



Article

A WebGL Serious Game for Practicing English Conversations in Public Places Using Speech Recognition

Rickman Roedavan¹, Iis Kunia Nurhayati², Kemas Muslim Lhaksana³, Muhammad Yudhi Rezaldi^{4,5}, Esa Prakasa⁵, and Sritenaya Geovani Putri⁶

¹Department of Multimedia Engineering Technology, School of Applied Science, Telkom University, Bandung, Indonesia;

²Departement of Communication, Faculty of Communication and Social Science, Telkom University, Bandung, Indonesia;

³Departement of Informatics, Faculty of Informatics, Telkom University, Bandung, Indonesia;

⁴Department of Visual Communication Design, Faculty of Creative Industries, Telkom University, Bandung, Indonesia;

⁵Research Center for Data and Information Sciences, National Research and Innovation Agency (BRIN), Indonesia;

⁶Department of Entrepreneurship and Strategy, Lancaster University Management School, Lancaster University, Bailrigg, Lancaster, LA1 4YX, England, United Kingdom

{rikman,iiskurnia,kemas.muslim}@telkomuniversity.ac.id; {esap001,myud001}@brin.go.id, s.g.putri@lancaster.ac.uk

Keywords:

Serious Game,
English Learning,
Speech Recognition,
WebGL

Received: October 2024

Accepted: April 2025

Published: May 2025

DOI: 10.17083/ijsg.v12i2.890

Abstract

Learning English as a Second Language (ESL), particularly for conversation, often presents challenges, especially due to learners' fear of judgment during public practice. Current tools lack the ability to provide private, practical, and interactive speaking environments. This study investigates how to create a private, practical, and interactive platform for English conversation practice. To address this gap, Langova, a WebGL-based online game leveraging Speech Recognition technology, was developed to facilitate English conversation practice in realistic public scenarios, such as airports, hotels, and malls. This approach enables learners to practice privately while building confidence. The study involved 57 university-level ESL students with varying levels of English proficiency, who participated in a three-month trial of Langova. The methodology includes system analysis using Word Error Rate (WER) and Average Word Error Rate per Sentence (Avg WER) to evaluate speech recognition accuracy. A trial with 57 students revealed WER values ranging from 0.23 to 1.22 and Avg WER values from 0.15 to 1.99, highlighting variability and feedback consistency. Compared to traditional learning methods, Langova provides an innovative, accessible, and interactive platform for language acquisition. By combining speech recognition technology with immersive learning environments, Langova empowers learners to build confidence and improve fluency. The platform also provides actionable data for educators to design targeted interventions, suggesting broader implications for integrating similar technologies into language education frameworks. Qualitative feedback from participants indicated that Langova's private and flexible environment can reduce their anxiety about speaking English, while also revealing technical challenges such as microphone sensitivity and accent recognition.

1. Introduction

English proficiency has become essential for academic and professional success. Despite its growing importance, many countries still struggle with low proficiency levels [1]. According to The EF English Proficiency Index, Indonesia ranks 79th out of 113 countries and 13th among 23 Asian countries, with a proficiency level below the global average. This indicates that a significant portion of the Indonesian population faces challenges in mastering English, hindering their academic and career opportunities.

English is undeniably the global lingua franca, playing a crucial role in education, business, and international relations. Proficiency in English is essential for students engaging in exchange programs, securing scholarships, and pursuing higher education abroad [2]. Mastery of the language opens doors to prestigious universities and diverse cultural experiences, significantly enriching both academic and personal growth. Moreover, English fluency enhances career prospects and fosters cross-cultural communication, making it a vital skill in today's interconnected world [3].

Students proficient in English can access a wealth of knowledge available in academic journals, books, and online resources, which are predominantly published in English. This linguistic capability allows students to stay updated with the latest research and developments in their fields of study [4]. Furthermore, English proficiency is often a prerequisite for enrolling in renowned institutions and participating in international conferences, where students can present their work and network with scholars from around the globe.

Beyond academia, English plays a pivotal role in professional settings [5]. Many multinational companies use English as their primary language for communication and documentation. Employees who are fluent in English can engage effectively with international colleagues, clients, and partners, facilitating smoother business operations and negotiations. In industries such as tourism, aviation, and hospitality, English is the standard mode of communication, underscoring its importance in providing quality service and ensuring customer satisfaction [6].

Culturally, English opens doors to understanding and appreciating global diversity [7]. It allows individuals to explore different cultures through literature, films, music, and other media. This cultural exposure fosters empathy and broadens perspectives, helping people develop a more inclusive worldview. Moreover, being able to communicate in English enables travelers to navigate foreign countries more easily, enhancing their travel experiences and making global exploration more accessible.

One significant factor contributing to low English proficiency in Indonesia, particularly in conversation, is language anxiety [8]. Language anxiety, often exacerbated by fears of public speaking, can severely impact a learner's ability to communicate effectively. This anxiety can stem from a variety of sources, including the pressure to perform well in social and academic settings, fear of making mistakes, and lack of confidence in one's language abilities.

Continuous practice is crucial for achieving fluency, ideally through conversations with native speakers or within environments where English is used regularly [9]. Such environments might include specific regions or workplaces that promote English-speaking days, where learners can immerse themselves in the language and practice without fear of judgment. However, many individuals feel embarrassed or anxious about speaking English in public, which highlights the need for private practice solutions using technology [10].

To address these gaps, we have developed Langova (previously named MetaSpeak), a serious game designed specifically for learning English in a thematic and structured manner. In this context, "thematic and structured manner" refers to organizing language learning around specific real-life scenarios that align with common public places, such as airports, train stations, and restaurants, to ensure contextual relevance. Studies in game-based learning emphasize that thematic approaches increase learner engagement by situating practice in meaningful contexts [11], while structured sequences of interactions scaffold learning and build fluency progressively [12]. By simulating real-life environments and interactions, Langova provides learners with relevant and realistic language practice scenarios. This thematic approach not only helps learners acquire language skills but also builds their confidence in using English in various everyday contexts.

The purpose of this study is to evaluate how Langova can help reduce language anxiety among university students in Indonesia, who often experience barriers to practicing English due to cultural and institutional factors. The findings aim to benefit not only this specific population but also broader demographics by providing an adaptable framework for technology-enhanced language learning. By demonstrating the effectiveness of game-based methods, this research has the potential to influence educational practices in other non-native English-speaking regions, addressing global challenges in English language acquisition.

To guide this research, we address the following research questions:

1. How can a WebGL-based game with Speech Recognition technology and simulated public settings help learners improve their English conversation skills and reduce language anxiety?
2. How effective are Word Error Rate (WER) and Average Word Error Rate per Sentence (Avg WER) in evaluating the system's ability to provide feedback on speaking performance?

This paper will present a comprehensive overview of the literature on the use of games in education, highlighting the benefits and challenges of game-based learning. Following this, a comparative analysis of existing games designed for English language learning. The design and implementation section will then detail the structural design of Langova, as well as the process and outcomes of its development. In the results and evaluation section, empirical data from the Langova trial will be presented, accompanied by a thorough discussion and evaluation of the findings. Finally, the paper concludes with an exploration of the broader implications of using serious games for language learning and suggests future directions for research and development in this field.

2. Background

2.1 Effective Methods for Language Learning

A holistic approach to language learning involves the integration of four core skills: listening, reading, writing, and speaking [13]. Each of these skills contributes uniquely to language acquisition, creating a comprehensive learning experience. Listening is particularly crucial as it helps learners internalize pronunciation, intonation, and contextual usage before speaking [14]. By focusing on listening first, learners build a foundation for accurate and natural speech, which is vital for engaging in meaningful conversations [15].

This sequence allows learners to internalize correct pronunciation, intonation, and contextual usage before they start speaking. Exposure to native speakers through audio materials or direct interaction significantly enhances speaking abilities. By first focusing on listening, learners can mimic the sounds and patterns they hear, leading to more authentic and fluent speech when they begin speaking themselves [16].

Reading expands vocabulary and contextual knowledge. It exposes learners to new words and phrases, as well as different ways of using them in context [17]. Reading various types of texts, from fiction and non-fiction to academic articles and news reports, broadens a learner's understanding of the language and its applications [18]. This skill not only enhances vocabulary but also helps learners understand cultural references and idiomatic expressions, which are crucial for fluency. Writing can enhance grammatical understanding and expression [19]. It allows learners to practice sentence structure, punctuation, and style, which are essential for clear and effective communication. Regular writing exercises help solidify these rules, making them second nature to the learner [20].

Speaking develops fluency and confidence. It is the practical application of all the other skills combined [21]. Through speaking, learners practice pronunciation, intonation, and rhythm, making their speech sound more natural. Engaging in conversations, whether with

peers or native speakers, allows learners to apply what they have learned in real-time, improving their ability to think and respond quickly [22]. Speaking practice builds confidence, reducing anxiety and making learners more comfortable using the language in various situations.

2.2 Educational Technology for English Learning

Given the importance of these holistic skills, integrating technology into language learning can further enhance the educational experience. Educational Technology like Duolingo, Rosetta Stone, and Babbel offer language learning opportunities through interactive lessons and activities. These platforms have gained popularity due to their accessibility and convenience, allowing users to practice at their own pace and on their own time. While these platforms offer valuable resources, they often lack thematic consistency, leading to sporadic and less effective language practice [23]. The lessons can feel disjointed, and learners may struggle to apply what they have learned in real-world scenarios. Comparative results of three different educational technologies are shown in Table 1.

Table 1. Comparative Analysis of Educational Technology Results

Element	Duolingo	Rosetta Stone	Babel
Methodology	Gamification	Immersion	Real-life context
Features	Point, Leaderboards	Speech Recognition	Grammar Explanation
User Experience	Fun, Simple	Immersive	Traditional, Detailed
Content Depth	Basic, Intermediate	Intermediate, Advanced	Detailed, Contextual
Grammar	Limited	Implicit	Comprehensive
Vocabulary	Repetitive	Contextual	Contextual
Speaking	Limited Context	Limited Context	Limited Context

The comparative analysis highlights the core methodologies and features of Duolingo, Rosetta Stone, and Babbel. While each platform offers distinct strengths such as gamification (Duolingo), immersion (Rosetta Stone), and real-life contextual learning (Babbel), they share a common limitation in developing conversational speaking skills. These platforms primarily rely on fragmented exercises or isolated short sentences, which lack the depth and fluidity necessary for real-life fluency. Furthermore, grammar instruction is either limited or implicit, and vocabulary exposure tends to be either repetitive or purely contextual, restricting learners from effectively constructing dynamic conversations.

Langova distinguishes itself by addressing these gaps through a scenario-based, adaptive learning environment that prioritizes active conversational engagement. Unlike Duolingo’s game-like exercises, which emphasize point-based progression, or Rosetta Stone’s passive immersion, Langova integrates AI-driven adaptive dialogue systems. This enables learners to engage in interactive role-playing within realistic scenarios, fostering real-time language production, contextual grammar application, and natural dialogue construction.

Several serious games have been developed to aid in English language learning, each focusing on different aspects of language acquisition. One such game is Lighthouse, an English assessment game that focuses on reading through a side scrolling format [24]. Lighthouse combines storytelling with educational content, enhancing players' reading comprehension. However, its primary focus on reading makes it less suitable for speaking practice. The game's strong entertainment element may also overshadow its educational purpose, limiting its effectiveness for comprehensive language learning.

Another example is Simulang, a game designed to improve listening skills within the context of English for business. Simulang uses thematic scenarios and multiple-choice questions to test and enhance listening abilities. While it is effective for its specific focus, the narrow theme of business English limits its applicability to general English learners [25]. Additionally, the

format of Simulang does not provide opportunities for speaking practice, which is crucial for holistic language acquisition. This limitation means learners may not fully develop the ability to converse fluently in English.

Bethe1Challenge is a game that targets English grammar proficiency through a choice-based format [26]. While it effectively addresses grammar learning, it lacks a focus on speaking practice and does not integrate seamlessly with a broader educational framework. The assessment in BethelChallenge relies on surveys rather than interactive feedback mechanisms. Although beneficial for grammar, it falls short in providing comprehensive language learning experience, particularly in speaking and real-time application.

The comparison of Lighthouse, Simulang, and BeThe1Challenge reveals that while these game-based learning platforms offer engaging and varied themes—such as adventure, simulation, and competition—their focus on the thematic elements of the game often overshadows the practical application of language skills in real-life scenarios. Table 2 shows the comparative results of three different games.

Table 2. Comparative Analysis of Serious Game Results

Element	Lighthouse	Simulang	BethelChallenge
Theme	Adventure	Simulation	Competition
Gameplay	Quest-based	Scenario-based	Challenge-based
Focus	Vocabulary	Conversation	Grammar, Vocabulary
Target Audience	Children, Teens	Adults	Teens, Adults
Difficulty	Basic, Intermediate	Intermediate, Advanced	Basic, Intermediate
Language Skill	Read	Read, Listening	Read

The material is heavily tied to the gameplay, which limits its effectiveness in preparing users for everyday conversations and real-world language use. This thematic emphasis, though beneficial for user engagement and motivation, tends to fall short in fostering practical language proficiency needed for daily interactions.

This gap highlights the need for a more integrated solution that provides learners with opportunities to practice conversational skills in immersive and contextually environments, addressing both the practical barriers to language learning. Langova's thematic and structured approach builds on existing research that emphasizes the role of context in language learning. By integrating storytelling and realistic simulations, Langova ensures that learners are not only exposed to the target language but also understand its practical applications. Langova introduces an innovative approach by combining simulated public settings and Speech Recognition technology, enabling users to practice speaking in realistic environments, such as airports or malls.

2.3 The Role of Context in Language Learning

Thematic and contextual learning are pivotal in bridging the gap between theoretical knowledge and practical application. Context allows learners to see how language functions in specific scenarios, making the learning process more relevant and engaging [27]. For instance, practicing English in contexts such as ordering food at a restaurant or navigating an airport gives learners practical vocabulary and sentence structures they can immediately apply.

Context helps learners understand the practical application of language, while storytelling engages them emotionally and intellectually. Story-based learning aids memory retention and makes practice enjoyable [28]. Serious games can incorporate storytelling to offer immersive environment so learners can practice language learning, in realistic scenarios [29]. These realistic simulations are crucial because they allow learners to experience and practice language in a safe and controlled environment before applying their skills in real-world situations.

The immersive nature of storytelling in games provides a rich context that helps learners internalize language patterns and vocabulary [30]. By placing language learning within a contextual framework, learners are more likely to remember and use what they have learned. This approach transforms language practice from a rote memorization task into an engaging activity that mirrors real-life interactions [31]. By simulating real-world conversations and scenarios, learners can practice speaking and listening in a low-pressure environment. This practice is invaluable for building the confidence needed to use the language in actual conversations.

The interactive nature of games also enhances the learning experience by shifting learners from passive observers to active participants. Traditional methods often rely on delayed feedback, which can hinder the learning process and reduce the learner's motivation. In contrast, the immediate feedback provided by games ensures that learners can quickly understand and correct their mistakes, leading to a more efficient and effective learning experience [32]. This adaptive nature of games allows them to cater to individual learning paces, making them an ideal tool for language acquisition where personalized attention to the learner's needs is vital [33].

The interactive elements of games play a significant role in enhancing the language learning experience [34]. Games require active participation from the learner, engaging them in tasks that demand the use of the target language in meaningful ways [35]. This dynamic and responsive feedback is essential for reinforcing correct language use and for helping learners internalize language rules more effectively. By involving learners in scenarios where they must apply language skills contextually, games help to build confidence in using the language. This active engagement is particularly beneficial in language education, as it shifts the learner from a passive recipient of information to an active participant in the learning process, thereby enhancing retention and understanding [36].

Moreover, the combination of context, storytelling, and simulation in language learning games creates a powerful and holistic educational tool [37]. Contextual learning allows learners to practice language in realistic situations, which is key to developing practical language skills. Storytelling adds a narrative element that not only makes learning more engaging but also helps learners remember language structures and vocabulary by associating them with memorable stories or situations [38]. Simulation provides a safe environment for learners to practice without the fear of making mistakes in real-life situations. Together, these elements create a comprehensive and immersive learning experience that can significantly enhance a learner's proficiency and confidence in using a new language. By integrating these components, language learning games offer a rich, multifaceted approach to education that traditional methods often lack, making them a valuable addition to any language-learning curriculum [39].

3. Design and Implementation

3.1 Langova Design Structure

This study employs a framework derived from prior research, known as the Educational Game Scenario Model Based on Imperative Game Goal Typology [40]. This framework adopts the Three-Act Structure model for the development of game levels, comprising three main phases: Setup, Confrontation, and Resolution. The Setup phase focuses on defining the learning subject and the game goal typology, serving as the foundation for the game's objectives and ensuring alignment with the intended educational outcomes. In the context of Langova, the learning subjects are English conversational skills, contextualized within specific public places to provide practical and relevant scenarios for learners. Figure 1 illustrates the Langova Scenario Design.

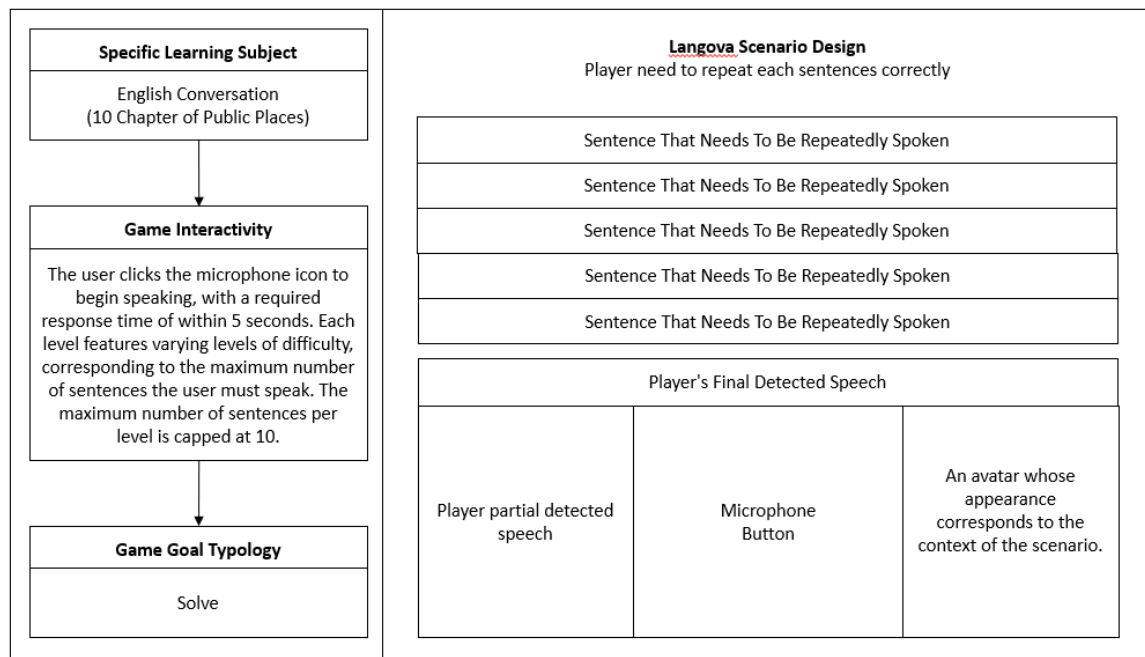


Figure 1. Langova Scenario Design

The core focus of Langova is to provide players with a platform to practice English conversation within the context of commonly encountered public places. The English learning material, developed by the Learning Center at Telkom University, is organized into 10 thematic chapters, each representing a distinct public location: Airport, Train Station, Bus Stop, Hotel, Restaurant, Street, Park, Hospital, Mall, and Office. These locations were selected for their relevance to situations that university students—particularly those with limited English proficiency—may encounter during academic activities such as learning assignments, international conferences, or student exchange programs abroad.

The Confrontation phase emphasizes the implementation of scenarios within the game by developing dynamic interaction algorithms. This involves designing user interactions through various input methods such as a mouse, keyboard, or, in the case of Langova, speech recognition. This phase also incorporates the game's win/lose logic. For Langova, the speech recognition system evaluates user input based on text similarity scores, making it a critical component of the gameplay mechanics and overall user experience.

Finally, the Resolution phase serves as the conclusion of the game. In Langova, this phase not only displays the user's score but also generates a comprehensive gameplay report for all participants. This report can be analyzed by administrators to gain valuable insights into user performance and engagement. Such data provides a foundation for further analysis and potential improvements to the system.

Each chapter in Langova includes 10 pairs of dialogues between a first and a second person, totaling 300 dialogue pairs across the entire game. These dialogues are designed to address the same focal problem using different vocabulary and expressions, providing players with alternative ways to express the same idea. Corresponding answers are provided with similar linguistic variety, encouraging players to expand their vocabulary and improve their conversational fluency. For example, players may encounter variations of a question about locating Gate 3 in an airport, such as:

"Excuse me. Could you tell me where Gate 3 is?"

"Pardon me, could you tell me the way to Gate 3?"

"Hi, sorry to bother you. Can you direct me to Gate 3?"

The overall design of Langova is structured into three main layers: Phase, Mission, and Exploration. Where phase layer is subdivided into three key segments—Start, Gameplay, and Finish. See Figure 2 for a Game Design Structure.

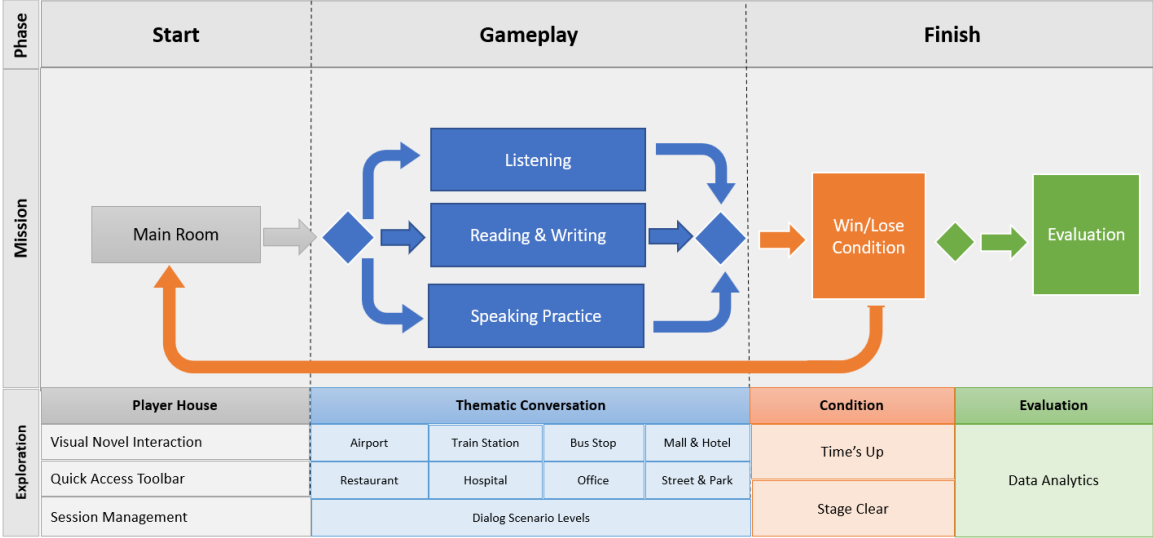


Figure 2. Langova Design Structure

In the Start Phase, players begin their journey in a main room that represents their personal space within the game. This room serves as the central hub where players can familiarize themselves with the environment before diving into the gameplay. The room is fully explorable using a point-and-click interface, reminiscent of the mechanics commonly found in visual novels. This familiar and intuitive interaction model allows players to navigate the space with ease, encouraging exploration and engagement right from the start [41].

The user interface (UI) is carefully designed to complement the immersive experience. A quick access toolbar is prominently featured, allowing players to swiftly jump to specific levels or areas of the game without needing to explore the entire room. This feature is particularly useful for players who wish to focus on specific learning tasks or revisit certain chapters. The UI design employs a blue color scheme, which is not only visually appealing but also chosen for its psychological effects [42]. The color blue is often associated with calmness, trust, and professionalism, making it an ideal choice for a serious educational game. Additionally, blue is also often considered a color that shows professionalism, so it can reflect the quality of the game and the developer's seriousness in creating a good experience.

The combination of the AI-generated visuals and the blue-themed UI creates an environment that is both inviting and conducive to focused learning. This setting helps establish a serious yet elegant atmosphere, encouraging players to engage deeply with the content while feeling as though they are part of a high-tech, educational journey. Figure 3 illustrates the implementation of the main room as starting point for the player's adventure.



Figure 3. Player’s Main Room

Meanwhile, during the Gameplay Phase, players are presented with a variety of interactive options designed to enhance their English language skills. They can choose to read and practice English through engaging text-based exercises, participate in listening activities that focus on building vocabulary, or practice speaking by selecting one of the 10 available chapters. Each chapter is tailored to a specific public place, offering choices such as whether to speak the lines of the first or second person and how many questions to tackle within each session [43]. This allows players to customize their learning experience based on their comfort level and specific language learning goals. The user interface is designed to be intuitive and user-friendly, with a clear and responsive selection interface as shown in Figure 4.

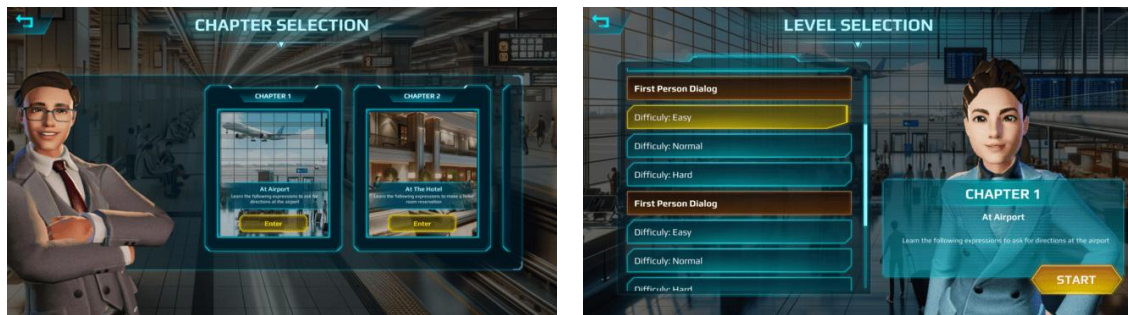


Figure 4. Langova Chapter & Level Selection

The visual elements of the game are particularly noteworthy, as all background environments are generated using AI to accurately reflect the context of the 10 public places, such as airports, train stations, bus stops, hotels, and more. These AI-generated environments incorporate realistic details, such as lighting, textures, and spatial arrangements, to closely mimic the appearance and ambiance of real-world locations. By providing such lifelike settings, the visuals help players immerse themselves in the scenarios, making language practice feel more natural and relevant to everyday life.

Additionally, players are represented by avatars that are generated using the Ready Player Me service. These avatars allow for personalized representation, enabling players to select features such as clothing, accessories, or hairstyles that align with their preferences and the context of each scenario. The combination of AI-generated visuals and personalized avatars transforms the game into an interactive and immersive platform, bridging the gap between digital learning and real-life conversational practice. Figure 5 illustrates the implementation of scenario levels.



Figure 5. Langova Scenario Level

The game concludes when the player successfully answers all questions within a 5-minute timeframe or when the allotted time expires. The player's score is determined based on the number of correctly repeat sentences, providing a straightforward metric to evaluate their performance. The assessment method for speaking relies on advanced speech-to-text technology, followed by a similarity check against the expected answers. This approach ensures that players receive immediate and relevant feedback on their spoken responses.

The similarity algorithm operates differently from traditional human-to-human speech detection thresholds, which define the minimum hearing level required for recognizing speech material 50% of the time. Unlike human interactions, the system must account for factors such as microphone noise, recognition errors, and varying levels of clarity in user speech. To accommodate these variables, different similarity thresholds are applied based on the game's difficulty levels. At the easy level, a response is considered correct if the similarity score is 60% or higher. For the normal level, this threshold increases to 70%, while the most difficult level requires a similarity score of 80% or higher for the answer to be deemed valid.

In addition to these scoring mechanisms, the game incorporates an empathetic response system that adds another layer of realism and engagement. When a player provides a correct or incorrect answer, specific expressions are triggered on their avatar, reflecting the accuracy of their response. These emotional cues are designed to deepen the player's emotional engagement, making the simulation feel more authentic and immersive. The change in avatar expressions—whether showing satisfaction for a correct answer or disappointment for an incorrect one—helps players to connect emotionally with the virtual environment.

Moreover, this dynamic emotional feedback serves an educational purpose beyond just language learning. By observing and responding to the avatar's facial expressions, players can practice interpreting and reacting to emotional cues, which can help reduce anxiety and build confidence as if they were conversing with a real person [44]. This aspect of the game allows players to experience and manage the emotional dynamics of communication in a controlled environment. Figure 6 illustrates the implementation of these correct and incorrect response conditions.

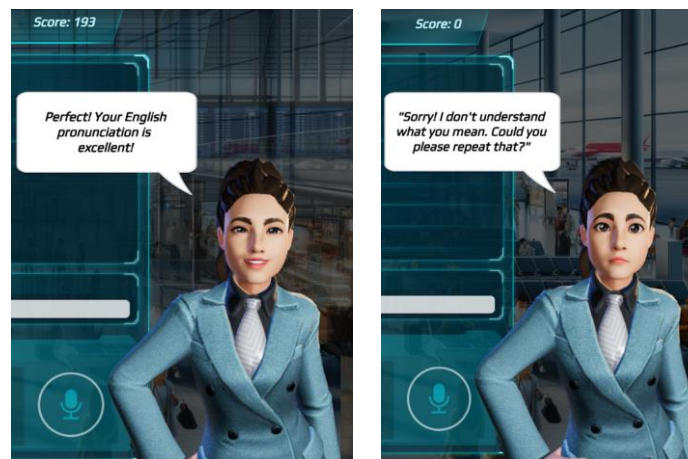


Figure 6. Langova Correct and Wrong Condition

3.2 Tools And Technology

The game was developed using the Unity Game Engine v2022.3.30f1 and deployed for WebGL, enabling seamless access through web browsers. For speech recognition, the system integrates the Kaldi plugin v5.5.636, an open-source speech recognition toolkit written in C++ for speech recognition and signal processing, available under the Apache License v2.0. This plugin allows both partial and full speech recognition to accurately detect spoken words. The

language model employed is the `small_english` model, which is optimized for lightweight deployment under the same Apache License.

The application is hosted on a server compatible with PHP 8.0–8.2 and uses CodeIgniter 3.0 to bridge the web application and the MySQL database. The MySQL database stores user data, including player responses and progress, which can be accessed and monitored via a dedicated dashboard. This infrastructure ensures the system's scalability and supports ongoing development and evaluation.

Each speaking interaction by a user is stored as a single transaction record in the database. Specifically, the system logs the words spoken by the player, the expected correct responses, and the similarity scores [45]. Ideally, completing two chapters (each consisting of 10 dialogues) would result in 20 transaction records. However, if a user generates significantly more transactions, it may indicate a hesitancy to speak or the need to repeat sentences due to unclear inputs or a lack of confidence. This pattern of increased transactions can be interpreted as reflective of initial language anxiety. Over time, as users engage more frequently with the application and complete additional levels, the repetition of speaking attempts can help build their confidence and fluency.

While the current research does not explicitly measure language anxiety, Langova's thematic and progressive structure, which includes multiple levels across diverse scenarios, inherently encourages continuous practice. Moreover, the ability to analyze transaction records allows for indirect inferences about users' comfort levels. For instance, users who engage in more speaking attempts than the expected number of transactions demonstrate a willingness to persevere despite initial hesitation. This behavior suggests that Langova not only facilitates frequent speaking practice but also gradually helps reduce language anxiety by encouraging users to speak more frequently, ultimately building their confidence in a private and supportive environment.

4. Results & Evaluation

4.1 Participants and Data Collection

The Langova platform was hosted online and tested with a group of 57 first-year and second-year students from Telkom University, consisting of 27 female and 30 male participants. The participants were recruited through invitations sent via WhatsApp group chats, which provided a convenient and accessible means of reaching them. Those who agreed to participate received a detailed instruction document outlining the steps for accessing Langova, creating an account, and navigating the application.

To ensure diversity in language proficiency, participants were selected from two academic cohorts: fourth-semester (Year 2) and second-semester (Year 1) students, with a distribution of 57% - 43%. While no formal English proficiency test was administered, their academic standing was used as a proxy indicator of their language abilities. Based on Telkom University's curriculum structure, fourth-semester students had undergone more English coursework than their second-semester counterparts, providing a reasonable assumption that their English proficiency level was generally higher. However, this assumption accounted for exceptions, such as second-semester students who might have had strong English skills prior to university enrollment.

Once participants received the instructions, they accessed Langova through the provided website link, where they were required to create an account to track their progress. The testing environment was flexible, allowing students to engage with the application from various locations—some from home and others from campus. After setting up their accounts, participants were required to complete at least two chapters, each consisting of 10 dialogues.

Throughout the testing period, participants who successfully completed the chapters reported their progress back to the WhatsApp group, where the admin monitored their activity. The admin then cross-referenced the reported data with the records stored on the server to track overall participation. This process allowed for monitoring of the testing phase, ensuring that all participants' interactions with the Langova were documented and analyzed. The combination of thorough instructions, flexible access options, and data monitoring provided a comprehensive and effective framework for evaluating the game's impact on English language learning among the students. The structured format of the chapters, along with the varied dialogue options, allowed students to practice real-world scenarios, making the learning process feel relevant and practical. Table 3 shows the results from a subset of the total 10,823 raw transactions or records.

Table 3. Subset of Raw Data Records

Student	Solution/Answer	Speech Detected Result
Student 1	Excuse me. Could you tell me where Gate 3 is?	Excuse me, could you tell me where Gabriel is?
Student 5	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where the tree is
Student 6	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where Gateway is
Student 7	Excuse me. Could you tell me where Gate 3 is?	Kiss me could you tell me where Get Rich
Student 9	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where the get rich
Student 10	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where Get Rich
Student 12	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where get there is
Student 16	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where get
Student 18	Excuse me. Could you tell me where Gate 3 is?	Excuse me, could you tell me where gate 3 is?
Student 23	Excuse me. Could you tell me where Gate 3 is?	Excuse me. Could you tell me where Gate 3 is?
Student 28	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where to get trip is
Student 30	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where to get three is
Student 34	Excuse me. Could you tell me where Gate 3 is?	Surfing Get Rich is just I have your the Advance
Student 36	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where get Three is
Student 43	Excuse me. Could you tell me where Gate 3 is?	Three three three three three three three three
Student 46	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where Gate 3 is
Student 49	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell
Student 50	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where get trees
Student 54	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where country is
Student 57	Excuse me. Could you tell me where Gate 3 is?	Excuse me could you tell me where Gateway is

The speech recognition results further illustrate the challenges faced by learners in replicating the expected sentences. For example, the expected phrase “Excuse me. Could you tell me where Gate 3 is?” often resulted in variations such as “Excuse me, could you tell me where Gateway is?” or “Excuse me, could you tell me where to get three is?” These deviations highlight issues such as pronunciation inaccuracies, speech recognition errors, and background noise interference. Interestingly, Student 43’s detected result, “Three three three three three three three three,” indicates repeated attempts, aligning with her need to retry multiple times due to anxiety and difficulty in articulation. While several students provided responses closer to the expected output, many struggled with maintaining accuracy, showcasing a broader pattern of speech-related challenges among participants.

In the Langova trial, each participants was required to complete at least two chapters, which theoretically results in a minimum of 20 transactions. However, the raw data reveals that the average number of transactions per participants reached 189, indicating a high level of participants engagement, with many participants attempting more than the required levels.

Langova offers 10 levels, each with two dialogue variations (first-person and second-person), resulting in a maximum of 200 transactions for participants who complete all levels. If a participant exceeds 200 transactions, it suggests a strong interest in practicing, frequent errors, or possibly experiencing language anxiety, prompting them to attempt more than the average number of times. Table 4 shows the results of the 20 users with the highest number of transactions.

Table 4. Overview of Student Highest Transactions

Student	Years	Total Transactions	Student	Years	Total Transactions
Student 43	2	1496	Student 20	2	231
Student 30	2	1187	Student 22	2	230
Student 50	1	603	Student 25	2	225
Student 34	1	495	Student 1	2	223
Student 36	2	392	Student 13	2	212
Student 6	2	375	Student 14	1	210
Student 46	1	351	Student 16	2	202
Student 37	2	332	Student 18	2	202
Student 49	1	331	Student 5	2	200
Student 7	2	237	Student 12	2	200

The findings demonstrate that Langova, as a WebGL-based game leveraging Speech Recognition technology, effectively supports learners in improving their English conversation skills and reducing language anxiety. By providing simulated public settings and enabling repeated practice with immediate feedback, Langova creates a private and flexible environment that allows learners to build confidence and overcome barriers related to public speaking.

For instance, the case of Student 43, a Year 2 (fourth-semester) student, who recorded 1,496 transactions, illustrates how Langova's design accommodates learners who experience language anxiety and require multiple attempts to master pronunciation and conversational fluency. This participant far exceeded the expected maximum of 200 transactions needed to complete all chapters in Langova. Reportedly, she is a shy learner who needed to repeat exercises multiple times before successfully completing each chapter. Despite these challenges, she exhibited remarkable perseverance and enthusiasm, ultimately completing all chapters. She acknowledged difficulties with pronunciation, often requiring continuous repetition to match the expected dialogue.

Interestingly, the research findings challenge initial predictions—it was expected that second-semester (Year 1) students would struggle more with language anxiety and require frequent repetitions. However, data analysis revealed that some fourth-semester students, like Student 43, were more likely to engage in excessive repetitions. This suggests that language anxiety may persist even among students with more academic exposure to English, possibly due to higher expectations, self-consciousness, or ingrained hesitation in speaking practices.

This case highlights Langova's effectiveness in supporting learners with language anxiety, providing them with a private, self-paced learning environment where they can practice without social pressure, gradually build confidence, and improve their speaking skills. Moreover, these findings suggest that advanced learners may still require structured support to overcome speaking-related anxiety, reinforcing the importance of adaptive, confidence-building learning tools in language education.

These observations emphasize the value of Langova as an educational tool. By enabling repeated practice and providing immediate feedback, Langova creates an environment that accommodates varying proficiency levels and individual learning paces. Overall, these findings demonstrate Langova's capacity to support learners in overcoming language-related barriers,

particularly for those who may feel hesitant or lack confidence in public speaking scenarios. This makes Langova a promising tool for larger-scale deployment in educational settings.

4.2 Game Analysis and Evaluation

In analyzing the effectiveness of Langova as a tool for practicing English, Word Error Rate (WER) was employed as a key metric for evaluating user performance [46,47]. WER is commonly used in measuring speech recognition systems. The metric quantifies how accurately spoken language matches the intended reference text. To perform the WER calculations and analyze the dataset, Google Colab was utilized as the computational platform. Google Colab offers a cloud-based environment that supports Python and includes pre-installed libraries commonly used in data analysis. By leveraging Google Colab, the analysis could be executed seamlessly, ensuring reproducibility and scalability of the computations across the dataset. It is calculated using the following formula:

$$WER = \frac{S + D + I}{N}$$

Where:

- S: Number of substitutions (incorrect words)
- D: Number of deletions (words that were omitted)
- I: Number of insertions (extra words that were added)
- N: Number of words in the reference text

Example Calculation:

For instance, if the reference sentence is 'I would like a cup of coffee' and the detected sentence is 'I will have a cup coffee,' the calculation would involve:

- S: 1 substitution (substituting 'would like' with 'will have')
- D: 0 deletions
- I: 0 insertions
- N: 6 words in the reference text

Thus:

$$WER = (1 + 0 + 0) / 6 = 0.167 \text{ (16.7\%)}$$

The value 0.167 or 16.7% represents the Word Error Rate (WER) for the spoken sentence. It indicates the percentage of words that were incorrectly recognized by the speech recognition system compared to the reference text. WER is used in this analysis because it provides a straightforward way to measure discrepancies between the intended reference text and the recognized output. Table 5 highlights the results from a subset of participants: users 1–20. The full dataset includes performance metrics for all 57 users and is available upon request.

Table 5. Performance Summary for Selected Users (1–20)

Student	Average Solution Word Count	Average Student Answer Word Count	Word Count Difference	Word Rate Error
Student 1	7.80	6.31	-1.49	0.83
Student 2	10.05	9.76	-0.29	0.60
Student 3	10.08	4.82	-5.26	0.89
Student 4	8.51	5.27	-3.24	1.05
Student 5	8.72	8.76	0.03	0.58
Student 6	10.50	7.50	-3.00	0.83

Student 7	9.70	8.21	-1.49	0.81
Student 8	9.72	6.68	-3.03	0.94
Student 9	11.00	8.00	-3.00	0.65
Student 10	9.50	9.51	0.01	0.23
Student 11	9.50	9.60	0.10	0.43
Student 12	7.63	6.95	-0.68	0.94
Student 13	10.00	10.33	0.33	0.63
Student 14	9.81	9.75	-0.06	0.65
Student 15	10.71	10.24	-0.48	0.73
Student 16	10.06	10.16	0.10	0.55
Student 17	9.30	9.25	-0.05	0.56
Student 18	10.75	11.00	0.25	0.75
Student 19	11.08	6.06	-5.02	1.02
Student 20	9.62	9.91	0.29	1.22

The WER value captures different types of errors (substitution, deletion, insertion), which are critical for understanding how well users can accurately pronounce English phrases, and also how well the system can recognize these phrases. The analysis, included a total of 10,823 records, demonstrates that both Word Error Rate (WER) are effective metrics for evaluating the system's ability to provide feedback on speaking performance. WER quantifies the absolute performance by measuring the total number of errors across all sentences, revealing significant variability among users. For instance, Student10 achieved the lowest WER of 0.23, indicating a strong alignment between their responses and the expected solutions, while Student20 recorded the highest WER of 1.22, reflecting substantial deviations. These discrepancies highlight the importance of word count alignment, as users who closely matched the provided solutions in word count, such as Student10, tended to achieve lower WER values.

In terms of word count, Student51 had a largest discrepancy, with an average difference of -5.72 words compared to the solution, which contributed to WER of 0.95. In contrast, Student10 had almost no difference in word count, with an average word count difference of only 0.01, which is closely aligned with the expected solution. Overall, students who maintained a similar word count to the provided solutions, such as Student10, tended to have lower WERs. This suggests that interventions focusing on reducing word count discrepancies could lead to better accuracy and understanding, as students who omit or add fewer words exhibit improved performance.

However, WER alone may not be sufficient for evaluating user performance comprehensively. This is due to the fact that users did not necessarily answer with the same number of sentences, which introduces variability in the total errors counted. Moreover, speech recognition systems are prone to certain inaccuracies that may cause some errors to be attributed to system limitations rather than user proficiency. To address these issues, we also used Average Word Error Rate per Sentence (Avg WER) to normalize the results across different users. AvgWER is calculated as follows:

$$\text{Avg WER per Sentence} = \frac{\sum_{i=1}^N \text{WER}_i}{N}$$

Where:

- N: Number of sentences spoken by the user
- WER_i: Word Error Rate for each individual sentence

Example Calculation:

If a user speaks three sentences with respective WER values of 0.3, 0.5, and 0.1, then the AvgWER is calculated as:

$$\text{AvgWER} = (0.3 + 0.5 + 0.1) / 3 = 0.3$$

Using Avg WER allows us to account for variability in the number of sentences spoken by each user. For example, if a user speaks fewer sentences but has a higher WER for each, AvgWER helps to normalize these differences by calculating the average error per sentence. This way, the impact of speech recognition errors and variability in sentence count can be better understood.

Figure 7 shows the Avg WER for each user. It highlights the effectiveness of Langova across different users by providing insights into average accuracy levels per spoken sentence. The lower the Avg WER, the better the system and user performance, indicating either improved recognition accuracy or more accurate pronunciation.

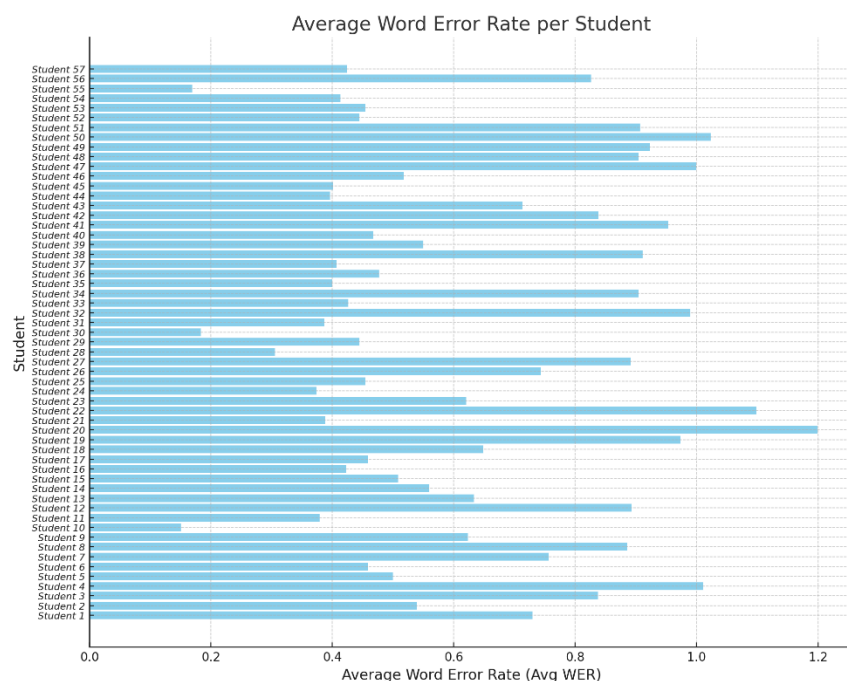


Figure 7. Average Word Error Rate (Avg WER) for each user

The lowest Avg WER value recorded is 0.150, which suggests that this user exhibited very few errors on average, indicating a high level of speaking proficiency or an effective match between the speech recognition system and the user's pronunciation. A lower Avg WER signifies that the system captured the user's speech with high accuracy. The highest Avg WER value is 1.199, indicating a higher average error rate per sentence. This suggests either greater difficulties in pronunciation or more frequent misrecognition by the system, potentially due to articulation issues, accent differences, or background noise. The range between 0.150 and 1.199 illustrates significant variability in the speaking performance and recognition accuracy across users. Some users achieved notably high accuracy with fewer errors, while others encountered challenges either from their pronunciation or from the speech recognition system's limitations.

Both WER and Avg WER are crucial metrics for evaluating the effectiveness of Langova. WER helps quantify the absolute performance of each user across all sentences and Avg WER provides a normalized view to evaluate user proficiency more fairly, especially when users contribute different amounts of data. Combining WER and Avg WER provides a more complete picture of user performance and system capabilities, revealing both the overall

effectiveness of practice sessions and the specific areas where system limitations or pronunciation difficulties may occur. The results from the chart demonstrate that while many users achieved a low average error rate, indicating effective practice outcomes, some users faced challenges—either due to personal speaking proficiency or limitations within the speech recognition technology itself.

The combined use of WER and Avg WER proves effective in assessing Langova's system capabilities and user proficiency. WER provides an overarching view of performance, while Avg WER normalizes the data to reveal consistent patterns of accuracy and challenges across users. The data shows that 43.86% of users achieved an Avg WER below 0.5, indicating relatively good English proficiency and successful outcomes from Langova's thematic practice approach. These metrics collectively demonstrate Langova's ability to provide meaningful feedback and facilitate improved speaking performance, while also identifying areas for potential refinement in the speech recognition system.

In addition to the quantitative metrics, qualitative feedback from participants provided valuable insights into their experience with Langova. Overall, students expressed a positive response to the private practice environment offered by the game. Many participants highlighted that the system allowed them to practice English without the fear of judgment or embarrassment, which they often felt when speaking in front of a classroom. This private and flexible approach was especially appreciated by students who struggled with confidence in public speaking.

However, several technical issues were reported during the trial. Some participants experienced delays in system responses due to unstable internet connections, which affected the flow of the practice sessions. Additionally, microphone responsiveness posed challenges, particularly when using built-in microphones on laptop devices. Participants noted that background noise often interfered with speech recognition, requiring the use of dedicated microphones for clearer input.

Another recurring issue was related to accents. Since the participants were non-native English speakers, their regional accents occasionally influenced pronunciation. While the system often recognized their speech as correct, there were instances where more complex accent variations led to unintended errors or inconsistencies. As a result, several participants suggested incorporating features to evaluate and adapt to different accents to improve the overall user experience. These qualitative insights not only complement the quantitative findings but also highlight areas for further refinement, particularly in addressing technical challenges and enhancing the adaptability of the system to accommodate diverse user needs.

5. Discussion and Conclusions

Langova has demonstrated considerable potential as a tool for enhancing English language learning, particularly in improving students' pronunciation and overall speech accuracy. The analysis of Word Error Rate (WER) from the trial highlights notable variations in student performance, with some students closely aligning their responses to the expected solutions, while others exhibited more significant deviations. For example, students who produced responses with minimal differences in word count tended to have lower WER, suggesting they were able to reproduce the provided sentences more accurately. This reflects Langova's ability to assist learners in reducing language errors when they engage closely with the material.

While Word Error Rate (WER) serves as a practical metric in speech recognition technology, its validity as a direct educational performance indicator warrants further discussion. Although lower WER values generally indicate better pronunciation alignment, they do not always capture a learner's comprehensive language proficiency, such as fluency, intonation, or contextual understanding. Therefore, future studies involving Langova may incorporate cross-referencing WER scores with human scoring assessments to examine correlations between automated metrics and actual language acquisition.

Moreover, the Avg WER across the trial demonstrated that students who kept their responses similar in length and structure to the expected solutions generally performed better. These findings indicate that Langova's immersive design, which incorporates realistic dialogue and repetition, plays a crucial role in helping students match their responses to native-like sentence structures. By providing an interactive and safe space for practicing English, Langova enables learners to explore various vocabulary and sentence forms without fear of making mistakes. Langova offers a private and flexible environment which significantly reduces performance anxiety, particularly for students who feel self-conscious or fear judgment in group settings. Participant feedback highlights this benefit, with many praising the ability to practice at their own pace and in a low-pressure environment.

In contrast, traditional classroom settings offer real-time, human feedback and spontaneous interaction, which can be beneficial for developing quick-thinking conversational skills [48]. However, such environments may not accommodate individual learning speeds and can sometimes hinder participation from students who are less confident [49]. Langova addresses these limitations by incorporating adaptive personalized learning environments, allowing learners to focus on their specific weaknesses, such as pronunciation or grammar, without the immediate social pressures typical in a classroom.

Many of the deviations observed in student responses may not solely reflect learner deficiencies but rather limitations inherent in the speech recognition system. For example, factors such as accent variation, background noise, and microphone quality significantly impacted recognition accuracy. This distinction is crucial because students who accurately repeated sentences in their regional dialects were sometimes penalized by the system due to misrecognition. To address this, future analyses should include error categorization distinguishing between system-induced errors and learner-originated errors. By implementing this differentiation, educators and developers can obtain a more accurate representation of actual learner capabilities and refine feedback mechanisms accordingly.

Despite all the advantages, Langova also has its limitations. The absence of face-to-face human interaction means learners may miss out on developing non-verbal communication skills, such as interpreting body language or facial expressions, which are crucial for effective communication. Additionally, while Langova provides scenario-based learning and adaptive feedback, it may not fully replicate the dynamic, unpredictable nature of real-world conversations that classroom discussions or live interactions offer. However, when used in conjunction with traditional classroom exercises, Langova can serve as a powerful supplementary tool in language learning, providing additional practice outside formal settings.

Future improvements for Langova will focus on addressing these technical limitations through several strategies. To mitigate microphone sensitivity issues, the most effective approach will be to establish a standard procedure recommending users to utilize specific headset models known for superior noise absorption. Providing clear guidelines on the optimal hardware setup will help reduce variability in audio input quality. Additionally, while the implementation of noise-cancellation algorithms can further minimize background interference, these algorithms may introduce processing delays when used in a game environment [50]. Therefore, careful evaluation of their performance impact will be essential before full integration to ensure they do not hinder real-time interaction within the application.

Regarding accent recognition, this remains a significant challenge across the field of Natural Language Processing (NLP), and there is no straightforward solution yet [51]. Recognizing and accurately interpreting diverse regional accents is a complex issue that even advanced speech recognition systems struggle to fully address. For the time being, Langova will focus on integrating the latest existing machine learning models specifically designed to improve accent detection and adaptability. By leveraging advancements in pre-trained speech recognition models and utilizing diverse linguistic datasets from these technologies, Langova aims to better accommodate a wide range of speech patterns. While this approach may not completely resolve the issue, it represents a practical and critical step towards making Langova more inclusive and effective for a global audience.

To offer more concrete future directions, Langova is exploring the integration of state-of-the-art speech recognition models such as Whisper by OpenAI and wav2vec 2.0 by Facebook AI, which have shown improved performance in handling noisy environments and diverse

accents. These upgrades aim not only to boost system accuracy but also to humanize the learning experience, making Langova a more immersive and emotionally resonant tool for language practice. Beyond technical refinements, future iterations of Langova could benefit from integrating adaptive learning mechanisms that adjust dialogue complexity based on user performance, providing a more personalized learning experience. By leveraging real-time analytics and user feedback, the system could dynamically tailor conversations to match the learner's proficiency level, gradually increasing complexity as their skills improve. This adaptive approach would not only enhance user engagement but also foster more effective language acquisition by addressing individual learning needs.

Acknowledgments

This research activity is supported by the Competitive Research and Innovation Funding for Advanced Indonesia (RIIM Kompetisi) program, funded by the Indonesia Endowment Fund for Education (LPDP) under the Ministry of Finance of the Republic of Indonesia, in collaboration with the National Research and Innovation Agency (BRIN). This support is provided under contract number 205/IV/KS/11/2023.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] A. Triastuti and M. Riazi, "Indonesian EFL teachers' content conceptualization and course organization: A portray of text-based teaching," *Indonesian Journal of Applied Linguistics*, vol. 9, no. 3, pp. 1-12, 2020, doi: <https://doi.org/10.17509/ijal.v9i3.23202>.
- [2] S. Dimova, "English language requirements for enrolment in EMI programs in higher education: A European case," *Journal of English for Academic Purposes*, vol. 47, pp. 1-10, Sep. 2020, doi: <https://doi.org/10.1016/j.jeap.2020.100896>.
- [3] S. A. Ramasamy, "Facilitating the construction of knowledge collectively through dialogic discourse: Teachers' perspectives and practices in English language teaching," *Teflin Journal*, vol. 31, no. 1, pp. 79-96, 2023, doi: <https://doi.org/10.15639/teflinjournal.v34i1/79-96>.
- [4] J. J. Ávila-Cabrera and A. C. Esteban, "The project SubESPSkills: Subtitling tasks for students of Business English to improve written production skills," *English for Specific Purposes*, vol. 63, pp. 33-44, Jul. 2021, doi: <https://doi.org/10.1016/j.esp.2021.02.004>.
- [5] J. Gimenez, "Integrating multi-communication research and the business English class," *English for Specific Purposes*, vol. 71, pp. 87-89, Jul. 2023, doi: <https://doi.org/10.1016/j.esp.2023.02.008>.
- [6] P. B. Thi and Y.-H. Cheng, "English needs analysis of food & beverage employees: A case study of a 5-star resort in Vietnam," *English for Specific Purposes*, vol. 76, pp. 74-105, Oct. 2024, doi: <https://doi.org/10.1016/j.esp.2024.07.001>.
- [7] T. Elyas, B. AlHashmi, and F. Fang, "Cognitive diversity among EFL learners: Implications for teaching in higher education," *Teflin Journal*, vol. 31, no. 1, pp. 44-69, 2020, doi: <https://doi.org/10.15639/teflinjournal.v31i1/44-69>.
- [8] N. H. Abdurahman and M. A. Rizqi, "Indonesian students' strategies to cope with foreign language anxiety," *Teflin Journal*, vol. 31, no. 1, pp. 1-18, 2020, doi: <https://doi.org/10.15639/teflinjournal.v31i1/1-18>.
- [9] S. Kirk and J. King, "EAP teacher observation: Developing criteria and identifying the forms of pedagogic practice they afford," *Journal of English for Academic Purposes*, vol. 59, pp. 1-12, Sep. 2022, doi: <https://doi.org/10.1016/j.jeap.2022.101139>.
- [10] S. Zhou, S. Zhao, and M. Groves, "Towards a digital bilingualism? Students' use of machine translation

- in international higher education," *Journal of English for Academic Purposes*, vol. 60, pp. 1-10, Nov. 2022, doi: <https://doi.org/10.1016/j.jeap.2022.101193>.
- [11] L. Shum, Y. Rosunally, S. Scarle, and K. Munir, "Personalised learning through context-based adaptation in the serious games with gating mechanism," *Education and Information Technologies*, vol. 28, pp. 1–32, 2023, doi: 10.1007/s10639-023-11695-8.
 - [12] S. Wang, B. Cohen, S. Yi, J. Park, N. Teo, and E. Andersen, "Goal-based progression synthesis in a Korean learning game," in *FDG 19: Proceedings of the 14th International Conference on the Foundations of Digital Games*, 2019, pp. 1–9, doi: 10.1145/3337722.3337745.
 - [13] J. M. Newton, D. R. Ferris, C. C. M. Goh, W. Grabe, F. L. Stoller, and L. Vandergrift, "Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking," Routledge, New York, 2018, 286 pp., ISBN: 978-1-138-64760-2.
 - [14] M. Vuković-Stamatović, "Suitability of science & technology documentaries for EAP and EST listening," *Journal of English for Academic Purposes*, vol. 58, pp. 101137, Jul. 2022, doi: <https://doi.org/10.1016/j.jeap.2022.101137>.
 - [15] R. Yeager, G. Park, and R. J. T. Liao, "Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment," *Journal of English for Academic Purposes*, vol. 68, pp. 101346, Mar. 2024, doi: <https://doi.org/10.1016/j.jeap.2024.101346>.
 - [16] M. Hinton, "An aptitude for speech: The importance of mimicry ability in foreign language pronunciation," in *Teaching and Researching English Accents in Native and Non-native Speakers*, E. Waniek-Klimczak and L. R. Shockey, Eds., Second Language Learning and Teaching, Springer, Berlin, Heidelberg, 2013, doi: 10.1007/978-3-642-24019-5_8.
 - [17] J. Geluso, "Grammatical and functional characteristics of preposition-based phrase frames in English argumentative essays by L1 English and Spanish speakers," *Journal of English for Academic Purposes*, vol. 55, pp. 101072, Jan. 2022, doi: <https://doi.org/10.1016/j.jeap.2021.101072>.
 - [18] M. Gui, X. Cao, and J. Zhao, "EAP reading ability development during the process of generating research questions: A CDST perspective," *Journal of English for Academic Purposes*, vol. 63, pp. 101249, May 2023, doi: <https://doi.org/10.1016/j.jeap.2023.101249>.
 - [19] E. Nikbakht and R. T. Miller, "The development of ESL students' synthesis writing through reading instruction," *Journal of English for Academic Purposes*, vol. 65, pp. 101274, Sep. 2023, doi: <https://doi.org/10.1016/j.jeap.2023.101274>.
 - [20] R. C. Anderson, E. A. Chaparro, K. Smolkowski, and R. Cameron, "Visual thinking and argumentative writing: A social-cognitive pairing for student writing development," *Assessing Writing*, vol. 55, pp. 100694, Jan. 2023, doi: <https://doi.org/10.1016/j.asw.2023.100694>.
 - [21] S. Yin, J. Fan, Y. Jin, and P. Stapleton, "Towards a framework of critical thinking for assessing EAP speaking," *Journal of English for Academic Purposes*, vol. 71, pp. 101426, Sep. 2024, doi: <https://doi.org/10.1016/j.asw.2024.101426>.
 - [22] T. Jakonen and D. Duran, "Learner-centered EAP practices: Managing agenda in tutorial interaction," *Journal of English for Academic Purposes*, vol. 68, pp. 101345, Mar. 2024, doi: <https://doi.org/10.1016/j.jeap.2024.101345>.
 - [23] Q. Xie, "Construct representation and predictive validity of integrated writing tasks: A study on the writing component of the Duolingo English Test," *Assessing Writing*, vol. 61, pp. 1-12, Jul. 2024, doi: <https://doi.org/10.1016/j.asw.2024.100846>.
 - [24] S. Egenfeldt-Nielsen, "Serious Games and Gamification: When Entertainment is not enough," in *Understanding Video Games: The Essential Introduction – 4th Edition*, S. Egenfeldt-Nielsen, J. Smith, and S. Tosca, Eds., Routledge, 2019. ISBN: 978-1138363052
 - [25] T. Beranič and M. Heričko, "The impact of serious games in economic and business education: A case of ERP business simulation," *Sustainability*, vol. 14, no. 2, p. 683, 2022, doi: <https://doi.org/10.3390/su14020683>.
 - [26] P. J. Aguilar-Cruz and H. A. Álvarez Guayara, "A serious game to learn English: The case of BethelChallenge," *International Journal of Serious Games*, vol. 8, no. 4, Dec. 2021, doi: <http://dx.doi.org/10.17083/ijsg.v8i4>
 - [27] B. Y. Cahyono and U. Widiati, "The teaching of EFL vocabulary in the Indonesian context: The state of the art," *TEFLIN Journal*, vol. 19, no. 1, pp. 1-17, 2018, doi:

- <https://doi.org/10.15639/teflinjournal.v19i1/1-17>.
- [28] H. Nesi and S. Alsop, "Stories and scenarios: Lecturers' use of fantastic hypothetical events," *Journal of English for Academic Purposes*, vol. 53, pp. 101022, Sep. 2021, doi: <https://doi.org/10.1016/j.jeap.2021.101022>.
- [29] C. Sandiford and M. Macken-Horarik, "Changing stories: Linguistically-informed assessment of development in narrative writing," *Assessing Writing*, vol. 45, pp. 100471, Jul. 2020, doi: <https://doi.org/10.1016/j.asw.2020.100471>.
- [30] V. Beltrán-Palanques, "Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP," *Assessing Writing*, vol. 60, pp. 100809, Apr. 2024, doi: <https://doi.org/10.1016/j.asw.2024.100809>.
- [31] F. Picavet, A. Henderson, and E. de Vries, "Storytelling in L2 English-medium engineering lectures: A typology," *Journal of English for Academic Purposes*, vol. 62, pp. 101214, Mar. 2023, doi: <https://doi.org/10.1016/j.jeap.2022.101214>.
- [32] N. Dinçer, "The effect of a serious game on aviation vocabulary acquisition," *International Journal of Serious Games*, vol. 8, no. 4, 2021, doi: <https://doi.org/10.17083/ijsg.v8i4.464>.
- [33] P. J. Aguilar Cruz, "Understanding students engagement with a serious game to learn English: A sociocultural perspective," *International Journal of Serious Games*, vol. 9, no. 4, 2022, doi: <https://doi.org/10.17083/ijsg.v9i4.554>.
- [34] J. Frome, "Interactive Works and Gameplay Emotions," *Games and Culture*, vol. 14, no. 7-8, pp. 715-728, 2019. doi: 10.1177/1555412019847907.
- [35] C. Cole, R. H. Parada, and E. Mackenzie, "Why and How to Define Educational Video Games?," *Games and Culture*, Jun. 2023. doi: 10.1177/15554120231183495.
- [36] E. Nimy and M. Mosia, "Modelling Student Retention in Tutorial Classes with Uncertainty—A Bayesian Approach to Predicting Attendance-Based Retention," *Education Sciences*, vol. 14, no. 8, p. 830, 2024. doi: 10.3390/educsci14080830.
- [37] J. Karapakdee and P. Wannapiroon, "Immersive Digital Storytelling Learning Experience with a Metaverse Gamification Game Platform to Enhance Game Developer Competency," *International Journal of Information and Education Technology*, vol. 13, no. 6, pp. 890-898, 2023. doi: 10.18178/ijiet.2023.13.6.1884.
- [38] M. Nowacki and A. Stasiak, "Identification of Memorable Game Experiences: Analysis of Escape Rooms Visitors' Experiences," *Games and Culture*, Aug. 2023. doi: 10.1177/15554120231190648.
- [39] M. Whitby, F. Mao, K. Stawarz, S. Marrero, and J. Walton, *Serious Games Cookbook: A Beginner's Guide to Using and Designing Serious Games*, 2024. doi: 10.31273/978-1-911675-13-6.a
- [40] R. Roedavan, Y. Siradj, and S. Stefany, "Educational game scenario model based on imperative game goal typology," *Journal of Games, Game Art, and Gamification*, vol. 8, no. 1, pp. 18-23, 2023. Doi: <https://doi.org/10.21512/jggag.v8i1.9497>
- [41] M. Consalvo and D. Staines, "Reading Ren'Py: Game Engine Affordances and Design Possibilities," *Games and Culture*, vol. 16, no. 6, pp. 659-680, Nov. 2020. doi: 10.1177/155541202009738.
- [42] C. Neale, A. Griffiths, L.S.C. Pui, S. Mendu, M. Boukhechba, and J. Roe. *Color aesthetics: A transatlantic Comparison of Psychological and Physiological Impacts of Warm and Cool Colors in Garden Landscapes. Wellbeing, Space and Society. Vol 2*, 2021. <https://doi.org/10.1016/j.wss.2021.100038>
- [43] M. Consalvo, T. Busch, and C. Jong, "Playing a Better Me: How Players Rehearse Their Ethos via Moral Choices," *Games and Culture*, vol. 14, no. 3, pp. 272-290, Nov. 2016. doi: 10.1177/1555412016677449.
- [44] A. Jerrett, P. Howell, and N. Dansey, "Developing an Empathy Spectrum for Games," *Games and Culture*, vol. 16, no. 6, pp. 681-700, Sep. 2020. doi: 10.1177/1555412020954019.
- [45] N. Gahman and V. Elangovan, "A Comparison of Document Similarity Algorithms," *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 14, no. 2, pp. 13-29, Mar. 2023. doi: 10.48550/arXiv.2304.01330.
- [46] A. Ali and S. Renals, "Word Error Rate Estimation for Speech Recognition: e-WER," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, I. Gurevych and Y. Miyao, Eds., Jul. 2018, pp. 20–24, Association for Computational Linguistics. doi: 10.18653/v1/P18-2004.

- [47] T. von Neumann, C. Boeddeker, K. Kinoshita, M. Delcroix, and R. Haeb-Umbach, "On Word Error Rate Definitions and their Efficient Computation for Multi-Speaker Speech Recognition Systems," *Electrical Engineering and Systems Science*, Jul. 21, 2023. doi: arXiv:2211.16112v2
- [48] M. Mohammed, "Investigating EFL learners' weakness in English as a foreign language," *Arab Journal of Humanities and Social Sciences*, vol. 2, no. 19, 2023. doi: 10.59735/arabjhs.v2i19.29.
- [49] C. Arranz-Barcenilla, S. Pavía, M. Saiz-Manzanares, L. Alameda, and S. Gutiérrez-González, "Developing an adaptive virtual learning environment for sustainable learning in individuals with Down syndrome," *The International Journal of Information and Learning Technology*, vol. 41, pp. 318–337, 2024. doi: 10.1108/IJILT-10-2023-0200.
- [50] L. Qian and Q. Zhu, "An improved microphone array noise reduction algorithm for speech recognition," in *Proceedings of the 2021 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy*, 2022. doi: 10.1007/978-3-030-89511-2_16.
- [51] R. Río, C. Miller, J. Profant, J. Drexler-Fox, Q. McNamara, N. Bhandari, N. Delworth, I. Pirkin, M. Jetté, S. Chandra, H. Peter, and R. Westerman, "Accents in speech recognition through the lens of a World Englishes evaluation set," *Research in Language*, vol. 21, pp. 225–244, 2023. doi: 10.18778/1731-7533.21.3.02.