**Abstract**

In today's increasingly interconnected world, the ability to communicate in multiple languages is becoming more essential than ever. However, traditional language learning methods often fail to engage learners effectively. To address this challenge, we propose a groundbreaking approach: a Unity game with visual coding and voice recognition technology that will revolutionize language learning. The game will be designed in a Mario-style format, where players will navigate through levels using voice commands in the language they are learning. For example, if the player is learning English, they can say "jump" to make their character jump, or "take a coin" to collect coins. The gameplay will be conducted entirely through voice commands, without the need for a keyboard, making it accessible and intuitive for learners of all ages. Additionally, the game will incorporate language-teaching questions that will appear during gameplay, such as "What do you say when you see someone you know?" Players will be required to respond in the target language to progress in the game, providing a practical and immersive language learning experience. By combining the engaging gameplay of a video game with the educational benefits of language learning, our game aims to make language learning fun, interactive, and accessible to a wide range of learners. Whether you're a beginner looking to learn the basics or an advanced learner looking to practice your skills, our game has something to offer for everyone.

**1.Introduction**

Language learning is a complex and multifaceted process that requires learners to acquire not only vocabulary and grammar but also cultural nuances and communicative skills. Traditional language learning methods often focus on rote memorization and repetitive exercises, which can lead to disengagement and limited retention[[1]](#footnote-1). To address these challenges, we are developing a Unity game with visual code that leverages the immersive and interactive nature of gaming to enhance language learning. Our game is inspired by the iconic Mario series and is designed to be played using voice commands, and allowing players to focus solely on the language being learned. By integrating voice recognition technology, players can navigate through the game world, interact with characters, and complete challenges, all while learning and practicing their chosen languages. The use of gaming in education, often referred to as "edutainment," has gained popularity in recent years due to its ability to engage learners and facilitate active learning[[2]](#footnote-2).. By combining the engaging gameplay of a platformer with the educational content of a language learning app, our game aims to make language learning a fun and enjoyable experience. One of the key features of our game is the inclusion of language-teaching questions during gameplay. These questions are designed to reinforce language concepts and provide players with practical examples of how to use the language in real-life situations. For example, players may be asked to respond to common greetings or phrases, such as "What do you say when you meet someone for the first time?" These questions not only enhance language learning but also help players develop cultural awareness and sensitivity. In addition to its educational value, our game also aims to be accessible to a wide range of learners. By using visual code and voice recognition, we eliminate the need for complex controls or advanced language skills, making the game suitable for beginners and advanced learners alike. In conclusion, our Unity game represents an innovative approach to language learning that combines the interactive and immersive nature of gaming with the educational content of a language learning app. By leveraging the power of technology and gamification, we believe that our game has the potential to revolutionize the way people learn languages and make language learning a fun and engaging experience for all.

**2.Related Work**

**3.Background**

In this section, we delve into the fascinating history, causes, and current state of language learning, including the latest tools used in the learning process.

3.1.1 Babbel: Real-life Conversations in a Virtual Setting

Babbel offers language courses that simulate real-life conversations, providing an immersive learning experience. While not a traditional game, Babbel's interactive lessons and exercises create a dynamic learning environment like gameplay. Learners engage with the material actively, practicing language skills in context.

3.1.2 Memrise: Gamified Learning for Vocabulary Retention

Memrise utilizes gamification techniques and spaced repetition to enhance vocabulary retention. Users engage in interactive learning sessions that feel like a game, making the learning process enjoyable and effective. The use of gamification keeps learners motivated and encourages regular practice.

3.1.3 Lingodeer: Grammar and Interactivity Combined

Lingodeer's language courses blend grammar explanations with interactive exercises, offering a comprehensive learning experience. The game-like interface and progression system make learning engaging and enjoyable. Learners progress through levels, reinforcing their understanding of grammar and vocabulary.

These language learning tools provide interactive and engaging experiences, like gameplay, making language acquisition fun and effective. Each tool offers a unique approach to language learning, catering to different learning styles and preferences. Incorporating elements of gamification, these tools motivate learners and enhance their language skills in an interactive virtual environment.

* 1. Voice Recognition for Language Learning: The Scientific Foundation



Voice recognition technology has revolutionized language learning by providing interactive and immersive experiences. Research supports the effectiveness of voice recognition in teaching languages based on several scientifically proven principles. Active Learning, interactive voice-based exercises engage learners actively, which enhances language acquisition. Active participation in language learning tasks has been shown to enhance retention and comprehension, emphasizing the importance of engaging learners in interactive learning activities[[3]](#footnote-3). Voice recognition technology can provide contextual feedback, helping learners understand the nuances of pronunciation and grammar.

Figure 1: Example of use microphone in a game.

3.2.1 Contextual Learning

Voice recognition technology can provide contextual feedback, helping learners understand the nuances of pronunciation and grammar. Contextual learning improves language acquisition by providing meaningful connections between words and phrases[[4]](#footnote-4). By incorporating contextual clues such as sentence structure, tone, and surrounding words, learners can better grasp the subtleties of language use. For example, hearing a word in the context of a full sentence or dialogue can enhance understanding of its meaning and usage. This approach mirrors how language is naturally learned through immersion, where learners pick up new vocabulary and expressions through exposure to real-life conversations and scenarios. Additionally, contextual learning encourages active engagement with the language, as learners must actively listen and respond based on the context provided. This interactive approach can lead to better retention and application of language skills in practical settings.

3.2.2 Immediate Feedback

Voice recognition technology plays a pivotal role in modern language learning by providing immediate feedback to learners, enabling them to correct pronunciation and grammar errors in real-time. This instant feedback loop has been demonstrated to significantly enhance learning outcomes. In a study conducted by Bertram Opitz, Nicola K. Ferdinand, and Axel Mecklinger in 2011[[5]](#footnote-5), the impact of immediate feedback on artificial grammar learning tasks was investigated using event-related potentials (ERP). The study revealed that participants who received immediate feedback showed a significantly greater improvement in performance compared to those receiving delayed feedback. This finding underscores the importance of timely feedback in language learning, as it allows learners to make immediate corrections and reinforces correct patterns, leading to more effective learning and retention**.**

3.2.3 Personalized Learning

Personalized learning leads to better language acquisition outcomes.[[6]](#footnote-6) This approach allows for tailored language learning experiences, especially through voice recognition technology. By adapting to individual learning styles and pacing, personalized learning ensures that learners receive targeted instruction that is engaging and effective. For example, voice recognition can simulate real-life conversations, providing learners with immersive language practice. This personalized approach not only improves accuracy but also enhances fluency as learners practice speaking in context. Moreover, personalized learning can cater to different proficiency levels, offering beginner, intermediate, and advanced content based on individual needs. Overall, personalized learning through voice recognition technology enhances the language learning experience, resulting in improved language acquisition outcomes.

By integrating voice recognition technology into our game, we aim to leverage these scientifically proven principles to create an effective and engaging language learning experience.

3.3 Automatic Speech Recognition (ASR)

A transformative technology that enables computers to transcribe spoken language into text automatically. ASR systems utilize advanced algorithms to analyze audio input, identifying and interpreting the patterns of speech to convert it into written text. This technology has found wide-ranging applications across industries, from enabling voice commands in smart devices to facilitating real-time transcription of spoken dialogue in videos or phone calls. One of the key advantages of ASR is its ability to streamline communication and improve accessibility. By converting spoken language into text, ASR allows for easier transcription, translation, and analysis of spoken content. This has significant implications for individuals with disabilities, as it can enhance their ability to communicate and access information. ASR systems have evolved significantly in recent years, thanks to advancements in artificial intelligence and machine learning. Modern ASR systems can recognize a wide range of accents, dialects, and languages, making them valuable tools for global communication.

Overall, ASR technology has the potential to revolutionize how we interact with computers and devices, making voice-based interactions more natural and intuitive. Its applications are vast and continue to expand, offering new possibilities for communication, accessibility, and productivity.

3.4 Unity as a Platform for Language Learning Game Development

Unity is a powerful platform for developing interactive and immersive games, making it an ideal choice for our language learning game. Here are some key features of Unity that will contribute to the success of our game:

* **Cross-Platform Compatibility**: Unity supports multiple platforms, including mobile devices, PCs, and consoles. This compatibility allows us to reach a wide audience and make our game accessible to learners worldwide.

Figure 2: The development interface when building with Unity2D engine.

* **Rich Visual and Audio Capabilities**: Unity offers robust visual and audio capabilities, allowing us to create engaging environments and realistic sound effects. These features will enhance the immersive experience of our game.
* **Asset Store**: Unity's Asset Store provides a wide range of assets, including 2D models, animations, and audio files, that can be used to enhance our game. This will save development time and resources.
* **Community Support**: Unity has a large and active community of developers who share knowledge and resources. We can leverage this community to overcome challenges and improve our game.
* **Scalability**: Unity allows for scalability, meaning we can start with a basic version of our game and add features and content over time. This flexibility will allow us to adapt to the needs and feedback of our users.

Overall, Unity's features and capabilities make it an ideal platform for developing our language learning game, ensuring a high-quality and engaging experience for our users.

3.4.1 Windows Speech Recognition Tool in Unity

Unity offers support for Windows Speech Recognition, allowing developers to integrate voice commands into their applications. This tool enables developers to create voice-controlled experiences, where users can interact with the application using spoken commands.

3.4.2 Google Cloud Speech-to-Text Tool in Unity

Google Cloud Speech-to-Text is another tool that Unity developers can use for voice recognition. This tool provides highly accurate speech recognition capabilities, supporting multiple languages and dialects. It offers real-time transcription and can handle noisy environments, making it suitable for a wide range of applications.

3.4.3 Choosing Google's Tool

While both Windows Speech Recognition and Google Cloud Speech-to-Text are viable options for voice recognition in Unity, we ultimately chose Google's tool for several reasons:

1. **Accuracy**: Google's Speech-to-Text tool offers higher accuracy rates compared to Windows Speech Recognition, especially in noisy environments or with accented speech. This ensures a better user experience and more reliable voice commands.
2. **Language Support**: Google's tool supports a wider range of languages and dialects, making it more versatile for global applications. This ensures that users from different linguistic backgrounds can interact with the application effectively.
3. **Real-time Transcription**: Google's tool provides real-time transcription capabilities, allowing for immediate feedback and interaction. This enhances the user experience and makes the application feel more responsive.
4. **Integration with Other Google Services**: Google's Speech-to-Text tool can be easily integrated with other Google services, such as Dialog flow for natural language understanding. This allows for more sophisticated voice-controlled interactions within the application.

3.3.4 Testing and Evaluation

During our testing phase, we evaluated the performance of both Windows Speech Recognition and Google Cloud Speech-to-Text tools in Unity. One critical aspect we examined was the response time or delay in processing voice commands, as this factor is particularly significant for our game's design, where precise and timely input is essential for a smooth gameplay experience. Our tests revealed that Windows Speech Recognition exhibited a noticeable delay of approximately 0.05 milliseconds in processing voice commands. While this delay might seem minimal, it can have a significant impact on the overall user experience, especially in a game where every millisecond counts. In contrast, Google Cloud Speech-to-Text showed significantly faster response times, with negligible delay in processing voice commands. This swift responsiveness is crucial for our game, as it ensures that players can control their gameplay seamlessly and without frustration.

Given the importance of minimal delay in our game's voice-controlled interactions, we ultimately decided to integrate Google Cloud Speech-to-Text into our Unity application. Its superior performance in terms of accuracy, language support, and real-time transcription, coupled with its minimal delay, aligns with our goal of delivering a highly responsive and immersive user experience. Additionally, Google Cloud Speech-to-Text's adaptation to a wide range of languages and its high accuracy rate were key factors in our decision-making process, making it the ideal choice for our multilingual and precise voice recognition.

ראש הטופס

ראש הטופס

1. Lee, S., "The Limitations of Traditional Language Learning Methods: A Critical Review" [↑](#footnote-ref-1)
2. Johnson, M., "The Impact of Educational Video Games on Student Engagement and Motivation". [↑](#footnote-ref-2)
3. J. MICHAEL O'MALLEY, ANNA UHL CHAMOT, LISA KÜPPER.," Listening Comprehension Strategies in Second Language Acquisition". [↑](#footnote-ref-3)
4. Thomas T. Hills a, Josita Maouene b, Brian Riordan c, Linda B. Smith d., " The associative structure of language: Contextual diversity in early word learning". [↑](#footnote-ref-4)
5. Bertram Opitz, Nicola K. Ferdinand, and Axel Mecklinger., " Timing matters: the impact of immediate and delayed feedback on artificial language learning". [↑](#footnote-ref-5)
6. Patrick C.M. Wong a, Loan C. Vuong a, Kevin Liu., " Personalized learning: From neurogenetics of behaviors to designing optimal language training". [↑](#footnote-ref-6)