https://doi.org/10.17323/jle.2025.19916

Intelligent Approaches to Computer Testing of Perception and Production Skills of Russian EFL Speakers

Marina Kolesnichenko [1], **Vitalii Kapitan** [2]

[1] Far Eastern Federal University, Vladivostok, Russia

[2] National University of Singapore, Singapore

ABSTRACT

Background: This study addresses a gap in applied phonetics by developing an evidence-based, neural network-driven computer phonetic test for Russian EFL learners. Integrating interdisciplinary methods, the system targets Russian-specific pronunciation deviations and delivers adaptive feedback, thereby advancing both perception and production skills while aligning technological innovation with pedagogical effectiveness.

Purpose: The purpose of the study is twofold: (1) to design and develop a computer-based system employing deep learning neural networks for objectively assessing Russian EFL students' perception and production skills, and (2) to evaluate the effectiveness and reliability of this system through repeated testing, statistical analysis of learner performance and user feedback.

Methods: A pre-test identified frequent segmental deviations, informing a targeted item pool. The software was developed in Microsoft Visual Studio 2022 (C#) using the Microsoft Speech Recognition Engine. The perception module used randomized audio stimuli (WAV files), while the speech recognition one recorded response via built-in microphones for automated accuracy evaluation. Twenty-five Russian EFL students (B1–B2 CEFR, aged 19–22) completed three test iterations at one-week intervals. Post-test questionnaire assessed usability and perceived learning gains. Data were analysed using descriptive statistics and correlation analysis.

Results: We designed a computer-based system employing deep learning neural networks and assessed its efficiency in Russian EFL learners. The study found a 14.5% overall improvement in participant performance, with results showing a clear linear increase supported by a high R^2 value. Students performed better in perception tasks than in production practice. Pearson correlation analysis indicated consistent performance between consecutive attempts, supporting robust test-retest reliability. Both modules showed high internal consistency ($\alpha = 0.90$ for perception, $\alpha = 0.88$ for production). Participants rated the tool as useful and interesting, although they suggested improving the speech recognition function due to minor technical flaws.

Conclusion: The module focused on testing perception skills can serve as an effective and engaging learning tool. While the pronunciation control component shows potential, its performance can be further enhanced through additional testing with high-sensitivity microphones to refine speech recognition accuracy. Overall, continued exploration of CAPT systems presents a promising direction for future research and innovation.

KEYWORDS

interdisciplinary approach, perception and production skills, speech recognition, neural networks, computer testing in phonetics, CAPT, Russian EFL speakers

Citation: Kolesnichenko M., & Kapitan V. (2025). Intelligent Approaches to Computer Testing of Perception and Production Skills of Russian EFL Speakers. *Journal of Language and Education, 11*(2), 75-93. https://doi.org/10.17323/jle.2025.19916

Correspondence:
Vitalii Kapitan,
kapitanvy@gmail.com

Received: March 14, 2024
Accepted: June 10, 2025
Published: June 30, 2025

INTRODUCTION

Computational linguistics emerged in the mid-twentieth century as a response to the growing challenges in information technology, with a primary focus on enabling computers to process and understand natural language (Luz, 2022). This specialized field has facilitated technological advancements, including the development of speech recognition systems, linguistic corpora, and machine translation tools. By leveraging interdisciplinary collaboration between linguistics, computer science, and engineering, computational linguistics has laid the

foundation for innovative educational systems and improved data interaction across scientific and educational domains (Omid, 2022; Urip, 2022). These developments have set the stage for further exploration of language processing technologies and their applications in language learning.

Artificial neural networks have played a pivotal role in advancing computational linguistics by modelling complex relationships between input and output data, mirroring the functioning of biological neural networks in the human brain. These networks consist of interconnected layers-input, hidden, and output-that enable the system to perform tasks such as classification, clustering, and forecasting (Ivanko et al, 2019). The integration of neural networks into language technologies has resulted in the creation of automatic speech recognition (ASR) and speech-to-text (STT) systems, which are now widely used in both research and practical applications (Belenko & Balakshin, 2017; Mehrish et al., 2023; Dovchin, 2024). As a result, they have become integral to the ongoing evolution of computational linguistics and its role in educational innovation.

At the same time the study of speech production and perception has significantly contributed to the field of applied phonetics, providing valuable insights into the mechanisms underlying spoken language (Crystal, 1970; Flege & Davidian, 2008; Vishnevskaya, 2014; Munro & Derwing, 2020). Research in this area has provoked the development of pronunciation training tools and methodologies, with a particular emphasis on the variability and complexity of natural speech. These findings have been instrumental in shaping computer-assisted pronunciation training (CAPT) systems, which aim to enhance learners' pronunciation skills through targeted, technology-driven instruction (Fouz-González, 2020; Alsuhaibani et al., 2024). The intersection of applied phonetics and computational tools has thus expanded the possibilities for effective language learning interventions.

CAPT systems are reported to aid in improving foreign language pronunciation (Wang & Munro, 2004; González & Ferreiro, 2024; Rogerson-Revell, 2021). Many of them leverage the principle of High-Variability Phonetic Training (HVPT), exposing learners to a wide range of speech samples to improve perceptual accuracy. Several studies propose innovative developments in this domain (Barriuso & Hayes-Harb, 2018; Thomson & Derwing, 2015; O'Brien et al, 2018). Notable examples include the "English Accent Coach"1, which uses gamification and multiple native speakers to increase learner motivation and effectiveness. Some researchers offer computer-based systems for pronunciation training in other languages save English. Blok (2019) developed a methodology for evaluating consonant pronunciation errors in German speech among Russian-speaking students. Similarly, Pashkovskaya (2010) created a flexible program that includes rhythmic-rhyming tasks aimed at improving the phonetics and intonation of Russian for students of various nationalities, with recommendations for pronunciation training in 17 languages. Other CAPT software integrate automatic speech recognition (ASR) and artificial intelligence to increase the potential for individualized learning outcomes (Rogerson-Revell, 2021).

Despite these technological advancements, the use of CAPT tools has often been criticized for placing greater emphasis on technological aspects, such as ASR and visual feedback, at the expense of sound pedagogical principles. This imbalance has created a gap between innovative technological tools and their educational value, with critics pointing to an over-reliance on repetitive drilling or mechanical exercises. Scholars (Pennington and Rogerson-Revel, 2019; Zou et al., 2024) have highlighted the need for a more thoughtful alignment between technology and teaching methodologies.

Additionally, many studies are noted for lacking robust designs, particularly the absence of control groups or delayed post-tests, which compromises the reliability and generalizability of the findings (Bliss et al., 2018; Agarwal, 2019). Moreover, the predominant focus on university-level learners limits the applicability of insights to other learning contexts, as noted by Mahdi and Al Khateeb (2019), and Rogerson-Revell (2021). Further criticism centers around the tendency of CAPT tools to adopt a one-size-fits-all approach. These systems often provide generalized feedback and learning materials, without sufficient customization to address individual learner needs. Derwing and Munro (2015) and Levis (2018) have consistently emphasized the necessity of more sophisticated and personalized feedback systems to enhance the effectiveness of pronunciation training.

Building on these insights, the present study aims to bridge the gap between technology and pedagogical effectiveness in computer-assisted pronunciation training for Russian EFL learners. By leveraging deep learning neural networks and incorporating principles from teaching methodologies, applied phonetics and psycholinguistics we seek to develop and empirically validate a computer-based system that addresses and evaluates Russian-specific pronunciation deviations.

Consequently, the aims of the present study were twofold: (1) to develop a computer-based system utilizing deep-learning neural networks to monitor Russian EFL students' perception and production skills; (2) to quantitatively evaluate the system's effectiveness through pre- and post-test comparisons, error rate analysis, and user feedback, ensuring accuracy, reliability, and practical value for EFL instruction.

[1] Thomson, R. I. (2017). English accent coach [Computer program]. Version 2.3. https://www.englishaccentcoach.com/

LITERATURE REVIEW

Linguistic, Psycholinguistic, and Cognitive Aspects in Modelling English Language Sound System for a Test

Phonetic and phonological competence is a critical component of communicative competence for EFL learners in non-English-speaking environments. This competence comprises internalized knowledge of the target language's sound system, perceptual and articulatory skills, and the ability to apply them effectively in communication (Goncharova, 2006). At each stage of knowledge formation, it is important to monitor acquired skills and analyze any potential deviations. Computer testing can be a useful tool in this process. However, designing such tests requires an understanding of skill formation processes, including cognitive, linguistic, psycholinguistic, and pedagogical aspects (Pennington & Rogerson-Revell, 2019; Flege & Bohn, 2021).

The motor theory of speech perception posits that phoneme acquisition relies on articulatory and acoustic cues rather than isolated auditory input (Stratton, 2025). Research supports this, demonstrating how resonant frequencies in the vocal tract, acoustic changes, and speech timing, influence speech quality (Leonov & Sorokin, 2007; Lam et al., 2012). Clear speech contributes more to identification accuracy revealing that information in the signal from the productions was crucial in facilitating word identification (Redmon et al., 2020; Sereno et al., 2025). Additionally, kinesthetic feedback plays a critical role in articulation control, allowing speakers to self-monitor and adjust their speech. For instance, for Russian EFL learners, contrasting native and English phonemes (e.g., apical-alveolar [s], [z] vs. dorsal-dental [c], [ɜ]) enhances awareness of phonetic differences. Effective foreign language acquisition thus relies not on imitation alone but on systematic analysis and comparison of speech signals (Pashkovskaya, 2010).

This systematic comparative practice should aid EFL learners in overcoming the side-effects of phonetic transfer. The latter is defined as the way where native-language sound system interferes with target-language perception and production (Mooney, 2019). To mitigate transfer effects, learners must develop a second phonological system through structured practice, reinforcing new auditory and articulatory patterns (Shevchenko, 2017). It is necessary to develop the movements for correct articulation of sounds (Stratton, 2025), using both auditory and motor analyzers through exercises that can restructure the speech functional system and develop new perceptual and articulatory images in the brain of a foreign language learner.

Phonological systems are hierarchically structured, with phonemes serving as mental prototypes for allophonic variations (Kulikov, 2005). Syllables, as the smallest functional units, provide perceptual cues for phonemic contrasts[2] (Bondarko, 1969). Cognitive linguistics extends this schema, demonstrating that speakers categorize sounds mentally, even without motor execution (Nesset, 2008). Modern cognitive phonology has shifted its focus from cataloging phonemic inventories to investigating the dynamic processes underlying phonological system formation and change (Ohala, 2013). The phonological system's schematic structure, rooted in structural and generative phonology, allows for the categorization and mental representation of phonological units such as phonemes, allophones, and syllables. These schematic relationships reinforce the cognitive processes involved in language learning and are enhanced through targeted exercises, gradually expanding the learner's phonetic and phonological repertoire, and enabling them to internalize L2 sound patterns algorithmically, a principle that also underpins computational speech recognition. The abstraction and algorithmization of these schematic relationships have facilitated the development of logical models for computer-based speech recognition systems, including those employing deep learning neural networks.

Deep Learning Neural Networks and Their Role in Speech Recognition

Linguistic analysis was founded on information theory techniques in the early days of computer technology when digital computers had just recently been introduced. These techniques fit in nicely with the philosophical and psychological ideas of the day. The application of deep neural networks has undergone a dramatic change in the last few decades (Mcshane & Nirenburg, 2021; Backus et al, 2023; Dovchin, 2024). With the use of this technology, performance on tasks involving natural language processing as well as numerous other speech, vision, and cognitive issues have improved dramatically. Tasks that were previously thought to be unsolvable can now be explored and solved due to advances in neural network technologies.

The concept of artificial intelligence and the era of formal language theory have arrived. This resulted in novel approaches for language processing and analysis while cognitive science was developing (Church & Liberman, 2021; Tikhonova & Raitskaya, 2023; Joshi et al, 2025). The application of deep learning neural networks, which have numerous layers of neurons and are based on principles of how the human brain functions, is one of the major ideas of modern times for the advancement of machine learning and artificial intelligence. Unlike previous methods that usually required significant manual feature engineering, these net-

2 Potapova, R. K. (1986). Syllabic phonetics of Germanic languages. Vysshaya shkola.

works learn hierarchical representations directly from data. For speech processing tasks, convolutional neural networks (CNNs) (Fukushima, 1980, p. 396; LeCun, 1998, p. 2278) and recurrent neural networks (RNNs) (Mikolov et al., 2010, p. 1045; Graves et al., 2013; Su & Kuo, 2022) have proven effective for speech processing tasks Li et al, 2024; Rudregowda et al, 2024).

CNNs, originally developed for image recognition, excel at extracting spatial features from spectrograms, visual representations of audio frequencies over time. The CNN architecture uses convolutional layers, which apply filters to all inputs to detect patterns regardless of their position. These networks process spectrograms through a series of operations: convolutions to extract features, activation functions such as ReLU to introduce nonlinearity, and pooling to reduce dimensionality while preserving important information (Fig. A.1, see Appendix A for complete proofs). This approach allows CNNs to identify critical acoustic features such as formant transitions and phonetic boundaries that distinguish speech sounds.

RNNs address the sequential nature of speech by incorporating memory mechanisms that retain information across time steps. Unlike traditional feedforward networks, RNNs incorporate feedback connections that allow previous outputs to influence ongoing processing, creating an internal state that functions as a dynamic memory. This architecture is particularly suited to modeling temporal dependencies in speech, where the interpretation of current sounds depends on prior context (Fig. A.2, see Appendix A for complete proofs). Advanced variants such as long short-term memory (LSTM) networks overcome the limitations of basic RNNs by selectively retaining information in extended sequences, making them particularly valuable for recognizing related speech patterns.

The CNN and RNN combination create hybrid systems that use the strengths of both architectures. In these systems, CNNS first processes a spectrogram to extract reliable acoustic features, which are then served in RNN, which simulate temporary speech dynamics. This approach was surprisingly effective, since it solves both the problems of extraction and the consistent nature of speech recognition in a single structure. Modern speech recognition systems often realize this hybrid approach along with attention mechanisms that help the network focus on the corresponding parts of the input sequence, further increasing the accuracy of recognition in different linguistic contexts and performance conditions (Soundarya et al., 2023; Mehrish et al., 2023).

Deep neural networks have advanced automatic speech recognition, evolving with improvements in hardware and algorithms. Modern systems use convolutional and recurrent networks, making deep learning vital in computational linguistics to improve recognition accuracy.

Through our literature review, we found that researchers use various methods and tools to address participant needs in learning environments. However, further investigation is needed to identify suitable CAPT tools for classroom integration. A key challenge is that many CAPT systems are designed for specific research goals and may not apply well to diverse learning contexts. This is understandable as scientists have different research goals and focus on different aspects (Nickolai et al, 2024). However, any effort in developing and implementing CAPT tools in modern education is valuable and relevant. Having examined linguistic, psycholinguistic, and cognitive aspects of phonetic skills formation and intelligent approaches to solving speech processing tasks, we set a goal to create a tool using neural networks that could effectively control the progress of Russian EFL students in acquiring auditory and pronunciation skills in English[3,4]. We, then, tested its effectiveness in Russian EFL learning environment.

METHOD

The following sections provide a detailed overview of the participants, context and approaches used in the study.

Participants

The study involved twenty-five first-year students enrolled in the Bachelor's Degree programs in Applied Linguistics or English Philology, majoring in English as a foreign language (EFL). The participants comprised eighteen female and seven male students, aged between 18 and 19 years (B1–B2 CEFR). All participants had recently completed an Introductory Phonetics Course designed specifically for Russian EFL learners. This sample was selected to represent typical learners at the initial stage of formal phonetic training within the Russian higher education context.

[3] Korshunova, Y. S., Kapitan, V. Y. & Kolesnichenko, M. A. (2020a) Komp'uterniy test dly kontroly sluhoproiznositel'nix navikov [Computer test for monitoring students' auditory pronunciation skills]. (Certificate of State Registration of Computer Program No. RU 2020612357). The Federal Institute of Industrial Property, Rospatent. https://www1.fips.ru/fips_servl/fips_servlet?DB=EVM&DocNumber=2020612357

[4] Korshunova, Y. S., Kapitan, V. Y. & Kolesnichenko, M. A. (2020b). Razrabotka sistemi komp'uternogo testirovania dly kontroly sluhoproiznositel'nix navikov u studentov [Development of a computer testing system for monitoring students' hearing and speaking skills]. *Materials of the regional scientific-practical conference of students, graduate students and young scientists in natural sciences* (pp. 82-83). Far Eastern Federal University.

Context

Russian EFL students follow an intensive (36 hours) one-month Introductory Phonetics Course (IPC) that presents a part of a broader Practical Phonetics curriculum aimed at developing their phonetic competence. The course curriculum includes a variety of activities intended to activate and deepen students' phonetic knowledge. These activities encompass understanding interference phenomena between Russian and English phonological systems, comparative analysis of the two languages' sound structures, auditory discrimination of English speech sounds versus native Russian sounds, error identification and correction, self-recording of speech samples, etc. The primary objective of the IPC is to facilitate the formation of a secondary phonological system in English, enabling students to recognize and overcome common pronunciation difficulties. Special emphasis is placed on training auditory and motor analyzers to develop both perceptual and productive pronunciation skills essential for mastering English phonetics. To additionally monitor the development of these skills by the end of the IPC a computer phonetic test was created.

Task

To address our goals, as outlined in the Introduction, we structured the research process into four sequential tasks presented below.

Pre-Test Preparation and Administration

The initial phase involved the preparation and administration of a pre-test to identify typical phonetic challenges faced by Russian EFL learners. This phase was based on a comprehensive analysis of the typology of Russian and English phonetic systems (Arakin, 2008), phonetic transfer effects and differentiation between typical and fossilized articulation deviations in English - Russian language pairs (Vishnevskaya, 2014). The pre-test aimed to reveal and verify among the participants the most prevalent difficulties in English phonetic acquisition. For instance, vowel and consonant substitution, challenges with vowel length, devoicing of final consonants, and syllable division. A detailed account of the specific test items and their corresponding phonetic phenomena is provided in Appendix B.

The next task of the study was to process the results of the pre-test and develop a computer-based system for testing auditory and production skills, particularly focusing on the sound and syllabic structure of English vocabulary.

Development of a Computer-Based Auditory Skills Testing Module

To create this module, we used the Microsoft Visual Studio 2022 development environment and the C# programming language (Schildt, 2010). This enabled us to create a Windows application with a user-friendly interface, using the latest version of the .Net Framework 4.8 for optimal performance on Windows OS. The Windows Forms technology, which is part of the .NET Framework, provides a set of managed libraries that simplify the development and implementation of applications, ultimately ensuring a high level of usability for end-users, such as teachers and students[5]. The application was designed to read audio files and task texts from a secure directory and to allow for the dynamic expansion of tasks without requiring source code modification or recompilation. To ensure the authenticity and accuracy of the auditory stimuli, recordings were produced by a native American English speaker (a 32-year-old male English teacher) using a Sony ICD-TX650 digital voice recorder in a traditional laboratory setting. All audio files were classified and securely stored within the computer system.

The first module focuses on assessing phonemic hearing. The test tasks are designed to evaluate the ability to perceive the meaningful units of the English language and to perform phonemic actions, such as phonemic differentiation, determination of the distinctive function of a phoneme, segmentation of words into phonetic components (syllables), sound analysis of words, and selection of sounds in a specific order.

Development of a Computer-Based Speech Recognition Module

The third task of the study was to develop and implement a function for English speech recognition to control Russian EFL subjects' pronunciation. To solve that issue, we utilized intelligent neural network technologies, specifically speech recognition tools based on the Microsoft Speech Recognition Engine. This technology allows for real-time conversion of audio streams into text using the Speech Recognition Engine object, which enables the application to recognize words spoken into the microphone[6].

In total, two modules comprise 60 questions, divided into 6 task blocks. Thus, the main directory contains 6 subdirectories with different types of tasks, and each of them contains audio (except for tasks for the second module) and text documents in the *.ssv format.

[5] Windows Forms overview. (2023). https://learn.microsoft.com/en-us/dotnet/desktop/winforms/windows-forms-overview?view=netframeworkdesktop-4.8.

[6] Get Started with Speech Recognition (Microsoft.Speech) - Microsoft Speech Platform SDK 11 Documentation. https://documentation.help/Microsoft-Speech-Platform-SDK-11/4ca93e5c-65c9-433a-95c7-4343d8db269c.htm.

Data Collection, Analysis and Post-Evaluation

The final task was to analyze and evaluate the results of testing our tool to ensure its accuracy and effectiveness. The respondents had three attempts (with a week's interval between them) to work with the test. Suggesting that the subjects should have three attempts we targeted the following: (1) checking whether there might be progress in students' performance (overall performance improvement across attempts); (2) finding out drawbacks in the test semantic structure and evaluating its technical characteristics, (3) evaluating test reliability and validity.

Following the final test administration, a 10-item feedback questionnaire was developed, incorporating both structured (tabular) and open-ended response formats (see Fig. C.1, Fig. C.2 with sample questions in Appendix C for complete proofs). This instrument was designed to provide qualitative feedback and user experience on such aspects as: time allocation, accuracy of responses (correct vs incorrect answers), usability of the interface, clarity of task instructions, perceived difficulty of the test, usefulness, and overall technical performance of the testing system.

Procedure

Participant Briefing

The participants were given a clear explanation of the goals of the study. Prior to their involvement, they provided verbal consent, signifying their willingness to participate as subjects in a testing procedure.

Testing Environment, Protocol and Administration

Testing took place in a computer lab and was integrated as in-class activity regarding the respondents' busy academic schedules. Each subject had their own individual workstation equipped with an Acer computer and headphones. This setup allowed for independent completion of tasks. The students were given three attempts to pass the test, with a one-week interval between each attempt. This repeated-measures design allowed for the assessment of progress and the reliability of the testing instrument over time. Following the final testing, they were asked to answer a questionnaire (see Fig. C.1, Fig. C.2 with sample questions in Appendix C for complete proofs). The participants completed 60 tasks in the phonetic test: 45 tasks in the first module, and 15 tasks in the second one. The results of the completed test were saved as a pdf file where student's first name, last name, and group number were included in the file name. Each student was to enter their information and begin the first test section. Figure D.1 displays the start screen (see Fig. D.1 Appendix D for complete proofs).

Security and Anti-Cheating Measures

To ensure protection against cheating, all test files are stored in encrypted and hidden archives and directories. The files are sorted by directories with the task number, and the audio files are converted into WAV format. A text document in the *.csv format is created for each directory with audio files. This document serves as an additional protection against cheating, as the CSV format is not recognized by default in Windows OS. It contains a formulated task, the number of audio recordings, and answer options. The order of displaying answer options is randomized in the program. The first part of the computer test begins after the program starts and the necessary data is filled in. It displays the tasks of the first block of questions, where subjects are asked to choose one correct answer from three options.

Test Interface and Task Flow

Here, we provide some tasks for illustration: *Click on the "Play" button and choose the word in which you will hear a short sound.* (see Fig. D.2 Appendix D for complete proofs).

The words are presented in audio form and are not visible to the respondents. Three versions of the audio recording are given, for example, with words such as: *do, two, good*. Each answer option has a "Play" button. After listening to each option, the listener must choose the correct answer, click the "Choice" button, and move on to the next question ("Next"). Test takers also have the option to skip the listening part by clicking on the "Skip listening part" button and proceed to the second part of the test, which checks pronunciation.

Here is one more example: *Click on the "Play" button to hear the word. Listen carefully, determine which sound you hear [w] or [v].* Audio is offered, and there are two possible answers. It is necessary to determine which sound is pronounced [w] or [v], (see Fig. D.3 Appendix D for complete proofs).

After completing the auditory section, the program automatically calculated the number of correct responses and prompted participants to proceed to the pronunciation assessment. In this section, participants were shown a word on the screen, given two seconds to prepare, and then instructed to pronounce the word clearly upon receiving a visual cue. The speech recognition module then displayed the recognized text, allowing for immediate feedback on pronunciation accuracy.

The task is formulated as follows: *Clicking on the "Start" button you will see the WORD for reading. You will have 2 seconds to read that word. After that, you will see the command «Speak». Please, pronounce the WORD distinctly. For the next word, please click on the «Next» button*.

This block contains 15 words to check pronunciation (see Fig. D.4 Appendix D for complete proofs). After 15 tasks are completed, there appears a window on the screen with the calculated number of correct answers.

Data Storage

Following completion of all test sections, participants were asked to complete a questionnaire developed by the authors. All responses and results were securely saved in PDF format. This systematic approach facilitated efficient data collection and subsequent analysis.

RESULTS

Overall Performance Improvement Across Attempts

We developed and tested a deep-learning-based software to monitor Russian EFL students' perception and production skills in an Introductory Phonetics course. Feedback was collected through a post-test questionnaire, then analyzed and visualized using MS Excel and Python. As shown in Fig. 1 and 2, students' performance improved by 14.5% between the first and third attempts, demonstrating the system's effectiveness.

To analyze the trend, the scores of students presented above were averaged over multiple attempts. R^2 analysis involves calculating the coefficient of determination, a statistical measure that assesses how well the regression line fits the data by quantifying the proportion of variance in the dependent variable explained by the independent variable. By performing the R^2 analysis on the average scores over multiple attempts, we assess how well the increase in scores can be explained by the number of attempts or the implementation of the proposed test. The results show a clear linear increase, which was further supported by a high confidence R^2 value (R^2 value measures the trendline: the closer R^2 is to 1, the better the trendline fits the data.). This upward trend suggests that the implementation of the proposed test has significantly improved the subjects' performance.

A positive correlation between attempts as well as clear linear trend and high R^2 value confirm that our solution is an effective pedagogical tool for enhancing Russian EFL students' phonological competence through systematic practice.

Next, we considered the two parts of the test separately.

Perception Skills Module: Performance Trends

When analyzing the first part of the test, which assessed auditory perception skills, students demonstrated consistently positive results across all three attempts. The data exhibited a clear upward linear trend, as depicted in Fig. 3 and 4.

Pronunciation Skills Module: Performance Trends

The second part of the test, focused on pronunciation and speech recognition, yielded heterogeneous results. Unlike the perception section, participants' performance did not follow a linear trend. Instead, polynomial approximation was necessary to model the data accurately (see Fig. 5 and 6).

Figure 1
Comparison of Students' Total Performance over Three Attempts

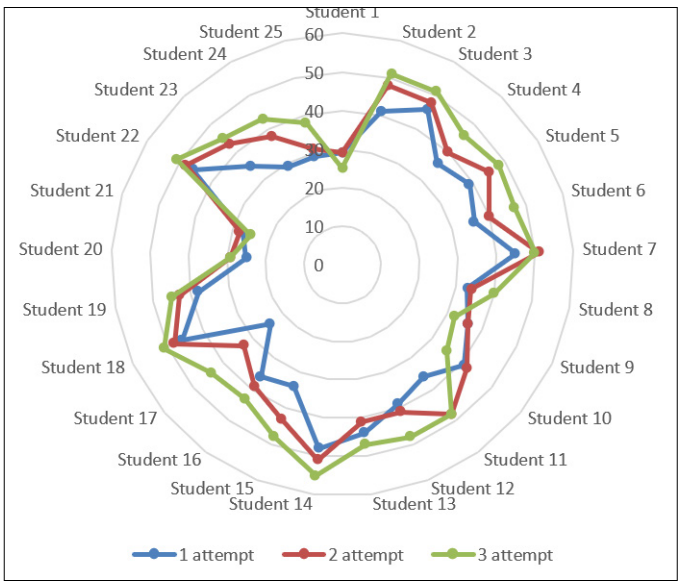


Figure 2
Trendline of the Average Score over Three Attempts

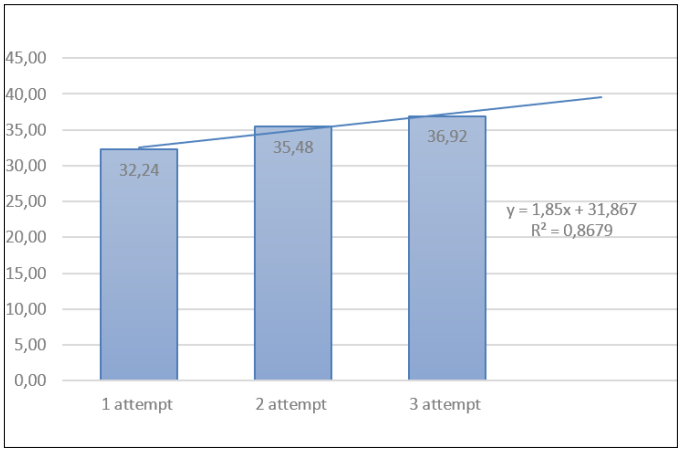


Figure 3

Comparison of Students' Performance in the Perception Part of the Test over Three Attempts

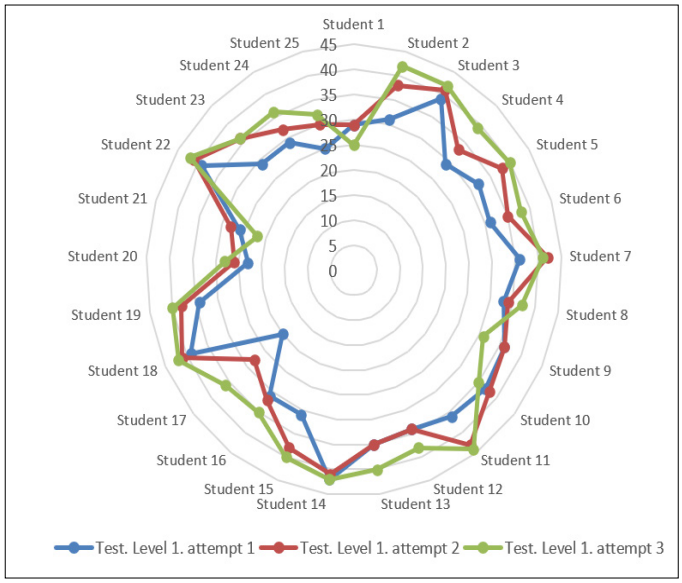


Figure 4

Trendline of the Average Score over Three Attempts after Passing the Perception Part of the Test

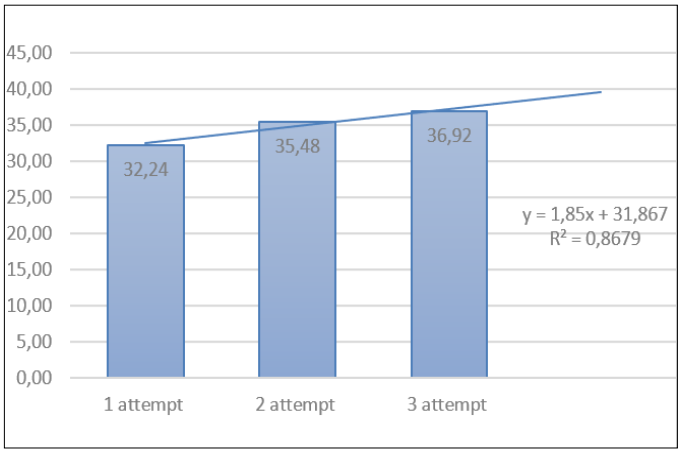


Figure 5

Comparison of Students' Performance in the Speech Recognition Part of the Test over Three Attempts

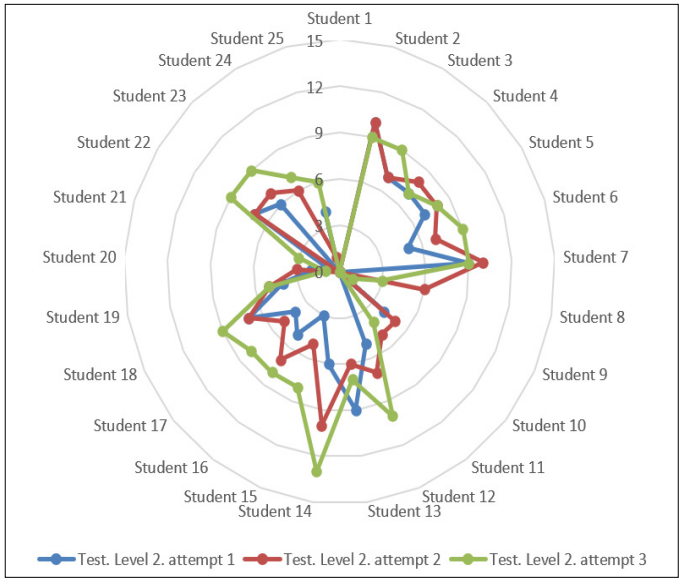
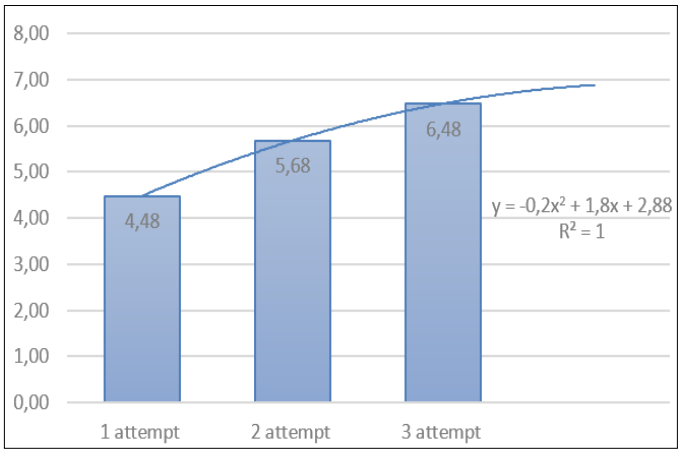


Figure 6

Trendline of the Average Score over Three Attempts after Passing the Speech Recognition Part of the Test



The analysis shows a tendency towards a smooth increase, which indicates a sufficient level of completion of the tasks of the second part of the test, as well as the gradual mastery of the material by students.

Test Completion Time Analysis

Analysis of the time required to complete the test across attempts revealed a non-linear pattern. While initial observations suggested a reduction in completion time, polynomial trend analysis (Fig. 7 and 8) indicated a gradual increase in time spent on subsequent attempts. Questionnaire responses clarified that increased time was often due to students' desire to double-check answers or further engage with challenging tasks.

Qualitative Feedback and User Experience

Post-test questionnaire responses provided valuable insights into user experience. In particular, the subjects' com-

Figure 7

Comparison of Time Spent on the Test across Three Attempts

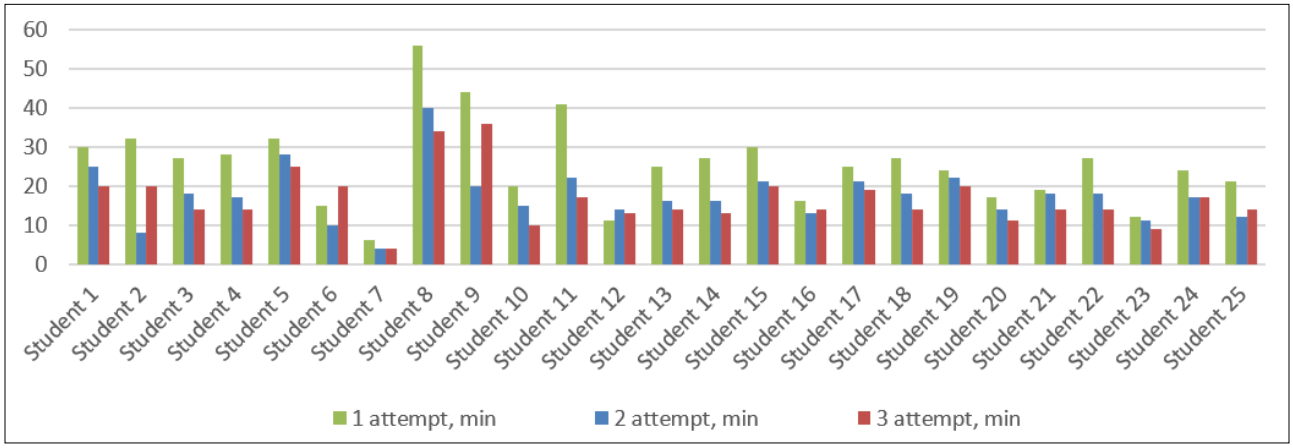
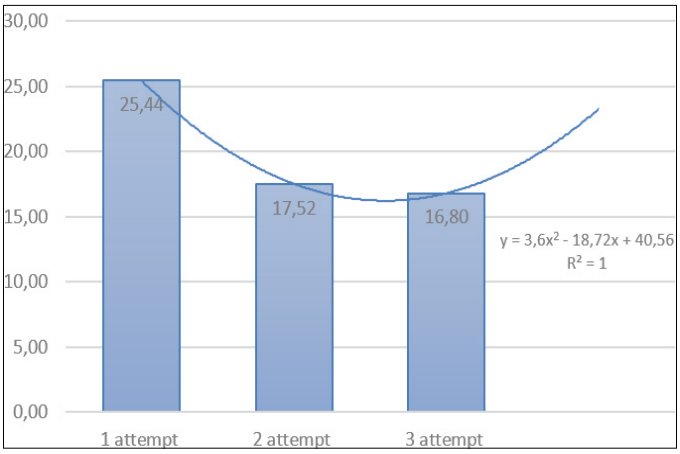


Figure 8

Trendline of Time Variation between Three Attempts



ments made it possible to understand why the test time increased during subsequent attempts. In some forms there were such answers as "*interesting, I went through several times because I had doubts about certain questions, decided to double-check myself*", "*useful, encouraged me to use a dictionary*". From a technical point of view, the first module of the Test has already been implemented in a pre-release form, which is why students found working with it more intuitive and useful, for example, we received the following comments: "*helps you assess your knowledge, test yourself*", "*required me to think, but it's interesting*".

The second set of tasks with speech recognition technology proved to be technically challenging for the subjects. They noted that the accuracy of pronunciation is affected by the sensitivity of the microphones. Some pointed out that the words appeared on the screen inconsistently, for example, some subjects wrote "*not all words were recognized accurately* ", "*it was unclear whether I should repeat the word again, but a new word was already appearing for reading*".

Other helpful feedback was the following: *"I would appreciate the option to go back to a previous question, I have to start over", "the system does not give comments, only a report on the number of correct answers".*

It should be noted that this is the first version of the application, designed to demonstrate its core functionality. Despite some technical issues in the second module of the test, the participants stated the absolute usefulness of this method of control, interest and increased motivation to achieve personal results in pronunciation.

Evaluation of Reliability and Validity of the Developed Test

We also assessed reliability and validity, metrics that are critical in social science research, to ensure that the measurement instruments developed capture the constructs the test is intended to measure with sufficient accuracy (Drost, 2011).

Test-Retest Reliability

Test-Retest Reliability is a method used to assess the consistency or stability of a measurement over time. It involves administering the same test to the same group of individuals several times. One way to calculate this is to calculate the Pearson correlation between scores from different trials (Attempt 1 vs. Attempt 2, Attempt 2 vs. Attempt 3 and Attempt 3 vs. Attempt 1 for Part 1 and Part 2 of the test). A higher Pearson correlation indicates greater test-retest reliability, meaning that the test produces consistent results when administered to the same individuals over time. In this analysis, correlation coefficients between 0.700 and 0.890 are considered to reflect strong reliability, while values between 0.500 and 0.690 indicate moderate reliability. The analysis of Pearson correlation coefficients revealed strong test-retest reliability across multiple assessment attempts. In Part 1 strong positive correlations were observed between Attempt 1 and Attempt 2 (0.860) and between Attempt 2 and Attempt 3 (0.811). Students who performed well in the first attempt also tended to perform well in the second attempt, and after that third one, suggesting reliability in their performance across these trials. A moderate correlation was found between Attempt 3 and Attempt 1 (0.597). Students' performance on the third attempt is not as strongly related to their performance on the first attempt. This drop in correlation suggests that some results changed significantly in students' performance between the first and third attempts, reflecting less consistency over time. Similarly, Part 2 demonstrated strong positive correlations between Attempt 1 and Attempt 2 (0.727) and between Attempt 2 and Attempt 3 (0.767), with a moderate to strong correlation between Attempt 3 and Attempt 1 (0.679). These findings indicate consistent student performance between consecutive attempts, with some expected variability over longer intervals. Overall, the assessment demonstrates robust test-retest reliability.

Internal Consistency (Cronbach's Alpha)

Cronbach's Alpha is a measure of internal consistency or reliability of a set survey questions that are supposed to measure the same construct. A higher Cronbach's Alpha indicates better reliability of the items. The 0.90 value of the Cronbach's Alpha for Part 1 falls in the "excellent" internal consistency range, while the 0.88 value for Part 2 is at the upper end of the "good" range. These high values suggest that the items within each part are measuring the same underlying construct consistently, indicating strong reliability of the assessment instrument. In social science research, these alpha values exceed generally accepted thresholds of sufficiency, representing very good indicators of reliability that satisfy methodological standards in this field.

Analysis of Construct and Criterion-Related Validity

Test validity is the extent to which a test accurately measures what it is supposed to measure. In the context of educational assessments in the social sciences, validity ensures that the test accurately reflects students' understanding of the subject matter. This analysis focuses solely on assessing the validity of a test based on student scores across three attempts.

Construct validity assesses whether the test measures the theoretical construct it is intended to measure. Let us examine the correlations between the attempts given above to prove construct validity. The strong positive correlations between Attempt 1 vs. Attempt 2 and between Attempt 2 vs. Attempt 3 suggest consistency in what the test measures across these attempts. This consistency supports the construct validity of the test and indicates that the test consistently measures the intended construct over time.

Criterion-related validity examines how well one measure predicts an outcome based on another measure, i.e. earlier attempts can predict performance on later attempts. Specifically, the strong positive correlation between Attempt 1 and Attempt 2 scores (0.860 and 0.727) indicates that students' performance on the initial attempt is a good predictor of their performance on the subsequent attempt. Similarly, the strong correlation between Attempt 2 and Attempt 3 scores (0.811 and 0.767) reinforces this predictive relationship between consecutive assessments. However, the weaker correlation between Attempt 1 and Attempt 3 scores (0.597 and 0.679) suggests that the test's ability to predict performance diminishes over non-consecutive attempts. This pattern highlights that while the test effectively estimates immediate future performance (supporting criterion-related validity in the short term), its predictive power over longer intervals is less pronounced.

DISCUSSION

The present study contributes to the expanding field of computer-assisted pronunciation training (CAPT), particularly as it pertains to the persistent challenges encountered by Russian EFL learners. Within the broader context of second language (L2) phonological acquisition and educational technology, our findings reinforce the growing consensus that targeted, technology-mediated interventions can meaningfully enhance learners' perceptual and productive pronunciation skills (González & Ferreiro, 2024; Alsuhaibani et al., 2024). By integrating established EFL phonetic instruction principles (Wang et al., 2025; Wei, 2025) into a CAPT system tailored for Russian university students, our research bridges a notable gap between theoretical frameworks and their practical application in language learning environments.

Our investigation employed a mixed-methods design, combining a pre-test for contrastive phonetic analysis, a computer-based assessment tool with perception and production modules, three iterative test administrations at weekly intervals, and a post-test questionnaire to capture user

experience. This methodological approach enabled both cross-sectional and longitudinal analysis of pronunciation development among 25 first-year Russian EFL students, providing a sound (Nickolai et al, 2024) foundation for interpreting our results.

The principal findings of this study show that overall participants' performance across attempts improved by 14.5% with the results showing a clear linear increase, which was further supported by a high confidence R2 value. Overall, students performed well on the tasks given, demonstrating confident results. Both the perception and production modules offered valuable insights into learner performance. Although some technical improvements were suggested for the production module, the system as a whole was rated as useful and efficient. The analysis of Pearson correlation coefficients indicates consistent student performance between consecutive attempts, with some expected variability over longer intervals. In total, the assessment demonstrates robust test-retest reliability. Both perception and production modules exhibited statistically significant internal consistency ($\alpha = 0.90$ and $\alpha = 0.88$ correspondently) confirming its reliability as a monitoring tool. Analysis of construct and criterion-related validity highlights that while the test effectively estimates immediate future performance (supporting criterion-related validity in the short term), its predictive power over longer intervals is less noticeable.

Interpreting these results, the marked improvement in perceptual discrimination supports the theoretical position that explicit training in articulation can facilitate perceptual gains (Pashkovskaya, 2010; Flege et al., 2021; Stratton, 2025). This finding aligns with the view that perception and production are mutually reinforcing processes in L2 phonological development, especially when training is tailored to learners' specific L1-L2 transfer zones. Our pre-test analysis and subsequent learner-oriented task design (Marefat et al., 2025) based on a thorough analysis of transfer zones and typical phonetic deviations were instrumental in targeting these areas, thereby enhancing the relevance and impact of the intervention.

Comparatively, the limited accuracy of the speech recognition system is consistent with recent literature highlighting the technical and contextual barriers to automated pronunciation evaluation (Souza & Gottardi, 2022; Shadiev, 2023; Nickolai et al, 2024). While some studies have used CAPT solutions based on more complex web-based AI modules (e.g., Dovchin, 2024), the performance of our system's production module was influenced by several contextual factors, such as variability in microphone quality in typical educational environments, limitations in acoustic modeling for non-native accents, and the absence of contextualized speech input. These challenges reflect the inherent complexity of reliably assessing L2 speech production and highlight the need for further development - potentially through the integration of higher-quality audio equipment or more advanced recognition algorithms.

The outcomes of this study were generally anticipated, given the theoretical and empirical foundations underpinning our intervention design. However, the magnitude of improvement in perceptual discrimination exceeded initial expectations, suggesting that even short-term, focused training can yield measurable gains. Conversely, the persistent challenges in automated production assessment highlight the need for continued innovation in CAPT technologies and methodologies.

Several limitations must be acknowledged. The study's sample was limited to first-year Russian university students, which may constrain the generalizability of findings to other learner populations or educational settings. Additionally, the test's focus on Russian university EFL learners, whose instructional goal is often native-like pronunciation, may limit its applicability in contexts with differing learner objectives (Hino, 2021). The production module's technical limitations, omission of capturing suprasegmental features and spontaneous speech, also limit the scope of our conclusions. Furthermore, the absence of a control group and the relatively short intervention period may limit the strength of causal inferences.

This research advances our understanding of CAPT's potential and limitations in Russian EFL contexts. The demonstrated efficacy of the auditory perception module provides a promising avenue for future development, while the challenges encountered in automated production assessment point to the need for further technological and pedagogical refinement. Expanding the system to address suprasegmental features and to accommodate a broader range of learner profiles represents a logical next step in optimizing technology-mediated pronunciation instruction.

CONCLUSION

This investigation has yielded some insights into the efficacy of computer-assisted phonetic assessment for Russian EFL learners. The study's principal findings demonstrate that technology-mediated training produces measurable improvements in L2 phonological perception, corroborating established psycholinguistic models and pronunciation teaching principles of EFL speech learning. The auditory discrimination module emerged as particularly effective, suggesting that structured perceptual training forms a foundation for phonological competence development.

The contribution of this research is threefold. First, it provides empirical validation for computer-based approaches

to monitoring phonetic challenges specific to Russian-English interlanguage. Second, it establishes a methodological framework for developing targeted pronunciation training protocols. Third, it identifies key technical limitations in current automated speech recognition applications for pedagogical contexts, regarding accented speech evaluation.

From a pedagogical perspective, these findings underscore the value of integrating diagnostic assessment tools within pronunciation curricula. The demonstrated effectiveness of iterative perceptual training suggests promising applications for autonomous learning environments. However, the technical constraints observed in production evaluation highlight the need for more sophisticated acoustic modeling approaches in CAPT systems.

Future investigations should prioritize: (1) longitudinal studies tracking the retention of training effects, (2) expansion of assessment parameters to encompass suprasegmental features, and (3) development of adaptive algorithms capable of processing non-native phonological variation. Such advancements would substantially enhance the validity and pedagogical utility of computer-assisted pronunciation training systems.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Y. S. Korshunova and A. O. Korol for their assistance in developing the computer test and preparing the materials for this article.

FUNDING

V. Kapitan conducted the study with financial support from the "MAPLE" project, grant number 22-5715-P0001, under the National University of Singapore Faculty of Science, Ministry of Education, Tier 1 grant "Data for Science and Science for Data collaborative scheme"

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Marina Kolesnichenko: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Vitalii Kapitan: Software; Data curation; Investigation; Methodology; Visualization; Resources; Funding acquisition; Writing – original draft; Writing – review & editing.

REFERENCES

Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies, 24*(6), 3731-3743. https://doi.org/10.1007/s10639-019-09955-7

Alsuhaibani, Y., Mahdi, H. S., Al Khateeb, A., Al Fadda, H. A., & Alkadi, H. (2024). Web-based pronunciation training and learning consonant clusters among EFL learners. *Acta Psychologica*, *249*, 104459. https://doi.org/10.1016/j.actpsy.2024.104459

Arakin, V. D. (2008). Comparative typology of English and Russian languages (4rd ed.). Fizmatlit.

Backus, A., Cohen, M., Cohn, N., Faber, M., Krahmer, E., Laparle, S., Maier, E., Van Miltenburg, E., Roelofsen, F., Sciubba, E., Scholman, M., Shterionov, D., Sie, M., Tomas, F., Vanmassenhove, E., Venhuizen, N., & De Vos, C. (2023). Minds: Big questions for linguistics in the age of AI. *Linguistics in the Netherlands*, *40*, 301–308. https://doi.org/10.1075/avt.00094.bac

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, *30*(1), 177-194. https://doi.org/10.5070/B5.35970

Belenko, M. V., & Balakshin, P. V. (2017). Comparative analysis of speech recognition systems with open code. *International Research Journal*, 4(58), 13-18. https://doi.org/10.23670/IRJ.2017.58.141

Bliss, H., Abel, J. & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation, 4*, 129-153. https://doi.org/10.1075/jslp.00006.bli

Blok, E. (2019). The planning and customization of introductory L2 phonetic courses on the basis of a numeric scale for assessing non-native speaker mistakes. *Rhema*, (4), 34-52. https://doi.org/10.31862/2500-2953-2019-4-34-52.

Bondarko, L. V. (1969). *Slogovaya struktura rechi i differentsial'nye priznaki fonem (eksperimental'no-foneticheskoe issledovanie na materiale russkogo yazyka)* [The syllabic structure of speech and the distinctive features of phonemes (experimental-phonetic research in the Russian language)] [Unpublished doctoral dissertation]. Leningr. gos. un-t im. A. A. Zhdanova. https://search.rsl.ru/ru/record/01010234177.

Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. *Frontiers in Artificial Intelligence*, *4,* 625341. https://doi.org/10.3389/frai.2021.625341

Crystal, D. (1970). Prosodic systems and language acquisition. *Prosodic Feature Analysis* (pp. 77-90). Didier Montreal and Paris.

Derwing, T. M., Munro, M. J. (2015) *Pronunciation fundamentals: Evidence-Based perspectives for L2 teaching and research*. John Benjamins.

Dovchin, S. (2024). Artificial Intelligence in Applied Linguistics: A double-edged sword. *Australian Review of Applied Linguistics*, *47*(3), 410–417. https://doi.org/10.1075/aral.24145.dov.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105-123. https://search.informit.org/doi/10.3316/.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning* (1st ed., pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002.

Flege, J. E., & Davidian, R. D. (2008). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, *5*(4), 323-347. https://doi.org/10.1017/S014271640000521X.

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation App. *Language Learning & Technology*, *24*(1), 62–85. https://doi.org/10125/44709

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193-202. https://doi.org/10.1007/BF00344251

Goncharova, N. L. (2006). *Formirovaniye inoyazychnoy fonetiko-fonologicheskoy kompetentsii u studentov-lingvistov: na materiale angliyskogo yazyka* [Forming foreign language phonetic-phonological competence of linguistic students based on the material of the English language] [Unpublished doctoral dissertation]. North Caucasus State Technical University. https://search.rsl.ru/ru/record/01003042566

González, M. D. L. Á. G., & Ferreiro, A. L. (2024). Web-assisted instruction for teaching and learning EFL phonetics to Spanish learners: Effectiveness, perceptions and challenges. *Computers and Education Open*, *7*, 100214. https://doi.org/10.1016/j.caeo.2024.100214

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). IEEE. https://doi.org/10.1109/ICASSP.2013.6638947

Hino, N. (2021). Language education from a post-native-speakerist perspective: The case of English as an international language. *Russian Journal of Linguistics*, *25*(2), 528-545. https://doi.org/10.22363/2687-0088-2021-25-2-528-545

Ivanko, A. F., Ivanko, M. A., Sizova, Y. A. (2019). Neural networks: General technological characteristics. *Scientific Review. Technical Sciences*, (2), 17-23.

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. (2025). Natural Language Processing for dialects of a language: A survey. *ACM Computing Surveys*, *57*(6), 1–37. https://doi.org/10.1145/3712060.

Kulikov, V. G. (2005). Phonological contexts and frames: Toward the unified methodology of cognitive linguistics. *Issues of Cognitive Linguistics*, (2), 28-40.

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of Clear Speech: Effect of Instruction. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1807–1821. https://doi.org/10.1044/1092-4388(2012/11-0154).

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Leonov, A. S., & Sorokin, V. N. (2007). K analizu rezonansnykh chastot rechevogo trakta [To the analysis of the resonant frequencies of the speech tract]. *Informacionnye Process, 4*(7), 386 - 400.

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

Li, Z., Basit, A., Daraz, A., & Jan, A. (2024). Deep causal speech enhancement and recognition using efficient long-short term memory Recurrent Neural Network. *PLOS ONE*, *19*(1), e0291240. https://doi.org/10.1371/journal.pone.0291240

Luz, S. (2022). Computational linguistics and natural language processing. *The Routledge handbook of translation and methodology* (pp. 373-391). Routledge. https://doi.org/10.4324/9781315158945-27

Mahdi, H. S., & Al Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education*, 7(3), 733-753. https://doi.org/10.1002/rev3.3165

McShane, M., & Nirenburg, S. (2021). *Linguistics for the Age of AI*. Mit Press. https://doi.org/10.7551/mitpress/13618.003.0003

Marefat, F., Hassanzadeh, M., Noureddini, S., & Ranjbar, M. (2025). Reporting practices in applied linguistics quantitative research articles across a decade: A methodological synthesis. *System*, *131*, 103627. https://doi.org/10.1016/j.system.2025.103627

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. *Information Fusion*, *99*, 101869. https://doi.org/10.1016/j.inffus.2023.101869

Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010). Recurrent neural network-based language model. *Interspeech*, *2*(3), 1045-1048. https://doi.org/10.21437/Interspeech.2010-343

Mooney, D. (2019). Phonetic transfer in language contact: Evidence for equivalence classification in the mid-vowels of Occitan–French bilinguals. *Journal of the International Phonetic Association*, *49*(1), 53-85. https://doi.org/10.1017/S0025100317000366

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*(3), 283-309. https://doi.org/10.1075/jslp.20038.mun

Nesset, T. (2008). Ronald W. Langacker, Cognitive Grammar: A basic introduction. Oxford: Oxford University Press, 2008. Pp. x+562. *Journal of Linguistics*, *45*(2), 477–480. https://doi.org/10.1017/S0022226709005799.

Nickolai, D., Schaefer, E., & Figueroa, P. (2024). Aggregating the evidence of automatic speech recognition research claims in CALL. *System*, *121*, 103250. https://doi.org/10.1016/j.system.2024.103250

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote J. A. & Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182-207. https://doi.org/10.1075/jslp.17001.obr

Ohala, J. J. (2010). The relation between phonetics and phonology. In *The Handbook of Phonetic Sciences* (pp. 653–677). Wiley Blackwell. https://doi.org/10.1002/9781444317251.ch17

Omid, M. (2022). Review of research on the use of Information and Communication Technologies (ICTs) in ELT-related academic writing classrooms. *Journal of Language and Education*, *8*(2), 165-178. https://doi.org/10.17323/jle.2022.13395

Pashkovskaya, S. S. (2010). *Differenciruyushaya model obucheniya russkomu proiznosheniyu* [Differentiating model of teaching Russian pronunciation] [Unpublished doctor dissertation]. Gos. in-t rus. iaz. im. A.S. Pushkina. https://search.rsl.ru/ru/record/01004949907

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-47677-7

Redmon, C., Leung, K., Wang, Y., McMurray, B., Jongman, A., & Sereno, J. A. (2020). Cross-linguistic perception of clearly spoken English tense and lax vowels based on auditory, visual, and auditory-visual information. *Journal of Phonetics*, *81*, 100980. https://doi.org/10.1016/j.wocn.2020.100980

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, 52(1), 189-205. https://doi.org/10.1177/0033688220977406

Rudregowda, S., Patilkulkarni, S., Ravi, V., H.L., G., & Krichen, M. (2024). Audiovisual speech recognition based on a deep convolutional neural network. *Data Science and Management*, *7*(1), 25–34. https://doi.org/10.1016/j.dsm.2023.10.002

Schildt, H. (2010). *C# 4.0: The complete reference*. McGraw-Hill.

Sereno, J. A., Jongman, A., Wang, Y., Tupper, P., Behne, D. M., Gu, J., & Ruan, H. (2025). Expectation of speech style improves audio-visual perception of English vowels. *Speech Communication*, *171*, 103243. https://doi.org/10.1016/j.specom.2025.103243

Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, *35*(1), 74–88. https://doi.org/10.1017/S095834402200012X

Shevchenko, T. I. (2017). Cognitive phonology: Theoretical and applied aspects. *Vestnik of Moscow State Linguistic University. Humanities*, *5*(776), 106-115.

Soundarya, M., Karthikeyan, P. R., & Thangarasu, G. (2023). Automatic speech recognition trained with convolutional neural network and predicted with recurrent neural network. In *2023 9th International Conference on Electrical Energy Systems,* (pp. 41-45). IEEE. https://doi.org/10.1109/ICEES57979.2023.10110224

Souza, H. K. D., & Gottardi, W. (2022). How well can ASR technology understand foreign-accented speech? *Trabalhos Em Linguística Aplicada*, *61*(3), 764–781. https://doi.org/10.1590/010318138668782v61n32022

Stratton, J. M. (2025). The effects of production training on speech perception in L2 learners of German. *Journal of Phonetics*, *108*, 101370. https://doi.org/10.1016/j.wocn.2024.101370

Su, Y., & Kuo, C. C. J. (2022). Recurrent neural networks and their memory behavior: A survey. *APSIPA Transactions on Signal and Information Processing*, *11*(1), e26 (1-38). http://dx.doi.org/10.1561/116.00000123

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344. https://doi.org/10.1093/applin/amu076

Tikhonova, E., & Raitskaya, L. (2023). ChatGPT: Where is a silver lining? Exploring the realm of GPT and large language models. *Journal of Language and Education*, *9*(3), 5-11. https://doi.org/10.17323/jle.2023.18119

Urip, S., Reli, H., Faruq, U. M., & Mujiyono, W. (2022). Determinants of Technology Acceptance Model (TAM) towards ICT use for English language learning. *Journal of Language and Education*, *8*(2), 17-30. https://doi.org/10.17323/jle.2022.12467

Vishnevskaya, E. M. (2014*). Metodika korrekcii fossilizacii foneticheskih navykov bakalavrov pedagogicheskogo obrazovaniya (na materiale anglijskogo yazyka kak vtorogo inostrannogo)* [Methodology for correcting the fossilization of phonetic skills of bachelors of pedagogical education (based on the material of English as a second foreign language)] [Unpublished doctor dissertation]. Mosk. gos. gumanitar. un-t im. M.A. Sholokhova]. University Repository. https://search.rsl.ru/ru/record/01007483765

Wang, J., Ahmad, N. K. B., Jamil, H. B., & Darmi, R. (2025). Resonating voices: Unpacking EFL teachers' beliefs regarding pronunciation instruction in Chinese tertiary context. *Journal of Curriculum and Teaching*, *14*(1), 30. https://doi.org/10.5430/jct.v14n1p30

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552. https://doi.org/10.1016/j.system.2004.09.011

Wei, Y. (2025). A study of non-native accent correction techniques combining phonetics, machine learning and biomechanics. *Molecular & Cellular Biomechanics*, *22*(1), 725. https://doi.org/10.62617/mcb725

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, *126*, 103497. https://doi.org/10.1016/j.system.2024.103497

APPENDIX A

SCHEMATIC DIAGRAMS: CONVOLUTIONAL AND RECURRENT NEURAL NETWORK

Figure A.1

Schematic Diagram of a Convolutional Neural Network

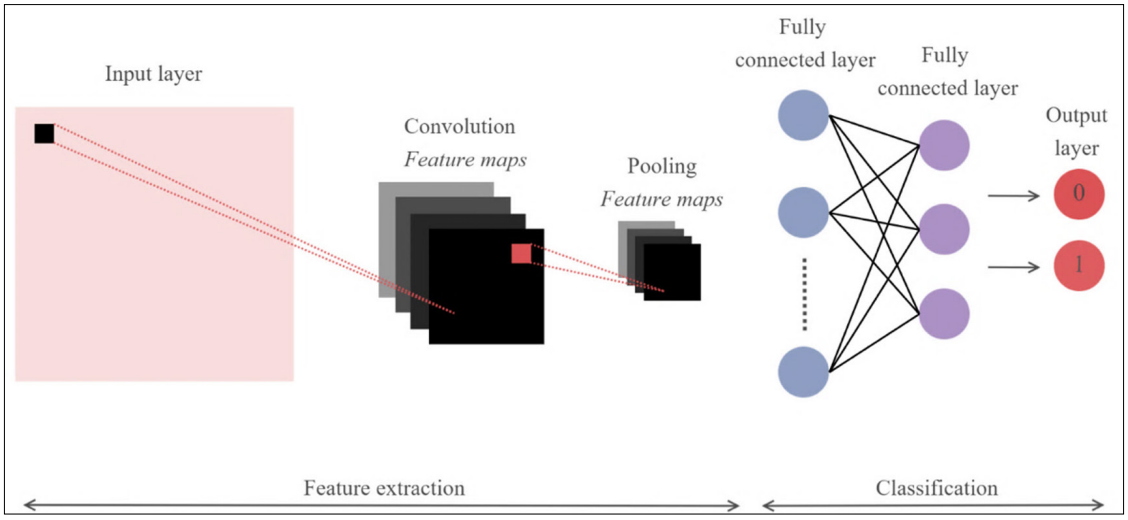
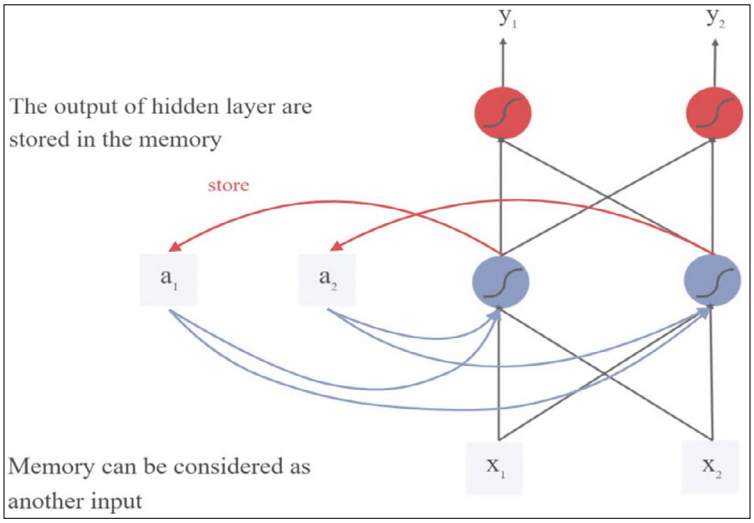


Figure A.2

Schematic Diagram of a Recurrent Neural Network



APPENDIX B

A DETAILED ACCOUNT OF THE SPECIFIC PRE-TEST ITEMS AND THEIR CORRESPONDING PHONETIC PHENOMENA

1. Incorrect articulation of vowels, such as replacing the English sound [ɜ] with Russian [jo] or [э], and English [ɑ:] with Russian [а].

2. Difficulty in distinguishing between vowel length and positional vowel length.

3. Violations in the implementation of English diphthongs, eg., monophthongization of [oʊ] with replacement by [о].

4. Difficulties in implementing the opposition of voiceless/voiced final consonants in English words.

5. Non-discrimination of phonemes [w] and [v], replacing English [s] with Russian [с], and English interdentals [θ] and [ð] with Russian [с], [ф], [т], and [д], [з], respectively.

6. Differences in syllable division, such as the addition of a subsequent consonant to a short vowel in a stressed syllable, which can cause difficulties in determining word stress and result in incorrect pronunciation of vowels in syllables.

APPENDIX C

QUESTIONNAIRE SAMPLES

Figure C.1

Sample Questions

1. Indicate the correct answers in percentages for each of the attempts:

First attempt %		Second attempt %		Third attempt %	
First part of the test	Second part of the test	First part of the test	Second part of the test	First part of the test	Second part of the test

2. Indicate the time to complete the test for each of the attempts:

First attempt	Second attempt	Third attempt

3. Did you use a dictionary to check pronunciation? (Underline the correct option):
   1. Yes
   2. No

4. Are the tasks clearly formulated? (Underline the correct option):
   1.Yes
   2. No

Figure C.2

Sample Questions

5. What recommendations could you suggest for improving this test? _____
6. Have you ever encountered situations where functions of the test were inconvenient? Which? _____
7. Do you think that completing such tasks helps you improve your listening and pronunciation skills in English? _____
8. Was it interesting to complete the test tasks in the first and second modules? What exactly sparked your interest? _____

APPENDIX D

GRAPHICAL USER INTERFACE SAMPLES

Figure D.1
Graphical User Interface (GUI) of Testing Software

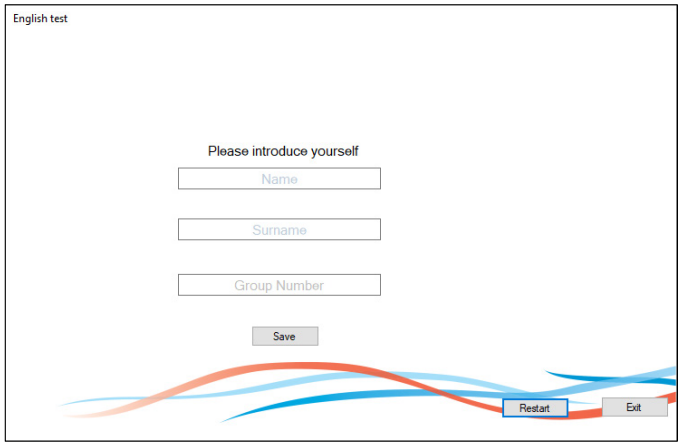


Figure D.2
GUI of the Computer Test for Vowel Recognition

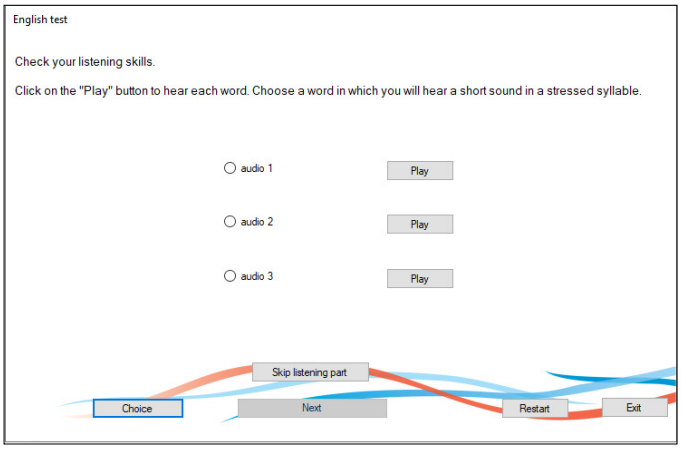


Figure D.3
GUI of the Computer Test for Consonant Recognition

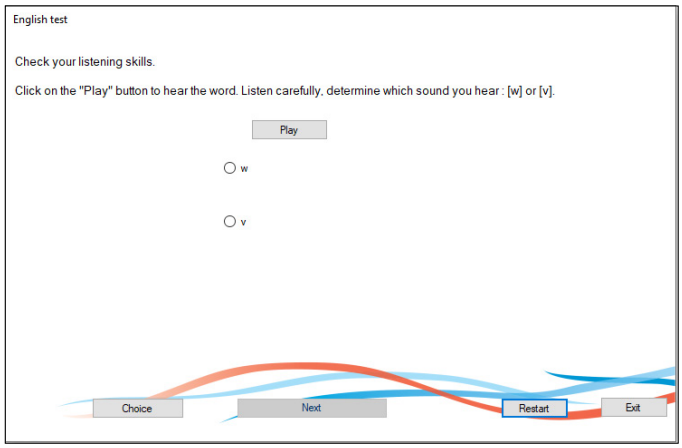


Figure D.4
GUI of the Computer Test with the Speech Recognition Part