Developing a listening English learning model using text2speech application

Majid Wajdi[1], I Nyoman Suka Sanjaya[2], I Made Sumartana[3]

Politeknik Negeri Bali[1,2,3]
Email: wajdi@pnb.ac.id[1]

Abstract - The research on the development of "Listening" practice using Text2Speech Applications aims to develop teaching of English, 'Listening' for teachers and students who are not experts in the field of information technology. This study tries to discuss how to adopt and adapt Text-to-Speech software / applications in teaching and learning of English. Text2Speech is a computer application (software) that makes it possible to convert written text into spoken text so that it can be heard. By using the Text2Speech application, written text in MS Word format can be converted into spoken text. Text2Speech allows any text to be read in a computer-synthesized voice. By utilizing Text2Speech, written text can be read and sounded so that it can be heard by the user. Text2Speech allows users to determine how to pronounce a word, phrase, sentence, and even text or parts of text. By utilizing the Text2Speech application, teachers and students get "listening" learning materials in native speakers' voices, so that students get used to hearing "listening" teaching materials from native speakers.

Keywords: how to teach listening, listening practice using Text2Speech, teaching listening in Indonesia, Text2Speech in English learning

1. Introduction

Principally, language learning, including English, both in non-English speaking countries and English-speaking countries, involves four skills that must be mastered. The four skills that are widely known in teaching and learning of English (and any other languages) are listening, speaking, reading, and writing. Among the four skills above, Listening often experiences problems in the learning of English processes, especially in non-English speaking countries such as Indonesia. Preparing, "Listening", English learning materials is the most difficult job and task for most English teachers, especially in non-English speaking countries such as Indonesia. Most English teachers are able to prepare written English learning materials, but they will have difficulty preparing "listening" learning materials for their students.

Teachers are accustomed to presenting written tests that will be given to students, but they will experience serious problems when they have to create and present unwritten teaching materials, namely "Listening" teaching materials. If teachers have to make their own "Listening" teaching materials, the instructors will experience two major obstacles. First, the teachers had to record their own voice or someone else's voice, but the drawback that emerged was that it was not a native speaker voice. Second, if teachers have to "hire" native speakers' voices, not all areas in Indonesia are easy to find native English speakers. For English teachers who live in areas other than international tourism areas where many foreigners visit, it is very difficult to realize the desire to present "Listening" teaching materials. Those two steps will definitely be burdensome for the teachers.

The development of information technology and computer software has made what was impossible in the past possible today. The two main obstacles stated above seem to be overcome by carefulness to take advantage of developments in information technology. The result of the development of information technology referred to in this study is a computer application called Text2Speech. Text-to-Speech is an application that can be run using a Macbook computer. By using the Text2Speech application, it allows written text to be converted into voice text and the resulting voice is also the voice of native speakers.

Text2Speech and Learning Arabic Tajweed. It is unavoidable that the Arabic as one of the international languages of the world is one of the most widely spoken languages in the world. In Indonesia, where the majority of the population is Muslim, Arabic is used as part of studying the Qur'an, which uses Arabic writing. However, based on BPS data in 2015, as many as 54 percent of Muslims in Indonesia are still illiterate in the Al-Quran / Arabic language. This study aims to design and compile a text to speech system that can generate spoken sounds for Arabic input text. In the process of normalizing the input text, illegal characters are eliminated and numeric characters are converted into their pronunciations. The basic pronunciation conventions are obtained using a rules-based approach. The rules applied are the rule of pronouncing Arabic alphabet of *nun* and *tanwin*, double vowels, long voice *(madd), tanwin, tasydid*, and *ta marbutah*. The identification of the rule of pronouncing *tanwin* is carried out by applying the concept of a mealy machine which is a development of the finite state automata. The sound generation process begins with creating a diphone code from a series of phonemes resulting from the text to phoneme module. Diphone codes are used to call and compose recorded diphone sound files. The resulting output is a sound file. The test was carried out by involving 13 respondents who understood the science of recitation. In testing using the legal input text reading of *idghaam bilaagunnah, idghaam bighunnah, iqlab, ikhfa 'adna, ta marbutah* 100% of respondents considered the sound produced was correct. Meanwhile, with the legal input text reading of *idzhaar,* the respondents who considered the sound produced were correct reached 92%. The lower percentage is found in the rule of double vowel reading (diphthong), *ikhfa 'ausath, tasydid, ikhfa' aqrab*, and long voice *(madd)* which get 81%, 77%, 73%, 69%, and 65%, respectively (Fauzan & Hartati, 2018).

The use of information technology in the learning process, especially in Indonesia, is not yet widely distributed. Likewise, research regarding the presence of information technology in the form of the Tex2Speech application has not been done much. Of the many searches regarding the use of Text2Speech in the learning process, there are still too few. There have only been three studies on the use of the Text2Speech application. The study found that the use of crossword puzzle media significantly improved students' Arabic vocabulary in the learning process. The results of this study suggest that teachers apply crossword puzzle learning media in learning Arabic (Fauzan & Hartati, 2018).

The study concluded that the World Englishes have not yet been taken into consideration in the present of Text-to-Speech tools for the sake of teaching and learning and in the future its present could be well considered. The uses of the Tex2Speech, based on the research, got the crucial need for a shift

in the design of such tools to get them adjusted to represent different types of English users (Karakaş, 2017). The study shows us that the English world has not considered yet the use of the Tex2Speech.

This study investigated the effectiveness use of the Text2Speech by taking the outcomes from four secondary the students with learning disabilities of oral reading fluency (ORF) and reading comprehension. The single-case A–B–A–B withdrawal design were applied by the researchers of the study in order to evaluate the effectiveness of the Text2Speech on reading outcomes. The study showed positive result in which all the participants of got their higher score in reading comprehension after the use of the Text2Speech in reading instructional passages and they were able to maintain the skills for 4 weeks. The results of the study on participants' ORF also indicated an increased level of words read per minute at the end of each accommodation condition. The results in presented in the comparison between the students' achievement in the pre- and post-test on the Lexile assessment showed that two of the four participants increased the students' reading scores. The main findings of the study were shown in the implications for the practice and recommendations for next research in order to increase the use of the Text2Speech in the classroom of English (Young et al, 2018; Yudhistiro, 2016; Manu, & Masan, 2020). Fifty percent (50%) of the participants improved in the reading comprehension scores but still needed further to ascertain the effect of the Text2Speech on the students' reading comprehension.

Text2Speech application and related read-aloud application or tools are widely implemented in an attempt to assist students' reading comprehension skills and students' listening skills especially in non-speaking countries. Read-aloud software or application, including text-to-speech application, is used to translate and convert the written text into spoken text, enabling one to listen to written text while reading along. It is not clear, so far, and the studies how effective text-to-speech is at improving reading comprehension have not been found. The study was showing the results of a meta-analysis on the effects of text-to-speech application and technology and related read-aloud application or tools on the reading comprehension of students with reading difficulties. The moderator effects of study were found to explain some of the variances. The study finally suggested that the use of the text-to-speech application and technologies may assist students in reading comprehension. However, for more definite results, more studies are importantly and really needed to further explore the moderating variables of text-to-speech application and read-aloud application and tools' effectiveness in improving reading comprehension. Implications and recommendations for future research are discussed (Wood et al, 2017; Jonathan & Suyanto, 2020; Thu & Zin, 2014). In general, it may be that the TTS cannot clearly help students in reading comprehension.

The experiment study aims to determine whether blind and visually impaired persons would accept the implementation of text-to-speech in the audio description of dubbed feature films in the Catalan context. The study was to observe sixty-seven (67) blind and partially sighted persons who assessed two synthetic voices (the Text2Speech voice) when applied to audio description, as compared to two natural voices (human voice). All the voices have been carefully and previously selected in a preliminary test. The analysis of the data, both the analysis of quantitative and qualitative, finally resulted that most participants accept Catalan text-to-speech audio description as an alternative solution to the standard human-voiced audio description. However, natural voices (not the Text2Speech voice), based on the study, got statistically higher scores than synthetic voices (the Text2Speech) and are still the preferred solution (Torne & Matamala, 2015; Arbie et al, 2013; Gelan, 2011; Hasanah & Jaroji, 2016). Natural voices from humans are easier to understand (high score) than Text2Speech voices.

This paper of the study demonstrated to convert the international language English text into speech sign. The exchange of text to speech was made by the speech synthesizer. The speech synthesis is an effort of the imitation technique of human speech. The text handling and speech generation are two main mechanisms of the Text2Speech system. In the Text2Speech system, spoken words are automatically formed from the text provided. The fundamental and vital talents of a synthesized speech are genuineness and fluency. Text to speech system finally could support in keeping and saving the information from the websites and documents in the different languages. Database formation, character recognition and text to speech conversion are the essential phases in the Text2Speech analysis (Kaladharan, 2015; Sahu, et al, 2012; Craig, 2018).

As the results of the study, the researcher through this study, presented the single Text2Speech system for the languages of Indian (Viz., Hindi, Telugu, Kannada etc.) to generate human voice or speech (text to a spoken waveform). In a Text2Speech application system, spoken utterances are automatically produced from the text provided. This paper study presented a corpus-driven Text2Speech application system based on the concatenative synthesis approach. The output generated by the proposed Text2Speech application and synthesis application system resembles natural human voice like.

Text2Speech accepts input in two forms: manual entry and from file (text or MS Word document). The proposed system supports multiple way of output; direct to computer speakers, Wav file, or MP3 file. Generated output will result and have different accent, tone based on selected languages. The study proposed Text2Speech application system will be implemented in C#.Net (Windows Form Application) and runs on Windows platforms. This paper of the study presents the examples for Hindi (North Indian) and Telugu (South Indian) languages' to elaborate proposed system. The study also elaborates what it is called inter-language text conversion (not translation). Therefore, the text in Hindi language can be converted into Telugu text and vice-versa. The research and development of this Text2Speech was done for the researcher's M. Tech major project (Sahu et al, 2012).

2. Method

Some of the steps used in this study are as follows. *First*, researchers downloaded a Text2Speech application on the App Store using a Macbook computer or laptop. The researchers open the App Store, in the "search" section, type the application they are looking for (Text2Speech). Currently (20 May 2021) there are many Text2Speech applications being offered, both paid application and free one. The researchers downloaded an application that is free of charged and has been done several years ago. *Second*, preparing English text that will be used as a sample to test the use of the Text2Speech application. The English text is typed in MS-Word format. *Third*, the researchers open the Text2Speech application and copy and paste the text that will be used.

3. Results and Discussion
3.1 Searching results of the application

The first step that the researcher do is searching and trying out the applications found. Trying out here means that which application is user friendly enough especially for the researchers who are not expert enough in using and operating technology innovations we face today. Although there a lot of applications we can find, but not all application will be simple enough for us to use in our daily life, both for official use or not formal one. The next figures is the final application the researchers finally used in the teaching and learning processes since it is user friendly and it is not complicated to use it in teaching and learning processes.
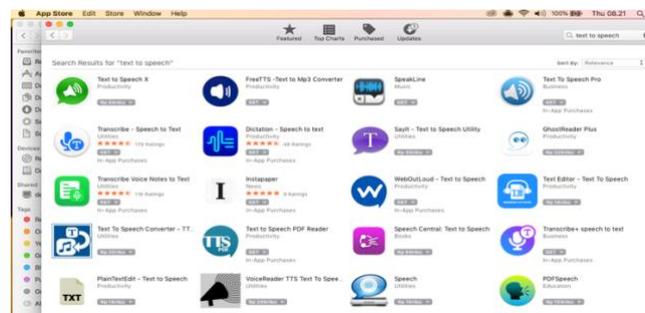

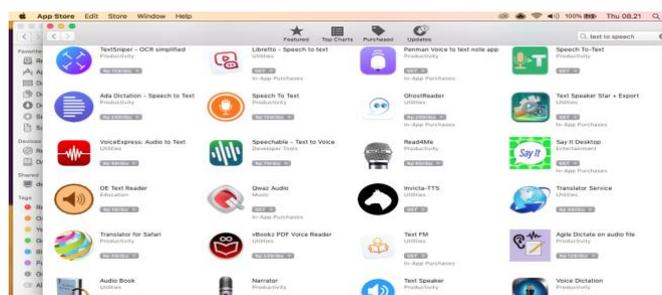Figure 1a Display of Text2Speech Application Search Results

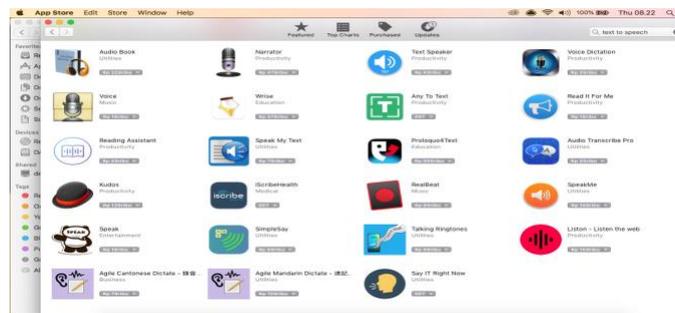Figure 1b Display of Text2Speech Application Search Results

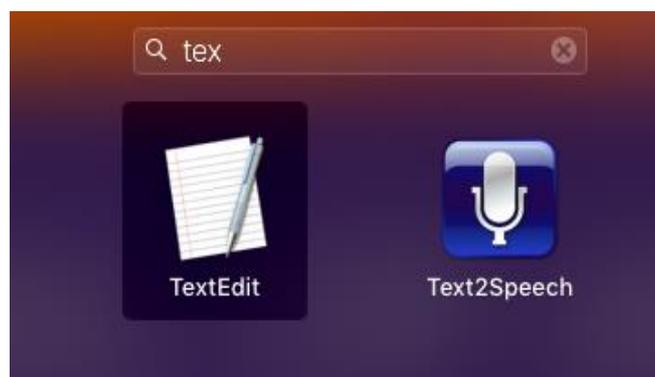Figure 1c Display of Text2Speech Application Search Results

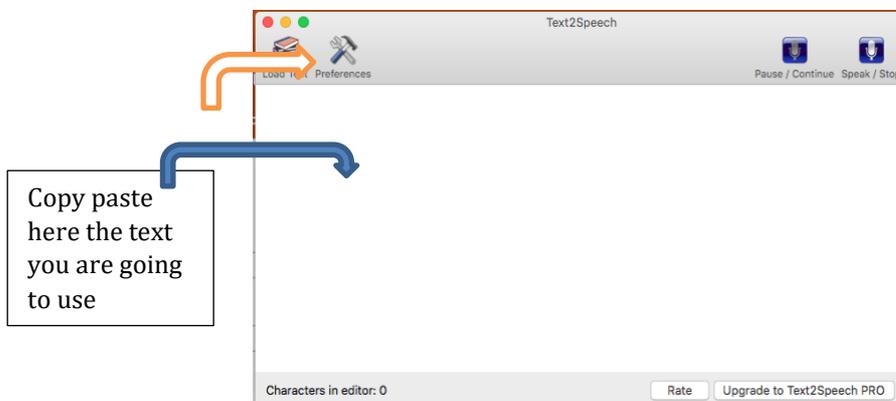Figure 1d Display of Text2Speech Application Search Results

Figure 2 TTS Screen Display on Computer

Load Text: the process of "loading the text" is by clicking the "load text" feature, we can load the text that will be used by being directed to search for files on the computer. We can also load it by copying and pasting it on the screen under 'load text'.

Preference: it is used to select the desired voice (sound) type. In the example Figure 3 shown is the voice of a woman (named: Samantha). We can test whether Samantha's voice is in line with our preference (choice of voice type) or not.
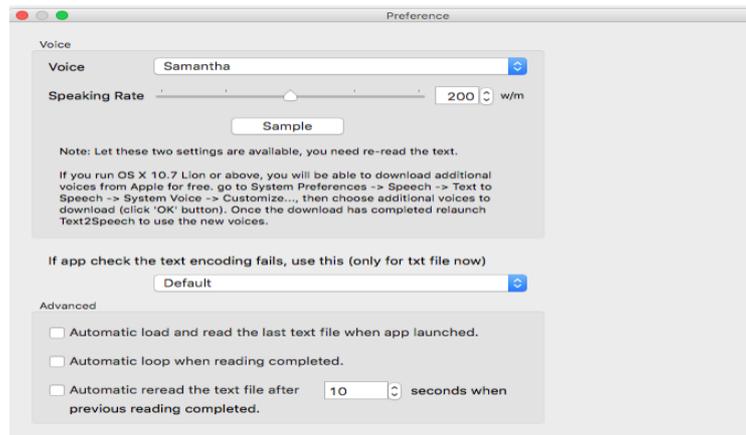
278

Figure 3 Preference (choice of voice type)

When you click on the part indicated by the arrow, several types of sound will appear that we can choose according to our wishes. There are 48 types of voices to choose from A (Alex) to Z (Zuzana).

Provided by 48 types of voice will enable the users to choose which sound or voice is relevant enough to use. We can make a decision which sound or voice is relevant and interesting enough to use. We can choose male or female voice so we can make some variations of using 'native speaker's voice.
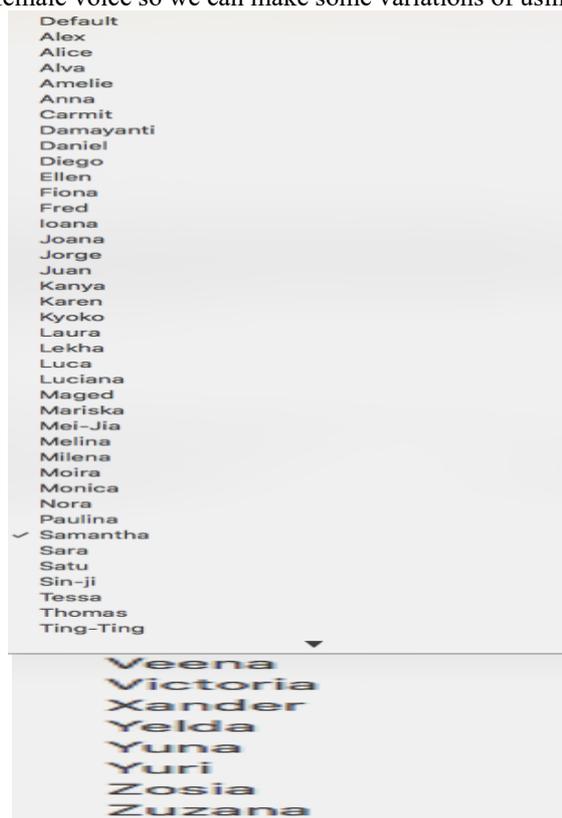
Figure 4 Choice of voice type

Pause/Continue: The voice (sound) will stop or continue when the "pause/continue" button is pressed to make sure we want to pause or to continue for example.

Speech/Stop: If the button of "speech/stop" is clicked or pressed, the sound of the voice will start sounding / or stop.
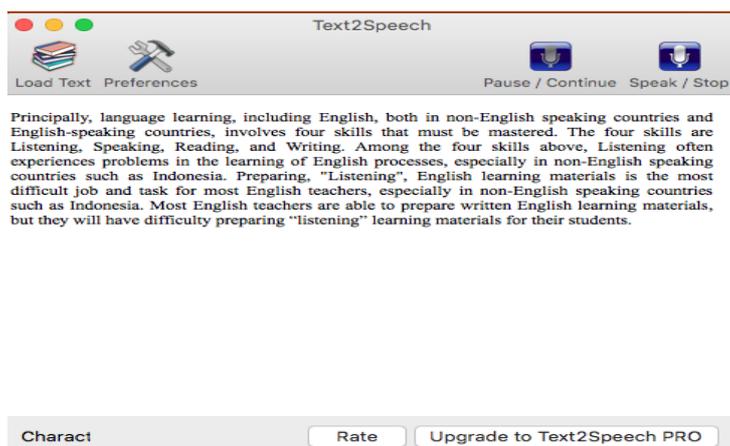
Figure 5 the sample of the text

Figure 5 is an example of the display of text that has been copied and pasted on the screen of a text-to-speech application that can produce sound according to the written text that is loaded. By clicking on the "speak/stop" button, a native speaker will be heard according to the written text that is loaded. The text can be replaced with other text according to the text that will be displayed and presented for the learning process "listening" or "dictating or dictation".

Using Text2Speech Application to present the listening practice for the students enable students to have listening practice of the native like speaker. The students also have an opportunity to practice listening anytime and anywhere since they could have listening practice and save it in their cell-phone or their smart phone. The teachers and the students have application in hand how to pronounce a certain word of English.

Speaking rate (adjusting the speed of sound): In Figure 5, click "preference", the display in Figure 6 will appear to set the speed of sound (speaking rate) presented. By shifting the sign on the speaking rate to the right or to the left, the sound speed will be adjusted according to the user's wishes. Figure 6 shows a speaking rate of 146w/m (146 words/minute).
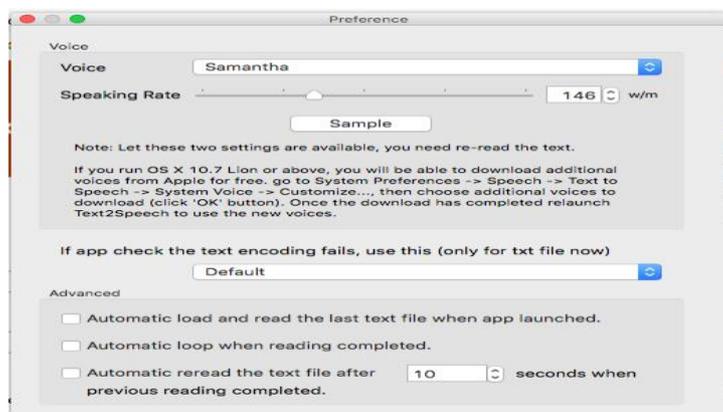

Figure 6 Speaking rate

4. Conclusion

By using the text2speech application, we can present a portable language laboratory in the grip of our portable computer or mobile phone that can be taken anywhere, making it possible for learners to learn anywhere and anytime. In the context of distance learning, listening material using the Text2speech application is ideal to use so as to facilitate students' learning as well as possible. Listening material in the text2speech application can be recorded using a computer/laptop so that it can be sent to students when needed. By using the Text2speech application, the English learning process can start from the listening stage so that the four skills that must be learned in English namely listening, speaking, reading, and writing could be easily carried out.

References

Arbie, S., Lumenta, A., Rumagit, A., & Wowor, A. (2013). Rancang Bangun Aplikasi Text to Speech Bahasa Indonesia. *Jurnal Teknik Elektro dan Komputer, 2*(3), 57-62. doi:https://doi.org/10.35793/jtek.2.3.2013.2534

Craig, SD. Schroeder, NL. 2018. Text-to-Speech Software and Learning: Investigating the Relevancy of the Voice Effect. Journal of Educational Computing Research. Available at: https://doi.org/10.1177/0735633118802877 https://journals.sagepub.com/doi/abs/10.1177/0735633118802877.

Fauzan, A., & Hartati, S. (2018). Text to Speech untuk Bahasa Arab Menggunakan Perangkaian Diphone. *JUITA: Jurnal Informatika, 6*(1), 9-14. doi:http://dx.doi.org/10.30595/juita.v6i1.1787

Gelan, A. (2011). Language and Text-to-Speech Technologies for Highly Accessible Language *(IJET), 6*(2), pp. 11-14. doi:http://dx.doi.org/10.3991/ijet.v6i2.1529.

Hasanah, U. & Jaroji, J. 2016. Aplikasi Dongeng Berbasis Text to Speech Untuk Platform Android. Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 7, Nomor 2, November 2016: 82-88. https://media.neliti.com/media/publications/53867-ID-aplikasi-dongeng-berbasis-text-to-speech.pdf

Hillaire, G., Iniesto, F. and Rienties, B. 2019. Humanising Text-to-Speech Through Emotional Expression in Online Courses. *Journal of Interactive Media in Education*, 2019(1), p.12. DOI: http://doi.org/10.5334/jime.519

Jonathan, J. & Suyanto, Y. 2020. Sintesis Suara Bernyanyi Dengan Teknologi Text-To-Speech untuk Notasi Musik Angka dan Lirik Lagu Berbahasa Indonesia. IJEIS Vol 10 No.1 available at https://doi.org/10.22146/ijeis.32131

Kaladharan, N. 2015. An English Text to Speech Conversion System. Semantic Scholar. https://www.semanticscholar.org/paper/An-English-Text-to-Speech-Conversion-System-Kaladharan/8ea77f80b06bf20e0242e633ea167aa3e7abacc5.

Karakaş, A. (2017). English voices in 'Text-to-speech tools': representation of English users and their varieties from a World Englishes perspective. *Advances in Language and Literary Studies, 8*(5), 108-119. doi:http://dx.doi.org/10.7575/aiac.alls.v.8n.5p.108

Manu, G., & Masan, P. (2020). APLIKASI TEXT TO SPEECH UNTUK MENINGKATKAN PEMBELAJARAN BAHASA INGGRIS BAGI SISWA DISABILITAS. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, *3*(2), 17-26. https://doi.org/10.37792/jukanti.v3i2.217

Sahu, L., Dhole, A., & Khadi, B. (2012). Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion. Semantic Scholar. https://www.semanticscholar.org/paper/Hindi-%26-Telugu-Text-to-Speech-Synthesis-(TTS)-and-Sahu-Dhole/c646b78896329d9126e8a536b4ac3d305b33c647.

Shetake, PS. , Patil, SA, Jadhav, PM. 2014. Review of Text To Speech Conversion Methods. International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, Aug.-2014.

Thu, CST & Zin, T. 20124. Implementation of Text to Speech Conversion. International Journal of Engineering Research & Technology (IJERT). Vol. 3 Issue 3, March 2014. https://www.ijert.org/research/implementation-of-text-to-speech-conversion- IJERTV3IS030548.pdf

Torne, AF & Matamala, A. 2015. Text-to-speech vs. human voiced audio descriptions: a reception study in films dubbed into Catalan. The Journal of Specialized Translation Issue 24 July 2015. https://www.jostrans.org/issue24/art_fernandez.pdf.

Wood, SG, Moxley, JH, Tighe, EL. 2017. Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis. Journal of Learning Disabilities. Available at: https://doi.org/10.1177/0022219416688170

Young, Mary & Courtad, Carrie Anna & Douglas, Karen & Chung, Yun-Ching. (2018). The Effects of Text-to-Speech on Reading Outcomes for Secondary Students With Learning Disabilities. Journal of Special Education Technology. https://journals.sagepub.com/doi/abs/10.1177/0162643418786047

Yudhistiro, Kukuh. (2016). PEMANFAATAN TEKNOLOGI TEXT-TO-SPEECH SEBAGAI MEDIA PEMBELAJARAN PADA LABORATORIUM BAHASA INGGRIS. Jurnal Teknologi dan Manajemen Informatika. 2. 10.26905/jtmi.v2i1.622.