A MOBILE CLOUD COMPUTING BASED INDEPENDENT LANGUAGE LEARNING SYSTEM WITH AUTOMATIC INTELLIGIBILITY ASSESSMENT AND INSTANT FEEDBACK

Imen M. Kasrani, Miteshkumar M. Vasoya, Ashutosh Shivakumar and Yong Pei
SMART Lab, Wright State University, Dayton, Ohio, USA

ABSTRACT

In this research paper, we present a novel language learning/training and assessment system that helps people to learn and practice a new language independently at low cost. To achieve an independent-learning workflow, we explore the use of real-time speech recognition, language translation, speech synthesis, and language intelligibility assessment technologies to provide automatic assessment and instant feedback of language-speaking performance. We also propose and adopt an objective assessment methodology that determines the intelligibility based on outcome of speech recognition. Our experimental results demonstrate that the proposed system can sufficiently analyze the intelligibility of one's speaking, accurately identify mispronounced words, and define a feedback that localizes and highlights errors for helping continuous practice toward perfection.

KEYWORDS

Mobile Learning, Language Learning, Speech Recognition, Language Intelligibility Assessment, Instant Feedback, Mobile Cloud Computing

1. INTRODUCTION

Recent advances in computer-assisted language-speaking learning/training technology have demonstrated its promising potential to improve the outcome of language learning in early education, special education, English as a Second Language (ESL), and foreign language (Krasnova and Bulgakova, 2014). The growing number of readily available mobile app-based solutions help encourage interest in learning to speak a foreign language, but their effectiveness is limited due to their lack of objective assessment and performance feedback resembling expert judgment. For example, in early education, it is a challenging task for students to extend the language learning at school to home without such feedback and intervention available at home.

In this research, our objective is to develop an effective and practical solution that will help people to learn and practice a new language independently at low cost. We have explored the use of real-time speech recognition, language translation, speech synthesis, and language intelligibility assessment technologies to develop a learning/training system that provides automatic assessment and instant feedback of language-speaking performance to achieve an independent-learning workflow.

1.1 Survey of Existing Language Learning Applications

High quality apps continuously arrive for both iOS and Android that help users to learn a new language effectively and efficiently. These apps cover almost all languages at little to no cost, and provide the private, virtual, and all-inclusive environment necessary for learning and perfecting a new language through reading, writing and speaking. They can be classified into 5 categories:

1. **Language courses**: Babbel, Duolingo and Busuu are among the most popular applications of language courses. These applications use translation and dictation to emulate traditional language classes. Learners read text and listen to videos, then interpret and answer questions. These apps are also used to help

memorize vocabulary. For speaking training, they use the pronunciation of a native speaker for every word and phrase. Unfortunately, this is an unorganized way of learning information because these apps start with complex words and tricky phrases, providing only a way for improving vocabulary rather than effective methods for enhancing conversational skills.

2. **FlashCards and SRS**: Memrise, Tinycards, and AnkiApp are popular examples of this category. They provide a way of practicing vocabulary using memorization of words and phrases, structured as a competitive game in which users are rewarded with points for every correct answer. It is worth noting that Memrise also has a unique feature that associates a new word with similar words from the user's native language to help make a link between words for better memorization.

3. **Educational games**: MindSnacks is such an educational game that helps users learn grammar and vocabulary and practice listening. In addition, this application teaches words and phrases by limiting the time in which to guess the correct answer. It is more applicable to children than to adults because it uses cartoon image.

4. **Q&A, chat and social**: The most popular chat and social applications used in learning new languages are HelloTalk, HiNative, and TripLingo. They use real-time conversation with unknown native speakers and a text-to-voice option to help pronounce received messages. TripLingo is different from HelloTalk and HiNative in that it provides the learner with information related to the place that he/she wants to travel. HiNative is a chat application that uses question and answer features so that the learner can ask the native about their language and culture. Hence, it is a place for one to introduce themselves more than it is a place to correctly practice a new language.

5. **Contextual reference**: Leaf is one of the contextual reference applications that explains the necessary words that the learner needs to know when encountering new situations. It is an application used only for learning English.

Clearly, current language training applications are limited in the following aspects:

1. Improve writing more than speaking: Most of the language training applications do not provide an efficient listening or speaking experience. Users can learn some new vocabulary and constructions, but unfortunately cannot carry on a deep conversation with a native speaker of the foreign language. Learning a new language is not only about learning new words and formulating new phrases with appropriate syntax; it is also about being understandable when pronouncing words (Heil, et al, 2016).

2. Lack of performance assessment and feedback: Current applications rarely evaluate speaking skills and language pronunciation quality. To make the learning of foreign language more efficient, applications need to deliver meaningful feedback that evaluates the quality of the user's speech. A successful application for learning new language needs to be able to make a real evaluation of mispronounced speech, and recognize an incorrect accent.

3. It is mostly about gaming: Applications that depend more on gaming than the actual fundaments of a language can be problematic in the long-term, as passing levels and scoring becomes more important than learning and practicing the language.

1.2 Why Automatic Intelligibility Assessment for Language Training?

People can learn vocabulary and grammar, and then read words and even sentences after practicing a new language, but oftentimes the challenge they are facing is speaking fluently in the language. For learners, either beginners or someone who needs a refresher, feedback on their level of speaking is necessary to correct and perfect their speaking. Learners need accurate feedback on their pronunciation practice because they are often unable to recognize the precise problem in their pronunciation by themselves. In traditional language classes, this is often achieved through practice sessions, such as roleplays, with assessment and feedback done by the instructor.

Clearly, instant assessment and feedback for improvement is key to effective independent language-speaking learning/training because they provide accurate evidence of the merits, progresses, and limitations of a learner's skills. The lack of timely, accurate, and consistent assessment capabilities that have acceptable operating complexity and affordable cost significantly limit the effectiveness of using mobile apps as a viable option to learn foreign language.

Thus, in this research, we will address this issue by focusing on the design and development of a performance assessment system that can offer the opportunity for language learners to have a more complete

picture of what they learn in pronunciation skills and what they need to enhance. The feedback needs to be modeled after human judgment and should be able to be easily interpreted by the learner. An automatic scoring system is appropriate, as it gives the learner instant information on overall result quality (Neri, et al, 2003). Moreover, providing instant feedback gives the learner an idea about their level of progress and gives them a motive to improve their skills over successive attempts.

2. SYSTEM DESIGNS

In this research, we intend to augment and enhance the existing computer-assisted language training idea, as evident by many existing language training applications, to enable independent learning workflow by building a critical new capability that provides automatic intelligibility assessment and feedback.

2.1 System Architecture

To enable a mobile device-centered solution, we decided to adopt a mobile-cloud computing solution that takes advantage of the processing and storage capabilities and capacities of cloud computing service, and the portability, availability and diversity of personal computing devices, particularly mobile devices, as illustrated in Figure 1.

Cloud computing allows fast processing of data from any place over the Internet in real-time without having to concern about the storage or computing power. It is a virtual on-demand delivery of service that does not require an investment in expensive local hardware and software. Moreover, it avoids the hassles to install, configure, and manage the hardware and the software. It provides low-cost and low-latency access to required resources at any time and from anywhere for the consumers. The most popular providers of cloud computing services are: Google Cloud, Amazon Web Services, Microsoft Azure, IBM Bluemix, and Aliyun.

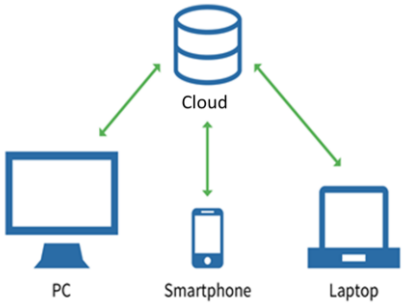

Figure 1. Cloud Computing

2.2 Overview of Our System

The main objectives of this research are: i.) augment the language learning system with accurate and automatic assessment of speaking of a language; and, ii.) enable independent learning workflow. To achieve these objectives, we identify the following key capabilities and features:

- **Recognize the speech**: The most important condition necessary for producing excellent results is accurate speech recognition.
- **Translate the speech**: The application must be able to translate the recognized text in source language (e.g., native language) to the target second language.
- **Synthesize the speech**: The application must be able to convert the translated text into the speech of the desired target language.
- **Assess the intelligibility**: The application must be able to evaluate the user's speech performance in a way that is comparable to human instructors. It should also give instant feedback to the user, e.g., present the overall score as a graded bar, and together with more specific/detailed feedback, such as highlighting the incorrectly pronounced words.

Figure 2 illustrates the workflow of our language learning and assessment system. We have explored the use of real-time speech recognition, language translation, speech synthesis, and language intelligibility assessment technologies to develop a learning/training system that provides automatic assessment and instant feedback of language-speaking performance to achieve an independent-learning workflow. Our prototype system demonstrates the feasibility and effectiveness of such a mobile cloud computing enabled independent learning/training solution. It can be easily used on a smartphone, tablet, computer, or other portable devices, and provides a new learning experience that is augmented and enhanced by objective assessment and significant feedback to improve the language-speaking proficiency of the user.

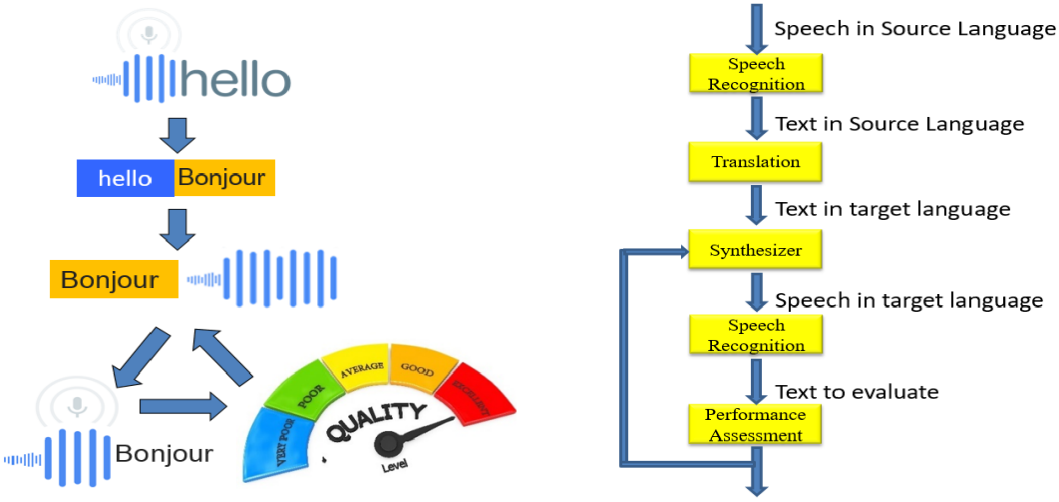

Figure 2. Overview of System Workflow

2.3 Enabling Technologies

In this section, we will discuss the related technologies that help enable the language learning system with accurate and automatic assessment of speaking of a language.

2.3.1 Speech Recognition

Speech recognition software can identify the words of a spoken language, captured from a microphone, and convert them to text. Today, it is used more and more in our daily lives, particularly with the mobile devices. For example, we can give a verbal command for phone calls on smartphones. There are several mature speech recognition services being offered through cloud services, e.g., *iOS* Siri, Amazon Alexa, Android speech to text, IBM Watson, and Google Cloud Speech.

In this research, we take advantage of the handily available cloud speech recognition capabilities offered by the Google Cloud Speech service. First, it produces accurate speech recognition results, even in a noisy environment, by using effective neural network modeling and machine-learning algorithm. Secondly, it recognizes over 110 languages, which fits perfectly with our language learning app. Furthermore, it can give word hints depending on the context provided as well as filter incorrect text results for certain languages. Finally, it is a low-cost service to the user.

2.3.2 Language Translation

Language translation services translate text input from one language to another language. Some of the most popular cloud-based translation services include IBM Watson language translator, Microsoft Translator, and Google Cloud Translation.

In this research, we make use of Google Cloud Translation API to translate the text from the source language to the desired target language in real-time. The Google Cloud Translation API translates text between 2 languages with high accuracy by using the state-of-the-art Neural Machine Translation. It supports over 100 languages, and makes translations between groupings of thousands of languages pairs.

2.3.3 Speech Synthesis

Speech synthesis, also called text-to-speech (TTS), converts natural language text into speech, so a computer, smartphone, tablet, or another device can read the produced audio stream aloud. Speech synthesis is the opposite of speech recognition, as it is a text to speech converter rather than a speech to text converter. The quality of a speech synthesizer is measured in two perspectives: naturalness and intelligibility. Naturalness is the ability to resemble the human voice, and intelligibility is the clearness of the output voice.

There are several available technologies that make conversion from text to speech, such as IBM Watson Text to Speech, FreeTTS, and MaryTTS (Modular Architecture for Research on Speech Synthesis). We use Mary TTS because it is an open source application written in Java and can synthesize many natural languages. MaryTTS was a shared project between DFKI's Language Technology Lab and the Institute of Phonetics at Saarland University. It supports many languages, including, e.g., German, British and American English, French, Italian, Russian, Swedish, Telugu, etc. It can also run in multiple platforms.

2.3.4 Speech Intelligibility Assessment

Speech intelligibility assessment is a complex and subjective process that may vary significantly from one human evaluator to another. In this research, we propose and adopt a more objective assessment methodology by determining the intelligibility based on outcome of speech recognition (Liu, et al, 2006). Following speech recognition, the assessment process is completed by an accurate comparison between speech-recognized spoken text and the original text. For instance, we need to compare the two texts to find the incorrect words that the learner spoke. Then, based on the result from the comparison, the learner will be given feedback of his/her intelligibility in speaking the language. Specifically, we take the following steps:

- Compare the recognized spoken text to the original given or translated text to identify the words that is missing or need to be replaced or removed.
- Assess the percentage of similarity between the two texts to determine the score of intelligibility.
- Highlight the words in green that need to be replaced in the recognized spoken text.
- Highlight the words in red that need to be removed in the recognized spoken text.
- Highlight the words in yellow that is missing in the original text.

To compare and identify the similarity/dissimilarity between two texts, we need to measure the distance between them. There exist different algorithms that count the number of operations needed to transform one string to another string, such as Levenshtein Distance, Hamming Distance, Longest Common Substring Distance and Jaro-Winkler Distance (Cohen, Ravikumar and Fienberg, 2003). In this research, we compare the recognized spoken text and the original text (as the control) word-by-word using the Levenshtein algorithm as illustrated in Table 1. It calculates the minimum numbers of change, including deletion (Missed), insertion (Removed), and substitutions (Replaced), required to transform one string to the other. One potential concern of adopting this algorithm for real-time mobile application is its complexity. The time complexity of the algorithm is $O(n*m)$, where n and m are the lengths of the two sentences being compared. The memory space complexity is $O(n*m)$ because it memorizes in matrix. However, it becomes less a concern nowadays as most of today's mobile devices can provide sufficient computing power and memory space for its operation, even for long sentences.

In Table 1, we illustrate the comparison between 2 sentences using the Levenshtein algorithm. For instance, the comparison between "the weather is nice today" and "the weather is it nice day" give 2 errors: 1 replacement between "today" and "day" and 1 deletion of the word "it".

The accurate analysis of learner speech makes it possible to provide instant feedback on what he/she did not observe otherwise. This feedback includes two parts: 1. The percentage score of intelligibility of the spoken language. 2. Highlight words that the learner needs to work on.

Instant feedback plays a crucial role in learning. It helps the learner clearly know the adjustment needed. Furthermore, it helps the learner to know whether he/she achieved the goal or not. Evaluation system of language learning may also help the trainer to develop training courses that concentrate better on identified weakness and provide highly personalized learning experience. After the Missed, Removed and Replaced words were identified by back tracing, we highlight them with different colors. The feedback of our language learning application provides the advantages of both Constructivist and Behavioristic theories of language learning. The application acts as a virtual facilitator by providing instant feedback that emulates constructivism. Further, it implements behaviorism by identifying errors pertaining to intelligibility and guiding the learner to practice on specific pronunciations (Heil, et al, 2016).

ISBN: 978-989-8533-76-0 © 2018

Table 1. Compare 2 Sentences Word-by-Word Using the Levenshtein Algorithm

		the	weather	is	nice	today
	0	1	2	3	4	5
the	1	0	1	2	3	4
weather	2	1	0	1	2	3
is	3	2	1	0	1	2
it	4	3	2	1	1	2
nice	5	4	3	2	1	2
day	6	5	4	3	2	2

We then calculate and report the percentage score of intelligibility use this formula:

Score of Intelligibility = 1- Number of Incorrect Words / Max (lengthSentence1, lengthSentence2);

Where:

Number of Incorrect Words = Missed Words + Removed Words + Replaced Words.

3. EXPERIMENTAL RESULTS

We have designed and implemented a successful prototype system that demonstrates the feasibility and effectiveness of such a mobile cloud computing enabled independent language learning/training solution, as shown in Figure 3. The following results present samples and offer validations for our system.

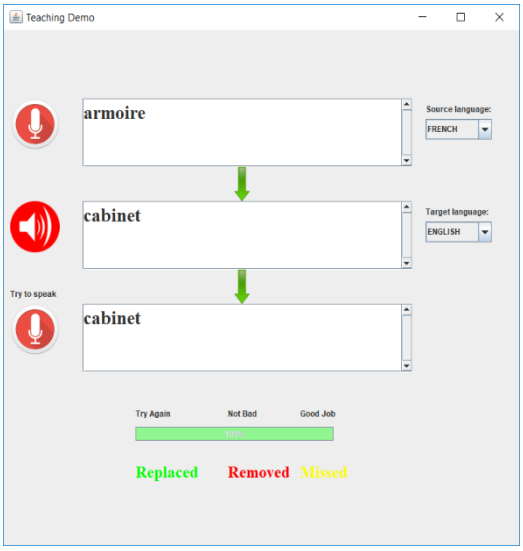

Figure 2. User Interface of the System

When starting the application, the user needs to select the source language and the target language. The user will then click the microphone button to start talking in the source language (e.g., the user's native language). As evident in Figure 3, the speech is successfully converted to text and displayed in the designated text area on the screen. Next, the text displayed in the first text area will be automatically translated to the target language, and then converted to audio for the user. The user can choose to listen again by clicking on the speaker button. When the user is ready to try, he/she needs to click the second microphone button and read the words inside the translated text area. The speech in target language will then be converted to text, but this time in the target language, and displayed in the specified text area. Finally, the system will compare the last recognized text to the translated text. The score of similarity will be displayed, and the missed words, wrong words, and the words that need to be substituted will be highlighted.

3.1 Testing with a Single Word

A user who wants to learn a foreign language will start by pronouncing a single word. So, we will first test with a single word as shown in Figure 4. At the left, the source language is French, and the target language is English. The word “armoire” is translated correctly to the word “cabinet”. The user pronounced the word “cabinet” correctly. Thus, the score of intelligibility is 100%. In the middle, the source language is French, and the target language is Italian. The user pronounced the word “scarpe” correctly, but added another word that was not supposed to be there. Consequently, the score of intelligibility is 50%. The word “le” should be removed, thus highlighted in red. At the right, the source language is English, and the target language is French. The user read the word “couverture” wrongly, resulting in a different word “converse”. Thus, they are both highlighted in green. The pronunciation is incorrect, so the score of intelligibility is 0%.

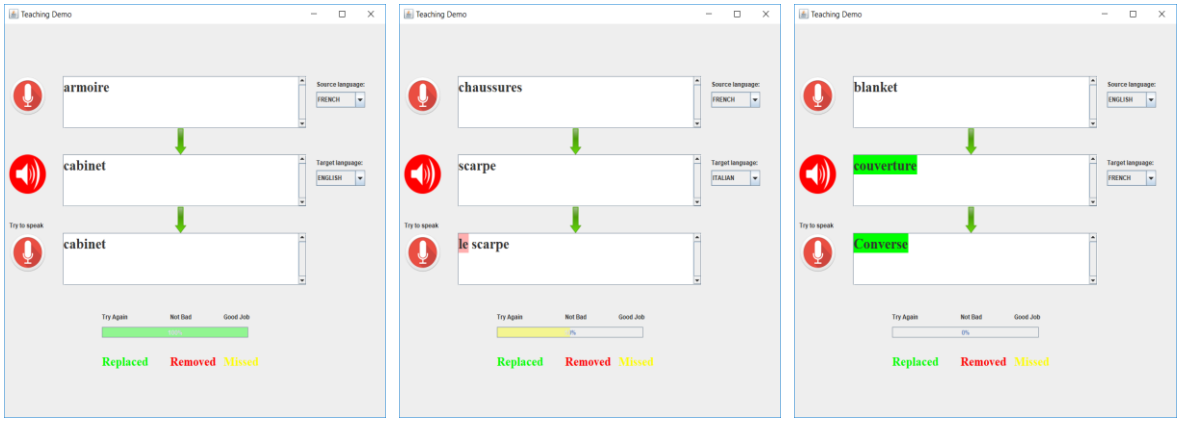

Figure 3. Tests with a Single Word

3.2 Testing with a Single Phrase

Next, we will test with a single phrase as shown in Figure 5. At the left, the source language is English, and the target language is French. The entire phrase is read correctly. Thus, the score of intelligibility is 100%. In the middle, English is the source language, and French is the target language. The feedback gives a score of intelligibility of 33% because only the word “bloc” is correct. The user mispronounced the word “autour” to “retour” and “du” to “de” which are both highlighted accordingly in green. At the right, English is the source language, and Italian is the target language. The user read the entire phrase “correre attraverso i boshi” incorrectly. Thus, the words that need to be replaced are highlighted in green, and the words that need to be removed are highlighted in red. The pronunciation is not correct, so the score of intelligibility is 0%.

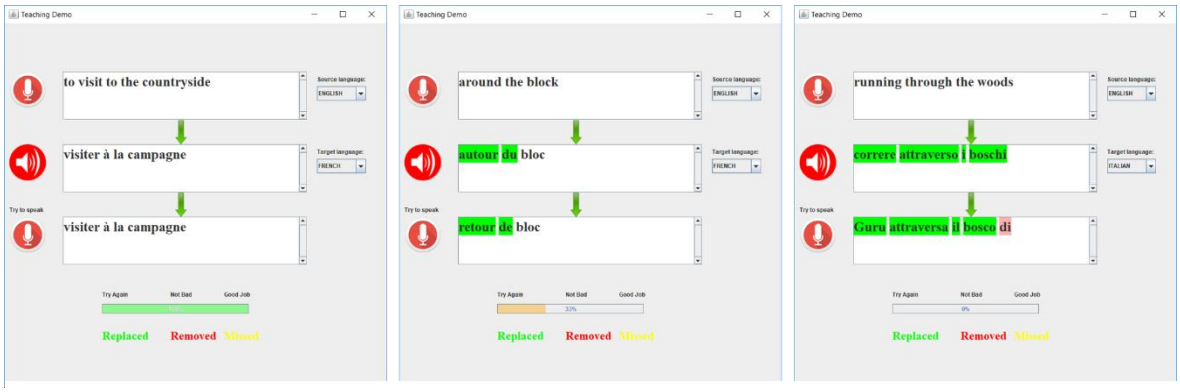

Figure 4. Tests with a Single Phrase

3.3 Testing with a Sentence

Finally, we will test with a sentence as shown in Figure 6. For all three tests, the source language is English, and the target language is French. At the left, the entire sentence is read correctly. Thus, the score of intelligibility is 100%. In the middle, the feedback gives a score of intelligibility of 66% because the user mispronounced the word “il” to “elle”, which is highlighted accordingly with green, and missed the word “si”, which is correctly highlighted in yellow. At the right, the user read the entire sentence incorrectly as highlighted in green, so the score of intelligibility is 0%.

Figure 5. Tests with a Sentence

4. CONCLUSION

An effective and practical solution that helps users to learn and practice a new language independently at a low cost will significantly improve learning efficiency and outcome, as well as encourage more people to learn new languages. In this research, we have successfully developed a mobile cloud computing based language-speaking learning tool that harnesses the latest advances in real-time speech recognition, language translation, speech synthesis and language intelligibility assessment technologies to produce automatic assessment and instant feedback of language-speaking performance to help achieve an independent-learning workflow. Our experimental results demonstrate that the proposed system can sufficiently and accurately analyze the intelligibility of one’s speaking, correctly identify the mispronounced words, and define a feedback mechanism that localizes and highlights errors for helping users to continuously practice towards perfection.

REFERENCES

Google Cloud Speech. Available at: https://cloud.google.com/speech/.

Google Cloud Translate. Available at: https://cloud.google.com/translate/.

Heil, C. R., Wu, J. S., Lee, J. J., & Schmidt, T. (2016). A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities. *The EuroCALL Review*, *24*(2), 32–50.

Krasnova E., Bulgakova E. (2014) The Use of Speech Technology in Computer Assisted Language Learning Systems. In: Ronzhin A., Potapova R., Delic V. (eds) Speech and Computer. SPECOM 2014. Lecture Notes in Computer Science, vol 8773. Springer, Cham

Liu, W. M., Jellyman, K. A., Mason, J. S. D., & Evans, N. W. D. (2006). Assessment of Objective Quality Measures for Speech Intelligibility Estimation. In 2006 IEEE ICASSP. https://doi.org/10.1109/ICASSP.2006.1660248

MaryTTS, Available at: http://mary.dfki.de/.

Neri, A., Cucchiarini, C. and Strik, H. (2003) Automatic speech recognition for second language learning: How and why it actually works. Speech Communication.

W. Cohen, W, Ravikumar, P. and E. Fienberg, S. (2003). A Comparison of String Metrics for Matching Names and Records. Proc of the KDD Workshop on Data Cleaning and Object Consolidation.