингвистика»

Астафьева Ирина Юрьевна

**Исследование консонантных изменений с набитым ртом**

Выпускная квалификационная работа студента 4 курса бакалавриата группы БКЛ153

|  |  |
| --- | --- |
| Академический руководитель образовательной программы | Научный руководитель |
| канд. филологических наук, доц. | Старший преподаватель |
| Ю.А. Ландер | Г. А. Мороз |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Москва 2019

Table of Contents

[Introduction 3](#_Toc10574850)

[Literature review 4](#_Toc10574851)

[Research question and hypotheses 6](#_Toc10574852)

[Methods 7](#_Toc10574853)

[Data analysis 11](#_Toc10574854)

[Gender 12](#_Toc10574855)

[Musical experience 13](#_Toc10574856)

[Linguistic education 14](#_Toc10574857)

[Number of learned foreign languages 14](#_Toc10574858)

[Number of syllables in stimuli 15](#_Toc10574859)

[Types of hampered conditions 16](#_Toc10574860)

[Summary 18](#_Toc10574861)

[Logistic regression 18](#_Toc10574862)

[Results 21](#_Toc10574863)

[Conclusion and implications 22](#_Toc10574864)

[Acknowledgments 23](#_Toc10574865)

[References 23](#_Toc10574866)

[Appendices 25](#_Toc10574867)

[1. One-syllable stimuli: 25](#_Toc10574868)

[2. One-syllable fillers: 25](#_Toc10574869)

[3. Two-syllable stimuli: 25](#_Toc10574870)

[4. Two-syllable fillers: 26](#_Toc10574871)

[5. Experimental lists. 26](#_Toc10574872)

[6. The experiment consent forms 30](#_Toc10574873)

*When I eat I’m deaf and dumb.*

*Russian proverb*

# Introduction

There are many studies that investigate the process of speech perception. Apparently, there is no single exhaustive reason for proper speech recognition. There are several significant aspects that can be highlighted.

Speech perception can be described through different approaches and the overall picture of this complex process is the following: semantic cues play an important role (Obleser, Wise, Dresner, & Scott, 2007), lexical units of the language compose into a sequence of segments Kenneth Noble Stevens (Pisoni, n.d., pp. 125–165), native speakers use all types of cues depending on context and situation in natural speech Lawrence J. Raphael (Pisoni, n.d., pp. 182–206). Another possible way to study speech perception is to conduct an experiment. According to the previous scientific experiments, it is known that native listeners perceive the information correctly in more cases than non-native listeners (Florentine, 1985) and that musical experience helps to understand acoustic information in noisy condition (Ito, Takeda, & Itakura, 2005) and emotions (Strait, Kraus, Skoe, & Ashley, 2009).

This study is focused on speech perception in daily life. Conditions for the generation and perception of speech in everyday life are not always perfect. Moreover, the conditions for listening to the interlocutor (listening condition) are often hindered by noise or side effects as, for example, eating (eating condition). To the best of our knowledge, there is no studies on phonetics of speech recognition in those conditions. This paper investigates the perception of velar and dental consonants under three hampered conditions. The first is mouthfuls of liquid condition (brushing teeth), the second and the third are eating conditions: soft food (sour cream or humus) and solid food (slices of apple).

During the experiment participants had to evaluate pairs of words from different speakers, answering the question whether they are the same word or two different words, and also guess the meaning. The main aim of the study was to investigate if velar and dental consonants are recognized as one sound in the challenging condition when some food or liquid are in mouth. Since speakers were forced to say the stimuli, we expected that they would similarly compensate lack of the ability to use the front part of the mouth using some sounds from the velar zone. This effect was tested on monosyllabic and disyllabic words of the Russian language.

Result of the experiment shows that Russian words with velar and dental consonants (e.g., *tort*[[1]](#footnote-2) (cake) and *kort* (court)) are not always perceived as the same word with velar sound (as *kort* (court)). Another expectation was that recognition of words will differ depending on the type of chosen hampered condition. The hypothesis was that the most difficult condition would be the solid one (with an apple), and the easiest condition would be the soft one (with humus or sour cream). It was partially confirmed; the soft condition really turned out to be the easiest (the number of answers 'different' greatly exceeds the answers ‘identical’). The hardest one was the liquid condition (with toothpaste foam). Participants answered 'identical' more often than in other conditions, and the number of ‘identical’ responses was greater than the number of ‘different’ answers. Based on several factors (linguistic education, type of difficulty, specific pair of incentives) logistic regression with mixed effects was built to predict the response ‘different’ or ‘identical’.

All statistical computations and visualization were conducted with R language (Team, 2016). We used the following R packages: tidyverse (H. Wickham, 2018) dplyr (Hadley Wickham, Francois, & Henry, 2018), ggplot2 (Hadley Wickham, 2016).

This study may be helpful for people with aphasia or children in the process of first language acquisition. On the material of the study exercises to practice articulation can be compiled. Furthermore, the data obtained during the work can be taken into account for creating algorithms aimed at automatic speech recognition.

# Literature review

Speech perception is a complex process which is not fully understood. Our everyday speech encoding is successful in spite of different listening conditions that are often hampered (for example, in noise or poor quality of telephone lines). Despite all noise the participants of the dialogue usually perceive the information unequivocally. It is believed that semantic context helps listener understand the message correctly (Obleser et al., 2007).  There are other explanations of speech perception.

For example, according to Kenneth Noble Stevens (Pisoni, n.d., pp. 125–165), natural words of any language are stored in memory as a sequence of segments. Each of the segments represents a set of distinctive features. Some factors may have influence on realization of distinctive features: prosodic context, additional articulate gestures (to enhance perceptual contrast) and an overlay of acoustic gestures of two neighboring segments. It allows us to make an assumption about the process of speech perception. Access to lexicon is a search for a sequence of a distinctive feature set in the lexicon. After this, by synthesis of the set there is verification of the result with the acoustic signal received at the input and the choice of the optimal variant. A person extracts a sequence of lexical units from a speech stream and encode it to words.

To encode the sequence of lexical units a person needs to highlight significant acoustic characteristics for speech perception and sound recognition. Lawrence J. Raphael (Pisoni, n.d., pp. 182–206) argue that many acoustic cues are involved in speech understanding and native speakers apparently use all of them depending on context and situation. For example, the cues are method of sound formation and place of sound formation, voicelessness or sonority of a sound, etc. Thus, it is hard to find a simple way to connect acoustic cues and speech perception.

Experiments are a natural method for studying speech perception. Consider several experiments that were aimed to discover how some sound combinations can be synthesized. Liberman et al. (1952) conducted an experiment in which the stimuli were artificially synthesized sounds. The stimuli formed combinations: an explosion at a certain height and a two-format vowel. It turned out that recognition of explosive consonants is context-dependent:

* 360 Hz explosion was perceived as [p];
* explosions above 3000 Hz were perceived as [t];
* explosions with a frequency of 1440 Hz were perceived as [p] before [i] and [u], but as [k] before [a].

Some experiments investigate the recognition of speech in difficult conditions. This is a common situation in our daily life, because ideal listening conditions are available for us to understand the message of the interlocutor very seldom. Florentine’s study (1985) found that native listeners can perceive the information correctly at significantly higher noise levels than non-native listeners.

There are a lot of studies that investigate speech perception in challenging conditions and lay stress on the different aspects of the normal speech. For instance, there are surveys that examine whispered speech. According to (Ito et al., 2005), in whispered speech the accuracy of speech is significantly reduced compared to the usual speech. Furthermore, in this condition the formant frequencies of vowels are shifted.

Selected studies focus on perception of consonants and vowels in noise. (Parikh & Loizou, 2005) finds that listeners rely on F1 frequency information and partial F2 information to recognize vowels. However, stop consonant identification is successful even in extreme noise. Authors suggest that listeners may depend on other cues, for example, formant transitions.

A large part of the phonetical research of speech perception in difficult conditions is devoted to influence of musical experience on sound recognition. It improves the ability to hear speech in difficult listening conditions; in addition, musical education enhance working memory and frequency discrimination (Ito et al., 2005). People with musical experience capture emotions in speech acts better than others (Strait et al., 2009).

According to Anderson et al. (2010), speech encoding is disrupted by background noise and in challenging listening conditions. Musicians show increased auditory attention which means shorter reaction time compared to non-musicians and ability for accurately repetition of sentences presented in noise. It demonstrates that musicians are more resistant for responses in hampered conditions (for instance, the noisy whisper to the background) and better perceive speech in higher levels of noise. The explanation is that the musical training forms neural mechanisms that underpinned selective attention to speech. (Strait & Kraus, 2011).

There is another specific but frequent situation in everyday life which is associated with speech perception in challenging conditions. It is speech perception in eating or mouthfuls condition: we often exchange cues with each other at breakfast or, for example, while we brush our teeth. It is significant that not only noisy condition can interfere with speech comprehension but also unclear speaker articulation. In such cases, the meanings of words may be partially dictated by the situation and semantics (Obleser et al., 2007), but it is not entirely clear whether something else affects the listener's understanding. To the best of our knowledge, there are no any published works that analyze speech perception in detail when one of the interlocutors eats. We chose this research niche for our research.

To sum up, there are many studies that investigate the process of speech perception. The one of the frequent conditions in our everyday life is speech perception in difficult conditions. Many studies created conditions of speech understanding designed in noisy environments, compared native speakers and non-native speakers, musicians and non-musicians, but no one explored speech in terms of eating condition at the phonetic level. The purpose of our research is to analyze the perception of individual words in different conditions when interlocutors eat.

# Research question and hypotheses

It can be observed that many background researches are focused on the mechanism of speech perception. Linguists are trying to explain principles by detailed describing of this process and usually scientists investigate difficult situations of speech perception, when there are some problems with the channel (noise (Florentine, 1985), (Ito et al., 2005), (Anderson et al., 2010), (Strait & Kraus, 2011), etc.) or with brain activity (aphasia (Blumstein, 1994), (Mummery, Ashburner, Scott, & Wise, 1999), (Hickok & Poeppel, 2000), etc.). For the first time we begin to investigate situations where problems are in the area of the articulators (speakers) themselves. How can a person be influenced to make him or her sure that different sounds are the same?

In this paper we would like to research speech perception in hampered with food and liquid conditions, and we base on Russian language material. With data from the experiment we aimed at determining in which cases people can perceive different sounds equally. All participants of the experiment listened to sounds produced in challenging conditions. In addition, we tried to determine if there is a connection between responses and the following factors:

1. The quality characteristics of the food and perception of stop consonants.
2. The presence and absence of linguistic education.
3. The presence and absence of musical experience.
4. Number of foreign languages that a person knows.
5. Gender.
6. Age.

The main hypothesis is that velar and dental consonants (for example, *t* and *k*) in Russian words (e.g., *tort* (cake) and *kort* (court)) in special condition can be distorted and both perceived as a velar sound (in case of *tort* (cake) and *kort* (court)they will be perceive as *kort* (court)). The full list of stimuli and fillers is available in the appendices (Appendices 1-4).

The present study may help to explain speech perception process and to compile phonetic exercises for people which have difficulties with articulation (for example, people with aphasia or foreign students that aimed at learning Russian language). Furthermore, results of the current research can be taken into account for creating speech recognition systems.

# Methods

For our study we conducted the following experiment. As a part of preparation (Figure 1, Figure 2) for the experiment we chose five men[[2]](#footnote-3), native speakers of Russian language with good pronunciation, neutral intonations and loud voices. We experimented with three different challenging conditions from everyday life: mouthfuls of liquid (brushing teeth), soft food (sour cream or humus) and solid food (slices of apple). Speakers read the list of stimuli and fillers four times (normally and in three hampered conditions). Based on their pronunciation we produced lists which were posted online.

Before running the experiment, a control groups were testing it. The first control group listened to stimuli in normal speech condition to make sure that in normal conditions speakers pronounce words unequivocally. The second control group tried to recognize stimuli while studying conditions and estimate how conveniently the task is constructed.

A group of people sitting at a table

Description automatically generatedA group of people sitting at a table

Description automatically generated

Figure 1. (left) The process of audio recording with Kirill Semenov.

Figure 2. (right) The process of audio recording with Stepan Mixajlov.

Initially, we planned to conduct a procedure of the experiment differently compared to the final version. Listeners were supposed to listen to certain single stimulus words mixed with fillers and write down the semantic meaning of the words they heard. However, when the control group were testing experiment procedure, nobody could identify all the words. Moreover, it was hard to make a decision about at least one word. We resolved to make the procedure easier and suggested that the participants evaluate stimuli as one (the same) or two (different) words. After this a person could offer a word pronounced on the recording.

There was no option with suggested answers for several reasons. First, stimuli differ from each other only in one particular sound: *t* or *k*. No matter how many options were offered for selection, for each pair of stimuli, two almost identical answers would have to be offered, although the person could not make out any of them. Secondly, in present work the most important for us is to know whether the stimuli are perceived as the same or different words. Finally, obligatory identification of words makes the procedure longer and more difficult. The experiment was conducted online; this meant that many of the participants would not complete the passage. Thus, the organization of the experiment was simplified.

The final experimental task consisted of:

1. listening to one pair of separate stimuli words;
2. valuing the couple of words as the same word pronounced by different speakers or two different words from two speakers;
3. trying to recognize them and writing down the participant’s own variant in a special field in the online questionnaire.

For this purpose, four pairs of two-syllable stimuli and three pairs of one-syllable stimuli, and also fifteen fillers were used. There were three types of one-syllable stimuli (one pair for one type); all stimuli pairs are the following:

* combination of consonant-vowel-consonant, in which the first consonant was tested for perception (*tot* (that) and *kot* (cat));
* combination of consonant cluster-vowel-consonant, in which consonant cluster is combination of fricative consonant and stop consonant, the stop consonant was tested for perception (*stol* (table) and (*skol* (chipped));
* combination of consonant-vowel-consonant cluster, in which consonant cluster is combination of liquid consonant and stop consonant, the first stop consonant was tested for perception (*tort* (cake) and *kort* (court)).

There were also three types of two-syllable stimuli (one pair for one type); all stimuli pair are the following:

* combination of consonant-vowel-consonant, in which the first consonant was tested for perception (*tomik* (book, volume) and *komik* (comedian), which contain palatalized consonant; *tonus* (tone) and *konus* (kone), which contain velarized consonant);
* combination of consonant cluster-vowel-consonant, in which consonant cluster is combination of fricative consonant and stop consonant, the stop consonant was tested for perception (*stolko* (so many/much) and *skolko* (how));
* combination of consonant-vowel-consonant, in which the first consonant is affricate or stop consonant and was tested for perception (*tsaplya* (heron) and *kaplya* (drop)).

In all two-syllable words (both in fillers and in stimuli) the stress fell on the first syllable. Fillers distracted participants from the real aim of the research; the first filler helped to understand the task and gave an example of the procedure.

The words for the experiment were selected according to the frequency dictionary of the Russian language by Lyashevskaya and Sharov (2009). All selected words are high frequency ones, except *skol* (chipped) and *tsaplya* (heron), which are not included in the first twenty thousand frequency words of the new frequency dictionary by Lyashevskaya and Sharov (2009); however, we believe that it is necessary to add them to our data. To the best of our knowledge, in Russian language *skol* (chipped) is the most frequent word which consists of one syllable, has a combination of fricative and another consonant before the stressed vowel, and is in a minimal pair (*skol* (chipped) and *stol* (table)). The word *tsaplya* (heron) was important to us in order to include affricate in our analysis.

Stimuli words are pairwise minimal pairs that differ only in one consonant (for instance, *tort* (cake) and *kort* (court)). A full list of words is available in the appendices (Appendices 1-4). Each participant receives a randomized sample of stimuli words when filling the questionnaire. We predicted that in hampered condition words with velar and dental sounds can be perceived identically because it was the most difficult condition for speakers to make a stop in the stimulus.

There were six different lists with stimuli and fillers (Appendix 5). The lists can be divided into two types. Type 1 (lists 1-3) differed from type 2 (lists 4-6) only by speakers who uttered words; the set of stimuli and fillers was the same. Each list included five pairs of words for each of the challenging conditions. Thus, each participant listened to fifteen pairs of words. On average, the experiment took about seven to ten minutes. Online participation in the procedure allowed us to interview a larger number of subjects.

Some principles for designing the experiment were described by Gries (2013). The order of stimuli was pseudo-random: using a randomizer, three sequences of stimuli and fillers were obtained, then they were transformed so that the stimulus never stood the first in any of the groups of five (for each condition), and two fillers did not stand in a row (wherever it was possible).

The experimental paradigm was programmed in online tool for surveys ‘Gizmo’ (“Online Survey Software & Tools,” n.d.). Participants could work in an experiment from any device (smartphone, tablet or desktop) because the program can automatically adjust to the screen size and device type. All subjects of the experiment were volunteers and participated of their own accord.

A screenshot of a cell phone

Description automatically generated

Figure 3. An example of page of survey.

At the beginning of the survey participants were asked to give personal information which included their age, gender, linguistic education, music experience and names of foreign languages spoken by a person. Thereafter, a subject entered the survey page which contained instructions and tasks. All questions were mandatory; if participant skipped at least one question, the program gave an error and did not allow going to the next page (Figure 3).

Data manipulation and statistics were made with stringr package (Hadley Wickham, 2015), tidyverse package (H. Wickham, 2018) and dplyr package (Hadley Wickham, Francois, & Henry, 2018)), graphing (ggplot2 package (Hadley Wickham, 2016)) and data analysis were performed in the R language (Team, 2016)). The code that reflects the progress of the statistical analysis is available online in special GitHub repository (i, 2018/2019).

# Data analysis

There were 105 volunteers, 37 men, 68 women, from 10 to 68 years old. We expected all participants of the experiment to consider the stimuli pairs as one word. If at least one subject decides that there are two different words among the stimuli pairs, we will consider our hypothesis incorrect.

The graph (Figure 4) consists of three columns, which contain the answers of participants in the experiment. The abscissa axis contains the response options, and the ordinate axis indicates the number of responses from the participants. The bar plot presents that in the total for all stimuli the ‘different’ answer option exceeds the ‘identical’ option. Some subjects could not determine whether a pair of stimuli was different words or the same. On the graph, this is the first column without a title; it will not be taken into account in further analysis.

A screenshot of a cell phone

Description automatically generated

Figure 4. Identical/Different Distribution in the total for stimuli pairs.

The following bar plot (Figure 5) displays a more detailed analysis of the separated stimulus pairs. The ordinate axis contains names of stimuli pairs, and the abscissa axis presents the number of responses from the participants. It is clearly seen that some rows vary greatly in values. However, responses for *tonus.conus* and *tot.kot* can be grouped together; response distribution for tort.kort does not have statistically significant difference from these two pairs of stimuli, and can be considered as part of the group. Answers for *tsaplya.kaplya* and *stol.skol* also constitute a common group. The row with distribution for *stolko.skolko* differs from all listed groups. Thus, there are three series of dissemination stimulus elements responses: ‘Identical’ slightly exceed ‘Different’ (two pairs; *tsaplya.kaplya* and *stol.skol*); ‘Different’ exceed ‘Identical’ (three pairs; *tonus.conus, tot.kot, tort.kort*); ‘Identical’ exceed ‘Different’ significantly (one pair; *stolko.skolko*).

By applying Fisher's Exact Test to the resulting contingency table, we obtained a low p-value (p-value = 6e-05 < 0.05). It allows to consider the difference between the columns statistically significant. However, if the test is applied to two random rows (for example, for *tonus.konus* and *tsaplya.kaplya*), then p-value will be high (p-value = 0.2191 > 0.05) and will not indicate statistical significance. Thus, it is required to connect the columns with similar values to check which data is actually different from each other.

A screenshot of a cell phone

Description automatically generated

Figure 5. Different/Identical Distribution for separated stimuli pairs.

The following several subheadings in this section describe the distribution of responses, depending on various factors that have been collected during the experiment procedure. They are gender, musical experience, linguistic education, number of learned foreign languages, number of syllables in the word (or length of the words) and type of challenging condition. The description is aimed at data observation in an effort to find out which factors influenced responses more and which factors affected them less and to construct the logistic regression.

## Gender

All participants demonstrated slight predominance of option ‘different’ over optional ‘identical’ regardless of gender. The graph (Figure 6) shows that responses of men (right column) and women (left column) look similar. The abscissa axis contains options of answer, and the ordinate axis presents the number of responses from the participants.

By applying Fisher's Exact Test to the resulting contingency table, we obtained a very high p-value (p-value = 1 > 0.05); therefore, the differences are not statistically significant, and there is no reason to discard the alternative hypothesis. Thus, according to the data, gender does not affect respondents’ answers.

A screenshot of a cell phone

Description automatically generated

Figure 6. Different/Identical Distribution by gender.

## Musical experience

The following bar plot (Figure 7) illustrates distribution of responses by musical experience. The instruction (Appendix 6) said that it was the main aim of current study. The abscissa axis contains options of answer, and the ordinate axis presents the number of responses from the participants.

A screenshot of a cell phone

Description automatically generated

Figure 7. Different/Identical Distribution by musical experience.

The graph presents slight difference between respondents with (left column) and without (right column) musical experience. Subjects with musical education evaluate stimuli pairs more as two different words; and the answers of the participants without musical education are distributed almost equally between the two columns.

By applying Fisher's Exact Test to the resulting contingency table, we acquired a high p-value (p-value = 0.1165 > 0.05); therefore, the differences are not statistically significant, and there is no reason to discard the alternative hypothesis. Thus, according to the data, musical experience does not affect respondents’ answers.

## Linguistic education

The figure (Figure 8) demonstrates distribution of responses by linguistic education. The abscissa axis contains options of answer, and the ordinate axis presents the number of responses from the participants. The graph indicates the difference between respondents with and without linguistic education. Subjects with linguistic education evaluate stimuli pairs more as two different words; and the answers of the participants without linguistic education present the opposite.

By applying Fisher's Exact Test to the resulting contingency table, we received a low p-value (p-value = 0.0004936 > 0.05); therefore, the differences are statistically significant. Thus, according to the data, linguistic education affects respondents’ answers.

A screenshot of a cell phone

Description automatically generated

Figure 8. Different/Identical Distribution by linguistic education.

## Number of learned foreign languages

The bar plot (Figure 9) represents distribution of responses by the number of foreign languages learned. As in the case with the musical experience participants believed that it was the studied factor for the present study. The abscissa axis contains options of answer, and the ordinate axis presents the number of responses from the participants. Despite the predominance of the answers of respondents with knowledge of one or two foreign languages, the graph presents difference between respondents with knowledge of zero or one foreign languages (the first and the second columns on the graph) and with knowledge of more than one foreign language (from third to sixth columns). The first group evaluates most the pairs under investigation as one word; the second group evaluates most of the investigated pairs as two different words. Besides, as the number of languages studied increases, the difference in the distribution of answers towards the option ‘different’ increases, too.

By applying Fisher's Exact Test to the resulting contingency table, we received a low p-value (p-value = 0.00047 < 0.05); therefore, the differences are statistically significant. Thus, according to the data, number of learned foreign languages affects respondents’ answers.

A screenshot of a cell phone

Description automatically generated

Figure 9. Different/Identical Distribution by number of learned foreign languages.

## Number of syllables in stimuli

The figure (Figure 10) shows distribution of responses by the number of syllables in the stimuli. The abscissa axis contains types of stimuli (one- or two-syllable words), and the ordinate axis illustrates the number of responses from the participants. The bar plot presents difference between respondents for monosyllable words (the first column) and for two-syllable words (the second column). The respondents evaluated most of one syllable words as two different ones and most of two-syllable words, as the same word.

By applying Fisher's Exact Test to the resulting contingency table, we obtained a low p-value (p-value = 0.00433 < 0.05); therefore, the differences are statistically significant. Thus, according to the data, number of learned foreign languages affects respondents’ answers.

A screenshot of a cell phone

Description automatically generated

Figure 10. Different/Identical Distribution by number of syllables in stimulus.

## Types of hampered conditions

The graph (Figure 11) indicates distribution of responses by the type of hampering condition. The abscissa axis contains types of challenging conditions (liquid, soft food and solid food), and the ordinate axis exposes the number of responses from the participants. The bar plot reveals the difference between respondents for liquid condition (the first column; in the experiment it was toothpaste), soft food (the second column; in the experiment it was humus or sour cream) and hard food (the third column; in the experiment it was an apple). According to the figure, the liquid condition was the most difficult for participants; it is the only condition in which the number of ‘Identical’ responses was higher than the number of ‘Different’ responses.

A screenshot of a cell phone

Description automatically generated

Figure 11. Different/Identical Distribution by type of hampered condition.

By applying Fisher's Exact Test to the resulting contingency table, we acquired a low p-value (p-value = 0.04941 < 0.05); therefore, the differences are statistically significant. Thus, according to the data, number of learned foreign languages affects respondents’ answers.

It is necessary to check whether the factors of word length (number of syllables; Figure 10) and of hampered condition type (Figure 11). They probably constitute a multi-factor, and they do not influence the responses separately.

By applying Fisher's Exact Test to the resulting contingency table of one syllable stimuli, we obtained a high p-value (p-value = 0.1128 < 0.05); therefore, the differences are not statistically significant and there is no reason to discard the alternative hypothesis. By applying Fisher's Exact Test to the resulting contingency table of two syllables stimuli, we received a high p-value (p-value = 0.00418 < 0.05); therefore, the differences are statistically significant. Нence, there is a connection between the length of words and types of hampered condition. These two factors constitute a single complex factor that can only be considered joint. There is no reason to argue that factors considered influence the responses separately.

There was one more issue about the type of challenging condition. The bar chart (Figure 12) illustrates subjects' assumptions about conditions. The abscissa axis contains reveals types of hampered conditions, and the ordinate axis exposes the answers of respondents about the condition. The figure presents that most often the response was ‘soft condition’; the respondents chose least of all option ‘liquid’ subjects. These data confirm the assumption that the liquid condition is the most challenging condition for understanding. Participants recognized the soft condition best, but there are also a lot of answers ‘hard’ there. Responses ‘hard’ and ‘soft’ for solid condition were almost equal.

A screenshot of a cell phone

Description automatically generated

Figure 12. Distribution of real types of hampered conditions and responses.

By applying Z-Test to the values, we compared responses and the random chance to guess the three conditions. For 247 coincidences from 640 cells we obtained very low p-value (p-value < 2.2e-16), which means that random probability and participants’ responses differ, and the difference is statistically significant. Thus, according to the data, number of foreign languages learned affects respondents’ answers. We interpret this in such a way that the participants' responses are more accurate than the chance to randomly answer correctly.

## Summary

To sum up, there are a number of distributions by different parameters which hypothetically influence the respondents’ responses. Some of them (gender, musical experience) do not significantly affect choices, but others (linguistic education, syllable amount in the word, number of foreign languages learned, type of studying condition) have a strong impact on responses. People guess the type of condition better than random probability does. Notwithstanding the results of the test, the graph (Figure 11) shows that in general, the participants rather poorly determine the type of experimental state.

## Logistic regression

Based on the factors described above, a model of logistic regression was built. Although all of the factors offered above might have potential influence over choosing the option ‘Identical’ or ‘Different’, our data do not allow us to build a model with all parameters. By mixing parameters up, we constructed several versions and chose the best one. By going through various factors, we have a model that includes four most significant parameters:

* + - 1. Number of foreign languages learned;
      2. Type of challenging condition (solid food, soft food or liquid);
      3. Specific pairs of stimuli;
      4. Interaction of conditions types (parameter 2) and separated stimuli pairs (parameter 3).

Figure 13 reflects the regression model output[[3]](#footnote-4). In this model, intercept is statistically significant. Thus, according to the data, for the speakers of the Russian language is such a ratio:

For the ‘same’ we took the probability that the respondent would rate a couple of stimuli as identical words, and for the ‘different’ (that is equal ***1 – same*** because there are only two options) the probability of evaluation the pair as two different words. The value of ***c*** is an intercept. Hence,

These results suggest that respondents who answered ‘different’, are times more. If the value of the interest were around zero, then it would turn out that, on the whole, the experimental one would not care which option to choose from the options. Statistically significant intercept allows us to estimate the likelihood overall for non-engaged groups.



Figure 13. The result of the regression model.

Stimuli pairs *stolko.skolko*, *tomik.komik* do not contribute to the model, though the interaction of *stolko.skolko* pair and liquid condition is the statistically significant parameter. The remaining stimuli contribute to the model. The negative weight for these parameters means that unlike the value in the intercept (this is a pair *stol.skol*), the number of "different" answers decreases. It is interesting that monosyllable *stol.skol* and almost all double syllable (except *tonus.conus* that has a low coefficient of significance) words perceive more like identical words compared to monosyllable words which are perceived as ‘more different’ (Figure 14).

Based on these results, it can be assumed that the more syllables there are in words, the more likely it is that the subject recognizes them as the same words (and studying sounds as identical ones). Moreover, if before the stressed vowel there is a combination of fricative and explosive, the sounds of *t* and *k* are perceived less correctly compared to standard pronunciation.

These findings may be explained by the following assumptions. First, in *stolko.skolko* and *stol.skol* quality of spirant consonant may effect subsequent explosive consonant. Fricative sound is formed by significant narrowing in the speech tract, causing air flow and fricative noise. Due to the lack of space in the front part of the mouth, it is impossible to create a flow of air that would create a clear fricative sound. Thus, in a such consonant cluster explosive consonants will be perceived less accurate. Second, in isolated disyllabic words the second syllable is used for recognition more (Cole & Jakimik, 1980).

Type of hampered condition also contributes to the model. The graph (Figure 14) demonstrates this explicitly. In the soft condition (the third column), subjects with the highest probability estimate of a pair of stimuli as two words. It may be connected with a higher ability to make a stop in an explosive consonant with such food in the mouth that is easy to crush. The most unintelligible words were in a liquid condition; compared to soft and hard conditions there were the least responses ‘different’ and the most of ‘identical’ ones. This result is consistent with the assumption that in the studied conditions with food (both soft and solid) the speakers are able to create something similar to the stop in explosives by putting tongue in a certain position. In the liquid condition, it is not possible: by opening the mouth the contents may spill. Thus, sounds are pronounced less clearly and legibly.

A close up of text on a white background

Description automatically generated

Figure 14. Ratio of answers ‘different’.

The most significant parameter in the resulting model is the number of foreign languages learned. The negative weight for ‘foreign’ parameter (Figure 13) means that with the increase in the number of foreign languages studied, subjects respond that a couple of stimuli are two different words more often. These results suggest that the more languages a person knows, the better his phonetic hearing is developed.

# Results

As is stated in the introductory, the purpose of this study was to analyze the perception of individual words in different conditions when interlocutors eat. The present study is the pilot study in the field. The subject was minimal pairs with explosive consonants [*t*]and [*k*] (and one pair with sounds [*t͡s*] and [*k*]).

The main research question of the study concerns the perception of explosive consonants [*t*]and [*k*] (There was one minimal pair in the experimental lists for sounds [*t͡s*] and [*k*]: *tsaplya* (heron) and *kaplya* (drop)). The assumption was that to compensate lack of the ability to use the front part of the mouth speakers would utter some sounds from the velar zone, and participants would hear the identical words. This hypothesis was rejected. It means that words *tomik* (book, volume) and *komik* (comic) are not perceived as *komik* in most of studied conditions (with the mouth full). In the experiments responses ‘Different’ slightly outreached the ‘Identical’.

Another expectation was that in different conditions the sound perception would vary. It is partly true: the most difficult condition for sound/word recognition is the one with liquid (the flam of toothpaste was used), and the easiest to understand is the soft food condition (humus or sour cream was used).

Nonetheless, the findings indicate factors with higher influence on the answer that were obtained by Exact Fisher’s test. These are linguistic education, the number of syllables in the word, the number of foreign languages learned, type of challenging condition. These results suggest that linguists evaluated stimuli as different words more often. Also, there is a significant difference in the perception of one-syllable and two-syllable words with [*t*] and [*k*]in the modern literary Russian language. Monosyllable words are more frequently perceived as two separate ones. Despite the fact that subjects often evaluated pairs of stimuli as different words, they did not always correctly determine the experimental state. In general, the rate of guessing the condition type is higher than the random chance, but it is quite low to say that the participants always understood the type of hampered condition well. The easiest condition to determinate was with soft food; the most difficult one was with liquid.

Based on some of these factors (number of foreign languages learned, type of hampered condition, pairs of stimuli and combination of two previous parameters) the logistic regression model was constructed. The results were obtained by logistic regression model and may be compared with the output of Exact Fisher’s test for the listed parameters. The results of the regression model confirmed and specified data. As was said before, one-syllable words are more often perceived as different words. It may have an explanation that in short isolated words the stop is expressed more clearly.

Based on these results, it can be assumed that the more syllables there are in words, the more likely it is that the subject recognizes them as the same word (and studied sounds as identical ones). It may be related to suggestion that in two syllable words the second syllable is more important for recognition (Cole & Jakimik, 1980). One more interesting finding concerns stimuli pairs *stolko.skolko* and *stol.skol*. According to the data, in such consonant cluster explosive consonants will be perceived less accurately because of the influence of unclear spirant.

The type of hampered condition also contributes to the model. The graph (Figure 14) demonstrates this explicitly. In the soft condition (the third column) there are subjects with the highest probability estimate of a pair of stimuli as two words. It may be related to higher ability to make a stop in an explosive consonant with such food in the mouth that is easy to crush. The most unintelligible words were in a liquid condition; compared to soft and hard conditions, there was the least number of ‘different’ responses and the most of ‘identical’ ones. We assume that this is because of the inability to create at least a semblance of a stop in the studied condition.

The most significant parameter is the number of foreign languages learned. With the increase in the number of familiar foreign languages, the listener has a more sensitive hearing and is able to detect the sound more accurately.

# Conclusion and implications

To the best of our knowledge, this paper is the first one to explore speech perception while eating and with mouths full. In addition, this paper is one of the first to investigate the process of speech recognition in challenging condition with the difficulties for speech generation. The assumption was that velar and dental consonants are both perceived as the same velar sound.  This hypothesis was not confirmed. The most difficult condition for respondents was the liquid one, the simplest one was with soft food. This hypothesis was partly confirmed.

The current research identified the factors that have influence on subjects’ responses by applying the Exact Fisher’s test and constructing the logistic regression model. They are linguistic education, number of foreign languages, type of hampering condition, and number of syllables in the word. The output of logistic regression confirmed and refined the results.

The experiment was conducted on the basis of Russian language material, and it is possible to investigate the effect of listed factors on the material of other languages. The study sheds a light on side effects that have an impact on speech perception and in what way they influence the hearer.

Our results are of interest to professionals involved in the remediation of language-based learning deficits, which are often characterized by poor speech perception in noise. Furthermore, the findings can be taken into consideration as additional data for natural speech recognition systems.

# Acknowledgments

A sincere thank you to George Moroz for assistance in recording material and for allowing me to grow as a research scientist; Olga Vinogradova for stylistic remarks; Inna Ziber for valuable advices on the structure of the experiment. Also, I want to express gratitude to Antonina Sinelnik for proofreading the article and Petr Grinko for statistical assistance.

Thanks to students of National Research University Higher School of Economics Stepan Mixajlov, Ivan Netkachev, Konstantin Filatov, Aleksandr Orlov, Kirill Semenov for the agreement to be recorded in the experiment, and to all the respondents for participation.

# References

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural Timing Is Linked to Speech Perception in Noise. *Journal of Neuroscience*, *30*(14), 4922–4926. https://doi.org/10.1523/JNEUROSCI.0107-10.2010

Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *346*(1315), 29–36.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words? *The Journal of the Acoustical Society of America*, *67*(3), 965–970. https://doi.org/10.1121/1.383939

Florentine, M. (1985). Speech perception in noise by fluent, non‐native listeners. *The Journal of the Acoustical Society of America*, *77*(S1), S106–S106. https://doi.org/10.1121/1.2022152

Gries, S. T. (2013). *Statistics for Linguistics with R: A Practical Introduction*. Walter de Gruyter.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138.

i. (2019). *Contribute to astafyevai/Graduate-work development by creating an account on GitHub*. Retrieved from https://github.com/astafyevai/Graduate-work (Original work published 2018)

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, *45*(2), 139–152. https://doi.org/10.1016/j.specom.2003.10.005

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The Rôle of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. *The American Journal of Psychology*, *65*(4), 497–516. https://doi.org/10.2307/1418032

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *The Journal of the Acoustical Society of America*, *106*(1), 449–457.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007

Online Survey Software & Tools. (n.d.). Retrieved May 26, 2019, from SurveyGizmo website: https://www.surveygizmo.com/

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear and Hearing*, *30*(6), 653. https://doi.org/10.1097/AUD.0b013e3181b412e9

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America*, *118*(6), 3874–3888. https://doi.org/10.1121/1.2118407

Pisoni, D. (2005). The Handbook of Speech Perception (ed.), 723.

Strait, D. L., & Kraus, N. (2011). Can You Hear Me Now? Musical Training Shapes Functional Brain Networks for Selective Auditory Attention and Hearing Speech in Noise. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00113

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *The European Journal of Neuroscience*, *29*(3), 661–668. https://doi.org/10.1111/j.1460-9568.2009.06617.x

Team, R. C. (2016). *R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria*.

Wickham, H. (2018). *Tidyverse: Easily Install and Load the’Tidyverse’. 2017. R package version 1.2. 1*.

Wickham, Hadley. (2015). Stringr: Simple, consistent wrappers for common string operations. 2015. *R Package Version*, *1*(0).

Wickham, Hadley. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Wickham, Hadley, Francois, R., & Henry, L. (2018). *Müller, K. dplyr: A Grammar of Data Manipulation. R package version 0.7. 6*.

 Lyashevskaya, O. N., & Sharov, S. A. (2009). Frequency dictionary of the modern Russian language: on the materials of the National Corps of the Russian language. The alphabet 

# Appendices

### One-syllable stimuli:

|  |  |
| --- | --- |
| *tot* (that) | *kot* (cat) |
| *stol* (table) | *skol* (chipped) |
| *tort* (cake) | *kort* (court) |

### One-syllable fillers:

*Vkus* (taste), *lob* (forehead), *hvost* (tail), *bol’* (pain), *dozhd’* (rain), *krov’* (blood), *grob* (coffin).

### Two-syllable stimuli:

|  |  |
| --- | --- |
| *stolko* (so many/much) | *skolko* (how) |
| *tomik* (book, volume) | *komik* (comedian) |
| *tonus* (tone) | *konus* (kone) |
| *tsaplya* (heron) | *kaplya* (drop) |

### Two-syllable fillers:

*Bomba* (bomb), *zamok* (castle), *sluzhba* (service), *druzhba* (friendship), *holod* (cold), *hohot* (laughter), provod (wire), povod (occasion).

### Experimental lists.

#### First list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| sluzhba/druzhba (service/ friendship) | Solid food | 4 | 5 | no |
| tort/kort (cake/court) | Solid food | 5 | 2 | yes |
| dozhd`/krov` (rain/blood) | Solid food | 2 | 3 | no |
| hvost/bol`(tail/pain) | Solid food | 3 | 1 | no |
| tsaplya/kaplya (heron/drop) | Solid food | 1 | 4 | yes |
| grob/lob (coffin/forehead) | Soft food | 2 | 3 | no |
| tonus/konus (tone/kone) | Soft food | 3 | 1 | yes |
| holod/hohot (cold/laughter) | Soft food | 5 | 2 | no |
| stol`ko/skol`ko (so many, much/how) | Soft food | 1 | 4 | yes |
| vkus/lob (taste/forehead) | Soft food | 4 | 5 | no |
| provod/povod (wire/occasion) | Liquid | 4 | 5 | no |
| stol/skol (table/chipped) | Liquid | 3 | 2 | yes |
| tomik/komik (book, volume/comedian) | Liquid | 1 | 3 | yes |
| bomba/zamok (bomb/castle) | Liquid | 2 | 1 | no |
| tot/kot (that/cat) | Liquid | 5 | 4 | yes |

#### Second list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| dozhd`/krov` (rain/blood) | Soft food | 1 | 4 | no |
| stol/skol (table/chipped) | Soft food | 2 | 5 | yes |
| tomik/komik (book, volume/comedian) | Soft food | 3 | 2 | yes |
| provod/povod (wire/occasion) | Soft food | 4 | 3 | no |
| tort/kort (cake/court) | Soft food | 5 | 1 | yes |
| sluzhba/druzhba (service/ friendship) | Liquid | 5 | 1 | no |
| tot/kot (that/cat) | Liquid | 4 | 2 | yes |
| bomba/zamok (bomb/castle) | Liquid | 1 | 3 | no |
| tsaplya/kaplya (heron/drop) | Liquid | 2 | 4 | yes |
| holod/hohot (cold/laughter) | Liquid | 3 | 5 | no |
| vkus/lob (taste/forehead) | Solid food | 3 | 5 | no |
| tonus/konus (tone/kone) | Solid food | 4 | 1 | yes |
| grob/lob (coffin/forehead) | Solid food | 5 | 2 | no |
| stol`ko/skol`ko (so many, much/how) | Solid food | 1 | 5 | yes |
| hvost/bol`(tail/pain) | Solid food | 2 | 3 | no |

#### Third list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| provod/povod (wire/occasion) | Liquid | 4 | 5 | no |
| vkus/lob (taste/forehead) | Liquid | 5 | 3 | no |
| stol`ko/skol`ko (so many, much/how) | Liquid | 1 | 4 | yes |
| grob/lob (coffin/forehead) | Liquid | 2 | 1 | no |
| tort/kort (cake/court) | Liquid | 3 | 2 | yes |
| bomba/zamok (bomb/castle) | Solid food | 1 | 4 | no |
| stol/skol (table/chipped) | Solid food | 2 | 1 | yes |
| holod/hohot (cold/laughter) | Solid food | 5 | 3 | no |
| tomik/komik (book, volume/comedian) | Solid food | 3 | 2 | yes |
| tot/kot (that/cat) | Solid food | 4 | 5 | yes |
| hvost/bol`(tail/pain) | Soft food | 3 | 1 | no |
| tsaplya/kaplya (heron/drop) | Soft food | 2 | 5 | yes |
| sluzhba/druzhba (service/ friendship) | Soft food | 4 | 3 | no |
| tonus/konus (tone/kone) | Soft food | 1 | 4 | yes |
| dozhd`/krov` (rain/blood) | Soft food | 5 | 2 | no |

#### Forth list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| sluzhba/druzhba (service/ friendship) | Solid food | 4 | 5 | no |
| tort/kort (cake/court) | Solid food | 4 | 3 | yes |
| dozhd`/krov` (rain/blood) | Solid food | 2 | 3 | no |
| hvost/bol`(tail/pain) | Solid food | 5 | 1 | no |
| tsaplya/kaplya (heron/drop) | Solid food | 2 | 5 | yes |
| grob/lob (coffin/forehead) | Soft food | 2 | 3 | no |
| tonus/konus (tone/kone) | Soft food | 5 | 2 | yes |
| holod/hohot (cold/laughter) | Soft food | 3 | 4 | no |
| stol`ko/skol`ko (so many, much/how) | Soft food | 3 | 1 | yes |
| vkus/lob (taste/forehead) | Soft food | 4 | 5 | no |
| provod/povod (wire/occasion) | Liquid | 5 | 3 | no |
| stol/skol (table/chipped) | Liquid | 4 | 5 | yes |
| tomik/komik (book, volume/comedian) | Liquid | 3 | 2 | yes |
| bomba/zamok (bomb/castle) | Liquid | 1 | 4 | no |
| tot/kot (that/cat) | Liquid | 2 | 1 | yes |

#### Fifth list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| dozhd`/krov` (rain/blood) | Soft food | 2 | 1 | no |
| stol/skol (table/chipped) | Soft food | 4 | 2 | yes |
| tomik/komik (book, volume/comedian) | Soft food | 1 | 5 | yes |
| provod/povod (wire/occasion) | Soft food | 4 | 3 | no |
| tort/kort (cake/court) | Soft food | 3 | 4 | yes |
| sluzhba/druzhba (service/ friendship) | Liquid | 5 | 1 | no |
| tot/kot (that/cat) | Liquid | 1 | 4 | yes |
| bomba/zamok (bomb/castle) | Liquid | 1 | 3 | no |
| tsaplya/kaplya (heron/drop) | Liquid | 2 | 4 | yes |
| holod/hohot (cold/laughter) | Liquid | 5 | 3 | no |
| vkus/lob (taste/forehead) | Solid food | 1 | 5 | no |
| tonus/konus (tone/kone) | Solid food | 2 | 3 | yes |
| grob/lob (coffin/forehead) | Solid food | 5 | 2 | no |
| stol`ko/skol`ko (so many, much/how) | Solid food | 3 | 4 | yes |
| hvost/bol`(tail/pain) | Solid food | 2 | 3 | no |

#### Sixth list.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couples of words (translation) | Type of hampered condition | First speaker | Second speaker | Stimilus |
| provod/povod (wire/occasion) | Liquid | 4 | 5 | no |
| vkus/lob (taste/forehead) | Liquid | 5 | 3 | no |
| stol`ko/skol`ko (so many, much/how) | Liquid | 5 | 2 | yes |
| grob/lob (coffin/forehead) | Liquid | 2 | 1 | no |
| tort/kort (cake/court) | Liquid | 4 | 3 | yes |
| bomba/zamok (bomb/castle) | Solid food | 1 | 4 | no |
| stol/skol (table/chipped) | Solid food | 5 | 3 | yes |
| holod/hohot (cold/laughter) | Solid food | 5 | 3 | no |
| tomik/komik (book, volume/comedian) | Solid food | 1 | 5 | yes |
| tot/kot (that/cat) | Solid food | 2 | 4 | yes |
| hvost/bol`(tail/pain) | Soft food | 2 | 5 | no |
| tsaplya/kaplya (heron/drop) | Soft food | 3 | 1 | yes |
| sluzhba/druzhba (service/ friendship) | Soft food | 1 | 3 | no |
| tonus/konus (tone/kone) | Soft food | 4 | 2 | yes |
| dozhd`/krov` (rain/blood) | Soft food | 5 | 2 | no |

### The experiment consent forms

**Original text:** Этот веселый опрос займет у вас около 5 минут. Мы исследуем, как музыкальное образование и знание иностранных языков влияет на восприятие слов на родном языке. Заполняя эту форму, вы заявляете о добровольном участии в лингвистическом исследовании и соглашаетесь с тем, что данные, полученные в ходе исследования, могут быть опубликованы и использованы в научных и учебных целях при условии соблюдения их анонимности.

**Translation:** This fun poll will take about 5 minutes. We explore how musical education and knowledge of foreign languages affect the perception of words in their native language. By filling out this form, you declare your voluntary participation in a linguistic study and agree that the data obtained in the course of the research can be published and used for scientific and educational purposes, provided that their anonymity is observed.

1. Most of transliteration in current paper is made in accordance with the transliteration rules according to GOST using the site https://transliteration-online.ru/. [↑](#footnote-ref-2)
2. The consent forms for the experiment from the first page of the survey are in the Appendix 6 (in Russian and in English). [↑](#footnote-ref-3)
3. Missing values in one of the rows in the table (NA) do not have an essential effect on the model results. [↑](#footnote-ref-4)