iLEAP: A HUMAN-AI TEAMING BASED MOBILE LANGUAGE LEARNING SOLUTION FOR DUAL LANGUAGE LEARNERS IN EARLY AND SPECIAL EDUCATIONS

Saurabh Shukla[1], Ashutosh Shivakumar[1], Miteshkumar Vasoya[1], Yong Pei[1] and Anna F. Lyon[2]
[1]*SMART Lab, Wright State University, Dayton, Ohio, USA*
[2]*Department of Teacher Education, Wright State University, Dayton, Ohio, USA*

ABSTRACT

In this research paper, we present an AR- and AI-based mobile learning tool that provides: 1.) automatic and accurate intelligibility analysis at various levels: letter, word, phrase and sentences, 2.) immediate feedback and multimodal coaching on how to correct pronunciation, and 3.) evidence-based dynamic training curriculum tailored to each individual's learning patterns and needs, e.g., retention of corrected pronunciation and typical pronunciation errors. The use of visible and interactive virtual expert technology capable of intuitive AR-based interactions will greatly increase student's acceptance and retention of a virtual coach. In school or at home, it will readily resemble an expert reading specialist to effectively guide and assist a student in practicing reading and speaking by him-/herself independently, which is particularly important for dual language learners (DLL) whose first language (L1) is not English as many of their parents don't speak English fluently and cannot offer the necessary help. Our human-AI teaming based solution overcomes the shortfall of conventional computer-based language learning tools and serve as a supportive and team-based learning platform that is critical for optimizing the learning outcomes.

KEYWORDS

Dual Language Learners, Mobile Learning, Human-AI Teaming, Language Intelligibility Assessment, Mobile Cloud Computing

1. INTRODUCTION

Learning English just like any other language can be equally challenging to dual language learners, both young and adults (Krasnova and Bulgakova, 2014). Dual language learners (DLL) whose first language (L1) is not English need many opportunities to speak and read English (L2) to achieve the English language proficiency needed for academic success, social and emotional competencies. Many schools offer programs during school time that assist such children in developing language proficiency. But those programs may not be enough due to restriction of time and staffing.

In this research, we have developed a mobile solution – iLeap, enabled by the latest artificial intelligence technologies, such as Machine Learning and Automatic Speech Recognition, that will support DLLs of young age. The iLeap learning tool offers them the option to practice accurate pronunciation with a virtual reading specialist and receive immediate feedback and instruction on how to correct pronunciation even when a native speaker is not available to assist. It will serve as a virtual assistant at school for the reading specialist since these students may require personalized attention which instructor cannot ensure due to limitation of staffing and practice time. Moreover, it helps address their biggest challenge in language learning - to extend the language practice and learning in school to home, as many of their parents don't speak English fluently and cannot offer the necessary help at home.

1.1 Background of Study

There are many learning apps already available, either web-based or on mobile platform, for Dual Language Learners that provide personalized language training (Heil, et al, 2016). Some of these applications use Flash cards, animation-based games (e.g. match words with pictures) to keep the learners engaged. They motivate the kids to memorize the vocabulary, but they hardly help in developing communication skills. There are some popular applications like Babbel, Duolingo that use translation and dictation to emulate traditional language classes. Learners read text, listen to videos and then interpret and answer questions. But most of these applications have a focus of improving vocabulary and writing skills than speaking and accurate pronunciation. The applications provide no tools to assess speaking skills and quality of pronunciation which is critical for student's practices (Neri, et al, 2003). Thus, there is a need of application that could assess the pronunciation of new learners, provide instant feedback on mispronounced words, pinpointing the mistake at the corresponding phonemes, and then be able to provide both audio and visual instructions on how to correct the pronunciation.

1.2 System Features of Proposed Solution

Our primary goal of this project is to support dual language learners for independent language learning. To achieve this goal, we identify the following key capabilities and features necessary for supporting effective pronunciation training/learning.

1.2.1 Emphasis on Reading and Pronunciation Skills

The iLEAP system insists on developing the reading and pronunciation skills of the learners. The learners work on various books reading sessions through the app and the system assess their performance in real time. Books are suggested to the learner intelligently based on the profile data. The application leverages speech recognition API provided by Google Cloud Speech services.

1.2.2 Intelligibility Assessment, Feedback and Phoneme Level Correction

The assessment of the performance is done in real-time. Learner gets to know immediately if he/she mispronounced any word. We make use of Android usability features Text highlighting, clickable spans to make the application easy and intuitive to use. The mispronounced word is compared with original word further at phoneme level. For this work, we have used the set of 39 distinct phonemes from CMU (CMU-Sphinx project). On summary view, when word playback is requested, only the phonemes that diverged on the recognized word from original word are uttered, with help of visual animation that show lip movements required to accurately pronounce that specific phoneme. For instance, if learner pronounce "LIFT" for original word "LEFT",

- Both words will be compared at phoneme level as:
  L EH F T → *L IH F T*
- The server returns mismatching phoneme "EH"
- The app will playback sound for "EH" with corresponding animation followed by utterance of original word "LEFT"

The accurate analysis of learner speech makes it possible to provide instant feedback on what he/she did not observe otherwise. 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. 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).

1.2.3 User Profiling and Learning Retention Assessment

The content server in the cloud also maintains user profile. After completion of each session, the app sends performance data (e.g., list of mispronounced words) during that session, which is updated by the server in database. This enables the cloud server to generate different insights into user profile, like most frequent mispronounced words, typical phonemes that the learner may have difficulty to pronounce, retention of learning over time, i.e., whether the learner's pronunciation improved for certain word. The scope of data collection and server-side capabilities can be conveniently extended as needed due to the use of cloud-based approach, once the basic framework is available. Thus, we may also enhance both app and the server in future for many other insights in user profile.

2. SYSTEM OVERVIEW

The iLEAP application focuses on the usability of the application, keeping a specific audience in mind: young kids of 4 - 8 age group. Hence the mobile application incorporates simple and intuitive ways to provide performance assessment on the reading session instantaneously.

2.1 System Architecture

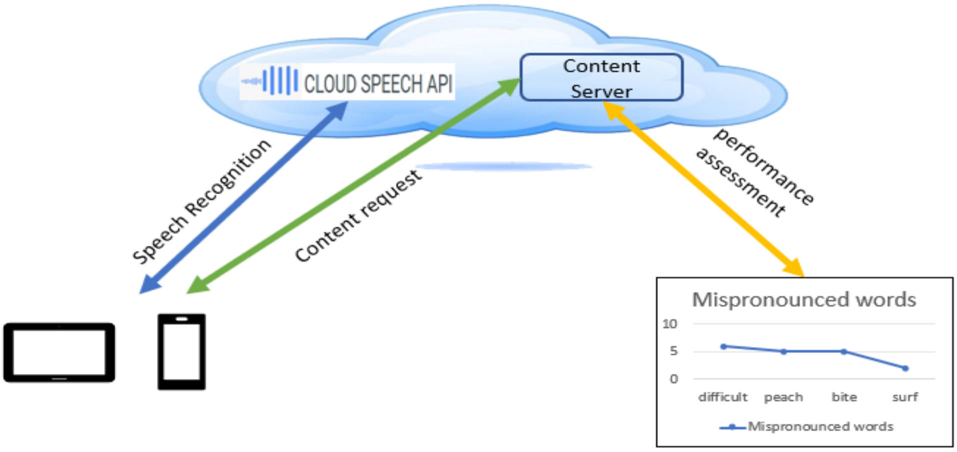


Figure 1. Overview of iLEAP System Architecture

Figure 1 illustrates the iLEAP system architecture. The application can be deployed on mobile device or tablet. The user is authenticated with content server and then the books that fits the authenticated account profile can be listed on the device. The title selected by the user will then be retrieved from server and the text content is displayed on the device. When learner starts reading the book, speech recognition service of android app captures audio stream and sends the audio data to Google cloud Server for recognition. When the recognition result is received from the cloud server, the recognized text is compared with source text from the book for word by word comparison. The learner will be given instantly the feedback of his/her intelligibility in speaking the language in terms of highlighted text as the reading progress –

- Green highlight indicates the word pronunciation was accurate
- Yellow highlight indicates the word was mispronounced

These mis-pronounced words can be rehearsed when the session ends. The content server also provides retention tracking. All the mispronounced words are updated in the database for learner's profile. This data can be used to perform analytics on the learner's profile and evaluate the user performance. The analytics may provide insights like words that learner persistently fails to read, or individual phoneme in different words that the student face most difficulty in pronouncing accurately. It may also provide pattern of retention in the learning; whether the learner improved on certain word that he/she faced difficulty in the beginning.

2.2 Components and Enabling Technologies

2.2.1 Speech Recognition

iLEAP uses Google Cloud Speech Streaming API to recognize AUDIO input. Streaming API enables it to perform speech recognition of continuous audio stream in real-time. Google cloud services provides gRPC stub for Android/Java platform. We implement speech recognition service using the gRPC stub APIs. For accounting purpose, the gRPC client stub needs authentication token in order to validate the account for the use of speech recognition service. Currently this service is available worldwide at $1.44 per hour, which is significantly lower when compared to hiring a personal language coach or tutor.

2.2.2 Content and Profiling Server

The contents for reading session are dynamically retrieved from the content server instance that is deployed on cloud for 24/7 availability. In our project, Amazon cloud service is used and the Content server is implemented in Flask/Python with MySQL as backend database. This server provides RESTful APIs such that android app will be able to request reading content, request phoneme level comparison of words, update user profile in MySQL database for mispronounced words or get analytics on user profile for reading patterns.

2.2.3 User Interface

The user interface of the prototype is the most critical part of any learning apps designed for children at young age, it has to be as simple as possible with the intention to avoid distraction due to unnecessarily complicated operations. Thus, in iLeap, most of the interactions are through intuitive components, such as buttons, layouts and views carry symbols that handily describe the objective of the interface. On completion of a reading session, it automatically summarizes all the mispronounced words from the session along with phoneme level intelligibility feedback, such that the system utters only individual phoneme that was mis-pronounced in case of homonyms. The coaching system simultaneously highlights correct way of lip gestures required to pronounce the phoneme accurately using visual animations through Emoji or Animoji.

2.2.4 Intelligibility Assessment

Speech intelligibility assessment is a complex 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.

To identify the similarity/dissimilarity between two texts, we need to measure the distance between them. This can be achieved using various minimum distance finding algorithm, 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 word-by-word using the Levenshtein algorithm. It calculates the minimum numbers of change, including deletion (Missed), insertion (Removed), and substitutions (Replaced), required to transform one string to the other. The time complexity of this 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. This could be a concerning factor considering we have to compare the sentence incrementally every time with speech recognized text if the sentence is uttered in multiple parts with pauses. However, it becomes less a concern nowadays as most of today’s mobile devices can provide enough 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 “five little monkeys jumping on the bed” and “five little monkey jumping the bad” computes similarity score of 71.

Table 1. Assessment through Levenshtein Algorithm

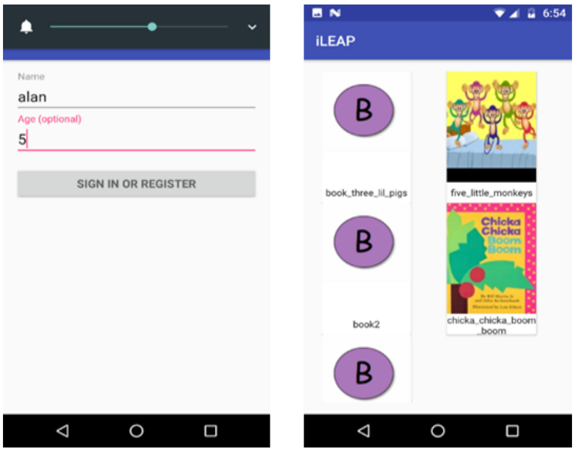
		five	little	monkey	jumping	the	bad
	0	1	2	3	4	5	6
five	1	0	1	2	3	4	5
little	2	1	0	1	2	3	4
monkeys	3	2	1	1	2	3	4
jumping	4	3	2	2	1	2	3
on	5	4	3	3	2	2	3
the	6	5	4	4	3	2	3
bed	7	6	5	5	4	3	3

3. EXPERIMENTAL RESULTS

The prototype has been developed to illustrate and evaluate the effectiveness of the mobile app enabled language learning. The following results provide validation for our approach.

3.1 User Interface

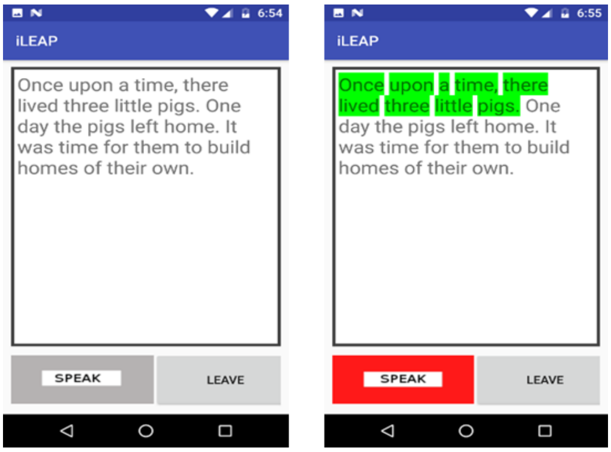
Once the app is launched on the device, user lands on login page as shown in Figure 2. The authentication process verifies user profile on the backend server. After authenticating the learner, the application lists book titles that are relevant to learner's profile. The profile level is derived at the server side based on learner's age and how he progresses through various reading sessions. Server maintains books in generic hierarchical structure so that random titles can be displayed to the learner in order to expose them to new content/vocabulary and avoid repetition.



Name: alan

Age (optional): 5

Figure 2. Launch the App



Once upon a time, there lived three little pigs. One day the pigs left home. It was time for them to build homes of their own.

Once upon a time, there lived three little pigs. One day the pigs left home. It was time for them to build homes of their own.

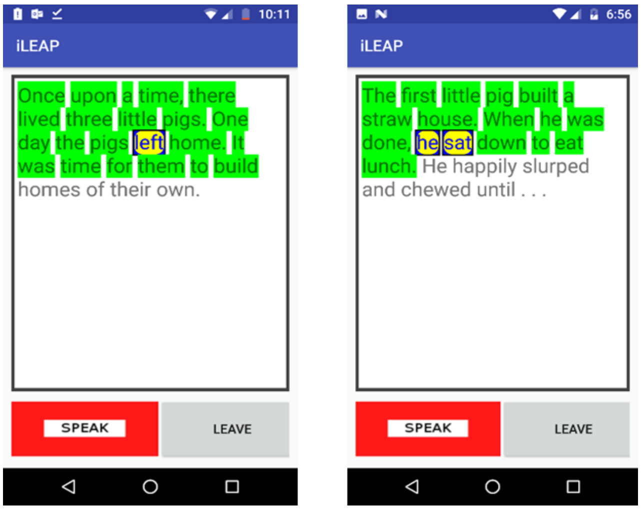
Figure 3. Reading progress without errors

3.2 Reading Progress with Accurate Pronunciation

When a book is selected by learner for reading, the contents are displayed as plain text. Once the audio recording is enabled with a button click, speech recognition results are matched with the original text in the background and text is instantly highlighted with appropriate color spans. As illustrated in Figure 3, if the recognized text matches with source text, the green background span highlights the portion of matched text.

3.3 Reading Progress with Dissimilarity Detection

If any word is mispronounced during the session, intelligibility assessment algorithm returns dissimilarity with original text. This dissimilarity is highlighted with yellow background on original text. The highlight also enables clickable interface on the word so that learner can click on the word to hear out correct pronunciation of the word using Android Text-To-Speech API. As illustrated in the Figure 4, when "left" was mispronounced as "lift", the intelligibility assessment detects the mismatch between recognized text and the text is highlighted accordingly.



Once upon a time, there lived three little pigs. One day the pigs left home. It was time for them to build homes of their own.

The first little pig built a straw house. When he was done, he sat down to eat lunch. He happily slurped and chewed until . . .

Figure 4. Reading Progress Errors

3.4 Session Review and Correction Coaching

At the end of the session, all dissimilar words are displayed for practice as shown in Figure 5. The dissimilarity is mapped at phoneme level such animation shows lip movement for the missing phoneme. As shown in the figure, learner pronounced "lift" for "left", the missing phoneme was identified as "EH". The animation mimics lip movements to pronounce "EH", along with Text-To-Speech utterance of the phoneme and entire word. The learner can practice again with the word that he/she failed to pronounce properly as shown in Figure 6. The mic button interface enables speech recognizer to accept audio input for speech-to-text translation. Intelligibility assessment feedback for the re-attempted word is also available in terms of background color of the mic button.

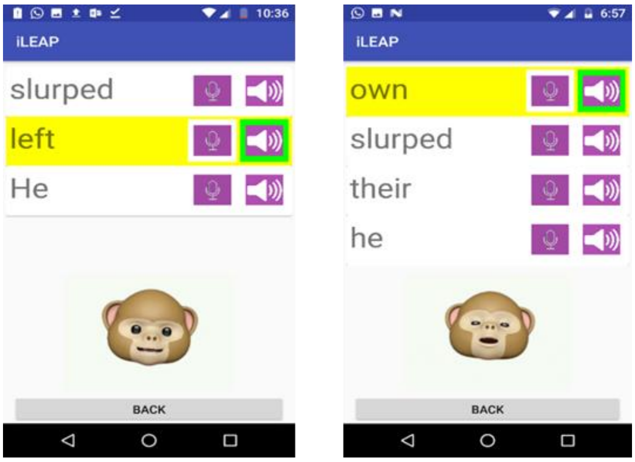


Figure 5. Correction Coaching

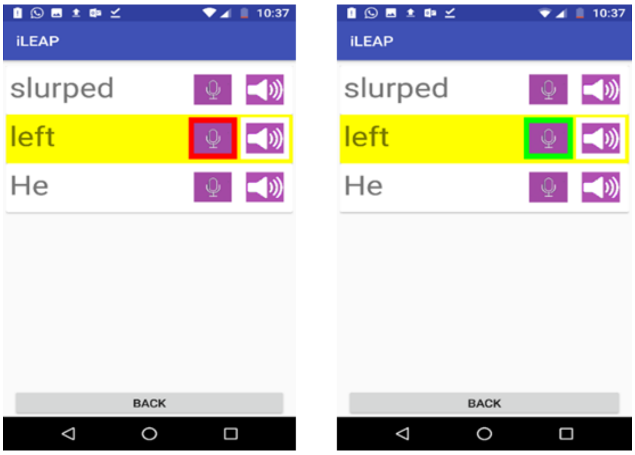


Figure 6. Correction Practice

3.5 Analysis of Retention Based on Learner's Profile Data

The backend server implements a comprehensive database to store profile data for each student. The tables retain information such as frequency count of mispronounced words, frequency count of phonemes that found to be mismatching in recognized words. The analysis results can be displayed to show the student's typical pronunciation errors at word and phoneme levels as illustrated in Figure 7. It can assist the classroom learning by providing the accurate and comprehensive list of assessment data to instructors. It is also used as evidence by iLEAP to automatically build dynamic training curriculum tailored to everyone's learning patterns and needs based on his/her typical pronunciation errors, e.g., by recommending books that have the same words or words with the same phonemes.

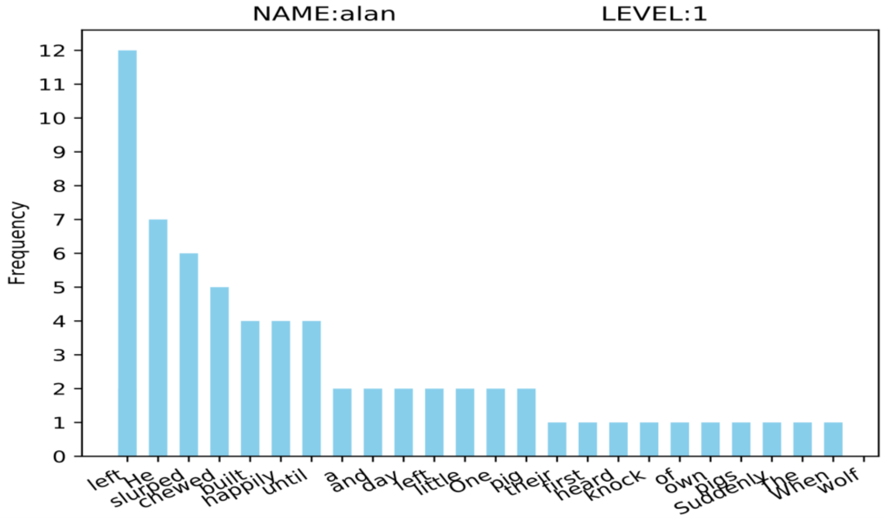


Figure 7. Performance Analysis, e.g., Frequency Count of Mispronounced Words

Furthermore, for individual word, iLEAP can also find pattern of retention, which can provide evidence that learner improved on the word over time as illustrated in Figure 8.

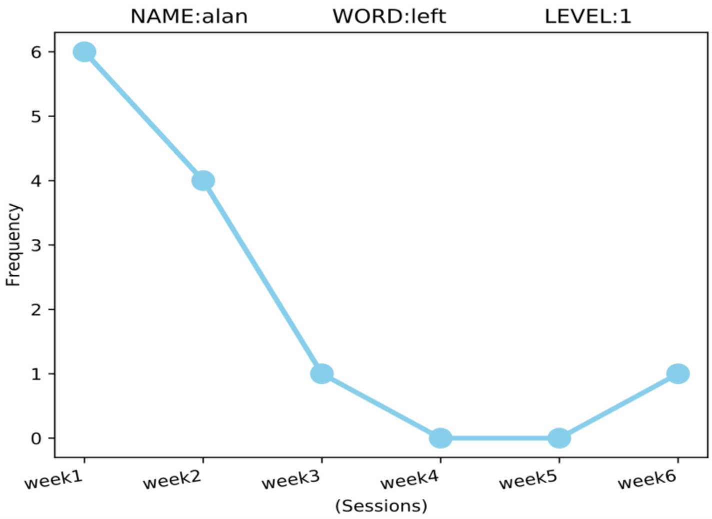


Figure 8. Performance Tracking, e.g., Retention of Corrected Pronunciation

4. CONCLUSION

The prototype iLEAP solution confirms that advanced technologies in speech recognition, AI and AR and mobile cloud computing can be leveraged to build a learning system for dual language learners. The system can provide a low cost, highly available and personalized tutoring with focus on reading and pronunciation skills of a learner who is attempting to learn English. Our experimental results demonstrate that the system is not only capable of providing immediate intelligibility assessment, but also tracking the learner's experience, which in long term can aid in improving the retention of the learning.

Even though the current system capabilities of iLEAP prototype are limited in terms of analyzing an individual's typical and atypical learning patterns, moving forward in future we could enhance backend system with No-SQL server, implement better analytics and profiling code that can generate a more detailed insight on learner's performance and trends in retention capabilities. Depending of those patterns, the system may better recommend a specific book that contains contents with a balance of learning new words and the retention of corrected words in a more engaging and supportive learning environment for young dual language learners.

REFERENCES

CMU-Sphinx project: http://www.speech.cs.cmu.edu/, https://cmusphinx.github.io/wiki/tutorial

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

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, Switzerland.

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

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.