The Application of Speech Recognition in Education and Teaching

Jiahao Mao*, Yanhan Wang[1], Xiangyu Wang[2], Luhan Yang[3] and Yibo Ding[4]

Corresponding author: HarryMaooxford @163.com*;
Email: LillyWangUCL@163.com[1], xiangyuwangoxford@163.com[2], LuhanyangUK@163.com[3], dingyibo2023@163.com[4]

Arts and Business College of London, London, British

Abstract. With the inclusion of middle school music exams in the middle school entrance examination, teachers and students are increasingly valuing music learning. Music teachers hope to have a software that helps students simulate exams to help them experience the atmosphere of music exams in advance and overcome the problem of students forgetting words in a panic. The Whisper model has been used frequently recently, making it convenient and fast. Developing a music simulation exam system based on Whisper speech recognition can help teachers solve this difficulty. This paper designs and implements a WeChat mini program for the middle school music simulation exam system. The system is dedicated to simulating random questions during real exams, identifying songs sung by students through speech recognition, and automatically scoring. Students sing songs at the exam end and submit their test papers after singing. The system automatically grades students based on their singing performance. Teachers can view students' exam records and scores. The music simulation exam system based on Whisper speech recognition, ultimately designed and implemented in this article, helps middle school music teachers and students learn better. It not only solves the problem of teachers organizing simulation music exams, but also helps students test their own level. It has been tested in middle school music teaching and has achieved excellent feedback effects, indicating that the system has certain application value.

Keywords: mock exam, random selection, song recognition, automatic scoring, whisper

1. Introduction

The rapid development of information technology has greatly enriched the people of the world. It has become an indispensable part for people. Education informatization is an important indicator for measuring the level of a country's education plan. It is developing towards the power of the internet and traditional education and challenges. The introduction of information technology in the classroom has made boring traditional music education and music activities more active. This can make students interested in Musicality. Introducing information technology into music exams not only reduces the workload of teachers, but also helps students overcome the problem of forgetting words on the spot. How to integrate information technology into modern music teaching and enable students to learn better is the task of music teachers in the new era.

ICMEIM 2023, September 08-10, Wuhan, People's Republic of China

DOI 10.4108/eai.8-9-2023.2340047

With the popularization of quality education, audio, sports, aesthetics, and information technology are becoming increasingly important in primary and secondary schools, and exams are also essential. However, in terms of music exams, the method of randomly selecting questions and singing lyrics and music scores on site is adopted. One sentence of lyrics from one song is selected and paired with the music scores of another song. Without providing song questions, the difficulty and flexibility of the exam questions are high, making it very difficult for teachers and students to prepare for the exam[9].

At present, most schools can only simulate and train test questions manually, and students often cannot remember the corresponding lyrics and notation after receiving the test questions, making it difficult to score[10]; During regular training, teachers can only print the lyrics and notation of each song to students for training, which wastes paper and cannot keep track of students' practice in real-time. Therefore, it is difficult to achieve high music scores. Especially after incorporating music into the middle school entrance examination, the requirements for music exams have become increasingly high. Therefore, it is very important to design and implement a program that simulates music exams for regular training[4].

2. Related work

2.1. Maintaining the Integrity of the Specifications

For simulation exam systems, scholars both domestically and internationally have conducted research on simulation exam systems and developed various feasible simulation exam products. In foreign countries, there are online teaching soft-ware such as Function Carnival, Fluencia, ClipConverter, etc[1]; In China," Rain Classroom" is a commonly used teaching soft-ware where you can view courseware, complete assignments, and conduct simulated exams. In terms of music, there are also a series of music software called "Sound Shell", and in terms of music exams, there is also a "Music Classroom". Combining WeChat mini programs with mock exams to develop a music mock exam mini program dedicated to practicing live singing. [2]. Compared with the" Music Classroom", this mini program can not only achieve two functions: voice recognition of songs and automatic grading, but also conduct daily simulation training.

The Whisper speech recognition model is an automatic speech recognition system open-sourced by artificial intelligence company OpenAI[11]. Whisper has strong speech recognition capabilities and can recognize over 90 languages, including tiny, base, small, medium, large[3].

2.2. Our contribution:

• Conduct correct needs analysis for teachers and students, and design and implement system functions based on the needs of teachers and students.

• We have designed and implemented three functions: random question selection, speech recognition, and automatic grading. Among them, we have fine-tuned the Whisper model to better recognize speech.

• Establish a database to store the results of simulated speech recognition exams, facilitating the management of exam results.

3. Method

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections III-A–IV-B below for more information on proofreading, spelling and grammar[5].

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads— LATEX will do that for you.

3.1. Requirement analysis

3.1.1 Analysis of Teacher Needs

As shown in Figure 1,the teacher's requirement is to maintain the student answering and simulation exam system. Teachers using the middle school music simulation exam system based on WeChat mini programs first need to register as administrators[6]. After logging in, they can create classes, manage students, manage question banks, publish questions, view students' grades, and view login logs in the background, in order to determine the learning situation of students at this stage and whether to enter future learning.

3.1.2 Analysis of Student Needs

As shown in Figure 2,the needs of students are to conduct simulated training, simulated exams, and view grades on the exam end. Students who use the middle school music simulation exam system based on WeChat mini programs must first scan the code to log in, fill in personal information, and save it before automatically entering the corresponding class. Students can choose" question bank training" in this system for daily learning and conduct tests in" simulated exams". After completing each training or exam, you can view the scores in the" Exam Record". If you encounter speech recognition or inability to view the scores, you can go to" My Customer Service" or" Contact Us" to report the problem.

3.2. Functional design

The main functions of this system are divided into three parts. They are randomly selected questions, identified songs, and automatically graded.

• Random quizzes. Teachers create a question bank in the background, where they can add, query, and delete questions. In the case of ensuring that there are exam questions in the question bank, students clicking on" simulated exam" will automatically generate a test question, which is automatically generated by calling a random function, and the generated test questions are different each time. This way of setting questions is the same as the current middle school music exam, which can help students feel the exam atmosphere in advance and adapt to the exam status.

• Identify songs. After obtaining the generated random test paper, students will click on the question and receive a prompt message. Students will sing the song based on the prompt message. This system stipulates that the time for students to sing is one minute. After the student's finish singing, click the" Confirm Submission" button. At this point, the student's operation is completed. After students submit their answers, the system will identify whether the songs they sing are the songs corresponding to the exam paper.

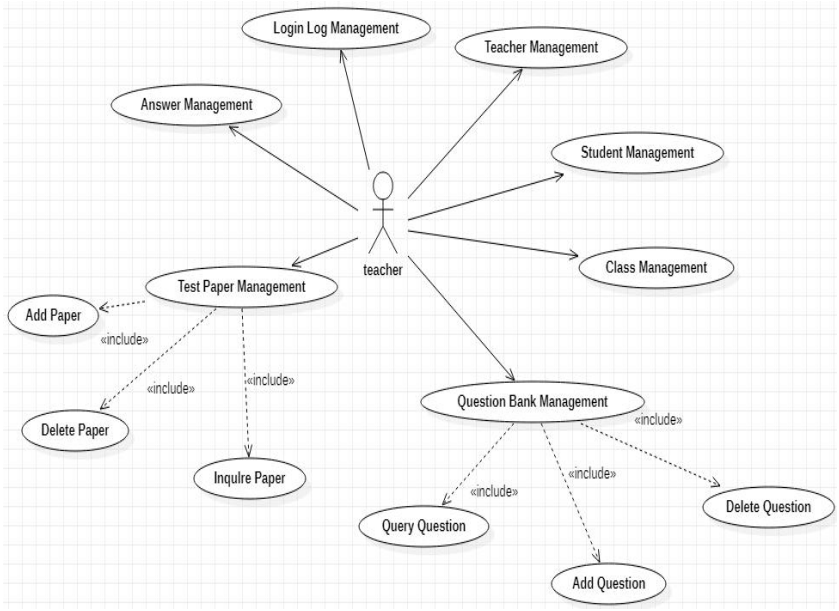

Fig. 1. Teacher Role Use Case Diagram.

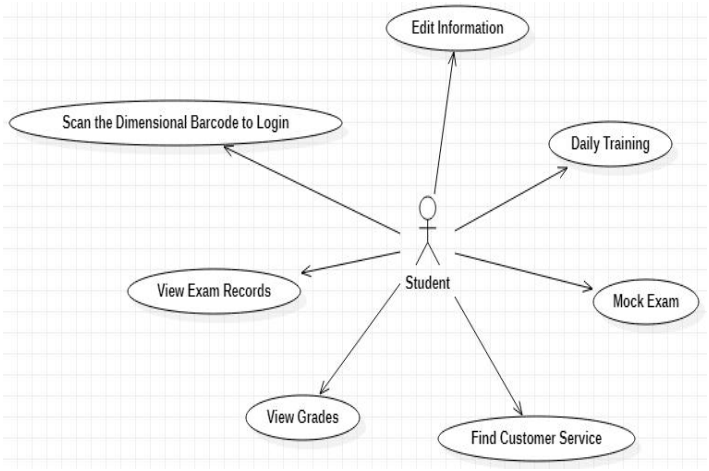

Fig. 2. Student Role Use Case Diagram.

• Automatic rating. Students can check their scores in the" Exam Record" after submitting their test papers. This score is automatically scored by the system. The system rating is specified based on the suggestions of the music exam judges and the examiner. There are two main criteria for system scoring: Is the song sung the song corresponding to the test question. If so, there is a basic score of 60, otherwise it is zero. Is the tone of the song being sung accurate and emotional. The system evaluates students based on their singing performance.

4. Experience

Main function implementation of the system

4.1. Functional design

The main functions of this system are divided into three parts. They are randomly selected questions, identified songs, and automatically graded.

• Automatic rating. The function implemented in this mod-ule is for the system to automatically rate students based on their singing performance. After the system collects students' audio, it begins to recognize ten answers. If the correct answer appears in the ten identified answers and the correct answer appears more frequently, the student's score will be higher. Based on this logic[7], in the process of coding, this paper adopts the method of adding basic scores, random scores, and matching scores to calculate students' singing scores[12]. The setting of random scores takes into account the subjective feelings of the judges in the simulated music exam, and calls a random function to randomly generate a number within an appropriate interval as a random score. As shown in Figure 3.

• Random quizzes. In the random question module, the first step is to determine the type of test paper generated, whether it is" question bank training" or" simulated exam". Define a type integer variable with values of "1" and "2". "1" represents" question bank training", "2" represents" mock exam". Before each test paper extraction, match the value of type to determine the type of test paper to be extracted.

According to the previously matched test paper types, the songs in the question bank and the number of questions to be generated are transferred as parameters to the class generating the random test paper[13], the random function is called to generate the random number and stored in the Hash array, and then the Hash array and the number of questions to be generated are compared. If the number stored is less than the number of specified generations, the recursive method is called to generate the random number of the remaining number, and so on until the specified number is reached[6].

This module is applied in daily training and simulation testing. The system configuration of this paper is to randomly select three song questions from the question bank training and one question from the simulated exam. As shown in Figure 4.

• Identify songs. This module implements the function of language recognition for songs. Firstly, it is necessary to collect the audio of students singing, obtain the per-mission for recording function[8], and save the audio after collecting it. When recognizing songs, divide them into frames, call back each frame, and take the first frame for recognition. This loop callback identifies[14]. This function is realized through the API of IFlytek and base64. As shown in Figure 6.

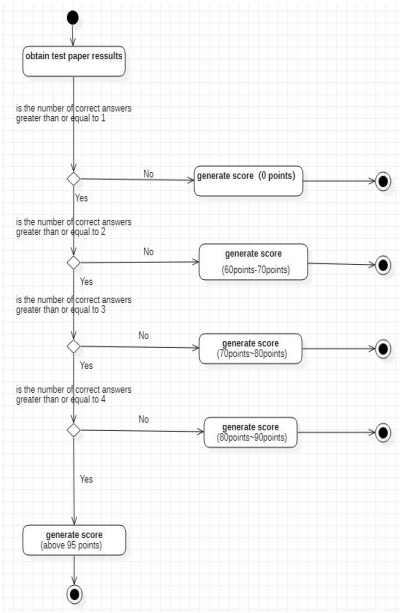

Fig. 3. Auto Score Activity diagram

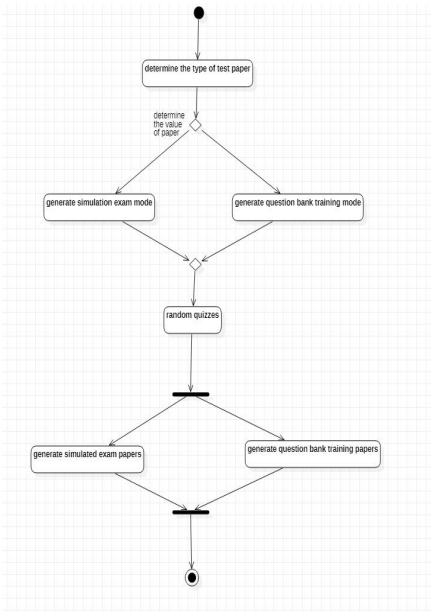

Fig. 4. activity diagram of randomly selected questions.

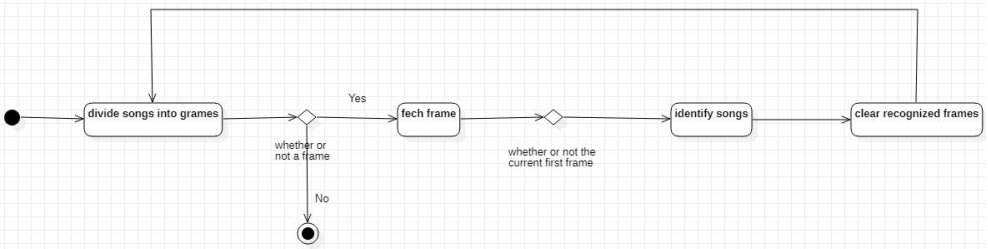

Fig. 5. speech Recognition Activity diagram

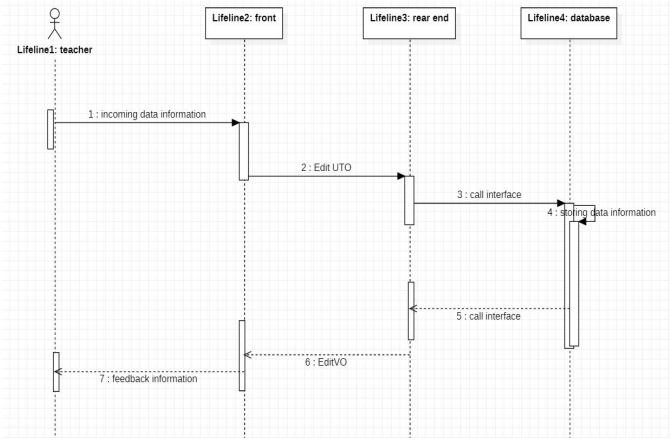

Fig. 6. Timing Chart of Teacher Transmission Information

The backend data such as songs, test papers, exam records, etc. from teachers, students, classes, and question banks are stored in the database. Implemented through interfaces. As shown in Figures 5.

4.2. Functional testing

• Run IntelliJ IDEA and WebStorm.

• Connect to the database, open the WeChat developer tool, and click Edit to display the following interface. Use your phone's WeChat" Scan" to scan the QR code that appears and enter the WeChat mini program. Click on" Development Debugging" to enter the following interface, as shown in Figure 7.

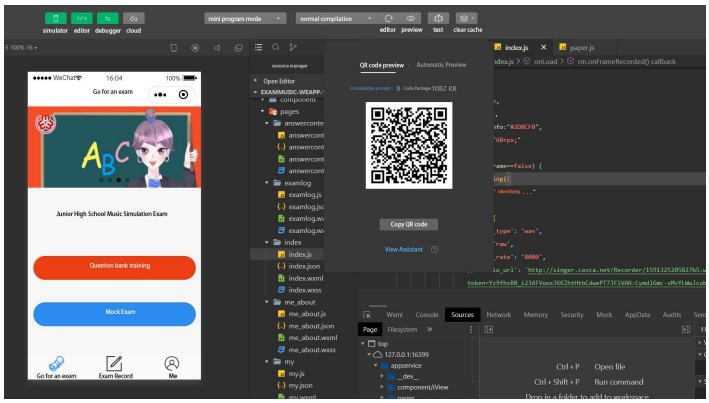

Fig. 7. Test interface display diagram

• Test the "question bank training and simulation exam" function. Click the "Question Bank Training" button to enter the training interface[15], Click on humming to identify the song, complete the training questions in sequence, and submit them after completing them, as shown in Figure 8.

• Click the "Simulated Exam" button to enter the simulated exam interface and obtain randomly generated test questions, after students sing and submit their test papers, the system will automatically generate a score, as shown in Figure 9.

Testing backend management functions, as shown in Table 1.

5. Conclusion

The music exam is officially included in the overall score of the middle school entrance examination. In order to improve the music score, teachers hope to provide students with a simulated exam atmosphere and help them adapt to the middle school music simulated exam as soon as possible. Students use this mini program for daily training and mock exams to experience the exam atmosphere in advance. Teachers can provide guidance to students based on their scores.

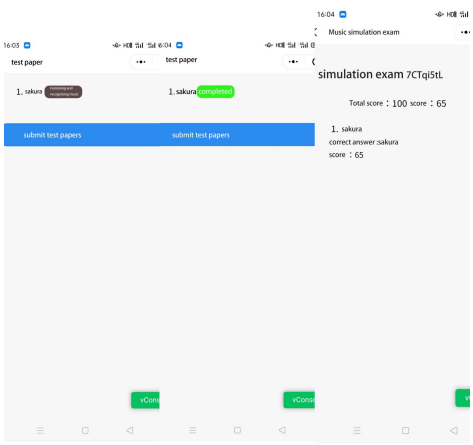

Fig. 8. Test interface display diagram

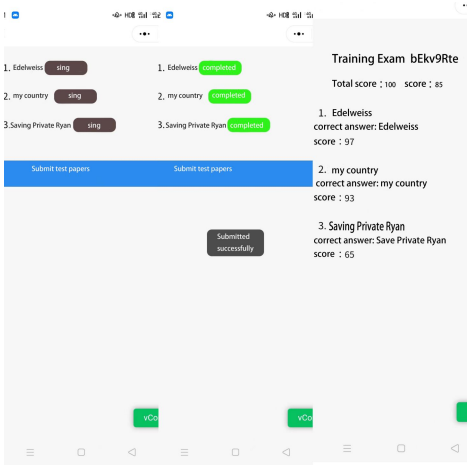

Fig. 9. Test interface display diagram

Table 1: Test database management function table

function	test	test result
Teacher Management Module	1. Add Teacher Information 2. Modify Teacher Information 3. Query Teacher Information 4. Delete Teacher Information	right
Student Management Module	1. Add Student Information 2. Modifying Student Information 3. Query Student information 4. Delete Student Information	right
Class Management Module	1. Add Class Information 2. Query Class Information 3. Delete Class Information	right
Question bank Management Module	1. Add Test Questions 2. Delete Test Questions	right
Test paper Management Module	Query test papers	right
Answer Management Module	Query answer sheet	right
Login Log Management Module	View login logs	right

This article introduces the research background and significance of the Whisper speech recognition model in education and teaching, as well as the current research status at home and abroad. Introduced the implementation and testing of a junior high school music simulation exam system. We have established a music simulation exam system based on the

Whisper speech recognition model. This system has completed three major functions: random question selection, voice recognition of songs, and automatic grading. The main features are as follows:

Firstly, the system randomly selects questions, which can realistically simulate the situation where students obtain test questions through drawing lots during music exams, and can help students adapt to this exam characteristic in advance.

Secondly, the system can recognize songs through voice recognition, which not only reduces the workload of teachers, but also provides grading services for students anytime and anywhere, giving them more time for simulation training and exams.

Thirdly, the system automatically scores, which can effectively avoid the fatigue effect and convergence effect of manual scoring. Automatic scoring has high efficiency, reliability, and validity. Provide students with a more realistic score.

Therefore, the music simulation exam system based on the Whisper speech recognition model has huge application space in education and teaching.

References

[1] Q. Zhang et al.," Transformer Transducer: A Streamable Speech Recognition Model with Transformer Encoders and RNN-T Loss," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 7829-7833, doi: 10.1109/ICASSP40776.2020.9053896.

[2] O. Hazrati, S. Ghaffarzadegan and J. H. L. Hansen," Leveraging automatic speech recognition in cochlear implants for im-proved speech intelligibility under reverberation," 2015 IEEE Inter-national Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 2015, pp. 5093-5097.

[3] M. Soleymanpour, M. T. Johnson, R. Soleymanpour and J. Berry," Synthesizing Dysarthric Speech Using Multi-Speaker Tts For Dysarthric Speech Recognition," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7382-7386.

[4] Z. Nian, Y. -H. Tu, J. Du and C. -H. Lee," A Progressive Learning Approach to Adaptive Noise and Speech Estimation for Speech Enhancement and Noisy Speech Recognition," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 6913-6917, doi: 10.1109/ICASSP39728.2021.9413395.

[5] Y. Li, P. Bell and C. Lai," Fusing ASR Outputs in Joint Training for Speech Emotion Recognition," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7362-7366, doi: 10.1109/ICASSP43922.2022.9746289.

[6] V. Mitra, W. Wang, C. Bartels, H. Franco and D. Vergyri," Articulatory Information and Multiview Features for Large Vocabulary Continuous Speech Recognition," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 5634-5638, doi: 10.1109/ICASSP.2018.8462028.

[7] A. S. Subramanian et al.,” Far-Field Location Guided Target Speech Extraction Using End-to-End Speech Recognition Objectives,” ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and
Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 7299-7303, doi: 10.1109/ICASSP40776.2020.9053692.
[8] C. Fan, J. Yi, J. Tao, Z. Tian, B. Liu and Z. Wen,” Gated Re-current Fusion with Joint Training Framework for Robust End-to-End Speech Recognition,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 198-209, 2021, doi: 10.1109/TASLP.2020.3039600.
[9] S. Karita, S. Watanabe, T. Iwata, M. Delcroix, A. Ogawa and T. Nakatani,” Semi-supervised End-to-end Speech Recognition Using Text-to-speech and Autoencoders,” ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 6166-6170, doi: 10.1109/ICASSP.2019.8682890.
[10] M. Yousefi and J. H. L. Hansen,” Speaker Conditioning of Acoustic Models Using Affine Transformation for Multi-Speaker Speech Recognition,” 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Cartagena, Colombia, 2021, pp. 283-288, doi: 10.1109/ASRU51503.2021.9688231.
[11] S. Shraddha, J. L. G and S. K. S,” Child Speech Recognition on End-to-End Neural ASR Models,” 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-6, doi: 10.1109/CONIT55038.2022.9847929.
[12] B. K. Sy and D. M. Horowitz,” A statistical causal model for the assessment of dysarthric speech and the utility of computer-based speech recognition,” in IEEE Transactions on Biomedical Engineering, vol. 40, no. 12, pp. 1282-1298, Dec. 1993, doi: 10.1109/10.250584.
[13] S. Amuda, H. Boˇril, A. Sangwan and J. H. L. Hansen,” Limited resource speech recognition for Nigerian English,” 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 2010, pp. 5090-5093, doi: 10.1109/ICASSP.2010.5495036.
[14] A. Asaei, M. Cernak and H. Bourlard,” Perceptual Information Loss due to Impaired Speech Production,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2433-2443, Dec. 2017, doi: 10.1109/TASLP.2017.2738445.
[15] K. Laurila, M. Vasilache and O. Vikki, ”A combination of discriminative and maximum likelihood techniques for noise robust speech recognition,” Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP ’98 (Cat. No.98CH36181), Seattle, WA, USA, 1998, pp. 85-88 vol.1.