

Educational Game Helping Hearing-Impaired People Improve Their Speaking

Jin You



Master of Science
Cognitive Science
School of Informatics
University of Edinburgh
2023

Abstract

Controlling pitch during speaking is crucial to conveying emotion. Hearing is the main way to get pitch information, and it is crucial for those who want to perceive and imitate pitch changes in speech accurately. People with hearing impairments have limited access to auditory information, which makes it challenging for them to perceive and control pitch changes in speech. Therefore, it is essential to assist the hearing impaired to perceive and learn pitches. This project focuses on aiding pitch learning for the hearing-impaired through a customizable, browser-based real-time voice interaction pitch visualization game.

The project follows the Waterfall Model, involving requirements gathering, system design, prototyping, frontend-backend development, and testing. By integrating machine learning and speech signal processing, the study calculates audio pitch and transforms it into visual elements, aiding pitch perception. Gamified education introduces personalized learning, enhancing engagement.

The research outcomes have broad applications, assisting not only pitch-impaired individuals but also contributing to speech education tools.

Keywords: Pitch learning, Real-time voice interaction, Speech signal processing, Gamified learning, Hearing-impaired.

Ethics approval

This project obtained approval from the Informatics Research Ethics committee.

Ethics application number: 262705

Date when approval was obtained: 2023-06-27

The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Jin You)

Acknowledgements

I would like to express my heartfelt thanks to many people for the completion of my dissertation. Without their help, the completion of this thesis would be impossible. First of all, I would like to extend my sincere gratitude to my supervisor, Dr Brian Mitchell for his valuable and enlightening suggestions for my thesis. In the process of completing the thesis, I am very grateful to him for his patient guidance on several technical problems at all stages of thesis writing. It would have been very difficult for me to complete this thesis work without his insightful comments and guidance. Furthermore, I extend profound gratitude to my parents whose unwavering encouragement, boundless patience, and unconditional love have been a constant source of strength throughout my life's journey. Their love and guidance have been the cornerstone of my growth, propelling me forward during moments of doubt and celebrating my triumphs with unfaltering pride. I am blessed to have such caring and devoted parents, and their immeasurable influence on my character and aspirations is beyond words.

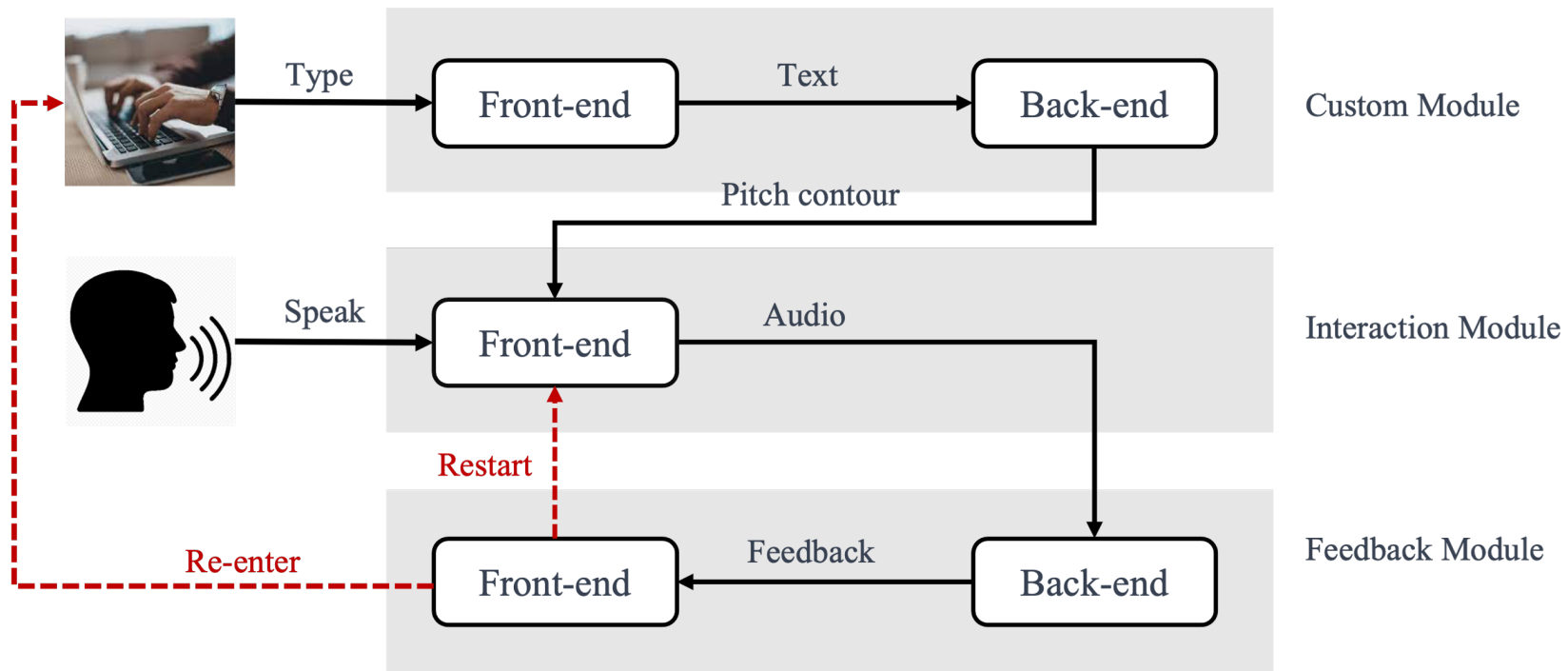


Figure 1 Project overview.

Figure 1 illustrates two interaction modes: user-front-end interaction through text and speech inputs, and back-end-front-end interaction via APIs. It also highlights the modular structure with customization, interactive gameplay, and feedback modules, allowing personalized learning and iterative practice.

Tables of contents

Abstract	i
Ethics approval	ii
Acknowledgements	iii
List of figures	vii
List of tables	viii
1 Introduction	1
1.1 Project motivations	1
1.2 Difficulties of Learning Pitches for Hearing Impaired People	2
1.3 The Importance of Learning Pitches for Hearing Impaired People	3
1.4 Project objectives	4
1.5 Report overview	5
2 Background	6
2.1 Essential concepts	6
2.2 Related work	10
2.3 Related technology	13
2.4 Summary	14
3 Design	15
3.1 Requirements capture	15
3.1.1 Requirements capture method	15
3.1.2 Analysis	15
3.2 System Architecture Design	16
3.3 User interfaces Design	18

3.3.1	Cognitive Walkthrough	20
4	Implementation	22
4.1	Front-end Implementation	22
4.2	Back-end Implementation	23
4.3	Interface Implementation	24
5	Results	25
5.1	Interface display	25
5.2	Interactive display	25
5.3	Performance display	25
6	Evaluation	28
6.1	User study	28
6.2	Usability testing	28
6.3	Questionnaire	29
6.4	Analysis	30
6.4.1	Usability testing analysis	30
6.4.2	Questionnaire analysis	31
6.4.3	Summary	31
7	Conclusions	33
7.1	Project summary	33
7.2	Future work	34
7.3	Reflections	35
7.3.1	Early Project Stage	35
7.3.2	Front-end Development Stage	36
	References	37
A	Participants' information sheet	41
B	Participants' consent form	44
C	API Document	45
D	First questionnaire	46
D.1	Results of first user study	47

E Results of usability testing	50
F Second questionnaire	53
F.1 Results of the second user study	54

List of figures

1	Project overview	iv
2.1	Comparison between time domain and frequency domain	7
2.2	Pitch Contour of “from home to the station”	8
2.3	Punctuation Ambiguity of “No dogs are here”	8
2.4	The speech exercise part in Lan et al. (2014)	12
2.5	The speech exercise part in Hair et al. (2018)	12
3.1	Custom Module	17
3.2	Flowchart of Module 1	17
3.3	Interaction Module and Feedback Module	17
3.4	Flowchart of Module 2	17
3.5	Flowchart of Module 3	17
3.6	UI design for User Input page	21
3.7	UI design for Game page	21
3.8	UI design for Game over page	21
3.9	UI design for Feedback page	21
5.1	User Input page	26
5.2	Game ready page	26
5.3	Game page	26
5.4	Game over page	26
5.5	Feedback page	27
6.1	A decision tree is used to define four severity levels by asking three questions	32

List of tables

2.1	Application Types and Development Technologies	9
2.2	Multimodal Human-Computer Interaction Techniques Related to Speech	9
5.1	Time to generate maps versus number of words	27
5.2	Relationship between audio duration and feedback retrieval time . . .	27

Chapter 1

Introduction

1.1 Project motivations

Intonation and pitch play extremely important roles in enhancing the clarity and fluency of language expression. In English, “pitch” refers to the variation in the highness or lowness of sound within a word, which can influence its part-of-speech and word sense. The pitch is different when “present” is used as a verb and as a noun. Intonation refers to the rising and falling variations in a sentence, related to conveying semantic, emotional, and syntactic information (Wells, 2006). The expression of intonation relies on changes in the pitch of sound, a phenomenon perceived by the auditory system and transmitted to the brain (Hall and Plack, 2009). Normal auditory function enables individuals to accurately recognize pitch differences in sounds, allowing them to understand and utilize pitches to differentiate between part-of-speech and word sense. However, for individuals affected by hearing impairments, the perception of pitch might be hindered. This impediment might further hinder the ability to imitate and produce accurate pitch and intonation.

People with hearing impairments can use a variety of technologies and methods to improve their perception of pitch and intonation. They might use cochlear implants to enhance auditory perception, thus partially restoring the ability to perceive pitch (Zeng, 2004; Lyxell et al., 2009). Utilizing sensory channels such as vision for lip-reading and seeing body language can compensate for auditory deficiencies. Under a speech therapist’s professional guidance, they can also undergo training in mouth shape and articulation, as well as targeted language and communication skills.

O’Callaghan, McAllister, and Wilson (2005) and Theodoros (2008) have shown that these aids and tools are often limited by geographic factors and monetary costs, and

that not all hearing-impaired individuals have access to them. Therefore, it is imperative to explore more effective methods to assist individuals with hearing impairments to help persons with hearing impairment overcome articulation barriers and improve their expressive language skills, while lowering the threshold and cost of learning pitch.

1.2 Difficulties of Learning Pitches for Hearing Impaired People

People with hearing impairments face the following challenges when learning pitch:

- 1. Difficulties in Speech Perception:** Hearing impairments adversely affect speech perception which impairs ability to accurately capture intonation and stress patterns. Due to the lack of direct auditory cues, it is difficult for people affected to recognise pitch and volume changes between syllables, resulting in less accurate perception. Methods of teaching auditory feedback such as lip-reading or auditory-verbal teaching have a limited effect on their intonation learning. This adversely affects their language learning and communicative competence, often creating barriers in interpersonal communication.
- 2. Difficulties in Imitation:** The lack of normal auditory feedback hinders people with hearing impairments to self-correct in verbal expression. Successful pronunciation usually depends on a process of continuous adjustment and refinement, yet the lack of auditory input impedes their awareness of pronunciation deviations. Lacking effective means of self-correction, they may perpetuate mispronunciation habits, thus increasing the difficulty of improving pronunciation.
- 3. Inadequate Resources and Support:** People with hearing impairments may face time, financial, and geographic barriers due to a lack of resources and professional support (Mashima and Doarn, 2008; Lee et al., 2016). Typically, speech therapists focus on large metropolitan areas, with expensive fees. These limitations may prevent accessing the necessary educational and training resources, thus affecting their ability to overcome the challenges associated with learning disabilities.
- 4. Lack of Motivation:** Struggling with pronunciation and communication disorders over a long period of time may lead to frustration and diminished self-confidence in speech expression for hearing impaired people. This diminished self-confidence may cause resistance when learning new pronunciation skills

as they fear that they may fail. As a result, their motivation and enthusiasm for learning may be affected.

1.3 The Importance of Learning Pitches for Hearing Impaired People

The importance of helping hearing-impaired people perceive and learn pitches includes:

Mitigation of Communication Barriers: As the importance of pitch in verbal communication continues to be emphasised, providing opportunities for people with hearing impairments to learn and master pitch helps them to overcome communication difficulties. Effective pitch learning enables people with hearing impairment to express themselves accurately, thereby enhancing interaction and understanding with others.

Cost-Effective and Efficient Digital Learning: Through the digital platform, people with hearing impairment can participate in learning according to their personal time and pace, and realise the full potential of self-directed learning. This has the potential to reduce the time and financial costs associated with traditional face-to-face teaching and learning, and to provide convenient access to learning for more persons with hearing impairment.

Effective Educational Approach: The project's innovation lies in the integration of gamification elements with pitch learning, providing individuals with hearing impairment a more engaging and pleasurable learning approach. By employing game-based educational tools, individuals with hearing impairment can acquire pitch skills within an enjoyable environment, thereby augmenting motivation and learning efficacy. This innovative educational approach holds the potential to offer insights and inspiration for educational paradigms in other domains.

Facilitation of Personal Development: Attaining proficiency in pitch not only bolsters the confidence of individuals with hearing impairment in everyday communication but also opens doors to expanded educational and career opportunities. Proficiency in speech expression enhances their competitiveness in the professional realm, facilitating personal growth and achievements. It is hard to overstate how important communication skills are to success in life. Similarly it is hard to overstate how damaging and limiting a lack of skills can be.

Promotion of Social Inclusivity: By providing innovative educational tools tailored to individuals with hearing impairment, society demonstrates greater inclusiveness and

attention to the needs of this population. This helps to reduce prejudice and discrimination against individuals with hearing impairment, promote diversity and inclusion in society and create a fair and harmonious environment.

1.4 Project objectives

The main objective of this project is to create an active, easy-to-use educational tool to help users with hearing impairments learn pitch and intonation. The research objectives can be divided into the following research investigations:

RQ1. Speech Perception Challenges:

- How can the concepts of intonation and stress be effectively translated into perceptible elements for users, facilitating an intuitive understanding and learning process for individuals with hearing impairment?

RQ2. Imitation Difficulties:

- How can individuals with hearing impairment perceive the pitch of their own vocalizations?
- How to design a feedback mechanism in the tool that allows hearing impaired people to notice the difference between the pitch they actually utter and the actual pitch of their voice?

RQ3. Design and Technology:

- How can personalized training levels be formulated based on user preferences to augment the tool's engagement and appeal?
- What technological support is requisite for the development of this tool?

RQ4. User Experience:

- What user interface design elements can enhance the user experience, ensuring interface and interaction modalities align with the needs of individuals with hearing impairment, thus providing a gratifying user encounter?
- Taking into account pronunciation disparities among various dialects, how can tool content and design be tailored to address distinct linguistic learning needs and application scenarios of diverse individuals with hearing impairment?

RQ5. Evaluation of Educational tool Effectiveness:

- How can the efficacy of the game be scientifically and objectively assessed, encompassing its potential to enhance the pronunciation ability and language expression proficiency of individuals with hearing impairment to a certain degree?

- Does the tool yield noticeable improvements in pronunciation difficulties among individuals with hearing impairment after a period of usage?

1.5 Report overview

The final outcome of this project was the development of a web-based, real-time speech-interactive pitch visualisation educational game. The highlight of the project is to address the difficulty of hearing impaired people in recognising pitch levels in a low-threshold, low-cost way, and to help them improve through real-time feedback.

The project was developed using the waterfall model because of the relative stability of the project goals. The specific flow of development is as follows: **Background** examines pertinent research endeavors within the field, revealing their constraints and deficiencies. Analysis of relevant technologies is conducted, accompanied by an assessment of their viability. **Design** designs the system architecture and user interface, in alignment with our captured requisites. Granular details of implementation are subsequently expounded upon in **Implementation**.

The tangible outcomes of the project materialize in the **Results**, unveiling an innovative educational tool that engenders the interest and motivation of hearing impaired individuals in learning pitch through a novel approach. In the **Evaluation**, a comprehensive depiction of project evaluation is presented, encompassing expert assessments and user feedback, thereby scientifically validating the efficacy and practicability of the endeavor. Lastly, the **Conclusions** serves as a comprehensive synthesis of the entire project, encapsulating not only a retrospective assessment of the completed work but also the proposition of future directions for refinement and advancement.

Chapter 2

Background

This chapter provides a comprehensive exploration of the basic concepts, techniques, and tools associated with the development of pitch tools.

2.1 Essential concepts

This section describes the core concepts associated with this project.

Time domain and frequency domain: [Figure 2.1](#) illustrates a comparison between the time domain and frequency domain plots of eight different waveforms.

Pitch: When we listen to sound, we are perceiving the frequency of vibration of the sound wave. Frequency represents the speed at which a sound vibrates and is usually measured in hertz (Hz) which is cycles per second.

From the perspective of English linguistics, the pitch can help people identify part-of-speech and word sense of the word. From a signal processing perspective, pitch can be mathematically expressed as the frequency of a sound. Mathematical tools such as the Fast Fourier Transform (FFT) ([Brigham, 1988](#)) can convert a sound signal from the time domain to the frequency domain, thus facilitating the analysis of frequency components and the calculation of specific values of pitch.

Pitch Contour: A pitch contour is a pattern of frequency changes in a speech segment, often used to describe the rising and falling patterns of a sound. [Figure 2.2](#) is an example of “from home to the station” where the diagram shows the audio waveform and the pitch contour corresponds to the sentence.

Punctuation Ambiguity is a linguistic phenomenon that refers to a situation in which the interpretation and meaning of a sentence changes due to different punctuation arrangements or sentence segmentation in a text. Specifically, the same passage may be

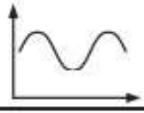
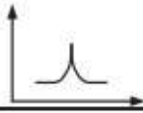
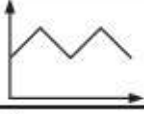

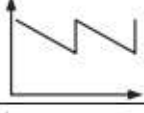
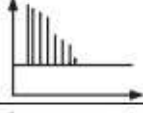
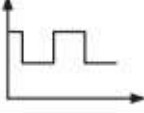


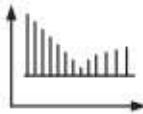
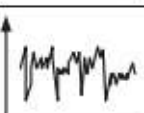

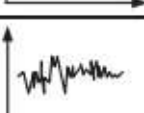

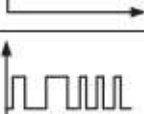

Waveform	Time domain	Frequency domain
Sinewave		
Triangle		
Sawtooth		
Rectangle		
Pulse		
Random noise		
Bandlimited noise		
Random binary sequence		

Figure 2.1 Comparison between time domain and frequency domain.

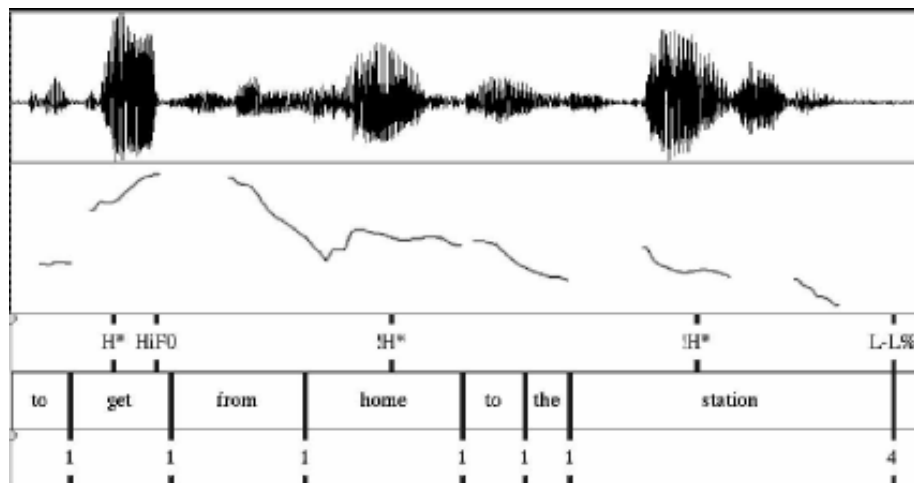


Figure 2.2 Pitch Contour of “from home to the station”.

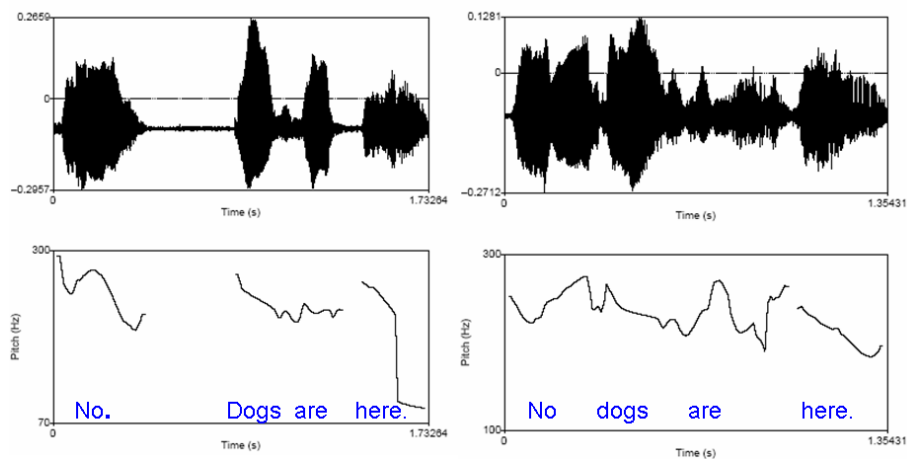


Figure 2.3 Punctuation Ambiguity of “No dogs are here”.

understood or interpreted differently due to different uses of punctuation, which in turn leads to changes in intonation. Figure 2.3 is an example of a change in the intonation of a sentence due to punctuation ambiguity.

Application Types: The ultimate objective of this project is to develop an application that assists individuals with hearing impairment in learning pitch. Table 2.1 illustrates common application types along with their corresponding development languages, frameworks, engines, and libraries.

Multimodal Human-Computer Interaction is a way of facilitating Human-Computer Interaction through the use of different technologies and means (Bourguet, 2003). In this type of interaction, “multimodal” means using multiple sensory inputs and outputs including visualisation, speech-to-text, acoustic feedback and haptics.

Application Types	Languages	Frameworks / Engines
Web Development	HTML, CSS, JavaScript	React, Vue.js, Django, Flask
App Development	Java, Kotlin, Objective-C, Swift	Flutter, React Native
AR & VR Development	C#, C++	Unity, Unreal Engine

Table 2.1 Application Types and Development Technologies.

Technique	Technology	Implementation
Visualizations	Audio Signal Processing	<p>Waveform Graph: Represent audio waveforms visually.</p> <p>Spectrogram: Visualize frequency components of audio.</p> <p>Waterfall Plot: Depict time-frequency characteristics of audio.</p>
Speech-to-Text Conversion	Machine Learning	<p>Natural Language Processing (NLP): Utilize language models and speech recognition engines.</p> <p>Recurrent Neural Networks (RNN): Process audio sequences to convert into text.</p>
Sonification and Tactile Feedback	Audio Signal Processing and Hardware	<p>Sonification: Transform data into sound signals with different attributes.</p> <p>Tactile Feedback: Convert sound vibrations into tactile sensations.</p>

Table 2.2 Multimodal Human-Computer Interaction Techniques Related to Speech.

Table 2.2 illustrates three typical methods of multimodal human-computer interaction related to speech.

Gamified Teaching: Gamification is a method of incorporating game design elements and mechanisms into the teaching and learning process, with the goal of enhancing the learning experience, increasing engagement and stimulating motivation. By introducing gamified elements, educators can design course content, activities and tasks in the form of games, such as setting up levels, reward mechanisms, competitions and role-playing, in order to capture students' interest and encourage them to participate more actively in learning.

2.2 Related work

Existing research has proposed numerous methods to assist individuals with hearing impairment in learning pitch. One research direction is to enhance the perception of pitch among individuals with hearing impairment by utilizing symbols, such as “Music Notation to Improve the Speech Prosody of Hearing Impaired Children” (Staum, 1987) in a treatment program to improve the verbal rhythmic and intonational accuracy of hearing impaired children. Another research approach is to employ assistive systems to aid individuals with hearing impairment in learning pitch. Yang et al. (2007) developed a computer-assisted music-learning system called CAMLS designed to aid individuals with hearing impairment in practicing musical melodies. CAMLS serves as a computer-supported learning tool facilitating the understanding of pitch and tempo for hearing impaired learners. Results indicate that in Staum's research, participants with reading skills successfully acquired rhythmic and inflectional patterns. Yang's research suggests that CAMLS could enhance hearing impaired students' learning performance in a music course. However, both of these two studies are pitch education based on the field of music, while there is still a gap in the field of learning pitch in speech and communication.

Speech and language impairment are basic categories that might be drawn in problems of communication involve hearing, speech, language, and fluency. Within speech-language impairment is a category called dysprosody. People diagnosed with dysprosody usually struggle with controlling the pitch of their voice (Sidtis and Van Lancker Sidtis, 2003), similar to how individuals with hearing impairment struggle with pitch control. We can refer to the methods in speech-language impairment therapy and apply them to pitch learning.

Practitioners in the field of speech and language impairment education have gained considerable experience in individualised teaching, multi-sensory teaching, and adaptive

approaches. For example speech educators often use visual aids (Monfort and Monfort-Juárez, 2001; Bernhardt et al., 2005) or tactile feedback (McCroskey Jr, 1958) when designing individualised education plans, and these methods have potential applications in pitch education. Therefore, the following further analyses common pedagogical approaches in the field of speech-language-impairment education.

The approach of using sound input alone as therapy was already proposed in 2014. “Flappy voice: an interactive game for childhood apraxia of speech therapy” (Lan et al., 2014) transforms a popular game Flappy Bird¹ into a therapeutic tool by changing the way interacts, replacing touch controls with sound controls, show in Figure 2.4. “Apraxia world: A speech therapy game for children with speech sound disorders” (Hair et al., 2018) proposed a video game using beautiful UI design, user-friendly interaction and speech recognition to help children improving their prosody skills, shown in Figure 2.5.

These innovative methods not only introduce novel perspectives to the fields of pitch education but also provide revelations: the integration of gamified elements into educational and rehabilitative practices has the potential to kindle learners’ interest and motivation, consequently enhancing their language proficiency and communicative efficacy. Utilizing real-time voice input as an interactive method presents another promising avenue for enhancing pitch education.

However, these assistive tools share a common limitation — they consist of predefined levels and lack a personalized touch. This deficiency in personalization hinders the adaptability of these tools to cater to the unique learning needs and preferences of individual users. Learners may consequently encounter challenges in engaging with the material effectively and maintaining sustained interest over time.

Designing a pitch education game that integrates real-time speech interaction and personalized customized levels, we can bridge the gaps in existing research and enhance our understanding of the potential applications of pitch education. This research direction not only holds the potential to offer innovative approaches to educational practices and rehabilitative strategies but also promises to provide fresh insights into interdisciplinary collaboration and technological innovation. By integrating real-time voice recognition technology into the game, users can immediately adjust and perceive changes in their pitch. This interaction provides prompt feedback guidance and gradual improvement in their pitch control. We plan to incorporate personalized algorithms that dynamically adjust the game’s difficulty levels based on each user’s abilities and progress.

¹<https://flappy-bird.io/>

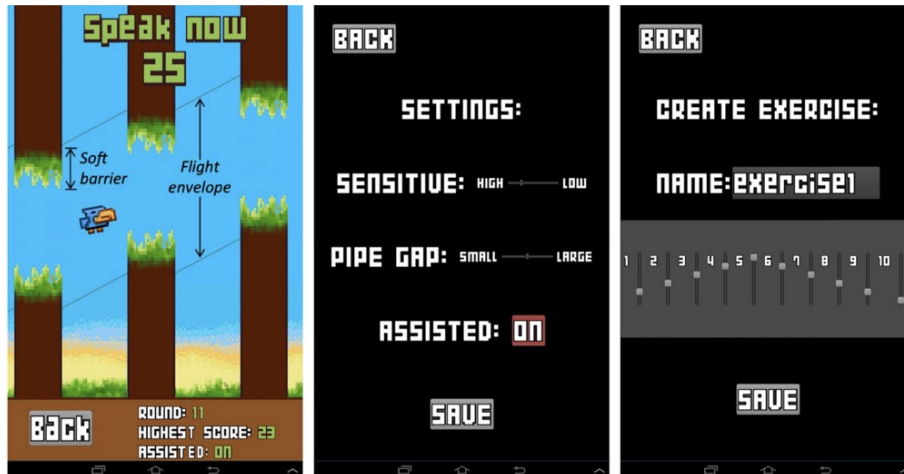


Figure 2.4 The speech exercise part in Lan et al. (2014).



Figure 2.5 The speech exercise part in Hair et al. (2018).

2.3 Related technology

Foremost among technological factors is the imperative evaluation of viable modalities for presenting the educational game. It is essential to deliberate upon the selection of suitable platforms for game deployment, with prevalent application development paradigms being delineated in [Table 2.1](#). VR (Virtual Reality) and AR (Augmented Reality) are immersive technologies that enhance the user's experience in differing ways. VR ([Park et al., 2019](#)) creates a simulated environment that users can interact with. It often requires the use of special headsets and controllers to fully immerse users in the virtual world. VR is extensively used in gaming and entertainment. AR ([Carmigniani and Furht, 2011](#)) overlays digital elements on the real world, usually viewed through devices like smartphones or AR glasses. AR enhances the user's perception of reality by adding computer-generated graphics, text, or other multimedia elements. However, VR and AR technologies require specific devices or physical spaces, limiting their accessibility and usability. Although Web development and App development are not as visually appealing as AR VR, the former does not require additional hardware configuration to be realised. For application development, it is necessary to consider the development platform, that is whether it is based on Android or iOS, whereas web development is more adaptable and can be easily adapted to a variety of terminals such as mobile phones and computers. To ensure ease of use for individuals with hearing impairment and to facilitate game development, basic front-end technologies such as HTML, JS, and CSS can be employed. Incorporating HTML for structuring content, JavaScript for interactivity and dynamic functionality, and CSS for styling enables a seamless and responsive user experience.

Another important factor is exploring pertinent methods, algorithms, and libraries. Given the focus on real-time voice interaction, implementing a reliable audio processing system capable of handling and manipulating sound input. This encompasses the conversion of audio signals into digital data, the elimination of potential background noise, and the extraction of pitch. Both front-end and back-end can process audio signals and extract pitch. On the front-end, audio signals can be processed using signal processing algorithms such as FFT to analyze audio. Audio processing can also be processed through third-party APIs like the Web Audio API ([Smus, 2013](#)). There are many natural language processing tools available to extract pitch in back-end, such as NLTK ([Bird, Klein, and Loper, 2009](#)), a widely used Python library for natural language processing tasks including text tokenization, part-of-speech tagging, named entity

recognition, sentiment analysis. CMU Sphinx (Huggins-Daines et al., 2006) supports multiple languages and audio formats. It provides a rich set of speech processing tools for building speech recognition and speech synthesis applications. To develop a large-scale speech recognition systems, HTK (Young et al., 2002) is a toolkit for speech recognition that supports various speech processing tasks, such as acoustic modeling,² lattice generation,³ and decoding.⁴ In natural language processing and text analysis, syllable tokenization is useful for handling pronunciation rules, rhythm, and prosody in certain languages. Librosa (McFee et al., 2020) is a powerful Python library specifically designed for audio and music-related tasks. It offers a wide range of functionalities, including audio feature extraction, time-domain and frequency-domain analysis, audio processing, and audio synthesis.

Real-time audio processing and pitch extraction are achievable on both the frontend and backend. Frontend methods are advantageous because they are real-time and allow for low-latency interactions, but they are limited by browser capabilities. Back-end has the advantage of possessing stronger computational power, allowing for the use of more complex algorithms, context analysis, and noise exclusion. The decision between frontend and backend processing should be based on practical factors such as specific requirements, system architecture, and performance demands.

2.4 Summary

Based on the analysis of the above research work and related technologies, we decided to develop a web-based, customisable learning objective, real-time voice-interactive game to help hearing impaired people learn pitch. This approach is user-friendly and has the potential to stimulate the interest of hearing impaired learners beyond traditional methods. Specifically, we will build an interactive front-end interface using HTML, CSS, and JavaScript to make the application easy to use and engaging. We will consider whether to choose the front-end or the back-end for pitch extraction according to the specific needs and complexity of the audio signal processing. Pitch extraction will consider whether to choose methods based on the front-end or the back-end, according to the specific needs and complexity of the audio signal processing.

²Acoustic modeling involves the statistical representation of sound characteristics to improve speech recognition accuracy.

³Lattice generation creates a graphical representation of possible word sequences, aiding in selecting the most probable transcription.

⁴Decoding refers to the process of determining the most likely sequence of words from the generated lattice during speech recognition.

Chapter 3

Design

A waterfall development methodology (Petersen, Wohlin, and Baca, 2009) was used for this project, which involved studying user requirements and designing the system architecture, user interface, and front-end-back-end development.

3.1 Requirements capture

3.1.1 Requirements capture method

The initial endeavour was to ensure that accurate and comprehensive data on user needs was obtained. However, given the limited depth of existing research in the relevant field, a questionnaire was purposely adopted as a proactive approach to gather user needs and insights that would address the identified knowledge gaps.

The questionnaire was primarily structured to investigate two key preferences of hearing impaired users: first for the interface and second for the game features. By understanding these users' perspectives on interactions, visual elements, and layout aesthetics within the interface, it informs both UI design and system architecture design. This work also sought to understand their requirements for tool functionality, personalised training modes and effective feedback mechanisms.

The specific questionnaire design is in [First questionnaire](#). The results of the questionnaire are in [Results of first user study](#).

3.1.2 Analysis

We have gathered 10 responses for this user study. Based on the findings from the user questionnaire, we have identified both functional and non-functional requirements.

Functional Requirements

1. Interface Design: We observed that seven users (70%) preferred a simple and clean design. This suggests that users prefer straightforward interfaces, possibly because they want to be able to find the information they need quickly.
2. Interaction Approach: Eight users (80%) chose visual interaction. This shows a preference for conveying information through visual elements.
3. Component Design: Nine users (90%) preferred simple layouts to minimise distraction. This may suggest that users have high expectations for layouts that allow them to focus on key information. Six users (60%) preferred larger buttons and icons, possibly because they wanted to be easily clickable and recognisable.

Non-Functional Requirements

1. Customising levels: Nine users (90%) would like to be able to customise levels. This indicates a demand for being able to adapt the learning objectives to their needs and preferences.
2. Feedback Mechanism: Seven users (70%) preferred after-the-fact feedback, possibly indicating that they preferred to receive overall feedback after the game was finished.

3.2 System Architecture Design

This study designs a real-time speech-interactive pitch educational game based on the user requirements in [Project objectives](#). The game allows the user to enter text, which is then used to dynamically generate a game map that reflects the characteristics of the entered text. The user can control the game with their voice and at the end of the game they receive feedback with suggestions for improvement, and at the end of the game they can practice as many times as they wish in the current level, or switch to the input screen to generate a new practice map. Therefore, the game process can be divided into three core modules: customisation, interaction, and feedback.

Custom Module: User input through the front-end interface initiates the process, as illustrated in [Figure 3.1](#). users input their desired exercise text, which is then processed by Module 1. This module is responsible for generating corresponding pitch contours, which are used to construct a game map that matches the customized text content. This

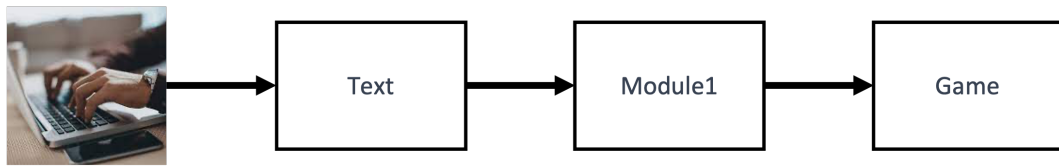


Figure 3.1 Custom Module.

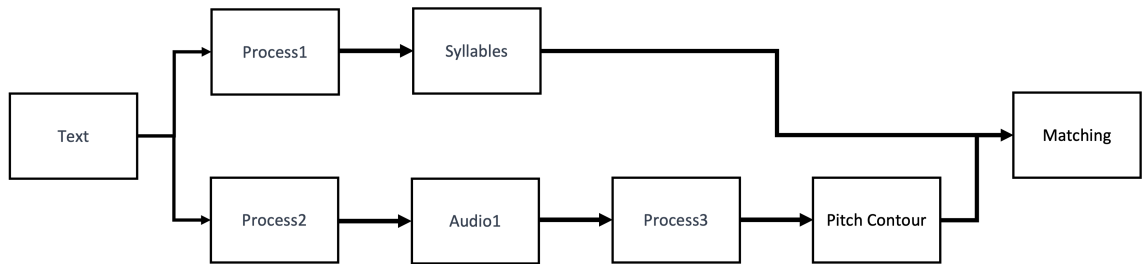


Figure 3.2 Flowchart of Module 1.

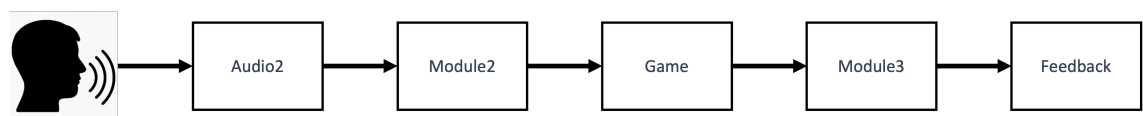


Figure 3.3 Interaction Module and Feedback Module.

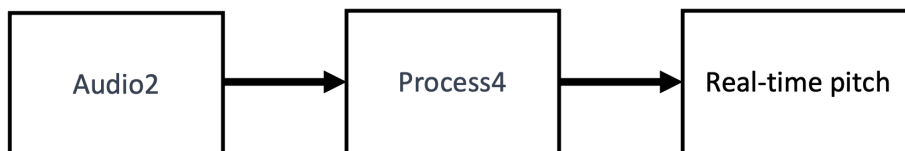


Figure 3.4 Flowchart of Module 2.

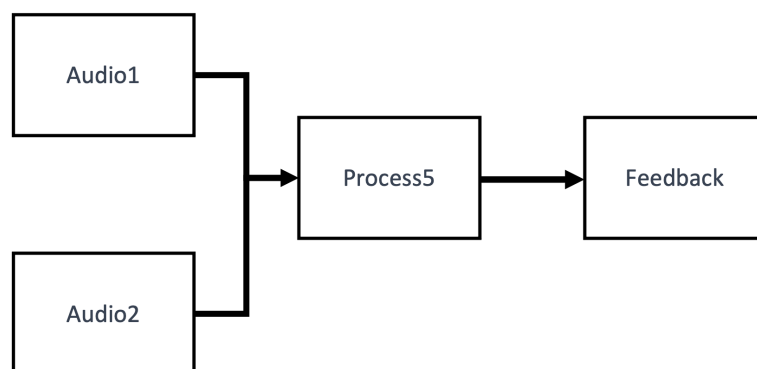


Figure 3.5 Flowchart of Module 3.

procedural framework encompasses the implementation of two key concepts: first, a method for converting the textual content into its corresponding pitch contour, and second, the establishment of an accurate mapping between each syllable in the text and its corresponding pitch. The detailed workflow is presented in [Figure 3.2](#).

First, the input text is split into different syllables through the step of “process1: natural language processing.” In parallel, the text is converted into corresponding audio1 by the “process2: text-to-speech” method. The pitch contour of the audio is then effectively extracted by the “process3: audio signal processing” technique. The next step calculates the duration of each syllable and establishing the correlation between the syllable and the pitch by averaging the pitch contour over the corresponding time.

Interaction Module: This module allows user to use sound to interact with the game ([Figure 3.3](#)). During this interaction, the sound signals generated by the user are processed by Module 2, detailed in [Figure 3.4](#). The sound signal is transformed into corresponding pitch information after processing in process4. Due to the real-time requirements, this processing is done at the front end of the system, using FFT to calculate real-time pitch values. Due to the real-time requirements, this processing is done at the front end of the system, using FFT to calculate real-time pitch values. These pitch data are then translated into a visual representation suitable for embedding in the game interface, thus allowing the user to perceive pitch changes in an intuitive means.

Feedback Module: After the user successfully completes the customised game level, the audio data generated during the gameplay Audio2 is carefully captured and recorded, and subsequently the user can get image feedback. See [Figure 3.3](#). Module 3 plays a central role, and its main function is to calculation of the pitch contour differences between the user-generated audio data (“audio2”) and the reference audio data (“audio1”). In order to accurately calculate and quantify the pitch accuracy, we ultimately adopted the Dynamic Time Warping (DTW) (Müller, 2007) algorithm. The application of the DTW algorithm in this scenario can capture not only the changes in pitch, but also the changes in the rhythm of speech, so as to provide users with an accurate and comprehensive evaluation of speech performance, the specific architecture is shown in [Figure 3.5](#).

3.3 User interfaces Design

User interface design endeavours to produce intuitive, user-friendly, visually appealing interfaces facilitating effective interaction between users and digital products or applica-

tions. User interface design involves not only visual aesthetics, but also considerations related to interface functionality, layout, interactive elements, and user perception.

User interface is pivotal in software development. It helps to organise the development logic before the front-end and back-end components are formally implemented. By carefully delineating interface layouts, elements, and interaction methods during the design phase, redundant rework and tweaking during development can be reduced.

Quickly identifying the optimal solution in UI design is often a difficult task. Through the application of experience and design principles, we can gradually approach the optimal solutions. *10 Usability Heuristics for User Interface Design* (Nielsen, 1994) gather experience and insights from past learnings and industry norms to assist designers in efficiently evaluating and enhancing their designs, enhancing the quality of their work and user experience. Making the design inclusive is important, and we can refer to some accessible design guidelines *Coblis — Color Blindness Simulator – Colblindor* (Colblindor, 2016), *Colour Blindness* (Colour Blind Awareness, 2022). The UI was designed in [Figma](#).

We have given due consideration to the requirements of our users and ensured that the page is visually friendly to colour-blind users. We have adopted a clean design style with grey, black and white as the main colour scheme.

[Figure 3.6](#) showcases a custom input interface designed for the game. This interface provides users with functions for entering text, adjusting speech speed, and toggling between languages. Clear labels and intuitive icons enable users identify and operate each function.

[Figure 3.7](#) illustrates the design of the game interface, where the canvas serves as the primary interaction area. There is a “Restart” button that enables users to practice the current level repeatedly and a “Re-enter” button that allows users to go back to the input interface and re-enter text. Within the canvas, musical notes represent real-time pitch, while a series of rectangles depict target pitches, with each rectangle corresponding to a syllable.

[Figure 3.8](#) depicts the appearance of the “get score” button at the bottom after the game ends, allowing users to view feedback by clicking the button.

[Figure 3.9](#) displays the feedback screen that appears when the user clicks “get score.” Users can review the game performance score and evaluation. The presence of a close button in the interface facilitates a convenient return to the game-ending page.

3.3.1 Cognitive Walkthrough

The cognitive walkthrough method is a usability inspection method used to identify usability problems in interactive systems to ensure the user experience during the game (Polson et al., 1992). We invited three testers familiar with game interface design to examine the following key processes in detail:

1. **Text input, speed adjustment, and language selection functions:** Testers found the design of these functions to be extremely intuitive and easy to use. It was easy to see how to operate the different functions, such as the text box for text input, the slider for speed adjustment, and the drop-down box for language switching. No testers were confused or made mistakes during the entire testing process.
2. **Text re-entry function:** Re-editing is very easy when entering text. The text input page enables the user to easily modify the entered text, the speed of speech and select the language. The re-enter button also enables the user to go back to the input page and re-enter the text when they are in the game page.
3. **Game restart function:** Testers reported that when they needed to restart a game, the restart button was clearly positioned and could be clicked to quickly return to the initial state without cumbersome steps. This intuitive design allows users to easily restart the game without wasting time and helps keep the game flowing.
4. **Feedback acquisition function:** Testers noted that the design of the “Score” button, which is presented at the end of the game, helps to make it clear to the user that the system state has changed. Once users clicked on the button, they could jump to the feedback page and return to the game page easily. This design provides users with a clear navigation path and ensures that they can easily transition from the game to the feedback and back again.

After a series of comprehensive tests, we concluded that the interface design is characterised by intuition and the absence of cumbersome and complex procedures. Consequently users are able to complete the above processes proficiently.



Figure 3.6 UI design for User Input page.

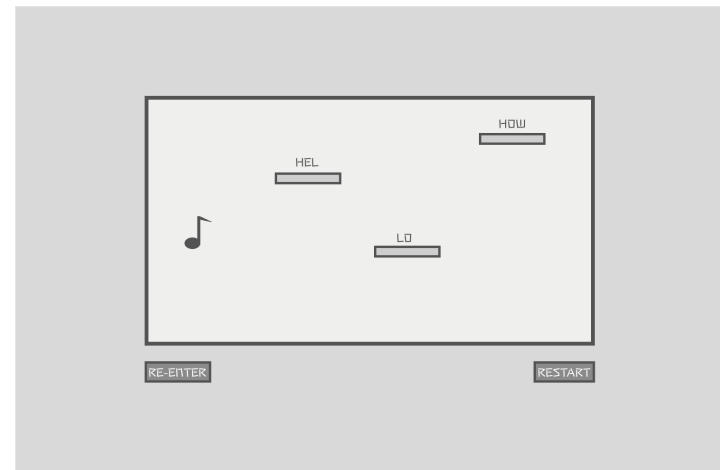


Figure 3.7 UI design for Game page.

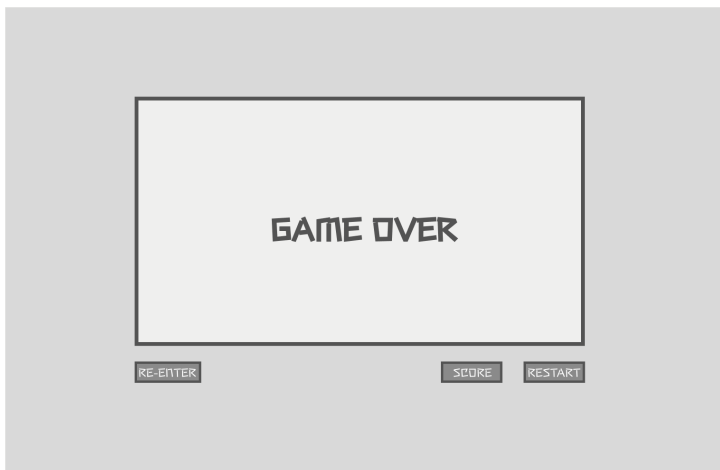


Figure 3.8 UI design for Game over page.

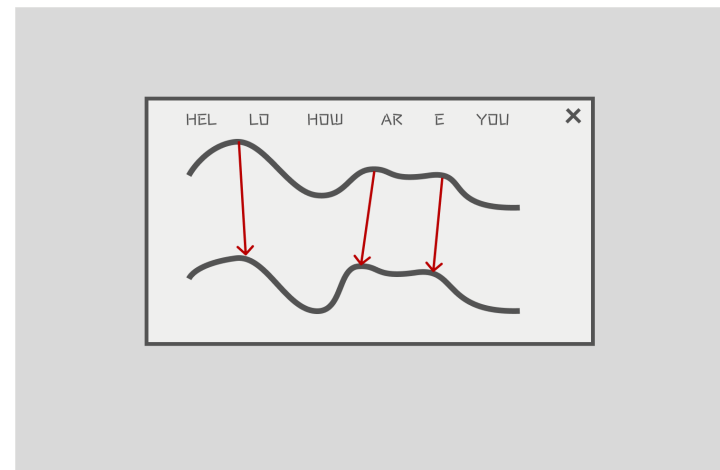


Figure 3.9 UI design for Feedback page.

Chapter 4

Implementation

4.1 Front-end Implementation

Based on the design of the user interface, this section explores the specific implementation of front-end development, which covers a number of key aspects, including technology selection, definition of core functionality, and precise implementation of the user interface.

Technology Selection:

The project is characterised by a clean interface without complex functionality, we have chosen a non-framework approach to maintain the integrity and flexibility of the design. Subsequent development will use JavaScript, HTML5 and CSS as core technologies to implement the front-end interface in a streamlined and efficient manner.

Implementation of Text Input Page:

The custom input page uses the pixel-inspired font “Press Start 2P” from Google Fonts to give the whole page a simple but fun look. Unlike in the UI design, we ended up using a tick box for the speed adjustment module instead of a slider because of the need to match the format of the data when interacting with the back-end data. The text input box and language selection button are consistent with the UI design, users can change the input text or switch between different languages or dialects. When the page is loaded, the window.onload event handler is triggered which binds a callback function for the start button click event. When users click the start button, front-end gets personalised content and uses the fetch function to make an POST request to the server.

Implementation of Game Interaction Page:

The game canvas is created using HTML and CSS, including the background, ground, and the note used to represent the user. The initial position of these elements

is determined by the data returned from the back-end. Implement a "Start Game" button and use JavaScript to listen to its click event. Once the user clicks the button, the microphone listening function is activated. Use the Web API to access the user's microphone and start capturing real-time audio data. Using FFT spectral analysis, calculate the major frequency components from which pitch information is extracted and the pitch is calculated. Adjust the vertical position of the notes based on the real-time pitch values to create a visual effect that matches the change in pitch.

Stop microphone listening when the game ends. Store the captured audio data as an audio file and send it to the backend via an HTTP POST request. Display a "View Results" button on the user interface. Once the user completes the game, they can click on the button. Clicking the button opens a new window. In the new window, a request is made to the backend to get the data from the game analysis results.

4.2 Back-end Implementation

This section describes the technology choices, libraries used, and specific implementations of basic functionality during back-end development. The back-end is divided into two phases: Module 1 is generating the pitch contour corresponding to the target text and Module 2 is processing the difference in pitch contour between the audio sent back from the front-end and the target audio, and returning it to the front-end.

The implementation of Module 1:

In the receiving request phase, the back-end creates routes using the Flask framework. As soon as the back-end receives a POST request from the front-end, it starts the processing and passes the data from the request to the later processing steps. Request method judgment is performed based on the request to ensure security and scalability. The system verifies whether the request method is OPTIONS, which is a common mechanism for handling cross-domain request preflight. When the request is OPTIONS, the back-end generates a response carrying empty JSON and sets the response header to allow cross-domain access from all sources, returning this response to the front-end. If the request is not OPTIONS, the back-end extracts data from the request in JSON format containing the textual content to be processed, the rate-of-speech adjustment requirements, and the selected language.

In the text processing phase, the back-end preprocesses the text content according to the request. The punctuation in the text is removed and the text is split into syllables using the SyllableTokenizer method in the nltk.tokenize module. We chose the

gttx library to implement text-to-speech because the gTTS library uses text parsing technology, which captures the precise meaning of the text and avoids the problem of punctuation ambiguity. And it has the advantage of adjusting the speed of speech and supporting the selection of different languages and dialects. The generated speech was temporarily stored in the system. The back-end extracts the pitch contour from the temporarily stored speech file using the pitch contour extraction function in the librosa library and calculates the duration and average pitch of the syllable.

In the response construction phase, the back-end constructs a response dictionary based on the processed syllable segmentation, pitch contour and average pitch data. This response dictionary is used to build the response object, which includes data in JSON format, status codes and corresponding response headers. In the response header, cross-domain access is explicitly allowed to ensure that the front-end can receive and process the response correctly.

The implementation of Module 2

The workflow in Module 2 involves processing requests, audio data processing, and returning results. Upon receiving a POST request, the system processes binary audio data by decoding it. Waveform generation is performed for audio1 and audio2, along with the setting of specific DTW parameters. The DTW algorithm evaluates spectrogram distances and optimized paths. The cumulative cost matrix is plotted, and the most favorable paths are labeled for mapping. The processed image is then returned to the front-end for display.

4.3 Interface Implementation

Due to the relatively small amount of data transfer and the emphasis on simplicity, we chose Flask API, a lightweight Python micro-web framework that allows for fast and flexible API development that effectively meets the requirements of front-end and back-end data interaction. The API comprises two key functionalities: text processing and retrieval of analyzed result images. The API documentation is shown in [API Document](#)

Chapter 5

Results

5.1 Interface display

This section shows the user interface of the Pitch Education game in detail, shown in [Figures 5.1 to 5.5](#)

5.2 Interactive display

This section demonstrates the actual interaction process of the audio processing system through a [video](#).

5.3 Performance display

The project can be explored in depth from three key perspectives. These two core perspectives relate to the time it takes to generate a map and the time to get feedback.

Time to Generate Maps: We focus on the performance of Module 1, which is responsible for generating maps based on user-defined inputs. The efficiency of this step is critical to the user experience, as prolonged waiting times can reduce satisfaction. We will conduct tests for different lengths of text inputs and record data on the time taken to generate the map to reveal trends in the performance of the system with different lengths of inputs. The results are shown in [Table 5.1](#).

Time to get feedback: Lastly we focus on the time it takes for the user to get feedback after completing the audio processing. The results are shown in [Table 5.2](#).

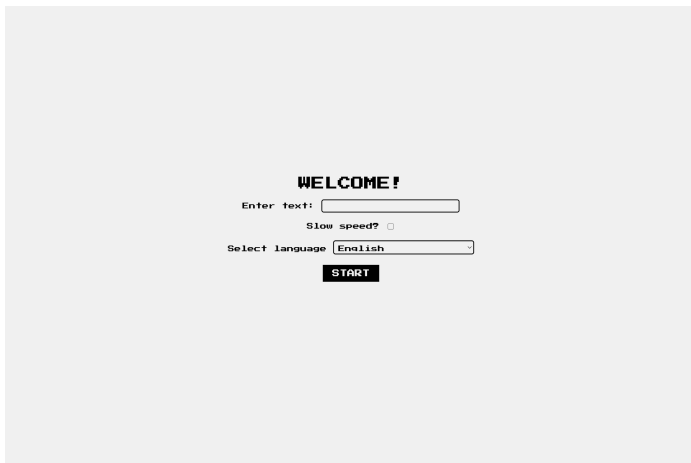


Figure 5.1 User Input page.

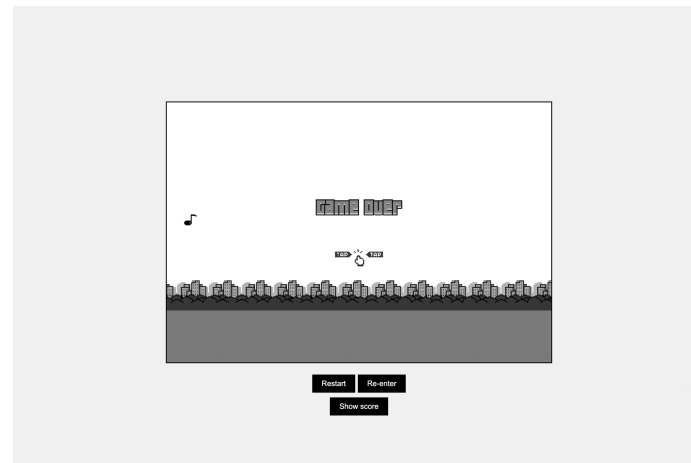


Figure 5.2 Game ready page.

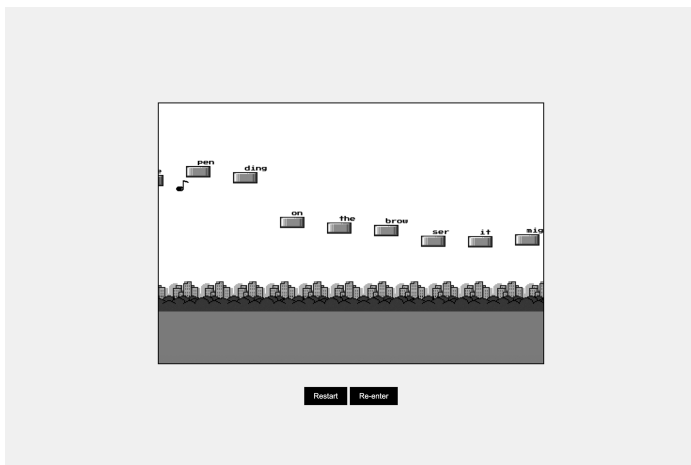


Figure 5.3 Game page.



Figure 5.4 Game over page.

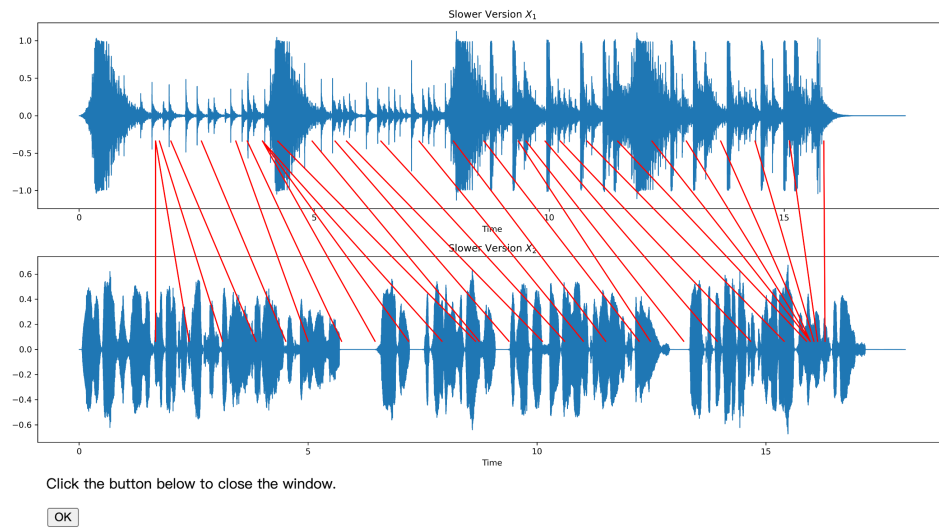


Figure 5.5 Feedback page.

Number of words	Average generation time (seconds)
1	0.2
10	1.2
100	7.9

Table 5.1 Time to generate maps versus number of words.

Audio Duration (seconds)	Feedback Retrieval Time (seconds)
10	4.7
30	5.3
60	5.8

Table 5.2 Relationship between audio duration and feedback retrieval time.

Chapter 6

Evaluation

6.1 User study

A user study is a useful approach for understanding users and improving our design. Designers often assume that users share the same understanding without realizing it. This availability bias needs to be eliminated through interaction with users. Common methods for user studies include face-to-face interviews, questionnaires, focus groups, and usability testing. We designed a user study that integrates usability testing and a questionnaire to identify usability problems and receive feedback from participants. The participants are individuals with hearing impairments who require learning pitch.

6.2 Usability testing

Usability testing is a useful method for identifying usability problems by observing representative participants as they perform tasks within the product. The usability testing process follows these 5-step:

- 1. Resource preparation:** A computer setup with a tested environment and equipped with a microphone.
- 2. Designing the tasks:**
 - Task 1: Participants input customized exercise text, adjust the speech speed, and select the language.
 - Task 2: Participants click the start button, engage with the interactive game.
 - Task 3: Restart current level.
 - Task 4: Returning to the text input page, participants enter a new customized exercise text.

Task 5: Participants get feedback when the game is over.

- 3. Recruiting participants:** Participants are hearing-impaired individuals who have difficulty perceiving pitch, possess basic knowledge of the English language, syllables and syllables pronunciation.
- 4. Conducting usability test:** Observe participants' facial expressions and body movements. Record usability levels, such as task completion status, task completion time, and task completion path.
- 5. Analysis:** Identify usability problems by summarising the usability levels. Define the severity of the usability problem, using three questions, according to the [Figure 6.1](#) process. Lastly, analyse the reasons behind the problem and find a feasible solution.

We recruited seven participants, usability levels results are in [Results of usability testing](#)

6.3 Questionnaire

This questionnaire was designed to cover four aspects of the system's usability evaluation.

- 1. Visual effect and interface design:** Users were asked to evaluate whether the overall appearance, colour scheme and layout of the system were satisfactory, and whether the interface was clear and easy to understand.
- 2. Interaction experience and fluency:** Users were invited to describe their interaction experience when using the system, including whether the operation was smooth and the interface responsive.
- 3. Functionality:** Users were asked to assess the usefulness of the functionality provided by the system for achieving their intended tasks. Particular attention was paid to the evaluation of user-defined learning objectives, real-time pitch calculation and visualisation features.
- 4. User Satisfaction:** Users were asked to rate their overall satisfaction, indicating their overall impression of the system and their willingness to continue using it.

Refer to [Second questionnaire](#) for specific questionnaire design; the relevant results are in [Results of the second user study](#).

6.4 Analysis

Seven participants were recruited, all of whom took no more than 25 minutes to complete the usability test and no more than 10 minutes to complete the questionnaire.

6.4.1 Usability testing analysis

All participants were able to quickly and without hesitation choose the correct path to complete the tasks in all five tasks. This shows that in the guidance and execution of the tasks, the users were able to accurately understand the goals and requirements of the game and make the right decisions quickly. The interaction design and execution logic of the game are clear and intuitive, enabling users to easily find the correct solution.

We noticed that during task execution, two participants showed nervous expressions in Task 2, and four participants showed puzzled expressions in Task 5. Although the users were able to complete the tasks efficiently and accurately, these changes in expression during the tasks reflected potential barriers or adverse emotional experiences in the game or interaction. This situation may be related to the design of the game content. We have analysed the possible problems in these two tasks separately:

For Task 2, participants showed nervous emotions that might stem from a lack of familiarity with the flow of the game. This situation might be reduced by repeated practice. It is also possible that there is a problem with the game content itself. In this case, we need to revise and improve the game content to better meet the users' requirements.

For Task 5, where confusion may stem from the user's difficulty in understanding the feedback content, is where the feedback content itself is problematic. In this case, the presentation of the feedback needs to be reworked to ensure that its message is communicated more clearly.

We asked participants with changed expressions to perform Task 2 and Task 5 again. The results showed that when executing Task 2, the participants' expressions eased, although still with some tension. When executing Task 5, the user's expression still showed confusion. This result suggests that Task 2 had problems with participants not being familiar with the flow of the game. Task 5 requires modification to address content problems.

6.4.2 Questionnaire analysis

An detailed analysis of the users' responses to the second questionnaire revealed that the participants' overall rating of the system's appearance averaged 4.1 in 5. This result suggests that in terms of visual experience, the system earns a high level of satisfaction. Three participants reported that they favoured the pixel style of the system, which further increased their interest in participating in this educational game.

Regarding the smoothness of the system's operation process, five participants were neutral about the flow of the interaction process. One participant felt that the flow of the interaction process did not feel smooth enough. Three participants expressed concerns about the speed at which the elements representing pitch moved through the interface. Especially at normal speech speed, they had difficulty in clearly identifying the syllables during the movement.

Six participants expressed interest in ways to customise the learning objectives. One participant noted that he lacked an idea of what should be entered. The participants gave a positive rating of 4.4 out of 5 for the entire game process. One participant felt that although he weren't certain if he could remember accurately the matching of pitch to syllables in words, they definitely felt an improvement in their personal ability to control pitch. All participants expressed their eagerness to continue using the educational game we created for pitch education.

In the concluding open-ended question of the questionnaire, regarding suggestions for the entire project, two participants suggested adjusting the speed of pitch components' movement during normal speech speed mode. Four users felt that the feedback was not intuitive enough and suggested that additional textual explanations or image processing may be necessary to help users better understand the feedback. Providing more detailed instructions or adding clarifying icons to images could help overcome this barrier and improve users' understanding of the system's feedback.

6.4.3 Summary

Combining the results of the usability testing and the questionnaire, we identified three usability problems. Based on the flow in [Figure 6.1](#), we categorised the severity of these problems. Two critical problems are listed below:

Game components moving too fast: The pitch component of the game moves too fast, making it difficult for users to recognise syllable text. Solutions may include

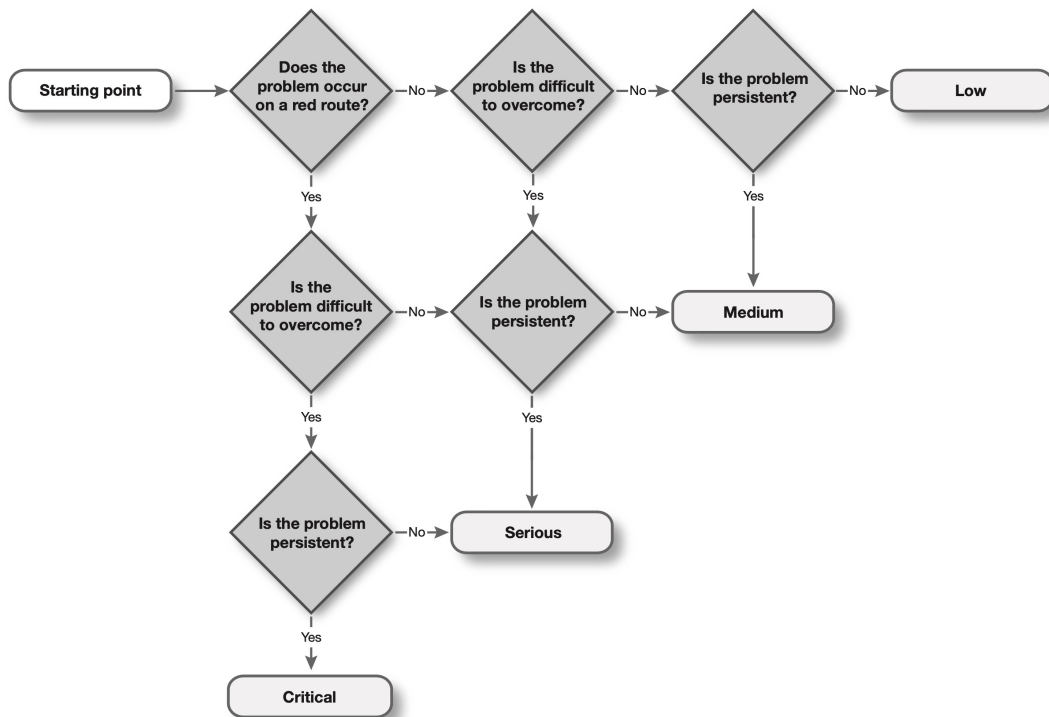


Figure 6.1 A decision tree is used to define four severity levels by asking three questions.

redesigning the game logic to adjust the movement speed of the pitch component, or reducing the movement speed by redesigning the distance between components.

The feedback provided is difficult to comprehend: There is an obvious problem with the readability of the feedback, and users are unable to quickly understand how to improve it. One way to achieve this is by redesigning the logic behind the feedback image. This could involve adding different syllable labels to the image or including a tabular background, which would help the user to more easily identify the high or low pitch.

One low problem is listed below:

Lack of Game Guidance: First-time users may lack sufficient guidance during game play, affecting their understanding and efficiency of the game flow. The solution could be to consider introducing methods such as video introductions to provide more guidance.

Chapter 7

Conclusions

7.1 Project summary

In this project, we have designed a real-time voice interaction pitch visualization educational game to help hearing-impaired people improve their speaking. We gather user requirements, followed by a detailed design of the entire game flow and user interface. We completed the development of both front-end and back-end parts, conducting tests. Throughout this process, we closely monitored user feedback, continuously optimizing and enhancing system functionality. In the concluding stages of the project, we adopted an interactive approach to finalizing our work.

Addressing the research questions posed in [Project objectives](#), we have formulated the following solutions:

RQ1: Speech Perception Difficulties We employed a method involving the computation of pitch contours from audio signals, thereby transforming tonal information into visual components. This approach aids individuals with hearing impairment in perceiving pitch variations more effectively.

RQ2: Imitation Difficulties We developed real-time pitch calculation and visualization functionalities to tackle difficulties in pitch imitation. These enable individuals with hearing impairment to promptly access their own pitch information and compare it with target pitch contours. We employed the DTW algorithm to accurately assess discrepancies between actual and target audio signals.

RQ3: Design and Technology In order to optimize process design and technological application, we introduced the concept of gamified education. Customized learn-

ing objectives are achieved through user-defined inputs, enhancing the appeal and engagement of the educational process.

RQ4: User Experience We devised a minimalist interface presented in pixel art style. Users are granted autonomy to define target languages and speech rates, catering to diverse regional dialects and pacing preferences.

RQ5: Evaluation of Educational tool Effectiveness In evaluating the effectiveness of the educational tool, we adopted two methodologies: expert evaluation through cognitive walkthroughs and user evaluate via user study.

7.2 Future work

Future work will focus on solving the two critical problems mentioned in [Analysis](#). Redesign the game logic to adjust the movement speed of the pitch components or reduce the movement speed by redesigning the distance between components. Redesign the feedback content, like adding syllable labels to the images, optimise the table background, and may even consider adding some text descriptions to assist users in better identifying areas for improvement.

Apart from optimizing the points of improvement mentioned above, we can also explore incorporating innovative functions to further enhance the user experience of the pitch education game. For example:

1. When designing the game, we did not take into account the common occurrence of liaison in English. This oversight may cause difficulties for users when trying to express themselves verbally in real-life situations. We intend to redesign the game's pitch-teaching approach to incorporate phenomena like liaison. This will help users better understand and adapt to phonetic changes that occur during actual conversations.
2. Our project currently focuses on practicing English pitch with different accents, but we plan to make it more versatile and globally applicable by exploring the unique speech characteristics of various languages. Based on these characteristics, we intend to improve the pitch segmentation algorithm in 'Module One' so that users worldwide can receive more pitch training beyond just English.
3. We can introduce more evaluative criteria for feedback, considering pitch accuracy, word and syllable pronunciation accuracy, and speech rhythm stability. This collective approach will enable us to provide more targeted training recommenda-

tions to help users with hearing impairments make progress in multiple aspects of their speech.

7.3 Reflections

In this project, I undertook the entire process, from gathering user requirements, UI design, front-end and back-end development, to integration and testing, for the first time on my own. Although it was full of challenges, it also provided me with valuable experience. Here, I will reflect on the difficulties encountered, the lessons learned, and the methods used to solve problems, especially the common problems in front-end and back-end development, as well as considerations when working with APIs.

7.3.1 Early Project Stage

1. **Lack of Clear Starting Sequence:** I lacked a well-defined plan and the necessary experience for project initiation, which caused confusion. I immediately started gathering user requirements and designing the system's architecture, but I unintentionally overlooked the UI design and interaction logic. This oversight resulted in problems during later stages of development, particularly concerning user input customization and front-end/back-end integration. I promptly added a complete UI design, focusing on layout, interaction, and navigation. This adjustment improved our later development process, streamlining the integration of interface functionalities.
2. **Ignoring Risk Assessment:** Neglecting risk evaluation and solely focusing on specific problems is a frequent trap in project development. While I confidently anticipated challenges, I neglected to consider uncertainties, resulting in unforeseen delays and risks. I underestimated technical unknowns, such as browser compatibility, which later made front-end-back-end interactions difficult. This experience taught me the significance of conducting detailed risk evaluation during the early stages to prevent such oversights in the future. This practice will enhance my ability to manage uncertainties, mitigate potential risks, and elevate both project quality and efficiency.

7.3.2 Front-end Development Stage

In front-end development, managing asynchronous requests is a significant problem, especially when dealing with interactions with back-end APIs. During the completion of this project, I faced some challenges related to asynchronous requests and gained valuable experience from them.

One mistake I made was inadvertently initiating multiple duplicate requests. This could burden the back-end server unnecessarily and lead to data inconsistency. To avoid this situation, I learned some effective methods:

- **Debouncing:** Using debouncing can limit the frequency of request triggering. After a user triggers a request, setting a timer and triggering subsequent requests before the timer ends will cancel the previous request, ensuring that only the last request takes effect.
- **Throttling:** Similar to debouncing, throttling limits the frequency of requests. However, throttling periodically triggers requests within a certain time interval instead of waiting for the last trigger.

Another problem was the need to include request headers in some situations, particularly when involving authentication, access control, or cross-origin requests. This was a bit challenging for me because I had to ensure that the requests included the correct headers for the backend to process them correctly. To address this problem, I learned the following methods:

- **Using Fetch API:** The Fetch API provides a convenient way to make asynchronous requests and allows setting request headers through options parameters. I can use the Headers object to define custom headers and pass them to the fetch function to ensure the necessary header information is included in the request.

Through the experience of completing this entire project independently, I acquired valuable knowledge and experience. I understand the importance of clear requirements, risk assessment, and communication during project development, as well as the solutions to common problems in front-end and back-end development and considerations when working with APIs. In future projects, I will focus more on the standardization of the development process and risk assessment to improve the quality and efficiency of project delivery. I will continuously learn and enhance my technical abilities to be better prepared for more complex projects.

References

Bernhardt, Barbara, Gick, Bryan, Bacsfalvi, Peter, and Adler-Bock, Marci, 2005. Ultrasound in speech therapy with adolescents and adults. *Clinical linguistics & phonetics*, 19(6-7), pp.605–617.

Bird, Steven, Klein, Ewan, and Loper, Edward, 2009. *Natural language processing with python*. O'Reilly Media.

Bourguet, Marie-Laure, 2003. Designing and prototyping multimodal commands. *Proceedings of human-computer interaction (interact'03)*, pp.717–720.

Brigham, E. Oran, 1988. *The fast fourier transform and its applications*. Prentice-Hall, Inc.

Carmigniani, Julien and Furht, Borko, 2011. Augmented reality: an overview. In: *Handbook of augmented reality* [Online]. Ed. by Borko Furht. New York, NY: Springer. Available from: https://doi.org/10.1007/978-1-4614-0064-6_1.

Colblindor, 2016. *Coblis — color blindness simulator – colblindor* [Online]. Colorblindness.com. Available from: <https://www.color-blindness.com/coblis-color-blindness-simulator/>.

Colour Blind Awareness, 2022. *Colour blindness* [Online]. Colour Blind Awareness. Available from: <https://www.colourblindawareness.org/colour-blindness/>.

Hair, Adam, Monroe, Penelope, Ahmed, Beena, Ballard, Kirrie J, and Gutierrez-Osuna, Ricardo, 2018. Apraxia world: a speech therapy game for children with speech sound disorders. *Proceedings of the 17th acm conference on interaction design and children*, pp.119–131.

- Hall, Deborah A. and Plack, Christopher J., 2009. Pitch processing sites in the human auditory brain. *Cerebral cortex* [Online], 19(3), pp.576–585. Available from: <https://doi.org/10.1093/cercor/bhn108>.
- Huggins-Daines, David et al., 2006. Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices. *Proceedings of the acm international conference on mobile computing and networking (mobicom)*, pp.3–8.
- Lan, Tian, Aryal, Sandesh, Ahmed, Beena, Ballard, Kirrie, and Gutierrez-Osuna, Ricardo, 2014. Flappy voice: an interactive game for childhood apraxia of speech therapy. *Proceedings of the first acm sigchi annual symposium on computer-human interaction in play*, pp.429–430.
- Lee, T., Liu, Y., Huang, P.-W., Chien, J.-T., Lam, W. K., Yeung, Y. T., et al., 2016. Automatic speech recognition for acoustical analysis and assessment of cantonese pathological voice and speech. *Acoustics, speech and signal processing (icassp), 2016 IEEE international conference on*. IEEE, pp.6475–6479.
- Lyxell, Björn, Wass, Malin, Sahlén, Birgitta, Samuelsson, Christina, ASKER-ÁRNASON, LENA, Ibertsson, Tina, MÄKI-TORKKO, ELINA, Larsby, Birgitta, and Hällgren, Mathias, 2009. Cognitive development, reading and prosodic skills in children with cochlear implants. *Scandinavian journal of psychology*, 50(5), pp.463–474.
- Mashima, Pamela A and Doarn, Charles R, 2008. Overview of telehealth activities in speech-language pathology. *Telemedicine and e-health*, 14(10), pp.1101–1117.
- McCroskey Jr, R L, 1958. The relative contribution of auditory and tactile cues to certain aspects of speech. *Southern journal of communication*, 24(2), pp.84–90.
- McFee, Brian et al., 2020. Librosa: audio and music signal analysis in python. *Proceedings of the 14th python in science conference*, pp.18–25.
- Monfort, M. and Monfort-Juárez, I., 2001. *En la mente: un soporte gráfico para el entrenamiento de las habilidades pragmáticas en el niño*. Madrid, Spain: Entha ediciones.
- Müller, Meinard, 2007. Dynamic time warping. In: *Information retrieval for music and motion*. Springer, pp.69–84.

- Nielsen, Jakob, 1994. *10 usability heuristics for user interface design* [Online]. Nielsen Norman Group. Available from: <https://www.nngroup.com/articles/ten-usability-heuristics/> [Accessed April 11, 2023].
- O’Callaghan, Anna M., McAllister, Lindy, and Wilson, Linda, 2005. Barriers to accessing rural paediatric speech pathology services: health care consumers’ perspectives. *Australian journal of rural health*, 13(3), pp.162–171.
- Park, Min Joo, Kim, Dooyoung J., Lee, Uichin, Na, Eunjung J., and Jeon, Hae-Jeong, 2019. A literature overview of virtual reality (vr) in treatment of psychiatric disorders: recent advances and limitations. *Frontiers in psychiatry* [Online], 10, p.505. Available from: <https://doi.org/10.3389/fpsyt.2019.00505>.
- Petersen, Kai, Wohlin, Claes, and Baca, Dejan, 2009. The waterfall model in large-scale development. In: Frank Bomarius, Markku Oivo, Päivi Jaring, and Pekka Abrahamsson, eds. *Product-focused software process improvement*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.386–400.
- Polson, Peter G, Lewis, Clayton, Rieman, John, and Wharton, Cathleen, 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International journal of man-machine studies*, 36(5), pp.741–773.
- Sidtis, John J and Van Lancker Sidtis, Diana, 2003. A neurobehavioral approach to dysprosody. *Seminars in speech and language* [Online], 24(2), pp.93–105. Available from: <https://doi.org/10.1055/s-2003-38901>.
- Smus, Boris, 2013. *Web audio api*. O’Reilly Media, Inc.
- Staum, Myra J., 1987. Music notation to improve the speech prosody of hearing impaired children. *Journal of music therapy* [Online], 24(3), pp.146–159. Available from: <https://doi.org/10.1093/jmt/24.3.146>.
- Theodoros, Deborah G, 2008. Telerehabilitation for service delivery in speech-language pathology. *Journal of telemedicine and telecare*, 14(5), pp.221–224.
- Wells, J. C., 2006. *English intonation pb and audio cd: an introduction*. Cambridge University Press.

Yang, Hsiang-Hui Jane, Lay, Yi-Ling, Liou, Yi-Ching, Tsao, Wei-Yun, and Lin, Chia-Kai, 2007. Development and evaluation of computer-aided music-learning system for the hearing impaired. *Journal of computer assisted learning* [Online], 23, pp.466–476. Available from: <https://doi.org/10.1111/j.1365-2729.2007.00229.x>.

Young, Steve et al., 2002. *The htk book (for htk version 3.2)*. Cambridge University.

Zeng, Fan-Gang, 2004. Trends in cochlear implants. *Trends in amplification*, 8(1), pp.1–34.

Participant Information Sheet

Project title:	A speech therapy game tool for hearing impaired people
Principal investigator:	Brian Mitchell
Researcher collecting data:	Jin You
Funder (if applicable):	

Participants' information sheet

This study was certified according to the Informatics Research Ethics Process, reference number 262705. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

Jin You and Brian Mitchell

What is the purpose of the study?

Speech disorder is a commonly occurring condition in both children and adults. Although there are many researchers and developers who have provided useful tools to assist in the clinical management of this condition, most of them cannot be directly applied to the speech therapy of people with hearing impairment. We decided to gamify speech recognition to develop a speech therapy support tool for the hearing impaired to help them pronounce words better.

Why have I been asked to take part?

individuals with congenital deafness who have regained their hearing through cochlear implants.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until 1 August 2023 without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

You will be using software, and I will collect your user experience through a survey. Based on the survey results, you may or may not be invited to participate in follow-up interviews.

The survey will consist of various options for judgment, ranking, and degree, as well as 2-3 open-ended questions. The completion time for the survey should not exceed 30 minutes. The duration of the follow-up interviews may range from 30 minutes to 1 hour.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

Are there any benefits associated with taking part?

No.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher/research team Jin You and Brian Mitchell.

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records



will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Jin You, s2450812@ed.ac.uk.

If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on <http://web.inf.ed.ac.uk/infweb/research/study-updates>.

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Jin You, s2450812@ed.ac.uk.

General information.

For general information about how we use your data, go to: edin.ac/privacy-research



Participant Consent Form

Project title:	Pitch Educational Game based on Real-time Visualization Speech Interaction
Principal investigator (PI):	Brian Mitchell
Researcher:	Jin You
PI contact details:	brian.x.mitchell@ed.ac.uk

Appendix B

Participants' consent form

By participating in the study you agree that

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1. I agree to being audio recorded.

--	--

Yes No

2. I agree to being video recorded.

--	--

Yes No

3. I allow my data to be used in future ethically approved research.

--	--

Yes No

4. I agree to take part in this study.

--	--

Yes No

Name of person giving consent

Date
dd/mm/yy

Signature

Name of person taking consent

Date
dd/mm/yy

Signature

Appendix C

API Document

1. Text Processing Functionality

- Endpoint: /process text
- Request Type: POST

Request Parameters:

- text (string): User-inputted text
- rate (boolean): Indicates whether slow speech synthesis is to be employed.
- language (string): The chosen language.

Response Data:

- syllables (array): Contains the list of syllables resulting from text conversion.
- syllable pitches (object): Maps syllables to their corresponding pitch contours.
- average pitches (string): Presents the matching of syllable pitch variations.
- syllable duration (float): Duration of each syllable.

Function Description:

- Parse frontend transmitted data including input text, speech rate, and language selection. Utilize NLTK for text tokenization and conversion into syllables. Leverage gTTS for text-to-speech synthesis, generating audio files. Extract pitch contour information from audio files. Construct and return response data encompassing syllables, pitch contours, syllable-pitch matches, and durations.

2. Functionality for Obtaining Analyzed Result Images

- Endpoint: /get score
- Request Type: POST

Request Parameter:

- audio-file (audio data): Audio data transmitted from the frontend.

Response Data:

- image (image data): Contains the image data depicting the outcome of audio analysis.

Function Description:

- audio data transmitted from the frontend. Invoke the audio analysis function, calculating the Dynamic Time Warping (DTW) distance between audios. Generate an image portraying the analysis outcome and provide the image data back to the frontend.

Appendix D

Questionnaire 1: Requirements Capture First questionnaire

I. Interface Design

1. When using a speech learning tool, which interface design do you find more user-friendly and understandable? (Please select one)
 1. Simple and clear interface, highlighting key functions
 2. Colorful and visually engaging interface
 3. Other (Please specify)
2. For interaction, which form do you prefer? (Please select one)
 1. Visual feedback, such as charts and animations
 2. Textual feedback, to compensate for hearing impairment
 3. Tactile feedback, through touch perception
 4. Other (Please specify)
3. How would you like the layout of the tool's interface to be?
 1. Simple and intuitive layout, minimizing distraction
 2. Customizable layout, tailored to personal preferences
 3. Large icons and buttons for touch accessibility
 4. Other (Please specify)

II. System Flow


4. Do you believe customizable levels in the game would enhance your motivation and interest in learning?
 1. Yes
 2. No
5. What options for customizable levels would you like in the game? (e.g., different languages, difficulty levels, specific pronunciation exercises)
6. How does the real-time feedback mechanism in the game impact your willingness to learn and its effectiveness? Please share your thoughts.
7. Would you like the option to save your learning progress in a database? (Please provide details about your thoughts)
8. Please share any other opinions, suggestions, or expectations you have for this web-based real-time speech interaction tone education game.

D.1 Results of first user study

Survey of Pitch Education Game Requirement for Hearing Impairments

10 responses

When using a speech learning tool, which interface design do you find more user-friendly and understandable? (Please select one)

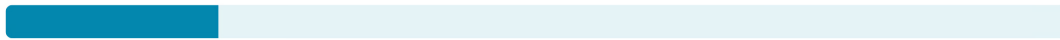
 Hide question

10 out of 10 answered

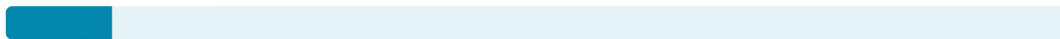
Simple and clear interface, highlighting key functions. 7 resp. 70%




Colorful and visually engaging interface. 2 resp. 20%



Other 1 resp. 10%



For interaction, which form do you prefer? (Please select one)

 Hide question

10 out of 10 answered

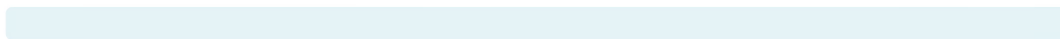
Visual feedback, such as charts and animations. 8 resp. 80%



Textual feedback, to compensate for hearing impairment. 2 resp. 20%



Tactile feedback, through touch perception. 0 resp. 0%



Other 0 resp. 0%



How would you like the layout of the tool's interface to be?

 Hide question

10 out of 10 answered

Simple and intuitive layout, minimizing distraction.

9 resp. 90%



Large icons and buttons for touch accessibility.

6 resp. 60%



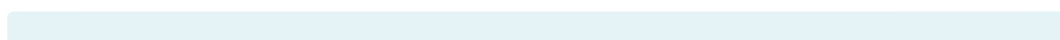
Customizable layout, tailored to personal preferences.

3 resp. 30%




Other

0 resp. 0%



Do you believe customizable levels in the game would enhance your motivation and interest in learning?

 Hide question

10 out of 10 answered

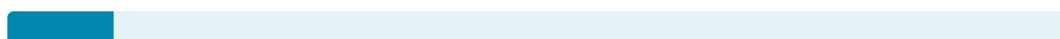
Yes

9 resp. 90%




No

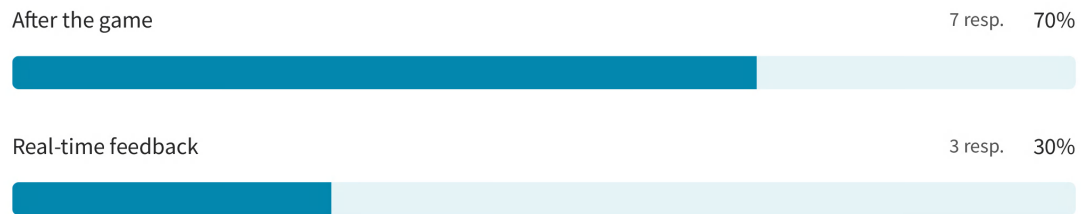
1 resp. 10%




When engaging in voice-based interactive games, do you prefer real-time feedback or receiving feedback together after the game concludes?

 Hide question

10 out of 10 answered



What options for customizable levels would you like in the game?


 Hide question

10 out of 10 answered

Latest Responses

Different speed

Would you like the option to save your learning progress in a database?

 Hide question

10 out of 10 answered

Latest Responses

I certainly hope that my learning progress can be saved, which is very helpful for me, I can know my stage through the progress, I can also know my own level through the progress, so as to facilitate future learning.

Appendix E

Results of usability testing

Task 1 completion status:

Completed	Completed with Assistance	Not Completed
7	0	0

Task 1 completion time:

Immediate	Hesitation
7	0

Task 1 completion path:

Correct Path	Alternative Paths
7	0

Nonverbal cues during Task 1:

YES	NO
0	7

Task 2 completion status:

Completed	Completed with Assistance	Not Completed
7	0	0

Task 2 completion time:

Immediate	Hesitation
7	0

Task 2 completion path:

Correct Path	Alternative Paths
7	0
	50

Nonverbal cues during Task 2:

YES	NO
2	5

Task 3 completion status:

Completed	Completed with Assistance	Not Completed
7	0	0

Task 3 completion time:

Immediate	Hesitation
7	0

Task 3 completion path:

Correct Path	Alternative Paths
7	0

Nonverbal cues during Task 3:

YES	NO
0	7

Task 4 completion status:

Completed	Completed with Assistance	Not Completed
7	0	0

Task 4 completion time:

Immediate	Hesitation
7	0

Task 4 completion path:

Correct Path	Alternative Paths
7	0

Nonverbal cues during Task 4:

YES	NO
0	7

Task 5 completion status:

Completed	Completed with Assistance	Not Completed
7	0	0

Task 5 completion time:

Immediate	Hesitation
7	0

Task 5 completion path:

Correct Path	Alternative Paths
7	0

Nonverbal cues during Task 5:

YES	NO
4	3

Appendix F

Questionnaire 2: Usability Testing

I. System Interest and Visual Appeal

1. How do you rate the overall appearance of the system? (Please select on a scale of 1 to 5, with 1 being unsatisfactory and 5 being very satisfactory)
2. What is your evaluation of the color scheme and interface layout of the system? (Please provide specific comments or suggestions below)

II. Interaction Experience and Smoothness

3. Did you find the system's operations to be smooth during use? (Select one option)
 1. Very Unsmooth
 2. Not Very Smooth
 3. Neutral
 4. Smooth
 5. Very Smooth
4. How would you rate the system's responsiveness to your actions? (Please provide specific comments or suggestions below)

III. Functionality and Practicality

5. How do you evaluate the functionality of the system's custom learning objectives feature? (Please provide specific comments or suggestions below)
6. Regarding the real-time pitch calculation and visualization feature, do you find it helpful for learning pitch? (Please provide specific comments or suggestions below)
7. In the game interface, do you find the visualization representation of notes and target pitch easy to understand? (Please provide specific comments or suggestions below)

IV. User Satisfaction

8. Overall, what is your impression of the system? (Please select on a scale of 1 to 5, with 1 being unsatisfactory and 5 being very satisfactory)
9. Would you be willing to continue using this system?
 1. Yes
 2. No
10. Please provide any additional comments or suggestions you may have about the system:

F.1 Results of the second user study

Survey 2 of Pitch Education Game for Hearing-Impairments

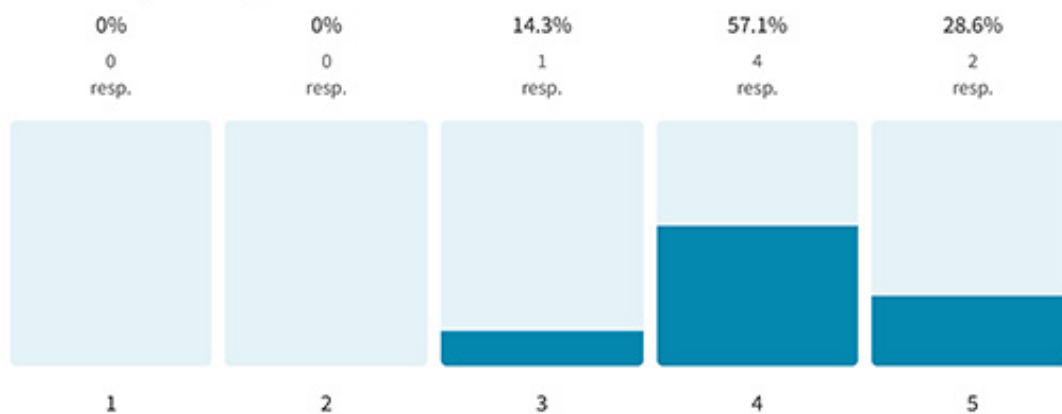
7 responses

How do you rate the overall appearance of the system?

[Hide question](#)

7 out of 7 answered

4.1 Average rating



What is your evaluation of the color scheme and interface layout of the system?

[Hide question](#)

7 out of 7 answered

Latest Responses

I like the pixel style and the whole layout is very simple.

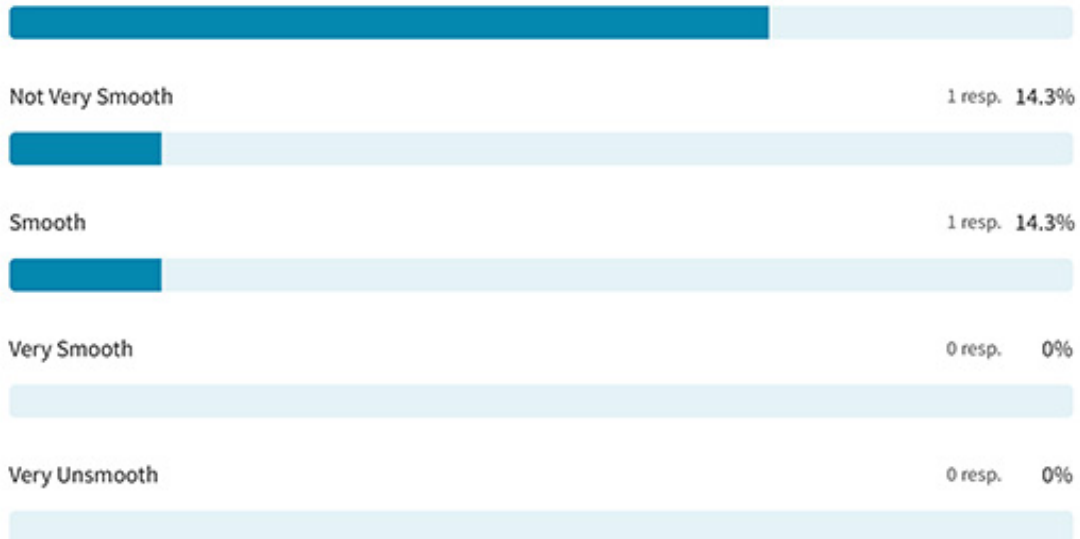
Did you find the system's operations to be smooth during use?

[Hide question](#)

7 out of 7 answered

Neutral

5 resp. 71.4%



How would you rate the system's responsiveness to your actions?

[Hide question](#)

7 out of 7 answered

Latest Responses

I just tried a few short sentences and the maps were generated right away.

How do you evaluate the functionality of the system's custom learning objectives feature?

[Hide question](#)

7 out of 7 answered

Latest Responses

It's very cool

Regarding the real-time pitch calculation and visualization feature, do you find it helpful for learning pitch?

[Hide question](#)

7 out of 7 answered

Latest Responses

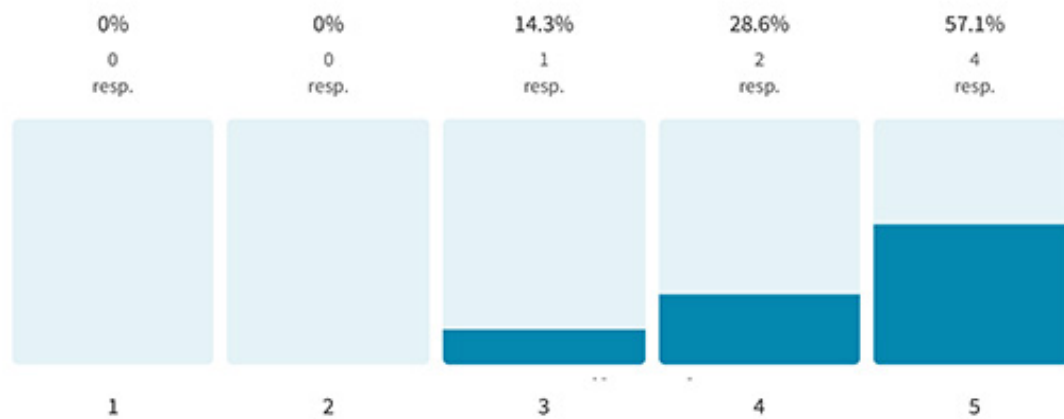
I just tried a few short sentences and the maps were generated right away.

Overall, what is your impression of the system?

 Hide question

7 out of 7 answered

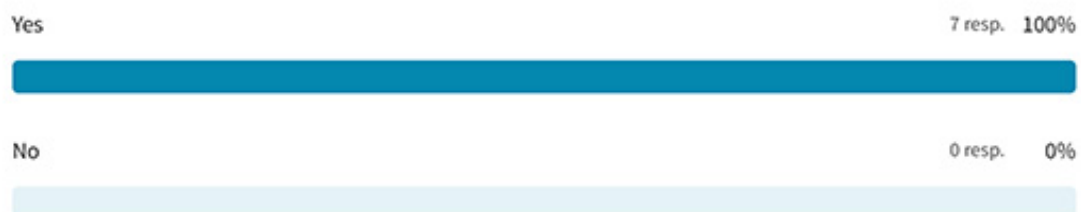
4.4 Average rating



Would you be willing to continue using this system?

 Hide question

7 out of 7 answered



Please provide any additional comments or suggestions you may have about the system

 Hide question

7 out of 7 answered

Latest Responses

It feels like the syllables move a bit fast and it's not very easy to read the words on the syllables