Università
Ca'Foscari
Venezia

Master's Degree Programme

in Scienze del Linguaggio

Final Thesis

Using Animation to Raise Phonetic Awareness of English Vowels Contrast in Russian L1 Schoolchildren

Supervisor

Dr. Pavel Duryagin

Assistant supervisor

Gianluca Lebani

Graduand

Anastasiia Matros

Matriculation Number 999149

Academic Year

2025 / 2026

Acknowledgements

I would like to express my warmest gratitude to my supervisor, Professor Pavel Duryagin, for his motivating guidance, constructive feedback and support. His profound knowledge and invaluable experience were indispensable for the completion of this thesis. I am also sincerely thankful to my co-supervisor, Gianluca Lebani, for his perspicacious suggestions and inspiration. Finally, I am immensely grateful to all who participated in the survey and shared the results. Without their essential contribution, this research would not have been possible.

Table of contents

Abstract

This study examines how Russian-speaking elementary school students perceive long and short English vowels [iː], [ɪ], [uː], and [ʊ], with a focus on the phonological challenges resulting from the differences between the English and Russian vowel systems. In English, vowel length and quality are phonemically distinctive, whereas in Russian, vowel length is not a contrastive feature. This discrepancy often creates difficulties for Russian-speaking learners in the early stages of English language acquisition.

To solve this problem, the study presents a specially designed educational animated cartoon as the main pedagogical tool. The animation combines clear instructions, visual input, primary audio input (highlighting contrasting vowels in English speech) and secondary audio input (a metalinguistic device in the Russian-language voice-over) to improve students' phonemic awareness. The narrative format and interactive elements aim to facilitate both the perception and productive assimilation of English vowel length. At the same time, students' phonetic perception — their ability to hear and distinguish vowel contrasts.

The experimental part of the study includes auditory discrimination tasks using minimal pairs. These tasks are designed to assess students' ability to distinguish long and short vowels in varying phonetic contexts. Responses were analysed to identify common perceptual patterns and errors.

Given the lack of exact equivalents for [iː], [ɪ], [uː], and [ʊ] in Russian, the articulatory and acoustic differences between the systems, Russian-speaking learners may misidentify or merge English vowel contrasts. The study explores how visual and auditory multimodal input can moderate these difficulties and promote accurate vowel perception.

The proposed approach highlights the importance of incorporating engaging multimedia resources into early language instruction. It emphasises the value of combining listening and speaking practice within meaningful contexts to strengthen learners' phonological competence. This research contributes to second language phonetics and offers

practical recommendations for educators seeking innovative ways to improve pronunciation among Russian-speaking learners of English.

Introduction

Speech perception is a complex cognitive process that allows people to extract linguistic meaning from the audio signal of spoken speech. Although all humans are endowed with the biological ability to process auditory information, the way speech sounds are classified and interpreted is largely determined by the phonological structure of the listener's language. As a result, the perception of unfamiliar or non-native sounds is often filtered through previously learned phonemic categories, which leads to systematic errors or reduced sensitivity to linguistically significant differences in L2.

One of the documented phenomena in this field is interlanguage perception of speech, which shows how phonological representations of L1 affect the perception of L2 contrasts. Students may not notice subtle differences that are not phonemic in their native language, or they may assimilate unfamiliar sounds into the closest equivalent of L1, even if these sounds function differently in L2. This process is a serious challenge for mastering accurate pronunciation at the L2 level and listening comprehension.

An example is the contrast between long and short English vowels — [iː], [ɪ] and [uː], [ʊ]. In English, these contrasts are both phonemic and functionally significant; they distinguish the meanings of different lexical units. However, in Russian, the length of the vowels is not a determining factor. Although Russian vowels may vary in length depending on stress, these differences have no lexical significance. Consequently, Russian-speakers often do not realise the communicative significance of vowel length in English and may not notice or reproduce such contrasts in perception and pronunciation.

These observations, supported by both empirical research and classroom experience, led to the development of the present study. The purpose of this study is to investigate the perception problems faced by Russian-speaking elementary school students when distinguishing between long and short vowels in English.

To investigate this phenomenon, an experimental study was developed combining pre- and post-test perception, and delayed post-test for the second experiment tasks with an

original animated training video. The video was created to highlight the contrast between vowels using a combination of words spoken by a native speaker, visual metaphors, and symbolic representations of individual phonemes. The design was based on modern theories of multimodal learning and adapted to the cognitive and motivational needs of students aged 9-11. It is important to note that the video included interactive elements and embedded metalinguistic explanations at the L1 level of students to support phonetic awareness and enhance contrastive features.

As part of the study, which included two experiments, a quiz format was used to assess the accuracy of phonetic perception of participants both before and after watching the educational animation video. In the first experiment, tasks were performed using the interactive educational platform *Interacty*, whereas in the second experiment, similar tasks were performed as part of individual classes. In both cases, students were asked to identify sound stimuli representing minimal pairs of English.

In both experiments, quantitative data was collected to determine whether perceptual accuracy improved after the intervention and whether problems with certain pairs of vowels remained despite targeted training. In the second experiment, the stability of the obtained effect was additionally assessed using a delayed post-test.

The theoretical framework of the study is informed by two influential models of L2 phonetic acquisition: the speech learning model (SLM) (Flege, 1995), which posits that L1 and L2 categories share a common perceptual space, and the perceptual assimilation model (PAM) (Best and Tyler, 2007), which explains initial L2 perception in terms of how novel sounds are assimilated to L1 categories. Both models highlight the difficulty of forming new phonetic categories when L2 sounds are perceived as sufficiently similar to existing L1 sounds. They also underscore the need for rich, explicit, and varied input to support the development of accurate phonetic representations.

Thus, this study aims to contribute both to the theoretical understanding of the phonetic development of the L2 language and to the practical development of learning

strategies to improve the perception of vowel sounds in the early stages of second language acquisition.

Chapter 1 provides a comprehensive theoretical overview of phonetic perception of a L2, with a particular focus on how vowel contrasts are perceived by students whose native language lacks equivalent oppositions. The chapter begins with an overview of the fundamental concepts in the field of speech perception and the development of this field since the 1950s. Special attention is paid to the phenomenon of interlanguage perception of speech and the role of phonological transfer of L1 in the formation of students' sensitivity to L2 differences. Next, I will look at the specific difficulties faced by Russian-speaking students in distinguishing between long and short vowels in English, especially the pairs [iː]–[ɪ] and [uː]-[ʊ], which are not contrastive in Russian. At the end of the chapter, an overview of pedagogical approaches to teaching phonetics is given, and the potential of multimodal and animated learning tools in developing phonemic awareness among young students is emphasised.

Chapter 2 presents the methodological plan of the study. It describes the participants in both experiments: 26 Russian-speaking schoolchildren aged 9-11 years and 10 schoolchildren aged 10-11 years, as well as the rationale for choosing this particular age group, which combines cognitive readiness and auditory flexibility. This chapter describes the structure of two experiments. The first experiment consisted of three stages: pre-test, an animated training video, and post-test. The second experiment consisted of four stages: pre-test, an animated tutorial video, post-test, and delayed post-test. A detailed description is given for each stage, including the nature of the test tasks, the design of the video clip. The data collection process is also governed by ethical research standards, including the procedure for obtaining informed parental consent.

Chapter 3 presents the results of an empirical study based on an analysis of the results before and after testing of both experiments. Quantitative data are provided indicating an overall improvement in the accuracy of students' perception after the intervention. A

comparison of the error rate in vowel pairs shows that the contrast [iː]–[ɪ] showed the most significant gain, while the pair [uː]–[ʊ] remained more problematic.

The final chapter summarises the main findings of the study, reflecting both its theoretical contribution and its practical application. Based on the results of two experiments, it is shown that purposeful animated learning can contribute to the development of phonetic perception in younger schoolchildren, even in the presence of contrasts that are absent in their L1. A comparison of the data obtained within the framework of the group and individual experimental formats allows for a more comprehensive assessment of the effectiveness of the proposed approach. The results of the second experiment, which included a delayed post-test, indicate a tendency to maintain the achieved effect, but require further analysis to confirm its stability. Although this intervention produced promising results, the chapter also recognises a number of limitations, including the relatively small sample size, the short-term nature of testing. In addition, differences in the experimental conditions (group and individual format, use of different platforms) may affect the comparability of the results. Recommendations are given for future research, such as conducting longitudinal studies to assess memorability, testing on larger and more diverse populations, and expanding the use of animated tools for other phonetic differences. Despite its limitations, the study demonstrates the pedagogical value of integrating speech perception theory into educational practice and offers a basis for further study of perception learning in L2 language acquisition.

Chapter 1. Second Language Speech Perception

One of the tasks of linguistics, psycholinguistics and phonetics is to study how L2 sounds are perceived, processed and assimilated. The influence of L1 on the perception and articulation of non-native sounds remains a particularly important aspect in this context. The differences between L1 and L2 can significantly complicate the formation of new sound categories.

1.1. Theoretical models of L2 speech perception

In recent decades, several theoretical models have been developed that explain how L1 affects the process of studying the phonetic features of L2. The purpose of this chapter is to review these patterns and identify their significance in explaining the effects of age, differences in perception, and how new phonetic categories are created.

In the 1950s, researchers focused on comparing the phonological systems L1 and L2. The contrastive analysis (CA), proposed in 1957 by Robert Lado (Lado, 1957), explained the difficulties of learning a foreign language using the differences between L1 and L2. This model showed that the more differences there are in the L1 and L2 sound systems, the more difficult it is to learn a new language. However, the CA was based only on articulatory differences, without considering the perceptual processing factor, so the model needed to be improved (Grosjean, 1998).

In the 1970s, researchers began to look more at the perspective of phonetics. Most of the work focused on the analysis of the voice time (VOT) parameter in the perception and creation of initial words in English by native Spanish speakers. Elman, Diehl, and Buchwald investigated how people recognise syllables that begin with consonants in which the VOT was either ahead of time, with a short delay, or with a long delay. The Spanish-speaking and English-speaking participants perceived such sounds differently: the Spaniards called the sound with a short delay [p], and the English speakers called it [b]. Bilinguals speaking both languages were asked to rate the same sounds in two different languages. For most of them, language did not significantly affect perception, but five of the 31 participants who were

described as balanced bilinguals were much more likely to perceive a short delay as a [b] in an English situation. This suggested that bilingualism can affect perception even at the initial level of language proficiency (Elman, Diehl, and Buchwald, 1977).

In 1982, Flege and Hammond tested the hypothesis of Lenneberg that there is a certain period when a person can learn a language. Lenneberg believed that this period ends by the age of 13 due to the natural development of the brain. He assumed that it was more difficult for adults to learn L2 because they could not automatically learn a new language as children with L1 did (Lenneberg, 1967). In order to test this hypothesis, Flege and Hammond conducted an experiment. English-speaking students who had previously heard a Spanish accent were selected for the experiment. They were asked to read an English text with an imitation of a Spanish accent. The experiment revealed that students were able to change their pronunciation under the influence of a Spanish accent, even without being bilingual speakers. The findings demonstrate that adults can notice and memorise the sound features of other languages, even after the age of 13 (Flege and Hammond, 1982). Later, Flege conducted an experiment with adults again, but focused on learning English at different ages. The results showed that early learners demonstrated phonetic characteristics close to native speakers, while late learners had intermediate values (Flege, 1991). This partially supports Lenneberg's hypothesis that learning decreases after the age of 13, but it is not completely absent. Some still had indicators closer to English standards, which puts into question the hypothesis of a strict age restriction.

Then a study was conducted by Flege, Munro and Skelton to find out how the duration of the stay affects the pronunciation of the sounds [t] and [d] in adults. Native speakers of Chinese and Spanish took part in the experiment. All participants started learning English as adults, but some of them have lived in the United States for less than a year, while others have been for several years (up to 9 years). These sounds were chosen because they do not occur at the end of words in either Chinese or Spanish, and, in theory, should have been perceived as "new" — and, therefore, easier to learn. After analysing the results, it showed that participants who had lived in the United States longer still often made mistakes when pronouncing sounds. Two primary explanations have been proposed for this phenomenon: it is difficult for

adults to form new speech patterns; participants do not surround themselves with correct pronunciation, even native-speaking children typically require several years to master the accurate production of [t] and [d] (Flege, Munro and Skelton, 1992).

Over time, researchers shifted their focus from phonemics to phonetics:

1) L1 affects the perception and pronunciation of L2 sounds;

2) Some L2 sounds are easier to learn than others;

3) Sounds without an analog in L1 are sometimes better absorbed;

4) The greater the amount of sound experience with the L2, the higher the likelihood of successful sound acquisition (Celce-Murcia, 2010).

By the 1990s, there was a need to create a more empirically based model of L2 perception that would take into account both articulatory and perceptual aspects, as well as the variability of phonetic systems in individual experience.

1.1.1. The Speech Learning Model

The speech learning model was developed by James Flege in 1995. The main goal was to explain how adults produce L2 sounds against the background of already formed L1 sound categories. SLM proves that adults retain the ability to perceive and even form new phonetic categories under certain conditions. SLM was one of the first models to offer a holistic approach to the study of L2 phonetics, combining perceptual, articulatory and cognitive aspects.

The SLM includes a number of fundamental provisions:

1) The assimilation of L2 sounds occurs in the perceptual system already formed by L1. That is, the sounds of the second language are perceived as variations or analogues of the already familiar sounds of the native language.

2) A new category can only be created if the L2 sound differs from the closest L1 sound so much that the listener can recognise this difference.

3) The formation of new categories is possible at any age, but the probability of this decreases with age, especially if stable L1 categories already exist.

4) The categories are formed on the basis of statistical distributions of sounds in the speech stream, which requires a significant amount and variety of phonetic input.

5) Perception and sound production are closely related: articulation relies on perception, and vice versa.

The model includes three interrelated levels: sensorimotor, phonetic, and lexico-phonological. The sensorimotor level is responsible for the primary analysis of the physical characteristics of the speech signal, such as the duration of the VOT, the position of the formants, and the presence of vocal vibration. The phonetic level groups sounds by perceptual similarity into categories that set articulatory goals and are involved in recognising sounds when understanding speech. The lexico-phonological level activates vocabulary units, where phonetic segments are recognised as part of specific words (Flege, 1995).

When adults start learning L2, they do not just memorise new sounds — they first compare them with already known L1 sounds. This process is called interlanguage identification. This process is typically automatic and occurs below the level of conscious awareness. If the sound of L2 is very similar to the sound of L1, the brain maps it onto an existing L1 category instead of forming a new one. For example, a Russian-speaking student may perceive [ɪ] and [i:] as one sound, since there is only one similar sound in Russian [i].

However, if the differences between the sounds of L1 and L2 are large enough and noticeable to students, a new phonetic category may form. This formation can occur gradually as learners gain experience in L2. According to SLM, the formation of a new category requires a significant difference between L2 and the closest L1 category, a sufficient amount of phonetic input with high frequency and a variety of use situations, as well as a stable perception between sounds.

Research on L2 has shown that it is important to distinguish between two types of tasks — identification and categorisation of sounds. In an experiment conducted by Bohn and Flege**,** Spanish-speaking participants had to determine which sound they were hearing: [d] or [t]. Although there is no English sound [t] with a long delay in their native language, they still chose it more often if the delay was pronounced. This was not because they had their own category for [t], but because they used an exclusion strategy: if the sound did not resemble their familiar [d], then it was [t] (Bohn and Flege, 1993).

SLM claims that L2 can affect L1. When students use L2 frequently, especially in communication, some L1 sounds may begin to change under the influence of new speech habits. Thus, in some cases, the opposite effect occurs — not only does L1 interfere with the study of L2, but L2 also affects the native system. If the L2 sound is very different from the L1 sound, then a hybrid or composite category may develop that combines the features of the L1 and L2 sounds. At the same time, the perceived connection between the new sound and the already familiar one from L1 remains, and a new, intermediate category arises based on the combined characteristics of both systems. The model predicts that when such a mixed category is formed, changes may occur in the L1 category itself — it may gradually shift towards L2. How much this shift will occur and whether it will be noticeable for L1 speakers depends on how the sound features of both language systems are distributed. In addition, SLM claims that over time, L2 students form the same clear phonetic categories for new sounds as in L1. These categories include not only the perception of sound, but also instructions for pronouncing it. Over time, the pronunciation of sounds in the second language should approach the ideal, that is, correspond to the internal phonetic image. However, the SLM does not specify how long it takes for perception and pronunciation to fully align.

1.1.2. The Revised Speech Learning Model

The revised speech learning model (SLM-r) was initially proposed by Flege in 2005, but it was not formally and comprehensively presented until 2021 (Flege and Bohn, 2021)**.** The new model retains the central ideas of SLM, but includes clarifications and new

provisions. Like the original SLM, SLM-r considers the situation when a person begins to study L2 with a full L1 level.

The main prerequisites of this model are:

1) Phonetic categories are formed based on the distribution of input data.

2) The mechanisms of learning L2 do not differ from those used in mastering L1 (children and adults use the same processes).

3) Differences between native speakers and L2 learners arise not because of a loss of the ability to learn, but because already formed L1 categories interfere with the formation of new categories for L2 sounds.

The SLM-r does not emphasise the distinction between "early" and "late" learners. Studies after 1995 have shown that the hypothesis of the critical period (CP), proposed by Lenneberg in 1967, does not reliably explain the age differences in the study of L2. Neuroplasticity has been shown to persist into adulthood, especially in speech-related areas (Zhang, Wang, 2007). A foreign accent is observed not only in adults, but also in those who started learning the language before the age of 13. The quality and volume of language experience is more important than age. If a student hears little L2 or hears it with an accent, he will not reach the native pronunciation level. Adults can also distinguish and reproduce the subtle sound features of L2, even without formal training, simply based on language experience (de Leeuw, Celata, 2019). Thus, SLM-r departs from the idea that after a certain age a person loses the ability to learn the sounds of a second language. Instead, the model explains the difficulties in learning L2 by how exactly L2 sounds are processed through the already existing L1 system.

SLM-r no longer aims to achieve native L2 proficiency. On the contrary, it recognises that students will almost never be able to reproduce and perceive the sounds of a second language in the same way as adult speakers. The SLM model put forward the idea that there is a certain limit to the accuracy in the perception and pronunciation of L2 sounds. This hypothesis has received support in several ways:

- Infants first begin to distinguish sounds by ear, and only later learn to reproduce them correctly (Kuhl, 2000). Errors in speech may be due to the fact that children do not recognise the difference between a correct sound and an erroneous one (Eilers, Oller, 1976).

- Accented L2 speakers are able to assess the strength of others' accents as accurately as L1 speakers (MacKay, Flege, Imai, 2006).

- Perceptual training can improve pronunciation even without special articulation training (Lengeris, Hazan, 2010).

The original SLM estimated the accumulation of phonetic knowledge by the number of years spent in the L2 environment. However, this turned out to be an unreliable indicator: the length of stay does not reflect either the actual amount of language contact or its quality. SLM-r is reviewing this approach. SLM-r defines phonetic input as a sensory experience associated with the perception of L2 sounds in real, meaningful communications, both heard and visually observed. SLM-r also maintained the assumption that L2 sounds automatically correlate with existing L1 categories, and the greater the difference between L2 and L1, the more likely the perceptual system is to initiate the formation of a distinct phonetic category (Flege and Bohn, 2021).

As mentioned above, SLM-r refuses to rely on the age of the beginning of the study of L2 as the main factor. SLM-r suggests replacing the age hypothesis with a new one, the hypothesis of the accuracy of phonetic categories L1. This hypothesis is based on the fact that the more precisely and stably the L1 categories are formed by the time of the first contact with L2, the easier it is to distinguish the L2 sound and form a new category for it.

The accuracy of categories is measured through the variability of acoustic parameters when the same sound is repeated. The less variability there is, the more accurate the category is. Auditory sensitivity, early auditory signal processing, and auditory memory can also affect accuracy. Kartushina found that naive listeners pronounced vowels much more accurately after learning than before. The most important thing was that the training did not affect the

accuracy with which the native French participants pronounced their native French vowels (Kartushina, 2016). This confirms that accuracy in the L1 category does not rely on language-specific phonetic inventory.

The SLM-r model assumes that the ability to form phonetic categories persists throughout life. However, new categories are not formed for all L2 sounds, even if they differ significantly from the nearest L1 sounds. New categories are formed based on statistical patterns revealed in the distribution of input data. Since this learning is gradual, SLM-r argues that the development of L2 categories is as slow a process as it was in childhood when learning L1.

The new phonetic category L2 is created in three stages:

1) The student of L2 must distinguish the phonetic difference between the realisations of the sound L2 and L1 of the sound that is closest to it in the phonetic space;

2) There should be a functional "equivalence class" of speech lexemes that resemble each other and are close to each other in phonetic space (Kuhl, 2000);

3) Moreover, the quantitative relationship between the L2 equivalence class and the L1 category remains poorly defined to this day. This separation can be accelerated by arranging the lexicon in the L2 language (Bundgaard-Nielsen, Best, Tyler, 2011).

As soon as the separation takes place, the development of a new L2 phonetic category will be based on statistical patterns of the distribution of L2 sounds, which are implicitly classified as instances of the new L2 phonetic category. In SLM-r, the formation of a new category is considered as a gradual process. Consider the results of the work of Thorin, Sadakata, Desain, and McQueen, who taught native Dutch students to pronounce and perceive the English vowels [æ] and [ə]. The result of the training was a slightly greater improvement in the pronunciation of English [æ] than [ə], which is consistent with the fact that of the two English vowels [æ] is located at a greater distance in the interval F1–F2 from the nearest Dutch vowel [ə]. This indicates the existence of incipient phonetic categories for

English [æ] before learning, which became "stronger" as a result of learning (Thorin, Sadakata, Desain, and McQueen, 2018). The SLM-r suggests that individual students should not form new phonetic categories for the L2 sound, which, in their opinion, is too phonetically similar to the closest L1 sound. However, it is important that students do not discard sound phonetic information in such cases. According to the hypothesis, the perceptual relationship between the sound L2 and the nearest sound L1 will continue to exist, and a composite phonetic category L1–L2 will be formed, determined by statistical patterns present in the combined distribution of perceptually related sounds L1 and L2. If the sound of L2 is perceived as too similar to L1, the new category may not appear. But this does not mean that the sound is completely ignored: in such cases, a composite category L1–L2 is created based on the general statistics of perceived sounds. McKay's research in 2001 confirmed the presence of such assimilation effects in the perception of consonants [b] in English and Italian.

The data also shows that learning L2 can change how a person perceives sounds in L1. Dmitrieva showed this using the example of Russian-English bilinguals. When perceiving final consonants, Russian monolinguals relied more on the pulsation of the glottis, while English speakers relied more on the duration of the preceding vowel. Russian bilinguals, who had lived in the United States for an average of 39 years, demonstrated perceptual patterns comparable to native English speakers when completing the English task, while they used an unusual strategy for Russians, attaching greater importance to the length of the vowel (Dmitrieva, 2019). This suggests that mastering L2 not only adds new sounds, but also changes the perception of L1 sounds, especially when both language systems interact in the same phonetic space. Consequently, SLM-r shows how a person's sound system can change throughout their life under the influence of the new speech they hear when learning a second language in a natural way.

1.1.3.The Perceptual Assimilation Model

The perceptual assimilation model (PAM) proposed by Best in 1994 was developed to explain how speech perception is shaped by the phonological system of the L1. The model

covers both the early period of development — infants' attunement to L1 sounds, and the manifestations of this attunement in adult perception (Antoniou, Best and Tyler, 2013). PAM makes predictions about how accurately L1 speakers will be able to distinguish pairs of non-native sounds based on how these sounds assimilate to the phonological categories of L1.

Researches based on the PAM model as a theoretical basis have identified six different ways of assimilating phonetic contrasts:

1) Two-Category (TC): Both sounds are perceived as representatives of different L1 categories. The distinction should be clear, similar to native phonological contrasts.

2) Single-Category (SC): Both sounds are perceived as equally good or bad representatives of the same L1 category. Discrimination is expected to be weak, although above the random level.

3) Category-Goodness (CG): Both sounds belong to the same L1 category, but one is perceived as more typical than the other. The accuracy of the distinction will be intermediate between TC and SC.

4) Uncategorised-Categorised (UC): One sound corresponds to the L1 category, the other does not. The difference depends on the degree of overlap: UC-N (non-overlapping): accurate discrimination, UC-P (partially matched): average accuracy, UC-C (completely matched): a slight difference.

5) Uncategorised-Uncategorised (UU): No sound falls into the existing L1 categories. The difference depends on the degree of overlap: UU-N > UU-P > UU-C.

6) Non-assimilable (NA): Sounds are perceived as non-verbal. The discrimination can be high if the acoustic differences are noticeable.

PAM assumes that listeners focus primarily on L1 phonological information, if it is available. Even if the acoustic distance between the sounds is large, it can be ignored if it does not correspond to the existing L1 categories. Only in the absence of phonological

reference points does attention shift to acoustic or articulatory differences (Idemaru, Holt, 2011).

Research in support of PAM evaluates the perceptual assimilation of individual non-native phonemes into the native phonological system using a categorisation task. This was initially done in the form of a post-test questionnaire in which participants provided descriptions of non-native phonemes using their native spelling (Best, McRoberts, Goodell, 2001). Later studies used the forced choice task, in which participants first chose the name of their native phonological category (which may include allophone variants), and then assessed compliance with this category on the Likert scale (Faris, Best, Tyler, 2016). If any label is selected above the threshold level, it is considered to correspond to the native phonological category, otherwise it is not assigned to the category. The types of contrast assimilation are determined by comparing the individual types of assimilation for each non-native phoneme. Quality ratings are taken into account only if both non-native phonemes belong to the same phonological category L1. In these cases, if there is a significant difference according to the same L1 category, then it is an assimilation by category, otherwise it is an assimilation by one category.

The model showed that even without prior training, it is possible to predict which contrasts will be difficult for L2 learners based on the configuration of their native language. PAM emphasises that the perception of non-native sounds is determined by how they "fit" into the already existing L1 system. This perception may limit sensitivity to new sound contrasts. However, research shows that L2 learners can learn to switch attention from native phonology to the phonetic features of L2 — especially if learning takes into account the functional significance of the features and uses methods that help refocus perception.

The PAM model primarily addresses the L1 phonological system and shows how L2 sounds are perceived through its prism. PAM offers a predictive framework for how unfamiliar phonemes are mapped onto native phonological categories. For instance, when two L2 sounds are assimilated into a single L1 category, perceptual discrimination becomes limited — a phenomenon classified by PAM as "Single Category" assimilation. PAM works

well as a diagnostic tool: it enables prediction of which phonemic contrasts are likely to present perceptual difficulty for L2 learners, depending on which language they speak from birth. In contrast, the SLM and SLM-r models consider the process of phonetic development in dynamics. They describe how new stable phonetic categories can be formed, even in adults, under the influence of input data, articulation, and cognitive processing.

The SLM, SLM-r, and PAM models represent three different but complementary perspectives on how the sounds of a L2 are perceived and assimilated. Together, they form a solid theoretical basis for understanding how pronunciation develops among language learners and what factors — such as age, mother tongue, language experience, and individual characteristics.

The speech learning model explains how the sounds of a second language are perceived and articulated against the background of already formed L1 categories. Its main strength lies in describing the mechanism of formation of new phonetic categories in adulthood and the relationship between perception and articulation. The model is widely used in studies on changes in the pronunciation of immigrants in a second-language environment (Flege, 1995).

The updated SLM-r model develops the ideas of SLM, focusing on the quality of phonetic input and the accuracy of already formed L1 categories. Unlike the original model, SLM-r departs from the strict age limit, emphasising that neuroplasticity persists into adulthood, and success in studying phonetics largely depends on the richness and representativeness of language experience.

The perceptual assimilation model is used in the early stages of perception and makes predictions about the difficulties of perceiving the sounds of a second language before learning.

Nevertheless, taken together, these models provide a solid foundation for analysing how L2 sounds are assimilated. In the next chapter, I will compare the English and Russian

vowel systems, focusing on the analysis of the two systems and the concept of duration in both languages.

1.2. Comparison of the phonetic vowel systems of English and Russian

The phonetic perception models (SLM, SLM-r, PAM) were discussed above, which explain why some L2 sounds are easier to learn, while others are not. This section will describe the differences between the phonetic systems of English and Russian, with an emphasis on vowel sounds.

English belongs to the Germanic language family, while Russian belongs to the Slavic language family. The differences in their phonetic systems are due to the historical and typological features of the development of each language. English has a complex system of vowels, diphthongs, and length contrasts. The Russian vowel system is characterised by strong positional variations, especially vowel shortening in unstressed syllables, which significantly affects the quality and duration of vowel pronunciation. Such systemic disparities necessitate the reorganisation of articulatory and perceptual strategies among L2 learners. Phonology is not just a list of sounds, but a structured system that controls how sounds are produced, perceived, and combined (Ladefoged and Johnson, 2011).

1.2.1. Russian vowel phonemes

In this thesis, the Russian vowel system is /i, e, a, o, u/. The status of /ɨ/ is still debated. While some researchers treat /i/ and /ɨ/ as separate phonemes, others, following Ruben Avanesov (1956), consider /ɨ/ to be an allophonic variant of /i/ occurring after hard consonants. This interpretation is adopted in the present study and is supported by their complementary distribution, with [ɨ] appearing after hard consonants and [i] after soft ones.

The vowel sounds of the Russian language are classified according to articulatory features reflecting the position and activity of the speech organs in the process of their formation (Figure 1).

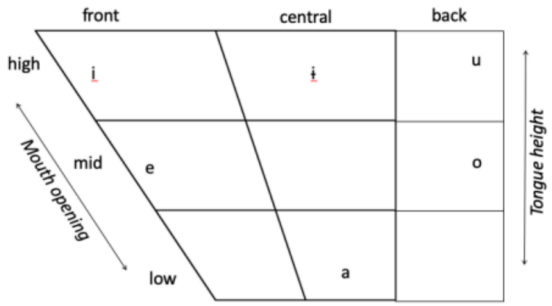

Figure 1. *A vowel chart showing the relative vowel qualities represented by some of the symbols used in transcribing Russian. (A. Letova, 2022)*

The classification is based on three main parameters:

1. The degree of elevation of the tongue.

This parameter reflects the vertical position of the tongue in the oral cavity during the pronunciation of a vowel. Depending on how high the tongue rises, there are:

- High vowels: the tongue is as close to the palate as possible: [i], [ɨ], [u];

- Vowels of the middle rise: the tongue occupies an average position: [e], [o];

- Vowels of the lower ascent: the tongue descends down into the oral cavity: [a].

2. Tongue lifting point (row).

The lifting point is the horizontal position of the tongue, the active zone of its movement. In Russian, the following are distinguished:

- Front row vowels: the front part of the tongue is active: [i], [e];

- Central vowels: the middle part of the tongue involved: [ɨ], [a];

- Back row vowels: the back of the tongue is active, the tongue is pulled back: [u], [o].

3. Labialisation.

Labialisation is the involvement of the lips in the articulation of vowels. At the same time, the lips take on a rounded shape and can "stick out" forward, which is especially noticeable when pronouncing sounds [u] and [o]. Russian vowels may be classified as:

- Labialised: lips are rounded: [u], [o];
- Non-labialised: lips do not take on a rounded shape: [i], [ɨ], [e], [a].

The vowels of the Russian language differ not only in articulatory characteristics, but also in the degree of tension, which is directly related to the presence or absence of stress. Stressed vowels are usually pronounced more clearly and with greater articulatory energy. In an unstressed position, this tension decreases, articulation becomes less precise, which leads to the phenomenon of reduction (Knyazev & Pozharitskaya, 2011). In other words, reduction can be considered as a result of a weakening of articulation.

1.2.2. English vowel phonemes

English has 20 vowel phonemes in its Received Pronunciation (RP) variety: 12 monophthongs and 8 diphthongs. Classification of English vowels:

1) By articulatory stability: monophthong or diphthong.

A monophthong is a vowel during the articulation of which the speech organs remain in a fixed position until the end of the sound: [iː], [ɪ], [e], [æ], [ɑː], [ʌ], [ɒ], [ɔː], [ʊ], [uː], [ə], [ɜː].

A diphthong is a combination of two vowel elements pronounced as a single whole within a single syllable: [eɪ], [aɪ], [ɔɪ], [əʊ], [aʊ], [ɪə], [eə], [ʊə]. When pronouncing a diphthong, the organs of speech first have the position of the first vowel, and then smoothly transition to the second vowel.

2) By tongue position: vertical or horizontal.

According to the position of the tongue in the vertical plane (tongue height), three groups are distinguished:

- High vowels: [iː], [ɪ], [ʊ], [uː].

- Mid vowels: [ə], [ɜː], [ɔː], [e].

- Low vowels: [æ], [ɑː], [ʌ], [ɒ].

When classifying vowels by the position of the tongue in the horizontal plane (the degree of frontness or backness), the following groups are distinguished:

- Front vowels: [iː], [e], [æ].

- Retracted front vowels: [ɪ].

- Central vowels: [ɜː], [ə], [ʌ].

- Advanced back vowels: [ʊ], [ɑː].

- Back vowels: [uː], [ɔː], [ɒ].

3) By lip position: rounded or unrounded.

- Rounded: [ɒ], [uː], [ɔː], [ʊ].

- Unrounded: [iː], [ɪ], [e], [æ], [ɑː], [ʌ], [ə], [ɜː].

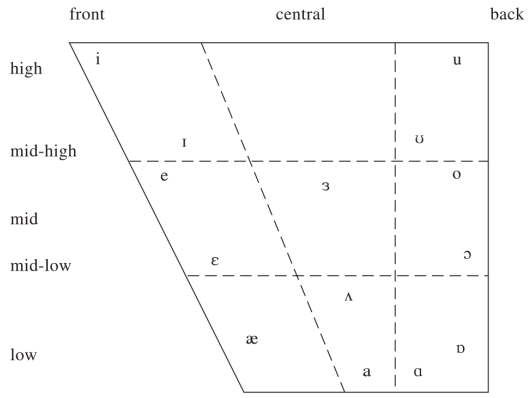

Figure 2. *A vowel chart showing the relative vowel qualities represented by some of the symbols used in transcribing English (Ladefoged and Johnson, 2011).*

The vowel articulation diagram (Figure 2) illustrates two main articulation parameters: the vertical position of the tongue and the horizontal position.

1.2.3. A comparative analysis of Russian and English vowel phonemes

There are about 20 vowel phonemes in English, while there are only five in Russian. This quantitative difference reflects the qualitative features of the two systems. The Russian vowel system does not include retracted front vowels or advanced back vowels — categories that are actively used in English. Furthermore, in English there are phonologically significant contrasts in the duration and degree of muscular tension during articulation. Such oppositions are not typical for Russian: all vowels are pronounced without marked tenseness. As for the parameter of labialisation, both language systems distinguish between rounded and unrounded vowels. However, the articulatory realisation of this feature differs: in English, rounding is typically achieved by shaping the lips into a rounded form without significant protrusion, whereas in Russian, by contrast, the lips are noticeably protruded forward when producing the sounds [o] and [u].

To compare the two systems, it is necessary to understand what reduction is. Reduction explains the process of sound attenuation in speech: we pronounce it less clearly, with less effort, which can shorten its duration and slightly change its quality. This happens not only because of neighboring sounds that affect each other, but also simply because in the flow of speech we do not control articulation equally strongly. This is especially noticeable in unstressed positions. In Russian, it plays a very important role: vowels almost always change depending on whether they are stressed or not. On the one hand, the vowel simply becomes shorter — this is a quantitative reduction. In such cases, the sound itself is generally preserved, but it sounds less clear. On the other hand, it can change more strongly and already differ in quality, approaching a different sound or a more "neutral" version. At the same time, quantitative and qualitative reduction cannot be considered as two completely separate processes. In practice, they are closely related: when a vowel is shortened in length, it often leads to a change in its quality (Knyazev & Pozharitskaya, 2011). The less time is devoted to

pronouncing a sound, the more difficult it is to maintain its "perfect" articulation. As a result, the vowel begins to deviate from its original version.

In English, reduction also exists, but it manifests itself somewhat differently and overall plays a smaller role. Most often, it is connected with the fact that in unstressed syllables vowels become weaker and shift to the neutral sound [ə] (schwa) (Roach, 2009). At the same time, such a complex and step-by-step system of changes as in Russian does not exist in English. This suggests that in Russian the description of vowels is largely built around reduction, whereas in English it is built around the differences between the phonemes themselves. This is important to take into account when it comes to length.

In different languages, duration may depend on different circumstances: in some cases it depends on the position of the sound in the word and the conditions of pronunciation, and in others it is included in the system of oppositions and helps to distinguish words. This difference is especially important when comparing Russian and English vowel systems.

In English, the length of a vowel cannot be reduced to just how long it sounds. Usually, along with the duration, other parameters change, such as sound quality, tension, and the position of the tongue. That is, long and short vowels differ not only in time, but also in how they are generally pronounced (Ladefoged & Johnson, 2011). Describing English vowels only through a row and a rise is not enough, because the system itself is more complicated and includes more contrasts. As a result, this suggests that duration in English is not just a consequence of the tempo of speech or position in a word, it is related to the language system itself and participates in the distinction of sounds (Giegerich, 1992).

In Russian, vowel length does not serve a phonemic function. In other words, there is no contrast comparable to that found in English, where vowel sounds may be distinguished by their length. When vowels occur in the same position with respect to stress, they are not contrasted as long and short. Only reduced vowels (often represented in Russian phonetic notation as [ъ] and [ь]) can be considered truly short, while full vowels cannot be classified as either long or short. Therefore, in Russian, vowel length is better understood as a feature of phonetic realisation rather than a property of the phonological system. It depends on stress

and other contextual factors rather than on the identity of the phoneme itself (Knyazev & Pozharitskaya, 2011).

This means that in Russian, vowel length primarily depends on its position within the word. Stressed vowels are usually the longest, as they are pronounced more clearly and with greater articulatory effort. The first pre-stressed syllable is already shorter, while the remaining unstressed syllables undergo even stronger reduction. Evidence for this can be found in the work of P.Duryagin, who demonstrated that vowel duration varies depending on stress: a stressed [а] lasts about 106 ms on average, the first pre-stressed [ɐ] about 60 ms, and a more reduced vowel, often represented as [ъ] in Russian phonetic notation, about 37 ms (Duryagin, 2018). In other words, this pattern is more systematic than random: the further away a vowel is from the stressed syllable, the shorter it becomes and the more noticeable the reduction.

In English, vowel length may also vary depending on factors such as stress or the surrounding sounds. However, this does not mean that length loses its functional role in the language system. For example, vowels are typically pronounced longer before voiced consonants and shorter before voiceless ones: thus, the vowel in *bad* is longer than in *bat* (Cruttenden, 2014). Such variations do not render vowel length random or entirely position-dependent; rather, they operate alongside an already established system of phonemic contrasts. Therefore, in English, vowel length cannot be explained solely by positional or contextual factors, as it remains an integral part of the phonological system.

Another important point to consider when comparing Russian with English is that Russian does not have diphthongs in the strict sense. As noted by Knyazev and Pozharitskaya (2011), individual vowels may exhibit some internal variation and may occasionally resemble diphthongs, but this does not imply the existence of a separate diphthongal system in the language. This is especially noticeable when pronouncing soft sounds, when a slight slip may appear in the vowel, similar to [i]. In English, however, the situation is different: diphthongs form an independent part of the vowel system. The internal changes within a vowel are

inherent in the sound itself, and are not primarily caused by positional factors, as is most often the case in Russian.

Vowel length functions differently in Russian and English. In Russian, it does not serve a phonemic contrastive function; rather, it reflects the way a sound is realised in a particular position. It is influenced by stress, reduction, and the position within the word. In English, by contrast, vowel length forms part of the system of distinctions between sounds and is closely related to vowel quality and tenseness. In other words, the key issue is not whether vowel length exists, but what role it plays in the language. This constitutes one of the fundamental differences between the two systems.

The typological differences between the English and Russian vowel systems discussed above are one of the sources of pronunciation difficulties for Russian-speaking learners of English. As Vasiliev indicates, pronunciation errors can be divided into two main types:

1. Phonemic errors — one phoneme is replaced by another. Such errors often affect the meaning of the utterance and hinder comprehension. Examples include confusion between short and long vowels in pairs like *sit* [sɪt] vs. *seat* [si:t], or *pot* [pɒt] vs. *port* [pɔ:t]. These substitutions may significantly obscure intended semantic content and can seriously prevent communication.

2. Non-phonemic errors — substitutions of one allophone with another within the same phoneme, or the use of sounds that are subjectively perceived as similar but do not exist in the target language system. Although such errors do not affect the meaning of the message, they make the speaker's pronunciation noticeably accented (Vasiliev, 1980).

The differences between the phonological systems of the two languages are not only of theoretical significance but also have a direct impact on the pronunciation aspect of acquiring English as a foreign language.

The most challenging vowel pairs for Russian-speaking learners of English are: [i:]–[ɪ], [u:]–[ʊ], [ɛ:]–[æ], and [ɑ:]–[ʌ]. These predictions are supported by empirical research. In the study by Kondaurova and Francis, it was demonstrated that Russian-speaking participants

experience significant difficulty in distinguishing the contrastive vowel pair [iː]–[ɪ]. In the experiment, participants were asked to listen to isolated word pairs and determine whether the words were the same or different. The results revealed a low level of accuracy in recognising distinctions between these sounds, especially when they appeared in unstressed positions or were presented with phonetic noise. This result points to insufficient perceptual discrimination of tense-lax contrasts, which plays a crucial role in the English vowel system but lacks phonemic status in Russian (Kondaurova and Francis, 2008).

A similar experiment was conducted by Makarova, which investigated both the perception and production of vowel contrasts that are particularly challenging for Russian-speaking learners —[iː]–[ɪ], [uː]–[ʊ], and [ɛː]–[æ]. In the study, students completed both identification tasks and production tasks, with their pronunciations recorded and analysed using the acoustic software Praat. The results showed that participants often failed to distinguish [uː] and [ʊ] by ear, confused [ɛː] and [æ] in production, and inconsistently produced the [iː]–[ɪ] contrast. These findings point to both insufficient perceptual differentiation and difficulty in forming accurate articulatory patterns, highlighting the dual challenge of perception and motor control in mastering English vowel distinctions (Makarova, 2010).

Both studies confirm that the difference between English vowels, especially in contexts where Russian does not offer similar oppositions, requires special efforts on the part of students and can become one of the main causes of phonemic errors in their pronunciation.

1.2.4. Articulatory difficulties in the pronunciation of English vowels [iː]–[ɪ] and [uː]–[ʊ]

Russian-speaking students often find it difficult to produce and distinguish certain English vowels because these sounds do not exist in their native language. To understand why these problems occur, it is helpful to take a closer look at some of challenging vowel pairs, such as [iː]–[ɪ] and [uː]–[ʊ], and explore what makes them hard to pronounce correctly.

At the beginning of the articulation of the English vowel [iː], the tongue position is less close and less front than that of the Russian [i]; however, by the end of the sound, it approaches the characteristics of [i], becoming nearly as close and fronted. This articulatory dynamic explains why Russian-speaking learners of English often substitute the English [iː] with the more familiar and stable [i] from their native system.

To correct this interference error, Vasiliev et al. (1980) recommends modifying the articulation of [iː] by keeping the front part of the tongue slightly lower than when producing the Russian [i], and by giving the vowel greater length and tenseness. The initial portion of the English [iː] should have a slight shade of the Russian [ɨ], which helps achieve a more accurate articulation. If the learner replaces [iː] with a clear [ɨ] or with the short English [ɪ] (Figure 3), they should be instructed to prolong the sound and bring its ending closer to the Russian [i].

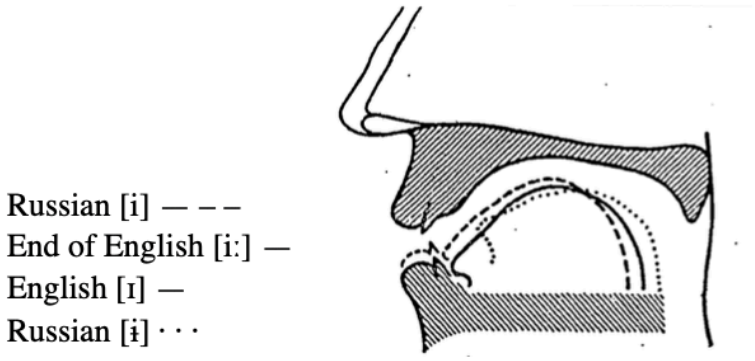

Figure 3. *Tongue positions for Russian [i] and [ɨ], English [ɪ], and the final position of English [iː] (Shevyakova, 1968).*

It is important to take into account that difficulties with the [iː]–[ɪ] pair are connected not only with articulation, but also with the perception of these sounds. In the study by Kondaurova and Francis (2008), it is shown that Russian-speaking listeners, when distinguishing these vowels, mainly rely on duration and almost do not take into account the qualitative characteristics of the sound in the same way as native speakers of English do. That is, the difference between [iː] and [ɪ] is often perceived by them primarily as a difference in

time, rather than as a combination of quantitative and qualitative features. This is important to take into account in teaching: work on articulation alone is not sufficient; it is necessary to develop the auditory discrimination of such contrasts separately.

Similar difficulties are observed in the acquisition of [ʊ] and [uː], which are often replaced by the Russian [u], despite significant articulatory and phonemic differences. Such substitutions frequently lead to word confusion, as they involve distinct phonemes.

The articulation of the Russian [u] differs significantly from that of the English vowels it is often confused with. When pronouncing [u], the front and central parts of the tongue are lowered, while the back of the tongue rises toward the soft palate. The lips are strongly rounded and protruded forward, making [u] a distinctly labialised sound with a low timbre. According to Vasiliev, when articulating the English vowels [ʊ] and [uː], the tongue is positioned noticeably closer to the front of the oral cavity than in the case of the Russian [u], and the back of the tongue is only moderately raised toward the soft palate. Vasiliev recommends giving the English [ʊ] and [uː] a slight shade of the Russian [ɨ], which helps achieve an intermediate quality between the familiar Russian [u] and the English vowel. This allows you to get closer to correct articulation without completely abandoning the usual articulatory attitudes of L1. (1980).

These differences are visually represented by Shevyakova, which shows the tongue positions for the English vowels [ʊ], [uː], and the Russian [u], further illustrating the importance of subtle articulatory shifts for accurate pronunciation (Figure 4).

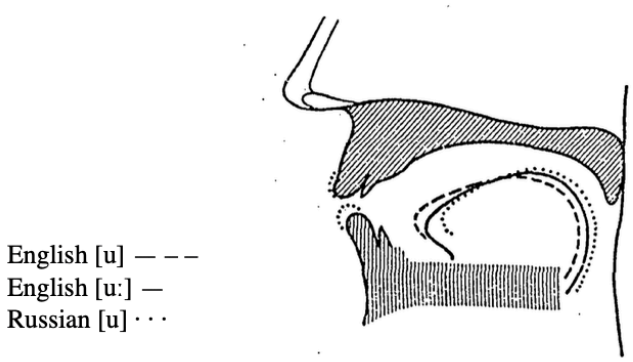

Figure 4. *Tongue positions for Russian [u], English [uː], and English [ʊ] (Shevyakova, 1968)*

A comparison of the articulatory characteristics of vowels in Russian and English allows us to conclude that their pronunciation is based on different articulatory strategies that form the unique articulatory bases of each language. These differences underlie the interference experienced by native Russian speakers when learning English, particularly in the production of vowel sounds.

In general, difficulties with English vowels for Russian-speaking learners are connected not only with the fact that some sounds are unfamiliar. More importantly, the systems themselves in the two languages are structured differently: in Russian, much depends on the position in the word and on reduction, whereas in English the differences between vowels — their quality, length, and tenseness — are significant. Given this articulatory distance between the L1 and L2 systems, the need to rely not only on auditory but also on visual and kinesthetic perception in phonetic instruction becomes evident. This is precisely where a multimodal approach gains particular significance, as it engages multiple perceptual channels simultaneously, helping to form a more accurate articulatory image and enhance phonological awareness.

1.3. Multimodal approaches to the teaching of English pronunciation

In online and hybrid learning environments, traditional explanations alone are often insufficient for effective phonetic instruction. Engaging multiple sensory channels through multimodal techniques helps sustain learner attention and supports more accurate pronunciation. As a result, multimodality has become an educational necessity rather than a supplementary enhancement. Learning is based on the simultaneous use of various modalities: auditory, visual, kinesthetic and cognitive. Traditional methods of teaching phonetics mainly rely on auditory exercises and explanations of the rules of articulation. The integration of visual and physical components, the efficiency of material assimilation increases. Visual cues, gestures, and the use of digital technology create a more understandable learning environment.

The multimodal approach makes it easier to distinguish between similar-sounding phonemes and helps to memorise them better and more consciously. This also helps Russian-

speaking students who have difficulty distinguishing between pairs of long and short vowels, in our case [iː] and [ɪ], [uː] and [ʊ]. Levis wrote: "Using software that displays vowel duration graphically can make abstract timing differences concrete and actionable for learners" (Levis, 2018). Visual tools such as Praat and Audacity allow students not only to hear, but also to see the length of a pronounced vowel, comparing it with the correct pronunciation.

The kinesthetic modality provides additional support. Kinesthetic supports such as hand gestures to mimic vowel length (for example, short tap for [ɪ], long sweep for [iː]) increase learner awareness and accuracy (Levis, 2018). In addition, the combination of auditory, visual and motor information contributes to the formation of new phonemic categories among students for whom a similar category is absent in their L1. Thus, multimodal input helps learners restructure their phonological space, leading to more stable and accurate representations of vowel contrasts (Levis, 2018).

This section is devoted to the theoretical foundations of a multimodal approach to phonetics teaching, the analysis of modern strategies and empirical research. Special attention will be paid to how digital technologies contribute to the formation of phonetic competence.

1.3.1. Theoretical foundations of multimodal interaction

Gunther Kress considers multimodality as a system of meaning production in which linguistic modality is only one of the components. In his approach, meaning arises as a result of the choice of modalities and their combinations, depending on the context, goals, social roles and available resources. He calls this process design of meaning, which shapes not just a message, but its interpretation and perception: "Meaning is made in many different ways, always, in culturally and socially shaped ways, using different modes; and the meanings made with each mode are different" (Kress, 2010). In the context of phonetics education, this means that articulation, intonation, facial expressions, gestures, and visual cues are equal carriers of meaning that support auditory information and make pronunciation more understandable and conscious for students.

A multimodal approach to phonetics teaching means that information is perceived and transmitted through several different channels, such as hearing, vision, and movement. To use this approach effectively, it is important to understand exactly which modalities are involved and how they work. This helps to consciously choose teaching methods and tools. When learning English pronunciation, the following types of modalities are especially important:

1) Auditory modality. The auditory channel plays a central role in teaching phonetics. Auditory modality provides the formation of phonemic hearing and allows students to recognise sound differences, such as the contrast between long and short vowels. As Mayer notes, auditory information is processed in a separate channel of working memory, which requires coordination with visual presentation for optimal assimilation (Mayer, 2005).

2) Visual modality. The visual channel allows to present audio information in a visual form — through spectrograms, articulation diagrams, vowel length graphs and video recordings of a speech act. For example, using programs like Praat allows students to see the difference in vowel length, which enhances perception and promotes the formation of correct articulation patterns. Visual modality also includes observing the teacher's facial expressions and articulation. According to Levis, the graphical representation of duration "makes abstract differences in time concrete and understandable" (Levis, 2018). In addition, Jewitt emphasises the importance of visual modality as an independent carrier of meaning, equal to language (Jewitt, 2009).

3) Kinesthetic modality. The kinesthetic channel involves the use of body movements and muscle memory to consolidate phonetic patterns. Jewitt argues that bodily interaction plays a key role in learning, especially when it comes to articulation (Jewitt, 2009).

4) Digital modality. Modern technologies complement traditional approaches, opening up access to digital learning tools: interactive simulators, speech recognition platforms, video tutorials, and sound visualisation programs. This

modality combines visual and auditory elements and creates a space for independent practice and feedback. As Gunther Kress emphasises, multimodality is not just a sum of channels of perception, but "a semantic interaction between different ways of representation" (Kress, 2010). This is the so-called technologically mediated multimodality. Platforms such as Audacity, Praat create a space for individual practice and feedback by combining visual, auditory, and motor channels.

Effective phonetics training requires a harmonious combination and interaction of all the modalities — this helps not only to master pronunciation, but also to consolidate it on a long-term level.

The rationale for the multimodal approach in teaching phonetics is based not only on pedagogical, but also on cognitive and neuropsychological principles. The use of multiple sensory channels in learning accelerates information processing, enhances memorisation, and improves the reproduction of phonetic elements (Mayer, 2005).

Therefore, when people hear and see the material at the same time, it helps to distribute the mental load and make it easier to assimilate information: "People learn more deeply from words and pictures than from words alone" (Mayer, 2005). In phonetics, this is especially important when learning vowel length, where auditory perception may not be reliable enough without visual supports.

The integration of information from different sensory channels activates multiple areas of the cerebral cortex, increasing the level of attention and the strength of memory traces (Shams and Seitz, 2008). This effect is called multimodal integration. In phonetics, when a student at the same time hears sounds, sees articulation and accompanies it with movement, a more stable neural trace is created. This is especially important in cases where it is necessary to form new phonemic categories, such as the difference between [iː] and [ɪ], [uː] and [ʊ] for Russian-speaking students.

Multi-channel learning promotes deep processing of information, which leads to better strengthening of long-term memory. Levis argues that even basic visual reinforcement, such as displaying the length of vowels, contributes not only to understanding, but also to more accurate memorisation and reproduction in the future (Levis, 2018). Thus, cognitive and neuropsychological data confirm the need and effectiveness of a multimodal approach in teaching phonetics. This approach is especially important in the formation of subtle auditory differences and stable articulatory skills, which makes it an indispensable tool in teaching English pronunciation.

In practice, multimodal phonetics training is implemented through a wide range of strategies that aim to simultaneously activate multiple channels of perception. It is important for the teacher to organise the lesson so that students can see, hear and reproduce the articulatory features of English speech with maximum accuracy and awareness. It is important to use visual tools such as articulation videos, mirrors, animations, infographics, and IPA tables contribute to the formation of correct motor patterns (Celce-Murcia et al., 2010).

These results clearly demonstrate how effective a multimodal approach (for example, using video) can be for teaching perception of speech as intonation, stress, and rhythm (Dohen and Loevenbruck, 2009). The integration of digital platforms expands the possibilities of a multimodal approach. Xodabande used Telegram to create a learning environment with video, voice messages, and texts (Xodabande, 2017).

Gestures and movements play a key role in multimodal learning. Kozlowska notes that the physical accompaniment of speech — for example, rhythmic clapping when pronouncing long vowels — helps students better distinguish sounds, especially those that are absent in their native language (Kozlowska, 2015). For Russian-speaking students, this is especially important when learning the difference between long and short vowels. Visual and tactile means such as gestures, facial expressions, graphic scales or spectrograms activate motor memory and make phonetic differences more visible and understandable. Visual elements such as gestures and facial expressions not only accompany speech, but also complement its

semantic content, helping to clarify the articulatory intention of the speaker (Kendon, 2004). Engaging multiple sensory channels increases motivation and reduces anxiety. **S**tudents feel more confident when they receive sound, visual, and body information at the same time. This is especially important in developing skills related to real-time sound reproduction.

According to Kress, multimodal literacy is not just knowledge of a language, but the ability to interpret and produce meanings through various modalities. This is especially important in phonetics, where students are often confronted with invisible aspects (for example, the position of the tongue, the work of the vocal cords, etc.). The use of multimodal resources helps students move from superficial repetition of sounds to conscious reproduction, which makes learning deeper and more effective.

1.3.2. Empirical research

The theoretical positions are also confirmed in empirical studies, the results of which give an idea of the real effect of multimodal approaches on pronunciation training. The empirical basis confirming the effectiveness of the multimodal approach in teaching English phonetics covers a variety of studies conducted in different countries and educational contexts. The Ganapathy and Seetharam experiment is a significant example of the successful application of a multimodal approach to L2 learning. The research involved 63 Malaysian schoolchildren aged 15 years, divided into experimental and control groups. The aim was to find out how the use of visual, auditory, kinesthetic and digital channels affects the understanding and reproduction of L2. The experimental group was trained using multimedia presentations, video materials, role-playing games and discussion of visual content, while the control group was trained using a traditional method based on working with texts. After four weeks of study, students from the experimental group showed a 35% improvement in phonetic accuracy, as well as demonstrated a high level of motivation and engagement. The authors emphasise that the combination of several sensory channels contributes to a deeper understanding of the material and facilitates the formation of stable phonetic skills (Ganapathy and Seetharam, 2016).

Kozlowska's research complements the findings of Ganapathy and Seetharam, confirming the high effectiveness of the multimodal approach in teaching phonetics. In her experiment with Polish students learning English as a L2, the author applied a holistic model. The students performed special phonetic exercises involving body movements aimed at strengthening the connection between auditory images and motor articulation patterns. As Kozlowska points out: "The tested procedure is both more effective than the traditional imitation tasks and more attractive to the participants, as shown in a post-test questionnaire study". The results of the study showed that the participation of the body in the process of assimilation of the sound system of language contributes to the activation of motor memory and the formation of more accurate articulatory skills, which is especially important when teaching adult learners with stable L1 skills.

Special attention should be paid to the Jingpeiyi study, which showed that the introduction of multimodality in phonetics courses at the undergraduate level increases not only the level of pronunciation, but also student engagement. The course involved 120 first-year students. The experimental group used multimodal methods — video tutorials, interactive platforms, exercises for articulatory imitation and visualisation of speech flows. The control group studied according to a traditional textbook. At the end of the course, a questionnaire was conducted in which 87% of the students in the experimental group noted an increase in interest in the subject. Also, the analysis of audio recordings showed an improvement in pronunciation. Multimodal teaching increases undergraduate interest and self-directed learning in phonetics, which makes this approach particularly promising in today's digital educational environment (Jingpeiyi, 2021).

A multimodal approach to teaching English phonetics is implemented not only through the choice of perception channels, but also through specific forms and methods of work in the classroom. The development of exercises involving visual, auditory and motor modalities allows not only to increase the motivation of students, but also to improve the accuracy of articulation and the stability of phonetic skills (Levis, 2018).

An effective practice is the simultaneous perception of video and audio, where students observe articulation, including the movement of lips, tongue, intonation contours. This can be done through:

- mirror training (the student repeats the sound while looking at himself in the mirror);
- slow motion videos of native speakers;
- author's animations where characters voice words with an emphasis on phonetic differences (Celce-Murcia et al., 2010).

A variety of approaches, from the use of digital tools and visualisation to bodily practices and authentic audiovisual content, contribute to the formation of a complex phonetic consciousness. This allows students not only to hear and distinguish sounds, but also to consciously reproduce them in a natural speech environment. The integration of various modalities — visual, auditory, kinesthetic and digital — ensures a more complete involvement of the student in the process of perception and reproduction of the sound side of speech.

The materials and examples provided in this chapter clearly show that the multimodal approach really works when teaching English phonetics. By combining auditory, visual, physical and digital formats, teaching becomes richer and more effective: it is easier for students to perceive and reproduce sounds, they understand the material better and are involved in the process with great interest.

Thus, multimodality in phonetics teaching can be considered as a mandatory component of modern language education. Based on the above, in the framework of this study, an attempt was made to test a multimodal approach in the context of school education. For this purpose, a Minecraft-style educational animation film was developed and implemented, aimed at developing Russian-speaking schoolchildren's skills in distinguishing between long and short English vowels. The use of visual, auditory and interactive

components within the framework of one educational media product allowed creating conditions for deep and conscious assimilation of phonetic material.

Chapter 2. The present study.

Chapter one provided a brief overview of the theory related to language models and multimodality. Although previous research has offered substantial insights into phonetic perception and the acquisition of English vowel sounds, relatively little attention has been paid to the specific question of how Russian-speaking schoolchildren perceive and distinguish between English long and short vowels — [iː]-[ɪ] and [uː]-[ʊ].

The present study includes two experiments. The first experiment consisted of three stages. At the first stage, an online quiz was conducted aimed at assessing the phonetic perception of schoolchildren. At the second stage, an educational animation was developed and presented, the purpose of which was to form students' understanding of the differences between long and short vowels in English. The third stage was a quiz, which was similar to the quiz from the first stage. and was aimed at assessing changes in vowel perception after watching the animation. The second experiment included four stages. The first three stages fully corresponded to the structure of the first experiment. The fourth stage was conducted a week after the completion of the third stage and was aimed at evaluating the long-term effectiveness of educational animation.

The aims of the present study were to examine the perception of English long and short vowels — [iː], [ɪ], [uː], and [ʊ] — by Russian-speaking primary school students, and to develop an original pedagogical solution designed to improve students' ability to perceive, distinguish, and reproduce the target English vowel contrasts. To achieve this goal, an educational animation was created.

The following research questions were proposed regarding the target vowel sounds:

1. To what extent are participants able to distinguish between target vowel pairs before any training or visual support?

2. Which specific pairs of vowels present the greatest difficulties for Russian-speaking schoolchildren: [iː]–[ɪ] or [uː]–[ʊ]?

3. Does phonetic perception improve after a multimodal intervention in the form of an educational cartoon?

4. What is the degree of preservation of the effect of multimodal intervention in the assimilation of contrasting vowel pairs based on the results of the delayed post-test?

2.1. Pre-test: initial assessment of vowel perception

2.1.1. Participants

The first experiment involved 26 subjects, both boys and girls, all of whom were native speakers of Russian. The gender distribution was equal: 13 girls (50%) and 13 boys (50%). The second experiment involved 10 children, including boys and girls who are native speakers of Russian. There were 4 boys (40%) and 6 girls (60%).

All participants for both experiments were primary school students enrolled in a general education program and were learning English as a foreign language as part of their regular school curriculum. At the time of data collection, their level of English proficiency corresponded to level A1 (Beginner) of the Common European Framework of Reference for Languages (CEFR). Learners at this level typically have a limited vocabulary, can understand and produce simple expressions, and are only beginning to acquire the phonological system of the English language, including the distinction between long and short vowel sounds.

Recruitment of participants for the first experiment was conducted through social media and personal networks to reach potential participants. The first experiment involved 12 children aged 9, 13 children aged 10, and one child aged 11. The recruitment process was facilitated by cooperation with school teachers and parents, who helped coordinate the schedule for the pre-test, the animated educational video, and the post-test. All research activities were carried out in a format appropriate to the developmental and psychological characteristics of the target age group, with an emphasis on maintaining the participants' comfort and engagement. An informed consent form was sent to the students' legal guardians stating that participation in the experiment was voluntary. It was also emphasised that the

participant has the right to terminate it at any time. After that, the children were instructed to complete the pre-test, which aimed to assess their perception of English vowel sounds prior to the introduction of the educational animation. They were then invited to watch the cartoon, followed by the post-test. These three components were presented to the participants as a single experiment divided into three phases. If I had sent them separate links for three individual experiments, the number of participants failing to complete one or more parts would likely have been higher, resulting in less data available for analysis. As an alternative, this integrated format also proved to be more time-efficient for both sides. The questionnaire was designed to assess the understanding of English vowel length perception among Russian primary school students. The quiz was conducted using the digital platform *Interacty*, which is widely used in educational settings for creating interactive learning materials and evaluating student performance (Interacty, n.d.). The quiz format consisted exclusively of multiple-choice tasks, in which participants were required to listen to short audio recordings and select the answer. This approach allowed for the quantitative measurement of perception accuracy and helped identify specific vowel contrasts that posed greater difficulty for the learners. The questionnaire maintained complete anonymity: no personal data were collected, nor was there any information linking individual responses to specific participants. The quiz was analysed based solely on the final answers submitted by all participants.

The participants in the second experiment were students of the author of the study. This experiment involved 5 children aged 10 and 5 children aged 11. Informed consent to the processing of personal data was obtained from the parents of all participants. To ensure anonymity and consistency in data analysis, each participant was assigned an individual code (P1–P10), which is used throughout the study. The quiz was conducted as part of individual lessons, which allowed me to obtain more detailed data and trace the process of mastering contrasting pairs of vowels at the level of individual students. The test tasks used in both experiments were identical. The difference was in the presence of a delayed post-test in the second experiment, conducted a week after the main test. In this regard, the description of the procedure below is the same for both groups of students studied.

2.1.2. Design and Materials

In order to determine the initial level of phonemic awareness among Russian-speaking schoolchildren regarding English long and short vowels, a pre-test was conducted. The goal was to assess how successfully the students could distinguish between the vowel pairs [iː] and [ɪ], as well as [uː] and [ʊ].

The test for two experiments consisted of eight questions, each requiring participants to listen to an audio clip and choose one of two given options that matched the sound they heard. The following words were selected for the pre-test to examine the understanding of English long and short vowels: *heat, full, slim, hit, school, food, beat, look*. These words were recorded by a native English speaker — a 20-year-old male born in Zimbabwe. English is one of the 16 official languages of Zimbabwe and serves as the language of education, administration, official documentation, and media. Although the vast majority of the population speaks it as a second language, for the speaker involved in this experiment, English is his first language. It is important to note that the target words for the experiment were not recorded in isolation but as part of full sentences:

1) Tell me about the heat in your country during summer.

2) Tell me about the time you felt completely full after a meal.

3) Tell me about how you stay slim and healthy.

4) Tell me about that little bit of advice you mentioned.

5) Tell me about your school and your favourite subject.

6) Tell me about the food you miss from your childhood.

7) Tell me about a song with a strong beat you love.

8) Tell me about the look on her face when she heard the news.

These segments were later extracted and processed using Adobe Audition.

The above table shows the words used in the pre-test (Table 1). They are arranged in the order in which they were presented to the participants during the audition. The table clearly shows that long vowels differ significantly from short vowels in terms of sound duration. For example, in the word *look*, the duration of the short vowel [ʊ] is 0.072 seconds, whereas in the word *school*, the duration of the long vowel [u:] reaches 0.149 seconds, which is almost twice as long. This confirms the acoustic difference between short and long vowels. Below are also screenshots (Figure 5 and 6) from the *Praat* program, illustrating the duration of the sound of the corresponding vowels in the sound files.

Word	IPA vowel	Duration of the sound (in seconds)
Heat	[i:]	0.200
Full	[ʊ]	0.087
Slim	[ɪ]	0.062
Bit	[ɪ]	0.090
School	[u:]	0.149
Food	[u:]	0.119
Beat	[i:]	0.170
Look	[ʊ]	0.072

Table 1. *Duration of target vowel sounds in the pre-test stimuli (measured in Praat).*

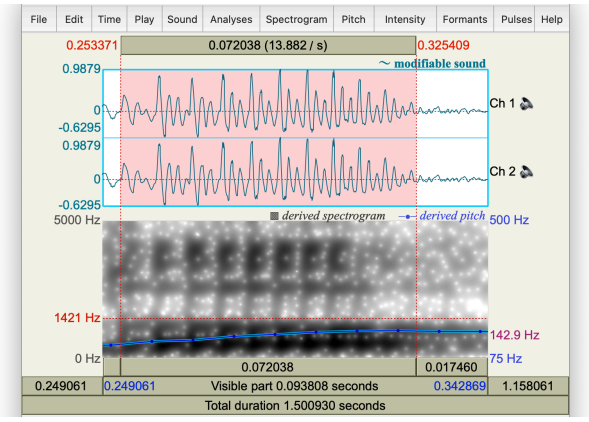

Figure 5. *Waveform and spectrogram of the vowel [ʊ] extracted from the word "look" are displayed in Praat. Its duration is 0.062 seconds.*

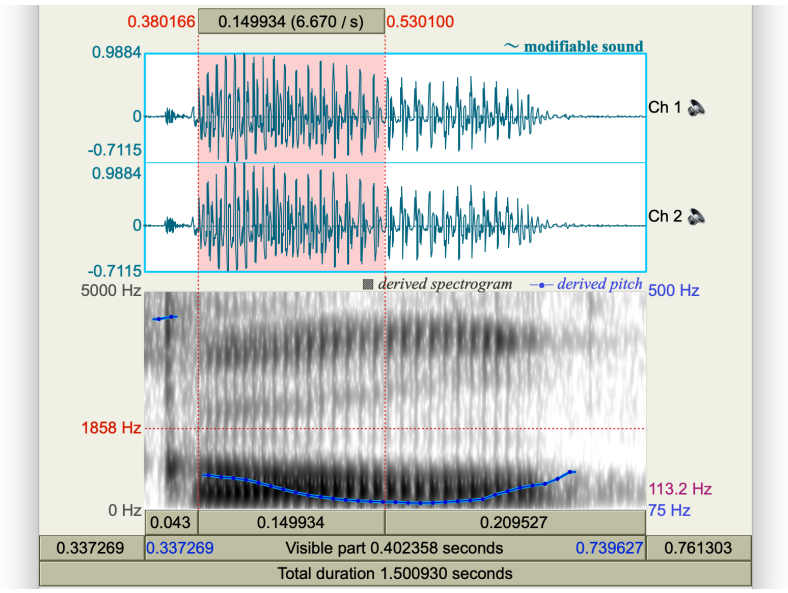

Figure 6. *Waveform and spectrogram of the vowel [u:] extracted from the word "school" are displayed in Praat. Its duration is 0.149 seconds.*

2.1.3. Procedure

The first experiment was created using the digital platform *Interacty* and accessed through a link sent to the participants' legal guardians. Upon clicking the link, parents were presented with detailed information about the structure and purpose of the experiment. They were then asked to complete a consent form for the processing of personal data, where they had to provide their full name, as well as the child's full name, age, and check a box indicating their consent. After that, the practice phase began, which included listening to sample sounds prior to the start of the quiz, followed by the actual pre-test. The pre-test aimed to examine the participants' perception of English vowel length. Each participant completed eight tasks, where they listened to an audio clip and chose between two answer options. After selecting an answer, they immediately saw whether it was correct and moved on to the next item. Figure 7 provides an illustrative example of the actual experimental procedure. It shows the screen presented to each participant during the sound pair task, where they had to click on one of the two answer options according to their perception.

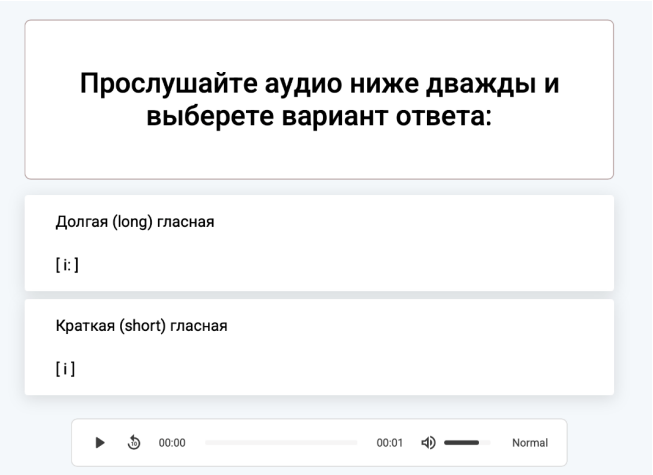

Figure 7. *Screenshot of the screen as it appeared during the pre-test. Participants listened to the sounds and were always presented with answer options.* Translation of the question: *Listen to the audio below twice and choose the correct answer.*

The second experiment was carried out with the participation of 10 schoolchildren as part of an English lesson under the guidance of the author of the study. The students listened to audio recordings and marked with a marker on the online board the answer they considered correct, without receiving any prompting from the teacher. Figure 8 shows the full experiment that was included in the lesson.

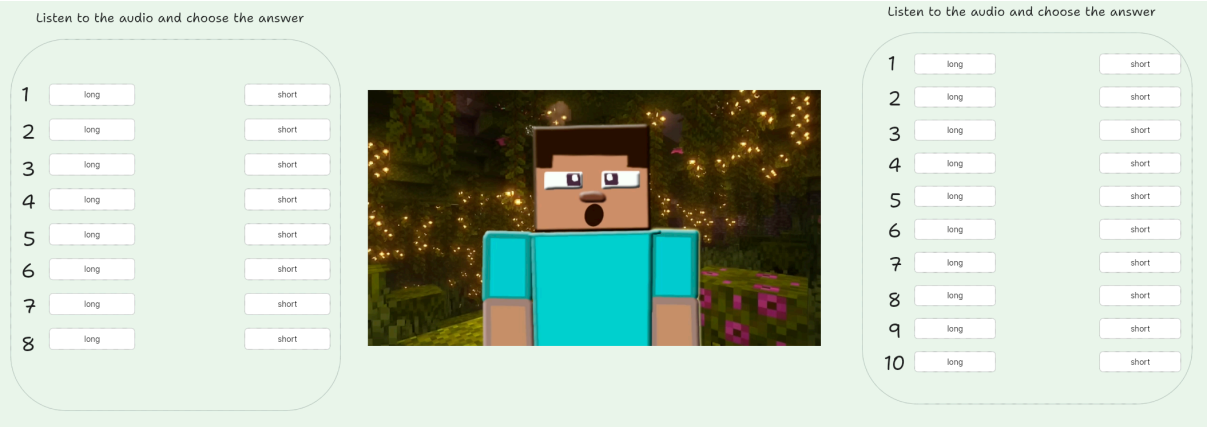

Figure 8. *Screenshot of the screen as it appeared during the test for the second experiment. Participants listened to the sounds and were always presented with answer options.*

2.1.4. Data analysis

In the following sections, I will comment on the collected data. The empirical component of this study is primarily based on quantitative data obtained through an online

test designed to assess the ability of Russian-speaking primary school students to distinguish between English long and short vowels — [iː]–[ɪ] and [uː]–[ʊ].

2.1.5. Results

The results of the pre-test of the two experiments showed that most students had a basic understanding of the difference between long and short vowels in English, but their perception of these distinctions was often unstable and uncertain.

During the first experiment, the following results were shown among 26 schoolchildren: in a number of tasks, the level of correct answers exceeded the probability level by 50%, in some cases reaching 73% (Table 2). However, in most of the tasks, the percentage of correct answers ranged from 38-40%, which indicates significant difficulties in auditory vowel discrimination. The confident distinction between long and short vowels was observed only sporadically, which emphasises the need for additional phonetic training.

Word	Vowel	Vowel type	Correct responses	Incorrect responses
heat	i	long	16 (62%)	10 (38%)
full	u	short	15 (58%)	11 (42%)
hit	i	short	11 (42%)	15 (58%)
slim	i	short	13 (50%)	13 (50%)
school	u	long	12 (46%)	14 (54%)
food	u	long	10 (38%)	16 (62%)
beat	i	long	19 (73%)	7 (27%)
look	u	short	15 (58%)	11 (42%)

Table 2. *Total number of data points collected from the pre-test.*

It is worth noting that in some cases the results were at the level of random guessing (50%) or even lower. This suggests that it was difficult for children to distinguish between long and short vowels by ear. This is especially noticeable in the pair [uː]–[ʊ]: for example,

the word food, which contains the long vowel [u:], the participants correctly recognised only in 38% of cases. Such indicators show how difficult it is to perceive this pair of sounds without additional visual or contextual support.

Comparing the results for the pair [i:]–[ɪ], you can see a completely different picture. So, the word beat with a long vowel [i:] was correctly recognised in 73% of cases, which is the highest result among all the words in the test. The word heat with the same sound scored 62% of correct answers. Most likely, this is due to the fact that the sound [i:] Russian-speaking schoolchildren perceive it as close to their native language, so it is easier to recognise it. But with the short [ɪ], everything turned out to be more complicated: the words hit (42%) and slim (50%) showed lower accuracy.

During pre-test involving 10 schoolchildren (P1–P10), the following results were showed (Table 3): the word *hit* was correctly recognised by all participants (100%), whereas for the words *look, slim,* and *heat*, were correctly recognised by 40%. The data obtained indicate that the students' auditory perception skills are not sufficiently developed; they require additional phonetic practice.

Word	**Vowel**	**Vowel type**	**Correct responses**	**Incorrect responses**
heat	i	long	4 (40%) P4, P6, P8, P10	6 (60%) P1, P2, P3, P5, P7, P9
full	u	short	5 (50%) P1, P2, P5, P6, P10	5 (50%) P3, P4, P7, P8, P9
hit	i	short	10 (100%) P1, P2, P3, P4, P5, P6, P7, P8, P9, P10	

Word	Vowel	Vowel type	Correct responses	Incorrect responses
slim	i	short	4 (40%) P5, P8, P9, P10	6 (60%) P1, P2, P3, P4, P6, P7
school	u	long	6 (60%) P3, P4, P5, P7, P8, P9	4 (40%) P1, P2, P6, P10
food	u	long	6 (60%) P2, P3, P4, P5, P6, P7	4 (40%) P1, P8, P9, P10
beat	i	long	6 (60%) P1, P2, P5, P6, P9, P10	4 (40%) P3, P4, P7, P8
look	u	short	4 (40%) P2, P3, P8, P9	6 (60%) P1, P4, P5, P6, P7, P10

Table 3. *Total number of data points collected from the pre-test.*

In some cases the results were at the level of random guessing (about 50%) or even lower. This indicates that the students had difficulty audibly distinguishing between long and short vowels. This is especially noticeable in the framework of the opposition [u:]–[ʊ]: the word *look*, containing a short vowel [ʊ], was correctly recognised only in 40%, and the word *full* — in 50%. Even when perceiving a long vowel [u:] the results remain unstable: the words *school* and *food* were recognised correctly only in 60% of cases. In the case of the opposition [i:]–[ɪ], a slightly different picture is observed. The word *hit* was correctly recognised by all participants (100%), but other words show significantly lower results: *slim* — 40% and *heat* — 40%, as well as *beat* — 60%. Despite some high rates, overall perception of both long and short vowels remains unstable.

The results of the pre-test demonstrate that the participants already have an initial understanding of the differences between long and short vowels, but at the same time, constant difficulties remain — especially in the case of the pair [u:]–[ʊ]. All this points to the need for more targeted phonetic training, preferably using various channels of perception - visual, auditory and emotional.

The combined results of the pre-test allow us to draw the following preliminary conclusions in the context of the research questions posed:

1) The first research question: to what extent are participants able to distinguish between target vowel pairs before any training or visual support?
- The results of both experiments showed that the level of discrimination remains generally low. In the first experiment even in the most successful cases, the accuracy did not exceed 73%, and for a number of words, the figures were significantly below 50%. This indicates an unstable phonemic perception of the target sounds. If I talk about the second experiment, I will say the analysis of individual words shows significant variability in the results: from 40% of correct answers to 100%. At the same time, most of the tasks were performed with an accuracy of about 40-60%, which indicates the instability of phonemic perception.

2) The second research question: which specific pairs of vowels present the greatest difficulties for Russian-speaking schoolchildren: [i:]–[ɪ] or [u:]–[ʊ]?
- In two experiments, the pair [u:]–[ʊ] became to be the most difficult to perceive. In the first experiment the words *food* — 38% and *school* — 46% gave the lowest results among all test units, which indicates significant difficulties in distinguishing this opposition compared to the pair [i:]–[ɪ]. In the second experiment, the average percentage of correct answers for the [i:]–[ɪ] pair was about 60%. In the case of the [u:]–[ʊ] pair, the average level of correct answers was about 55%. Words with a short vowel [ʊ], such as *look* — 40% and *full* — 50%, caused the greatest difficulties.

The results of the study show that Russian-speaking schoolchildren face difficulties in distinguishing the contrasts of English vowels, especially when they rely solely on hearing. This underscores the need for targeted development of phonemic awareness using multimodal

strategies and the gradual introduction of distinguishing features, particularly for vowel pairs that have no direct equivalents in the learners' native language.

2.2. Development of the educational animation

The creation of the educational cartoon was a key element of the experimental design in the present study. This multimedia product was developed individually from scratch and was specifically intended to support the formation and development of skills in distinguishing English long and short vowels ([iː]–[ɪ], [uː]–[ʊ]) among Russian-speaking primary school students.

The main goal in creating the educational cartoon was to design a clear, accessible, and engaging tool to help Russian-speaking children learn to differentiate between long and short English vowels. The difficulty in perceiving phonetic contrasts is one of the key challenges in learning English, especially in a limited language environment. The cartoon was designed to address this issue by combining visual imagery with interactive gameplay. Its goal was not only to convey phonetic differences but also to make the learning process enjoyable, thereby reducing anxiety associated with unfamiliar sounds. It was assumed that the emotional engagement of the learners would be a crucial factor in the successful reinforcement of the material.

The cartoon was created to support multimodal learning by combining visual and auditory input with simple interactive tasks that kept children actively involved. It also incorporates the principles of an interactive approach, where the viewer is not a passive recipient of information but an active participant in the process. The cartoon includes tasks and prompts to which the viewer is expected to respond either mentally or aloud, encouraging engagement and reinforcing learning through active involvement.

The initial goal was to create a multimedia tool that would be engaging for young school-aged children, incorporate the target phonetic pairs in a lively context, rely on multimodal perception (audio+visualisation), use storytelling as the foundation for

pedagogical impact, and matching the cognitive and emotional characteristics of the target age group.

The storyline followed a boy named Steve on a journey through two magical forests — the forest of short vowels and the forest of long vowels. Based on the phonetic aims and target age group, a complete script was written:

1. Introduction: the hero finds a map and receives a mission — to enter two forests and uncover the secret of vowel sounds.

2. Plot summary: the main character travels through the forest of short vowels ([ɪ], [ʊ]) and the forest of long vowels ([i:], [u:]), meeting vowel characters along the way. Each character has a name, a distinct voice, and a personality. After each viewing, the viewer is asked to identify the vowel sound they heard.

3. Conclusion: after successfully identifying all the sounds, the hero receives a medal, and the viewer is told that they have become a "Sound Explorer".

The script was written in Russian with the inclusion of English vowels as phonetic anchors. This approach ensured that the content was understandable for the learners while simultaneously providing exposure to the target sounds. Each vowel sound corresponds to a character with pronounced features, both visual and verbal:

• Uggie ([ʊ]) is a green, slightly awkward and funny character who speaks with a short, cheerful, and fragmented intonation pattern.

• Ikki ([i]) is a square character, he is represented by a cheerful and sharp hero who pronounces words clearly and quickly.

• Niiki ([i:]) is more strict and smooth, his speech has a length and softness, he is also taller than the main character and stretches out when pronouncing a remark.

• Nuuki ([u:]) is a huge ghost that also stretches at the moment of speaking, as well as his speech is smooth and drawling.

This personification of sounds allows students to associate phonemes with specific features, which contributes to better memorisation and discrimination.

All visuals were hand-drawn on a tablet using Procreate, then assembled and animated using Adobe After Effects. The visual style was inspired by the aesthetics of the popular videogame Minecraft (pixelated graphics, blocky architecture) because it is familiar to most children and easily adaptable for animation. Each scene was drawn as a separate PSD file with layered components (background, characters, environmental details). This layering was essential for subsequent animation in Adobe After Effects.

The following structure was used to organise the project:

- Each scene was organised as an individual composition with layered assets for animation.
- The characters are placed on separate layers.
- The sound is played on a separate audio track.
- Keyframes were used to animate attributes such as position, opacity, scale, and rotation.

In addition, specific animation techniques were employed:

1) Puppet Pin Tool — to create "realistic" character movements (e.g., head tilts, swaying motions).

2) Fade In/Out — for the appearance and disappearance of objects.

3) Trim Paths — for animating movement along a path (e.g., Steve walking through the forest).

To create the effect of a "talking character", manual lip-syncing was used. Two separate drawings were created — one with an open mouth and one with a closed mouth. The mouth shapes were alternated frame by frame using masks and opacity adjustments.

The cartoon also included visually simulated interactive segments — for example, buttons labeled "[u:]" and "[ʊ]" would appear on the screen, prompting children to mentally choose or say aloud the correct option, after which the correct answer was revealed. This approach was designed to allow viewers to actively engage in the learning process, even though the video itself was not technically clickable.

All voice lines were recorded using an external USB condenser microphone, with the initial recording done through the voice recorder app on a smartphone. The lines were voiced using different character voices — representing the main character, the map, and the vowel characters. To emphasise the distinctions between the vowel sounds, various techniques were applied: changes in timbre (raising/lowering pitch), the use of filters (equalisation, resonance enhancement), and time-stretching or compression (to highlight vowel length). The voice acting was performed by four Russian-speaking students from Ca' Foscari University. The audio files were then processed in Adobe Audition, where background noise was removed, volume levels normalised, silence trimmed, and the final files exported in .WAV format.

The final cartoon was exported in .MP4 format with a resolution of 1920×1080 and a frame rate of 30 frames per second. The video was uploaded to Google Drive, where it was embedded into the online quiz as the main instructional component between the pre-test and the post-test.

Stage	Duration	Content
Introduction	44 sec.	Steve finds the map
Forest of short vowels	1 min. 12 sec.	Characters [ɪ], [ʊ] + tasks
Forest of long vowels	1 min. 31 sec.	Characters [i:], [u:] + tasks
Final	29 сек.	Finding the chest
Quiz	1 min. 58 sec.	Quiz
Conclusion and praise	14 sec.	Receiving the medal and summary
Total:	5 min. 28 sec.	Full educational cycle

Table 4. *Structure and timing of the educational animation.*

The cartoon was built on key didactic principles: repetition (each sound is introduced at least twice), comparison (contrasting sounds both visually and aurally), interactivity (simulated tasks and feedback), association (sounds are represented as characters with specific visual traits), and immersion (the storyline and visual atmosphere create a cohesive "world of sounds").

As a result, a multimedia resource was developed that not only conveyed phonetic material but also promoted active participation, emotional engagement, and effective learning. Its use within the experiment led to a significant improvement in the perception of English vowels among Russian-speaking schoolchildren, making this format a promising tool for further integration into primary education.

2.3. Post-test: measuring the effect of multimedia exposure

The post-test involved the same 36 participants who had previously completed the pre-test. The previously described procedures for participant selection and informed consent remained unchanged. The post-test participants of the first experiment completed the final stage in the same online format using the *Interacty* platform and the same 10 participants of the second experiment also completed a survey during an individual English lesson. All testing conditions were identical to those used in the pre-test, ensuring maximum comparability of the data.

2.3.1. Design and Materials

After watching the educational cartoon, the participants of the two experiments were offered a final test consisting of 10 tasks, each aimed at assessing their ability to distinguish between English long and short vowels ([iː]–[ɪ] and [uː]–[ʊ]). The format of the test remained the same: participants listened to an audio clip and selected one of two answer options. The list of words used in the post-test included the following units: *set, rule, look, cook, sheep, beat, school, slim, book, leave* (Table 4). Some of these words coincided with those that had already been presented in the pre-test (look, school, slim, beat), which made it possible to compare the dynamics of perception of the same sounds before and after multimodal learning.

As in the pre-test, the tasks were presented sequentially in a unified structure. All audio fragments were recorded by the same native English speaker (a 20-year-old male from Zimbabwe), then extracted from full sentences and processed in Adobe Audition to ensure acoustic consistency:

1) Tell me about the last time you had to sit through a really boring meeting.

2) Tell me about a time you didn't want to leave but had no choice.

3) Tell me about a strange or unfair rule you had to follow in school.

4) Tell me about the look on her face when she heard the news.

5) Tell me about the first time you tried to cook something completely on your own.

6) Tell me about the first time you saw a sheep — was it real or just in a picture book?

7) Tell me about a song with a strong beat you love.

8) Tell me about your school and your favourite subject.

9) Tell me about how you stay slim and healthy.

10) Tell me about a book that changed how you think about something important.

Word	**Sound**	**Length of the sound (in seconds)**
Sit	[ɪ]	0.069
Leave	[iː]	0.149
Rule	[uː]	0.122
Look	[ʊ]	0.072
Cook	[ʊ]	0.077

Word	Sound	Length of the sound (in seconds)
Sheep	[iː]	0.158
Beat	[iː]	0.170
School	[uː]	0.149
Slim	[ɪ]	0.062
Book	[ʊ]	0.077

Table 5. *Duration of target vowel sounds in the post-test Stimuli (measured in Praat).*

2.3.2. Procedure

After watching the educational cartoon, the participant of the first experiment returned to the previous tab, back to the *Interacty* platform, to continue the test. The post-test was designed to analyse the perception of English vowel length following the instructional animation. The listener was presented with ten questions, each containing an audio clip and two answer options. After responding to all the post-test questions, the participant could simply close the browser window — the results were submitted automatically without the need for confirmation of completion.

The post-test was presented to the participants of the second experiment on the same page as the other stages of the study, and was conducted by analogy with the pre-test. The students marked the correct answer with a marker.

2.3.3. Data Analysis

This section presents and analyses the data collected at the third stage — after the participants had watched the educational cartoon. This part of the study relies on quantitative data collected through the post-test, administered as an online quiz. The purpose of this test was to evaluate the effectiveness of multimedia effects on the ability of Russian-speaking primary school students to distinguish between pairs of long and short vowels in English — [iː]–[ɪ] and [uː]–[ʊ].

2.3.4. Results

An analysis of the results obtained after two experiments showed a significant improvement in the ability to distinguish both long and short English vowels following the viewing of the multimedia educational resource. The percentage of correct responses increased across nearly all items. The post-test indicated qualitative changes: after watching the cartoon, students became more confident in distinguishing words, the number of errors observed earlier decreased by more than a third, and the level of random guessing declined — participants were more likely to choose the correct answer consciously rather than by chance.

The table presents data on how well the 26 Russian-speaking students in the first experiment were able to perceive English long and short vowels after watching the cartoon (Table 6). Overall, the data demonstrate a substantial improvement in phonemic discrimination compared to the pre-test.

Word	Vowel	Vowel type	Correct responses	Incorrect responses
sit	i	short	24 (92%)	2 (8%)
rule	u	long	22 (85%)	4 (15%)
look	u	short	24 (92%)	2 (8%)
cook	u	short	19 (73%)	7 (27%)
sheep	i	long	23 (88%)	3 (12%)
beat	i	long	18 (69%)	8 (31%)
school	u	long	21 (81%)	5 (19%)
slim	i	short	24 (92%)	2 (8%)
book	u	short	25 (96%)	1 (4%)
leave	i	long	18 (69%)	8 (31%)

Table 6. *Total number of data points collected from the post-test.*

Response accuracy ranged from 69% to 96%, with no single word causing major difficulty for most participants. The highest result was recorded for the word *book*, which

contains the short vowel [ʊ], with 96% correct answers. A high level of accuracy was also observed for the words *sit*, *look*, and *slim*, each correctly identified by 92% of participants. These words contain the short vowels [ɪ] or [ʊ], and their confident recognition suggests that students successfully internalised the characteristics of short vowels following the multimedia intervention. At the same time, words containing long vowels — which had previously posed difficulties — were also recognised much more accurately. For example, the word *rule* ([u:]) received 85% correct responses, and *school* 81%, whereas in the pre-test, similar words with the long vowel [u:] had not even reached 50% accuracy. When examining the average scores by vowel type, it becomes clear that for words containing [u] (including both [ʊ] and [u:]), correct answers were given on average in 85.4% of cases, while words with [ɪ] scored 82% of the points. Thus, even the previously challenging [u:]–[ʊ] pair was perceived with high accuracy in the post-test. Moreover, the distinction between long and short forms became less pronounced, indicating the development of less stable phonemic representations.

The results of the second experiment were also analysed. Data processing was carried out on an individual level for each participant, which made it possible to trace the dynamics of assimilation of contrasting vowel pairs. The table (Table 7) below shows the results obtained after post-test.

Word	**Vowel**	**Vowel type**	**Correct responses**	**Incorrect responses**
sit	i	short	9 (90%) P1, P2, P3, P4, P5, P6, P7, P9, P10	1 (10%) P8
rule	u	long	8 (80%) P1, P2, P3, P4, P5, P8, P9, P10	2 (20%) P6, P7

Word	Vowel	Vowel type	Correct responses	Incorrect responses
look	u	short	6 (60%) P2, P3, P4, P6, P9, P10	4 (40%) P1, P5, P7, P8
cook	u	short	6 (60%) P2, P3, P6, P7, P9, P10	4 (40%) P1, P4, P5, P8
sheep	i	long	9 (90%) P1, P2, P3, P4, P5, P6, P7, P9, P10	1 (10%) P8
beat	i	long	8 (80%) P1, P3, P4, P5, P6, P7, P9, P10	2 (20%) P2, P8
school	u	long	10 (100%) P1, P2, P3, P4, P5, P6, P7, P8, P9, P10	
slim	i	short	10 (100%) P1, P2, P3, P4, P5, P6, P7, P8, P9, P10	
book	u	short	7 (70%) P2, P3, P5, P6, P7, P8, P10	3 (30%) P1, P4, P9

Word	Vowel	Vowel type	Correct responses	Incorrect responses
leave	i	long	9 (90%) P1, P2, P3, P4, P5, P7, P8, P9, P10	1 (10%) P6

Table 7. *Total number of data points collected from the post-test.*

The accuracy of the answers in the final test ranged from 60% to 100%. The highest results were recorded for the words *school* and *slim*, which were correctly recognised by all participants (100%). Such a high result may be due to the fact that these words were in the educational cartoon, which contributed to their better memorisation. A high level of accuracy was also observed for the words *sit, sheep*, and *leave* (90% each), as well as *rule* and *beat* (80% each). These data indicate a significant improvement in students' ability to distinguish between both short and long vowels after learning. Words with the vowels [u] were recognised on average in about 74% of cases, while words with [i] — in about 90% of cases. This indicates that, despite the overall improvement, the opposition [u:]–[ʊ] remains somewhat more complex compared to [i:]–[ɪ].

Above, I analysed the overall picture of the second experiment. Comparing the results of the pre-test and the post-test also makes it possible to track the individual dynamics of each participant (P1-P10).

P1: completed 3 of the 8 tasks in the pre-test. The main difficulties were related to the pair [u:]–[ʊ]. In the post-test, the result increased to 7 out of 10. However, there was still a problem with the short sound [ʊ]: the errors remained in the words *look, cook and book*.

P2: completed 5 out of 8 tasks in the pre-test. In the post-test, the result increased to 9 out of 10. All the words with [u:]–[ʊ] were recognised correctly, and the only mistake remained in the word *beat* with a long [i:].

P3: completed 4 of the 8 tasks correctly in the pre-test. The most noticeable problems were with the [iː]–[ɪ] pair: errors in *heat, slim and beat*. In the post-test, the participant gave 10 out of 10 correct answers.

P4: completed 4 out of 8 tasks in the pre-test. Errors were observed in the words *full*, *slim*, *beat*, and *look*, meaning difficulties were associated with both [iː]–[ɪ] and [uː]–[ʊ]. In the post-test, the result increased to 8 out of 10. All words with [iː]–[ɪ] were recognised correctly, however, errors were preserved in the words *cook* and *book* containing the short [ʊ].

P5: answered 6 out of 8 tasks correctly in the pre-test. In the post-test, the result was 8 out of 10. The errors remained in the words *look* and *cook*, meaning the difficulty with the short [ʊ] remained.

P6: gave 5 correct answers out of 8 in the pre-test. The errors were in the words *slim*, *school* and *look*, which indicates difficulties with the short [ɪ] and with the pair [uː]–[ʊ]. In the post-test, the result increased to 8 out of 10. The errors remained in the words *rule* and *leave*.

P7: in the pre-test, the result was 3 out of 8. Errors were observed in the words *heat*, *full*, *slim*, *beat* and *look*, that is, the participant had difficulties with almost all target oppositions. In the post-test, the result increased to 8 out of 10. All the words with [iː]–[ɪ] were recognised correctly, but the pair [uː]–[ʊ] still needs to be fixed.

P8: in the pre-test, the result was 5 out of 8. The errors were in the words *full, food* and *beat.* In the post-test, the result was 5 out of 10, meaning there was no marked improvement. This is the only participant whose dynamics looks unstable. Perhaps the result is related to individual factors: inattention or lack of concentration during the test.

P9: completed 5 out of 8 tasks correctly in the pre-test. The errors were in the words *heat, full, and food,* meaning they mostly affected long vowels and short [ʊ]. In the post-test, the result increased to 9 out of 10. The only mistake left was in the word *book* with the short [ʊ].

P10: in the pre-test, the result was 5 out of 8. The main problem was related to the pair [uː]–[ʊ]. In the post-test, the participant gave 10 out of 10 correct answers.

After conducting multimodal training in the post-test, there is a significant improvement in the performance of most schoolchildren. The most pronounced progress is observed among the participants who made the most mistakes in the pre-test: in the post-test, their results become more balanced, and the number of correct answers increases. This indicates the effectiveness of the training conducted. At the same time, participants who showed relatively high results already at the pre-test stage also show improvement or maintain a high level of accuracy.

The results of two post-tests confirm the third research question, which stated that the multimedia format can improve vowel recognition, especially for previously problematic contrasts. The data demonstrate that the multimedia educational video contributed to the development of phonemic awareness among Russian-speaking primary school students. The improvement observed in both long and short vowel recognition indicates the formation of more stable auditory representations of English vowels.

2.4. Delayed post-test: measuring the long-term effect of multimedia exposure

The delayed post-test was attended by 10 students from the second experiment only who had previously passed the pre-test and post-test. The name of the participants also remains unchanged throughout the experiment — P1-P10. The procedures for selecting participants and obtaining informed consent remained unchanged.

2.4.1. Design and Materials

The delayed post-test was performed one week after the main test in order to assess the effect of the multimodal intervention. The experimental format and conditions fully corresponded to the post-test.

2.4.2. Procedure

The procedure for conducting the delayed post-test corresponded to the format of the post-test of the second experiment. The testing was conducted as part of individual classes. The participants listened to audio recordings and marked the chosen answer with a marker without any help from the teacher.

2.4.3. Data Analysis

The main purpose of this stage was to assess the degree of preservation of the formed perceptual skills, as well as to determine the stability of the effect of multimodal intervention in the development of the ability of Russian-speaking children to distinguish between contrasting pairs of vowels [iː]–[ɪ] and [uː]–[ʊ].

2.4.4. Results

The results obtained after the postponed post-test showed that most of the skills formed after multimodal training were preserved. The students continued to demonstrate higher recognition skills compared to the pre-test, which indicates a partial retention of phonemic skills.

The table presents data on the results of perception of long and short vowels of the English language by 10 Russian-speaking schoolchildren a week after showing the cartoon (Table 8).

Word	Vowel	Vowel type	Correct responses	Incorrect responses
sit	i	short	8 (80%) P1, P2, P4, P5, P6, P7, P9, P10	2 (20%) P3, P8
rule	u	long	9 (90%) P1, P2, P4, P5, P3, P7, P8, P9, P10	1 (10%) P6

Word	Vowel	Vowel type	Correct responses	Incorrect responses
look	u	short	7 (70%) P2, P3, P4, P6, P8, P9, P10	3 (30%) P1, P5, P7
cook	u	short	9 (90%) P1, P2, P3, P4, P6, P7, P8, P9, P10	1 (10%) P5
sheep	i	long	10 (100%) P1, P2, P4, P5, P3, P6, P7, P8, P9, P10	
beat	i	long	5 (50%) P5, P6, P7, P9, P10	5 (50%) P1, P2, P3, P4, P8
school	u	long	9 (90%) P2, P3, P4, P5, P6, P7, P8, P9, P10	1 (10%) P1
slim	i	short	5 (50%) P1, P4, P6, P7, P9	5 (50%) P2, P3, P5, P8, P10
book	u	short	8 (80%) P1, P2, P4, P5, P6, P7, P8, P9, P10	2 (20%) P3, P8

Word	Vowel	Vowel type	Correct responses	Incorrect responses
leave	i	long	9 (90%) P1, P2, P3, P5, P6, P7, P8, P9, P10	1 (10%) P4

Table 8. *Total number of data points collected from the delayed post-test.*

The highest results were recorded in the word *sheep*, which was recognised by all participants of the experiment (100%). At the same time, some of the contrasts were still unstable: the words *slim* and *beat* were correctly recognised only in 50%, which is lower than the rest. Despite a separate decrease in performance compared to the post-test, the results of deferred testing significantly exceed the data of the pre-test. If in the pre-test many answers were in the range of probabilistic choice — 40-60%, then after a week most of the stimuli retained a recognition level of 70%. This indicates that the multimodal intervention had not only a short-term, but also a relatively stable effect on the development of perception skills of English long and short vowels in Russian-speaking students.

Comparing the results of the post-test and the delayed post-test allows to track the individual dynamics of each participant (P1-P10).

P1: in the post-test successfully recognised most of the stimuli, including *cook, sheep, school* and *book*, but in the delayed post-test had difficulties with the words *look, beat* and *school*. The results were still significantly higher compared to the pre-test.

P2: successfully distinguished almost all contrasting pairs in the post-test, and after a week showed a high level of recognition of most words.

P3: showed positive dynamics compared to the pre-test, but in the delayed post-test had difficulties with some words containing [ɪ] and [ʊ] sounds.

P4: showed less stable results. Although he successfully recognised most of the stimuli in the post-test, had errors in the words *beat* and *leave* in the delayed post-test. This may indicate that distinguishing long vowel [i:] required a longer fixation.

P5: demonstrated a high level of skill retention and made a minimal number of errors in the delayed post-test.

P6: showed a significant improvement compared to the pre-test, but a week later, separate errors in the words *rule* and *leave* reappeared.

P7: demonstrated stable perception of most stimuli, minor difficulties were observed only in individual words with [I:] sound.

P8: most of the stimuli were recognised correctly both in the post-test and a week later.

P9: successfully distinguished almost all target vowels, which indicates that phonemic contrasts are well fixed.

P10: after cartoon, correctly recognised most of the stimuli, and in the delayed post-test made errors, mainly related to short back vowels.

In answer to the fourth research question, the delayed post-test results remained higher than the pre-test results. Multimodal visual and auditory support not only promotes short-term improvement in English vowel perception but also helps Russian-speaking students develop more stable phonemic awareness.

Chapter 3. General results and discussion

The present study was aimed at identifying how Russian-speaking primary school students perceive and distinguish between long and short vowels in English [iː], [ɪ], [uː] and [ʊ]. Given the phonological differences between English and Russian, the aim of the experiment is to assess the difficulties faced by students and identify effective strategies for improving vowel perception using multimodality. Phonetic features of the English language, in particular quantitative and qualitative oppositions, pose a significant challenge for Russian speakers, where differences do not play a phonemic role.

The experimental part of the study focused on answering three research questions:

1. To what extent are participants able to distinguish between target vowel pairs before any training or visual support?

2. Which specific pairs of vowels present the greatest difficulties for Russian-speaking schoolchildren: [iː]–[ɪ] or [uː]–[ʊ]?

3. Does phonetic perception improve after a multimodal intervention in the form of an educational cartoon?

4. What is the degree of preservation of the effect of multimodal intervention in the assimilation of contrasting vowel pairs based on the results of the delayed post-test?

During the two experiments, the pre-test and the post-test were conducted, and the educational cartoon was shown, the delayed post-test was also conducted for the second experiment. A comparison of the results showed weaknesses in phonetic perception before the start of the experiments and the improvements achieved through multimodal intervention, as well as the maintenance of results after a week.

Analysis of the pre-test results in two experiments revealed difficulties in the perception of long and short English vowels by Russian-speaking students. Participants were not always able to confidently distinguish between both the length of a vowel and its qualitative characteristics, especially in cases with contrasts [i:]–[ɪ] and [u:]–[ʊ].

In the first experiment, the lowest results were recorded when recognising the long vowel [u:]. The word *food* was correctly identified only in 38% of cases, and *school* — in 46%, which indicates a weak perception of long back vowels. In the second experiment, despite the smaller number of participants, the words *school* and *food* were recognised correctly in 60% of cases. The indicators of the second experiment turned out to be higher.

In contrast, the results for [i:]–[ɪ] were inconsistent. In the first experiment, the word *beat* showed the highest percentage of correct answers — 73%, however, even this indicator cannot be considered completely reliable in terms of stable phonological differentiation._In the second experiment, the recognition accuracy of this word decreased to 60%. At the same time, the short vowel [ɪ] in the word *hit* showed significantly better results: in the second experiment, all participants correctly identified this stimulus (100%), whereas in the first experiment, only 42% had correct answers. This is explained by the more recognisable articulation pattern of this word.

It should be noted that a significant part of the results in both experiments were in the range of 40-60%, which can be interpreted as a level close to a random choice of the answer. This is especially true for the words *heat, slim, full* and *look*, where participants demonstrated unstable results and a lack of a confident strategy for distinguishing sounds, for example, in the second experiment, the words *heat* and *slim* were correctly recognised only in 40% of cases.

The results of two pre-tests allow us to conclude that Russian-speaking students have significant difficulties both in perceiving the length of English vowels and in distinguishing their qualitative characteristics. The data obtained confirm the first research question and indicate the need for targeted phonetic training aimed at developing skills in perceiving English vowel contrasts.

The second research question examined which contrasting pairs of English vowels caused the greatest difficulties for Russian-speaking students. To answer this question, a comparative analysis of the results of preliminary and post-test testing in two experiments was carried out. For each sound — [i:], [ɪ], [u:], and [ʊ], the average correct response rates were calculated based on all stimuli containing the corresponding vowel (Diagram 1 and Diagram 2).

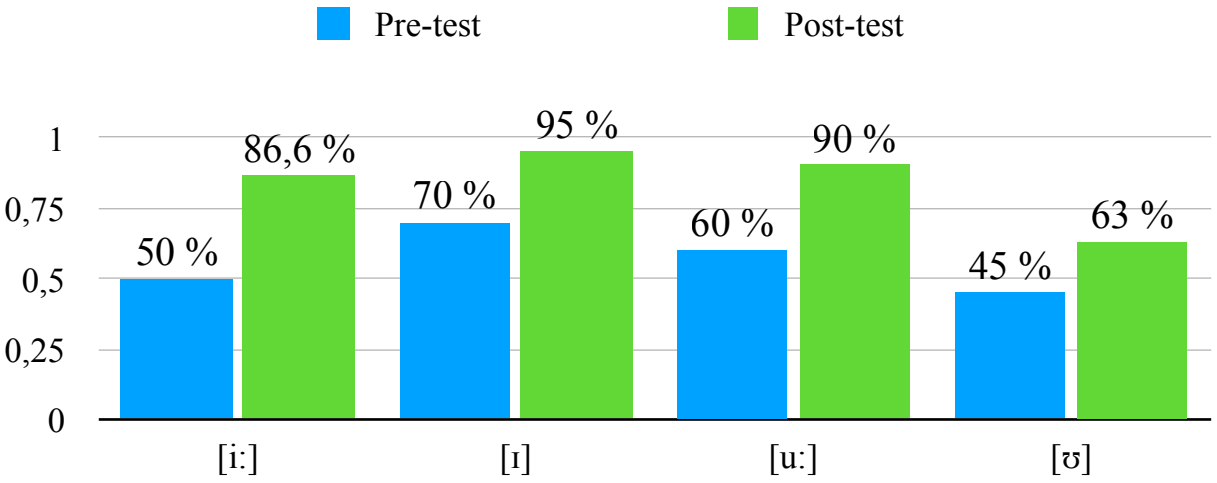

Diagram 1. *Pre-test and post-test results for [i:]–[ɪ] and [u:]–[ʊ] of the first experiment.*

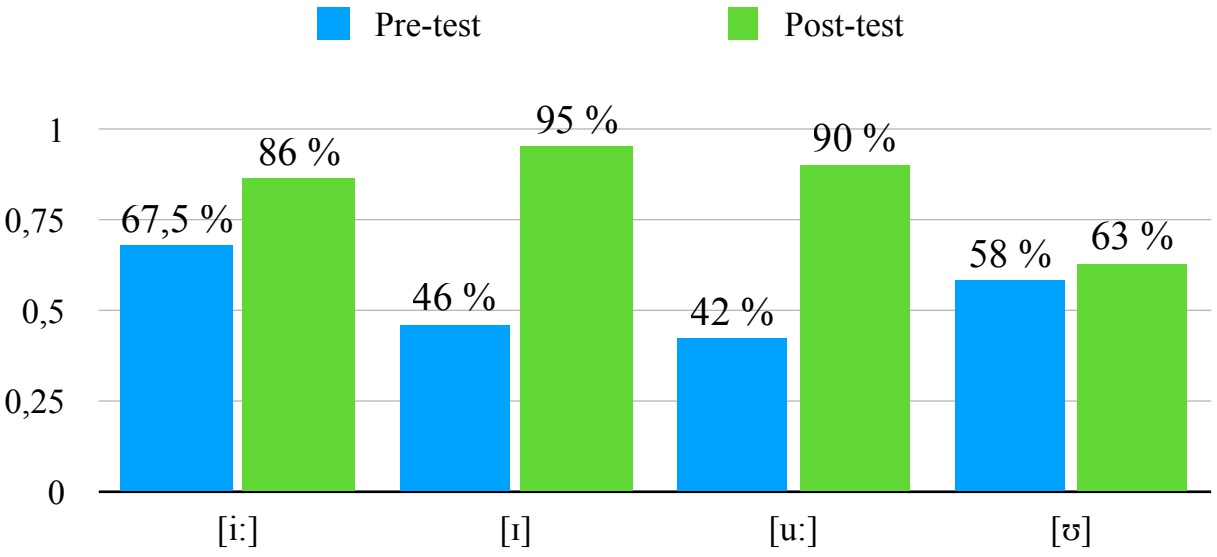

Diagram 2. *Pre-test and post-test results for [i:]–[ɪ] and [u:]–[ʊ] of the second experiment.*

An analysis of the results of the pre-test showed that the contrast pair [u:]–[ʊ] caused the greatest difficulties for the participants of the two experiments. The pair [i:]–[ɪ]

demonstrated a higher level of recognition compared to [u:]–[ʊ], however, significant difficulties were also observed here. The long vowel [i:] showed relatively high results in the word *beat* — 73% in the first experiment and 60% in the second. The short vowel [ɪ] was also perceived ambiguously: if the word *hit* was correctly recognised by all participants in the second experiment, then in the first experiment this indicator was only 42%. The word *slim* showed low results in both experiments (50% and 40%), which indicates instability in the perception of the short front vowel.

After watching the educational cartoon, the results of post-test in both experiments showed a noticeable improvement in the perception of all the vowels studied. A significant increase was observed in the recognition of the [u:]–[ʊ] pair, which caused the greatest difficulties at the pre-test stage. In the first experiment, the recognition rates [u:] increased to 81-88% (*rule, school*), and [ʊ] reached 73-96% (*cook, book*). In the second experiment, positive dynamics was also recorded: *rule* — 80%, *school* — 100%, *book* — 70%. This indicates the formation of a more stable phonological differentiation between long and short back vowels after multimodal intervention. The [i:]–[ɪ] pair also showed a marked improvement. In the first experiment, the words *sit* and *slim* were correctly recognised in 92% of cases, and long vowels in the word sheep — in 88%. In the second experiment, the results were even higher: *sit* — 90%, *slim* — 100%, *sheep* — 90%. Despite the fact that the word *beat* retained relatively low values, the overall dynamics demonstrates a significant improvement in the perception of the contrast pair [i:]–[ɪ].

The pair [u:]–[ʊ] initially presented the greatest difficulties for Russian-speaking students, probably due to the lack of a similar phonological opposition in the Russian language. This pair showed the most noticeable progress after multimodal training, which indicates the high effectiveness of visual and auditory support in developing phonemic perception skills of English vowels.

The third research question concerned the effectiveness of an educational cartoon on the development of phonemic perception in Russian-speaking schoolchildren. The effectiveness of animation is determined by a specific combination of visual, auditory, and

cognitive-emotional elements, each of which contributed to the formation of stable phonological representations of L1 schoolchildren. Firstly, the visual representation of sounds in the form of characters with well-defined characteristics created an associative link between sound and image. Secondly, the sound component of the cartoon was recorded taking into account the phonetic features of the hero sounds: the stretching of long vowels ([i:], [u:]), changing the timbre. Thirdly, the built-in interactive tasks and the game structure (passing through two "forests", receiving a medal) contributed to cognitive engagement and emotional memory. According to Mayer (2005), learning through story and active participation promotes deep information processing rather than superficial recognition. This is confirmed by an increase in the number of informed choices and a decrease in guessing, recorded when comparing the pre-test and post-test.

The cartoon played not just an auxiliary, but a central pedagogical role, acting as a means of purposeful phonetic learning, in which each modal component reinforced the other. The effect of multimodality in this case was manifested not in a general increase in interest, but in a systematic effect on the auditory, visual and associative systems. This effect is consistent with the provisions of the SLM model, according to which visually and acoustically emphasised L2 sounds are more likely to form a new phonemic category.

The results of both post-tests showed a clear and significant improvement across all target words. Particularly high rates were observed for those lexical units that were directly related to cartoon characters and were accompanied by vivid visualisation. In the first experiment, the highest results were obtained for the words *book* — 96%, *look* — 92%, *sit* — 92% and *slim* — 92%. Significant progress was also observed in the perception of long vowels: the word *rule* was correctly recognised in 85% of cases, *school* — in 81%, and the words *beat* and *leave* — in 69% of cases. In the second experiment, the highest results were shown by the words *school* and *slim*, which were correctly recognised by all participants. High rates were also recorded for the words *sit*, *sheep* and *leave* — 90% each, and the word *rule* reached 80% of correct answers. Even those stimuli that continued to cause some difficulties, such as *book* (70%) and *look* (60%), showed better results compared to the pre-test stage.

This contrasts with the results of the pre-test, where the same words often did not exceed 40-60% accuracy. The results indicate that multimodal learning does have a profound effect on the processes of phonological differentiation in L2 settings. Consequently, multimedia learning based on visualisation, story structure, character representation of sounds and active involvement of students is a powerful and effective tool for developing phonemic perception in children learning English as a foreign language. This is especially true in cases where target phonemes are absent from the learners' native language system and require additional mechanisms of comprehension and discrimination. Based on the results obtained, it is possible to recommend the inclusion of similar multimodal solutions in the school curriculum of English pronunciation. This technique can be adapted to different phonological contrasts and age groups, which makes it promising for wide-scale application in the field of primary language education.

Visual and bodily images accompanying sound contribute to the formation of stable articulatory memory (Kozlowska, 2016). The results obtained during the study confirm the effectiveness of the multimodal approach in teaching English phonetics. The combination of visual, classroom and interactive components implemented in the format of an educational cartoon has demonstrated a significant impact on the development of phonemic hearing in Russian-speaking schoolchildren, especially in relation to those sounds that are absent in L1.

To answer the fourth research question: to what extent is the effect of multimodal intervention on the assimilation of contrastive vowel pairs maintained based on the results of the delayed post-test, a delayed post-test was conducted. The results of the delayed post-test showed that the effect of multimodal learning was largely preserved after a week. Based on the data obtained (Diagram 3), I can conclude that the phonemic skills have been partially consolidated.

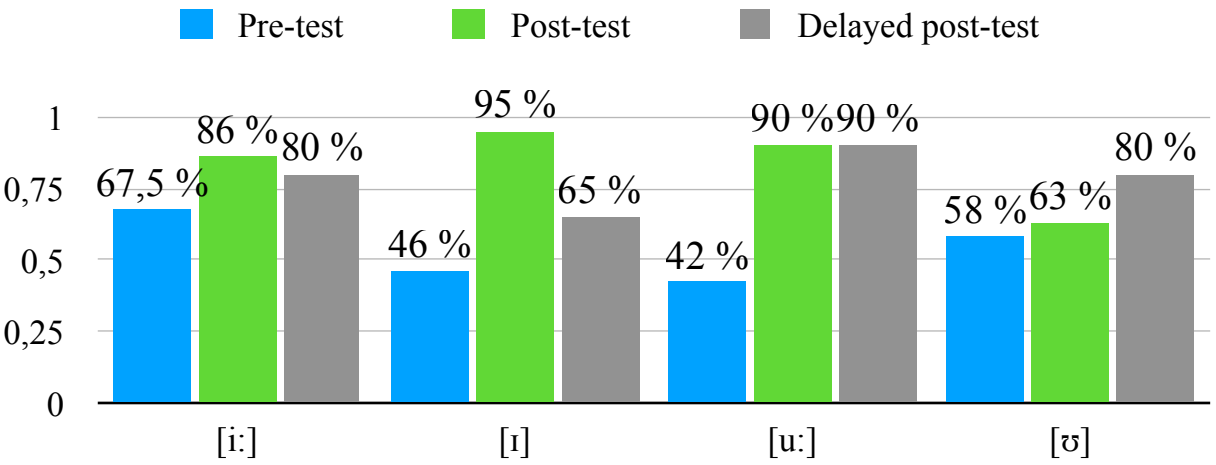

Diagram 3. *Pre-test, post-test, delayed post-test results for [i:]–[ɪ] and [u:]–[ʊ] of the second experiment.*

The most stable results were recorded for the [u:]. At the pre-test stage, this sound showed the lowest recognition rates (42%), after multimodal intervention, the accuracy increased to 90% and remained at the same level in the delayed post-test. The [i:] also showed steady positive dynamics. If the average correct answer rate in the pre-test was 67.5%, then after training it increased to 86%, and after a week it remained at 80%. For the short vowel [ʊ], a positive trend was noted at all stages of the study. In the pre-test, the average was 58%, in the post-test it was 63%, and in the delayed post-test it increased to 80%. I concluded that the indicator increased due to the preservation of visual-auditory associations. The least stable results were recorded for the short vowel [ɪ]. After a significant increase from 46% in the pre-test to 95% in the post-test, the indicator decreased to 65% in the delayed post-test. Despite this decrease, the results still exceeded the initial level of the pre-test, which indicates a partial preservation of the learning effect.

Answering the fourth research question, I can say that multimodal intervention had not only a short-term, but also a relatively stable effect on the development of perception skills of English contrasting vowels. The highest level of effect retention was recorded for the pair [u:]–[ʊ], while the contrast [i:]–[ɪ], especially the brief [ɪ], turned out to be less stable over time.

Chapter 4. Conclusions, shortcomings and promising directions

This study is one of the first attempts to systematically analyse the perception of English long and short vowels [iː], [ɪ], [uː], [ʊ] by Russian-speaking primary school students while simultaneously introducing an original multimedia learning resource. Special attention was paid to the extent to which the lack of opposition in length and quality in the Russian language affects the ability of schoolchildren to distinguish the target sounds of the English language.

As the results of the pre-test showed, Russian–speaking schoolchildren perceived pairs of English vowels [iː]–[ɪ] and [uː]-[ʊ] as similar, which is confirmed by a high proportion of errors and guesses in classroom isolation. This observation is consistent with the provisions of the SLM model, according to which sounds of a L2 that do not have clear phonological analogues in L1 hardly form new phonemic categories and are perceived as acoustically close. However, the data also demonstrate that perceived similarity does not always lead to unsuccessful identification: with multimodal support in the post-test, participants were able to successfully distinguish previously problematic pairs, despite their phonetic proximity. This indicates the importance of the conditions for presenting the material and supporting visual information, which can facilitate the formation of stable phonemic representations. The process of identifying L2 sounds in younger schoolchildren is determined not so much by their phonological description in terms of their native language, but rather by a complex combination of perception, cognitive load, and multimodal reinforcement. To further confirm these conclusions, it is advisable to conduct a more detailed survey of participants on the nature of their language experience and an extended experiment on the perception of vowels in various contexts, both in isolation and in speech units.

The experimental design of the present study has a number of limitations that should be taken into account when interpreting the results. First of all, the perception of vowels was assessed at the level of individual words presented in isolation, outside of a coherent context. Although this approach allows you to control variables and focus on the sound difference, it does not reflect the difficulties of perception in real speech, where intonation, prosodic and

phrasal characteristics play an important role. Within the framework of the SLM and PAM models, it is emphasised that the perception of L2 sounds depends not only on their acoustic and articulatory proximity to L1 sounds, but also on how they integrate into the speech stream, where there is an additional cognitive load. In the future, it seems advisable to expand the study and analyse the perception of the target vowel sounds [iː], [ɪ], [uː], [ʊ] as part of words located in various positions — initial, medial, and final — taking into account the phonetic context, prosodic features, and coarticulation effects. This will allow to better understand how Russian-speaking schoolchildren perceive the length and quality of vowels in more natural speech conditions, close to real communication.

The volume of the test material was limited to eight words in the pre-test and ten in the post-test. This decision was made taking into account the age of the participants and the need to maintain their attention and motivation during the testing process. Nevertheless, further expansion of the set of stimuli could contribute to a more reliable assessment of perception strategies and identify possible patterns in the nature of errors.

The first experiment of the study involved 26 primary school students in Russian secondary schools. All of them were native speakers of Russian, were at the A1 level of the CEFR and had not previously received systematic phonetic training in English. Despite the homogeneity of the group, a number of methodological limitations affect the interpretation of the results obtained. Firstly, the testing was conducted in an online format, without direct supervision of the participants. Thus, it is not possible to reliably determine whether they passed the test in an isolated and quiet environment, without distractions and outside help. These conditions could affect the level of concentration and the correct perception of sound material. Secondly, due to the complete anonymity of the procedure (the system did not record personal data, age, gender, or individual responses to each stimulus), the researcher did not have access to piecemeal error analysis and could not track individual progress trajectories. This significantly limits the possibility of more subtle qualitative analysis and comparisons between participants. Only generalised data on the correctness of recognition for each word was available for analysis, which, on the one hand, ensures ethical transparency,

but on the other hand, reduces the research depth and limits the statistical power of conclusions.

In the second experiment, these limitations were partially eliminated. The study involved 10 students between the ages of 10 and 11, who were tested as part of individual online classes. This format made it possible to directly monitor the process of completing tasks, control the testing conditions and minimise the influence of external factors. The results were recorded individually, which made it possible to track the dynamics of each participant, identify the most difficult phonetic contrasts, and analyse changes in the perception of vowels after a multimodal intervention. But the second experiment also had certain limitations. First of all, the sample was significantly smaller, which limits the possibility of generalising the results. Secondly, the participation of the researcher's own students may affect the objectivity of the data, since the participants were already familiar with the format of classes, the style of explanation of the material and the teacher's voice. This could have contributed to higher engagement and motivation, as well as a positive effect on the results of the post-test. In addition, the criteria for inclusion in the sample were based only on age, native language, and current English proficiency. At the same time, there was no preliminary assessment of phonemic awareness, which could explain the differences in results between the participants.

It is also important to check the stability of the results over time. In the first experiment, post-testing was performed immediately after the multimodal intervention, which did not allow conclusions to be drawn about the long-term effect of learning. In the second experiment, this limitation was partially overcome by conducting a delayed post-test one week after completing the training. This allowed us to assess the degree of preservation of the formed phonemic differences and the stability of the effect of multimodal intervention. Nevertheless, a longer follow-up period could provide more accurate data on long-term retention of English vowel discrimination skills.

Future research may focus on the long-term effects of such interventions, including delayed testing 2-3 weeks after exposure to assess the resilience of acquired skills. Taken together, this study highlights the importance of integrating multimodal strategies into

English phonetics teaching as an effective means of overcoming cross-linguistic difficulties and improving the quality of perception and reproduction of second language sounds.

Finally, a promising direction is to compare different types of multimodal effects and their impact on the formation of phonemic awareness. I hope that this study will serve as a starting point for further developments in the field of teaching phonetics and will contribute to the development of effective methods of teaching English pronunciation at the initial stage.

References

Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America, 133*(4), 2397–2411.

Avanesov, R. I. (1956). Phonetics of modern Russian literary language. *Moscow: Academy of Sciences of the USSR.*

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775–794.

Bohn, O.-S., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics, 21*, 267–290.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*, 433–461.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Chew, P. A. (2003). *A computational phonology of Russian.* Doctoral dissertation, dissertation.com.

Cruttenden, A. (2014). Gimson's pronunciation of English (8th ed.). *London: Routledge*.

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics, 83*, 128–143.

Dohen, M., & Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech, 52*(2–3), 177–206.

Duryagin, P. V. (2018). Duration and formant values of unstressed vowels in Russian as acoustic cues for segmentation: A perceptive experiment based on nonce words. *Russian Language Studies,* 16(3), 322–343. https://www.researchgate.net/publication/

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America, 62*(4), 971–974.

Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America, 139*(1), EL1–EL5.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89*, 395–411. https://www.ling.ohio-state.edu/research/groups/lacqueys/readings/earlier/flege1991_VOT.pdf

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). York Press.

Flege, J. E., & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition, 5*(1), 1–16.

Flege, J. E., Munro, M. J., & Skelton, L. (1994). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America, 92*(1), 128–143.

Ganapathy, M., & Seetharam, S. (2016). The Effects of Using Multimodal Approaches in Meaning-Making of 21st Century Literacy Texts Among ESL Students in a Private School in Malaysia. *Advances in Language and Literary Studies, 7*(4), 132–139.

Giegerich, H. J. (1992). *English phonology*. Cambridge University Press. https://archive.org/details/englishphonology0000gieg

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*, 131–149.

Idemaru, K., & Holt, L. L. (2011). Word Recognition Reflects Dimension-based Statistical Learning, *37*(6), 1939.

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge. https://archive.org/details/routledgehandboo0000unse_a9b3

Jingpeiyi, X. (2021). The Development of English Phonetics in Multimodal Teaching. *Journal of English Teaching Research, 5*(2), 45–59.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics, 57*, 21–39. https://www.sciencedirect.com/science/article/pii/S0095447016300080

Knyazev, S. V., & Pozharitskaya, S. K. (2011). Modern Russian literary language: Phonetics, graphics, orthography, orthoepy. *Moscow: Akademicheskiy Proekt*.

Kendon, A. (2004). Gesture: Visible Action as Utterance. Cambridge University Press.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America, 124*(6), 3959–3971. https://doi.org/10.1121/1.2990709

Kozlowska, K. (2016). Verifying a holistic multimodal approach to pronunciation training of intermediate Polish learners of English. *Research in Language, 13*(2), 123–138.

Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. Routledge.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(2), 11850–11857.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (4th ed.). Heinle & Heinle.

Lado, R. (1957). *Linguistics across cultures: Applied Linguistics and Language Teachers.* University of Michigan Press, Ann Arbor.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America, 128*(6), 3757–3768.

Lenneberg, E. H. (1967). Biological foundations of language. Wiley.

Letova, A. A. (2022). Comparative analysis of the North American and Russian vowels. *Young Scientist*, 44(439), 379–385. https://moluch.ru/archive/439/95889

Levis, J. (2018). Intelligibility, oral communication, and the teaching of pronunciation. Cambridge University Press.

MacKay, I. R. A., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics, 27*, 157–183.

Makarova, A. O. (2010). Acquisition of three vowel contrasts by Russian speakers of American English [Doctoral dissertation, Harvard University].

Mayer, R. E. (Ed.). (2005). The Cambridge handbook of multimedia learning. Cambridge University Press.

Roach, P. (2009). English phonetics and phonology: A practical course (4th ed.). Cambridge University Press.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*, 411–417.

Shevyakova, V. E. (1968). Corrective Phonetics Course of the English Language. Nauka Publishing.

Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *Journal of the Acoustical Society of America, 144*(1), 92–103.

Vasiliev, V. A., Katanskaya, A. R., Lukina, N. D., Maslova, L. P., & Torsueva, E. I. (1980). *English Phonetics: A Normative Course* (2nd ed., rev.). Vysshaya Shkola Publishing.

Wayland, Ratree (Ed.). (2021). Second Language Speech Learning: Theoretical and Empirical Progress. Cambridge University Press.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences, 12*(4), 144–150.

Xodabande, I. (2017). The effectiveness of social media network Telegram in teaching English language pronunciation to Iranian EFL learners. *Teaching English with Technology, 17*(2), 67–86.

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition, 10*(2), 147–160.

Appendix 1

Università Ca'Foscari Venezia
Dipartimento di Studi Linguistici e Culturali Comparati
BemboLab
BimbiLab
—
Ca' Bembo
Dorsoduro 1075
30123 Venezia
T +39 041 2345738
bembolab@unive.it

Информированного согласие на обработку персональных данных

Уважаемые родители/лица, осуществляющие родительские обязанности, Настоящее исследование проводится **Матрос Анастасией Сергеевной** под руководством **Павла Дурягина** из Департамента сравнительных лингвистических и культурных исследований Университета Ка' Фоскари в Венеции через онлайн-платформу Interacty. Приняв данную форму, вы даёте своё согласие на участие ребёнка в исследовании и связанных с ним активностях.

Участие в исследовании является добровольным. Вы можете прекратить участие в любой момент без каких-либо негативных последствий. Подписывая согласие, вы разрешаете исследователям хранить ваши личные данные и данные ребёнка в бумажном/электронном виде и обрабатывать их конфиденциально на протяжении всего проекта. Все собранные данные будут обезличены и не позволят установить личность ребёнка, в соответствии с этическим кодексом и действующими нормами Университета Ка' Фоскари в Венеции.

Обработка данных будет анонимной в соответствии с Регламентом ЕС 2016/679 и Законодательным декретом № 196/2003. Результаты анализа могут быть опубликованы в дипломных работах, книгах или научных статьях в обобщённом и анонимном виде. Вы можете в любой момент изменить, уточнить или отозвать согласие, обратившись к ответственному за сбор данных.

Настоящее исследование и формы, которые вас просят заполнить, получили одобрение Этической комиссии Университета 05.02.2020, протокол № 1/2020.

Методология исследования:
Данное исследование предназначено для детей в возрасте от 7 до 10 лет. Главная цель — изучить способность восприятия долгих и кратких гласных английского языка российскими школьниками. Могут быть использованы изображения, анимации, аудиозаписи, слова, фразы и тексты. Все задания проводятся в непринуждённой обстановке, и вы можете присутствовать рядом с ребёнком. Также может быть

предложено заполнить короткую анкету о языковом и образовательном фоне.

Контакты:
По любым вопросам о процедуре или для изменения/отзыва согласия на участие можно связаться с:

- Ответственным за сбор данных: Матрос Анастасия Сергеевна, +79117190094, nastya.matros@mail.ru
- Другие контакты: bembolab@unive.it, тел.: 041/2345738 - 041/2345748

Информация об обработке персональных данных в рамках проекта «Using Animation to Raise Phonetic Awareness of English Vowels Contrast in Russian L1 Schoolchildren» в соответствии со статьёй 13 Регламента ЕС 2016/679.

Ответственным за обработку является Университет Ка' Фоскари Венеция, юридический адрес: Dorsoduro 3246, 30123 Venezia.

Университет назначил ответственного по защите данных. Связь: dpo@unive.it или по адресу Dorsoduro 3246, 30123 Venezia.

Обрабатываемые категории данных:
Анкетные данные, контактные данные, а также особые категории (например, сведения о здоровье, расовое/этническое происхождение, политические взгляды и др.). Сбор данных может проводиться с помощью форм, аудио- и видеозаписей, наблюдений, интервью и т.д.

Данные будут обезличены и использоваться исключительно в рамках проекта. Правовая основа обработки: статья 6.1.e) и 9.2.a) Регламента ЕС. Вы можете отозвать согласие в любое время.

Срок хранения данных:
Данные хранятся в течение всего проекта, после чего они <будут обезличены/удалены/сохранены на срок X лет и затем обезличены>. Анонимные данные могут использоваться в других исследованиях.

Данные могут обрабатываться исследователями и техническим персоналом, а также передаваться в агрегированной форме другим университетам и организациям. Документация может быть проверена этическими комитетами и аудиторами.

Вы можете воспользоваться правами, предусмотренными Регламентом: доступ, исправление, удаление, ограничение или возражение против обработки. Запрос можно направить ответственному за сбор данных или по адресу dpo@unive.it.

Если вы считаете, что ваши данные обрабатываются с нарушениями, вы можете подать жалобу в надзорный орган.

Раздел для родителей: заполнение персональных данных и подписи с согласием или отказом от участия ребёнка.

Дополнительное согласие на обработку особых категорий персональных данных: инвалидность, происхождение, убеждения и др.

Appendix 2

Consent to Data Processing

Dear parents/legal guardians,

This study is conducted by Anastasiia Sergeevna Matros under the supervision of Pavel Duryagin from the Department of Linguistic and Comparative Cultural Studies at Ca' Foscari University of Venice, through the online platform Interacty.

By accepting this form, you consent to your child's participation in the study and related activities.

Participation is voluntary. You may withdraw from the study at any time without any negative consequences. By signing this consent form, you authorise the researchers to store your and your child's personal data in paper and/or electronic format and to process it confidentially for the entire duration of the project. All data collected will be anonymised and will not allow identification of the child, in accordance with the ethical code and current regulations of Ca' Foscari University of Venice.

Data processing will be conducted anonymously in accordance with EU Regulation 2016/679 and Legislative Decree No. 196/2003. The results of the analysis may be published in a thesis, books, or scientific articles in aggregated and anonymous form. You may modify, clarify, or revoke your consent at any time by contacting the person responsible for data collection.

This study and the related consent forms were approved by the University Ethics Committee on 05/02/2020, protocol no. 1/2020.

Research Methodology

The study is aimed at children aged 7 to 10. The main objective is to investigate the perception of English long and short vowels by Russian-speaking schoolchildren. Images, animations, audio recordings, words, sentences, and texts may be used. All tasks will be carried out in a relaxed environment, and the presence of an adult next to the child is allowed.

A short questionnaire about the child's linguistic and educational background may also be requested.

Contacts

For any questions regarding the procedure or to modify/revoke consent, please contact:

- Data Collection Supervisor: Anastasiia Sergeevna Matros, +7 911 719 0094, nastya.matros@mail.ru
- Additional contact: bembolab@unive.it, tel.: +39 041 2345738 / 2345748

Information on Personal Data Processing

As part of the project *"Using Animation to Raise Phonetic Awareness of English Vowels Contrast in Russian L1 Schoolchildren"*, pursuant to Article 13 of EU Regulation 2016/679.

The Data Controller is Ca' Foscari University of Venice, registered office: Dorsoduro 3246, 30123 Venice.
The University has appointed a Data Protection Officer. Contact: dpo@unive.it or at the office Dorsoduro 3246, 30123 Venice.

Categories of processed data: personal data, contact details, and special categories of data (e.g., health information, ethnic origin, political opinions, etc.). Data may be collected via forms, audio and video recordings, observations, interviews, etc.

All data will be anonymised and used exclusively within the scope of the project.
Legal basis for processing: Articles 6.1.e) and 9.2.a) of the EU Regulation. Consent may be revoked at any time.

Data Retention Period

The data will be retained for the entire duration of the project, after which it will be <anonymised/deleted/retained for X years and then anonymised>. Anonymous data may be used in other studies.

The data may be processed by researchers and technical staff and shared in aggregated form with other universities and institutions. Documentation may be reviewed by ethics committees and auditors.

You have the right to exercise the rights provided by the Regulation: access, rectification, deletion, restriction, or objection to processing. Requests may be sent to the data collection supervisor or to dpo@unive.it.
If you believe that your data is being processed in violation of the law, you may lodge a complaint with the supervisory authority.

Parent/Guardian Section

Completion of personal information and signature to consent or refuse the child's participation.

Additional consent:
to the processing of special categories of personal data: disability, origin, beliefs, etc.

Appendix 3

Educational cartoon

Cartoon Script

Scene 1:

Steve: Hi, guys! My name is Steve. I love searching for treasures and solving mysteries. What's this? A map?

Scene 2:

Map: Hello, travelers! Today you're in for an amazing adventure! We're going to visit two magical forests — the Short Vowels Forest and the Long Vowels Forest. Help the forest dwellers figure out which sounds are short and which are long. Pay attention, because only those who can tell the difference will be able to open the Magical Sound Chest.

Scene 3:

Steve: Wow! Two forests with different sounds? That sounds super cool. I'm ready, are you? -Pause-

Steve: Then let's go!

Scene 4:

Steve: Whoa… Here I am in the Short Vowels Forest. It's so beautiful here! Who lives in this place?

Characters appear

Uggie: Hi! I'm Uggie! I'm the /ʊ/ sound, like in "cool".

Scene 5:

Uggie: Let's play a game! Listen carefully and repeat the words after me. Try to hear the sound and remember — my sound is short. That's really important for the journey! Ready? Let's go!

Scene 6: -Sound repetition activity-

Scene 7:

Ikki: Hello! I'm Ikki! A fast and short /ɪ/ sound — like a grasshopper's jump! Let's practice together!

-Sound repetition activity-

Scene 8:

Steve: Thank you, friends! Now I understand short vowels in English much better. You were amazing helpers. But now we have to go to the next forest!

Scene 9:

Map: Well done! You've made it through the Short Vowels Forest. Where are we going next? -Pause-

That's right — to the Long Vowels Forest! Good luck!

Scene 10:

Steve: Whoa! It feels totally different here! The trees are huge! Looks like we're in the Long Vowels Forest.

Nuuki: UUU, UUU, UUU Hellooooo! I'm Nuuki, the /uː/ sound. Long and smooth, uuuu!

Niiki: And I'm Niiki, the /iː/ sound. I'm light and long, like the song of the wind. IIII! Let's play together!

Scene 11: -Practice session-

Scene 12:

Steve: Thank you, Nuuki and Nikki! Thanks to you, we learned what long vowels sound like. But it's time for us to go! See you soon!

Scene 13:

Map: Great job! Kids, you and Steve have made it through both forests — Short Vowels and Long Vowels. The next stop is the Magical Sound Chest!

Scene 14:

Steve: Phew, I think we're here. To enter the house, you must loudly shout the word "Vowels"… Let's do it one more time.

Scene 15:

Steve: There it is — the Magical Sound Chest. It only opens for those who listened well and remembered the sounds. Listen carefully to the words with familiar sounds and choose the right answer!

Scene 16: -Quiz-

Scene 17: -Medal opens-

Steve: Wooow!

Scene 18:

Steve: We did it! Now we're real Sound Explorers. Thanks to everyone who helped on this journey. Now I know for sure — listening means understanding. And that's true magic.