

REAL TIME AUTOMATED SPEECH RECOGNITION TRANSCRIPTION AND SIGN LANGUAGE CHARACTER ANIMATION ON LEARNING MEDIA

Komang Kurniawan Widiartha¹, Ketut Agustini², I Made Tegeh³,
I Wayan Sukra Warpala⁴

^{1,2,3,4}Pascasarjana Program Studi Teknologi Pendidikan, Universitas Pendidikan Ganesha

email: komang.kurniawan@instiki.ac.id¹, ketutagustini@undiksha.ac.id², im-tegeh@undiksha.ac.id³,
wayan.sukra@undiksha.ac.id⁴

Abstract

Inclusive education for deaf students requires a technology approach to address communication and comprehension challenges. This study aims to develop innovative learning media that integrates real-time ASR (Automated Speech Recognition) transcription technology and sign language character animation to improve accessibility and comprehension of materials for deaf students. This learning media receives input from live voice, voice from learning videos, and text inputted by teachers. Using the Google Cloud API-based ASR transcription module, voice and text are converted into written text, broken down into vocabulary for sign language animation search. The search is carried out using an interpolation algorithm in the sign language animation asset database, allowing the display of animations relevant to the transcribed vocabulary.

The development process follows the ADDIE instructional design model, starting with needs analysis and ending with implementation and evaluation. The analysis stage includes data collection through teacher interviews, classroom observations, and curriculum reviews. The media design is designed to meet the specific needs of deaf students, while development and implementation focus on technology integration and effective material delivery. Evaluation is carried out to assess the effectiveness of the media in improving student understanding and participation. The study's results showed that this learning media can improve deaf students' understanding of the material and increase their involvement in the learning process. ASR technology and sign language animation contribute significantly to making learning materials more accessible and understandable, supporting the goals of inclusive education.

Keywords : Learning Media, ASR Transcription, Sign Language Animation, Deaf Education, Educational Technology, Educational Inclusion

Received: 17-09-2024 | **Revised:** 31-10-2024 | **Accepted:** 05-11-2024

DOI: <https://doi.org/10.23887/janapati.v13i3.85065>

INTRODUCTION

Special Education or Special Schools (SLB) is education for students who have difficulty following the learning process due to physical, emotional, mental, and social disorders but have the potential for intelligence and unique talents [1][2]. Special education will be appropriate only if the needs of students cannot be accommodated in the general education program, where students require the use of unique materials, equipment, services, and teaching strategies. This allows for a learning design specifically designed to meet the unique needs of students.

In general, deaf students have difficulty expressing their thoughts and desires through speech and understanding what others say. This is due to limited language proficiency, an obstacle to communication with their social

environment. One of the factors is due to limited hearing function. Teachers need to fully understand students' unique needs, especially when utilizing technology as a learning aid. Meanwhile, in the digital era, many teachers still need to be convinced to use information and communication technology in learning due to a lack of technology or supporting policies [2]. Technology in education can be a solution to overcome gaps and ensure inclusive and supportive learning for deaf students.

Verbal communication and language comprehension difficulties can lead to limited social interaction and low academic achievement [3]. Deaf students experience low self-confidence and gaps in participation in student activities at school, along with limited access to technology [4]. Difficulties in verbal communication and language comprehension

can hinder students' active participation in class discussions, make it difficult to understand lesson materials, and impact lower academic achievement. These are some of the impacts of the lack of access to and use of technology in learning.

The issuance of the Independent Learning policy by the government, which is a follow-up to the President's direction on efforts to improve the quality of human resources (HR), provides the broadest possible opportunity for educational units to design the implementation of education that allows teachers (educators) to create learning creatively, innovatively, and enjoyably so that students can enjoy a comfortable learning atmosphere in developing their potential according to the principles of independent learning [5]. Referring to Law Number 8 of 2016, every person with a disability has the right to receive quality education, which can be done through special or inclusive education [6].

By looking at the characteristics and diversity, education is not only limited to academic education but also non-academic education, including skills education and education for special programs in special schools. This makes it essential for innovation in learning media. Freedom to know also requires the participation of educators to provide the broadest possible freedom of learning to students so that they can choose learning materials. Of course, this conflicts with special deaf students who are very dependent on teachers at school who understand sign language when communicating [7]. The development of learning media with real-time ASR (Automated Speech Recognition) transcription and sign language character animation is essential to be developed as a supporter in delivering material to deaf students with sign language animation that students understand.

The aim of developing learning media with real-time ASR (Automated Speech Recognition) transcription and sign language character animation is so that educators with diverse knowledge can convey their understanding to deaf students.

STUDY LITERATURE

A. Sign Language

Sign languages have complex grammar and rules, each with vocabulary, idioms, and conventions [8][9]. Each sign language has a complex and distinct grammar system; not all sign languages are related to a specific spoken language. Sign languages are vital to the deaf

community and those interacting with them [10]. Sign languages enable people with hearing loss to communicate effectively and strengthen their cultural identity. In addition, sign languages also help people without hearing loss learn about the perspectives and lives of deaf people and raise awareness about the importance of inclusion and accessibility [11].

B. Transcription

Transcription converts text from one form to another into readable written text, primarily when written or translated from sound, speech, or other audio sources [12][13]. According to the Language Development and Fostering Agency (2016), transcription is the transfer of speech (in the form of sound) into written form, writing words, sentences, or text using sound symbols [12].

C. Automated Speech Recognition (ASR)

Automated Speech Recognition (ASR), or Speech-to-Text (STT), is a technology that converts human speech into written text [14][15]. ASR is one of the essential applications of natural language processing (NLP) and is an integral part of many speech and language-based applications [16]. The working of ASR involves a series of processing stages, including audio recording, pre-processing, feature extraction, ASR model, decoding, and correction [17]. ASR technology advances with increasing accuracy thanks to advances in deep learning and more extensive training data. ASR enables more natural interactions between humans and computers and increases efficiency in industries that rely on speech recognition [13].

ASR is based on searching for language in big data and finding matching patterns in the data to convert audio into written text [15]. However, it does not analyze the audio semantically, and its output cannot assess meaning or coherence. Unlike Natural Language Processing, which parses and analyzes language, ASR only transcribes speech and converts spoken language into written language using complex statistical models and language analysis [15]. ASR converts speech into written text using a 'speech recognition engine.' A speech recognition engine is a software system that takes the audio output produced by a user's speech (i.e., a digital audio signal from a microphone or recording) and processes the speech stream into individual sounds, vowels and consonants, lexical items and produces a written transcription of the speech [15][18].

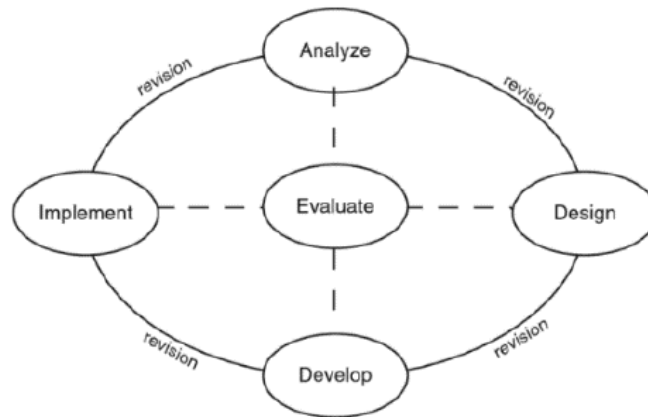


Figure 1. ADDIE Model Design Stages

D. Interpolation Search

Interpolation Search is an algorithm for searching for a given key value in an array at an index ordered by key values [19]. The search process is almost similar to searching for words in a dictionary, namely searching for data by estimating the location of the data.

METHOD

The model used in this study is the ADDIE instructional design model (Analyze, Design, Develop, Implement, and Evaluate). Branch, R. M. describes the stages of the ADDIE instructional design model as follows [20].

Analysis Phase The researcher understands the problem or educational need that will be addressed with learning media with real-

time ASR (Automated Speech Recognition) transcription and sign language character animation, which involves identifying the purpose of developing learning media, users of teaching media, and the context in which learning will take place. It begins with collecting relevant data and information, including interviews with educators, classroom observations of learning taking place at school, and analysis of the existing curriculum and materials. Furthermore, the researcher details how learning media will be built based on the findings and objectives collected during the Analysis phase. An overview of the design of learning media with real-time ASR (Automated Speech Recognition) transcription and sign language character animation is depicted in the following figure.

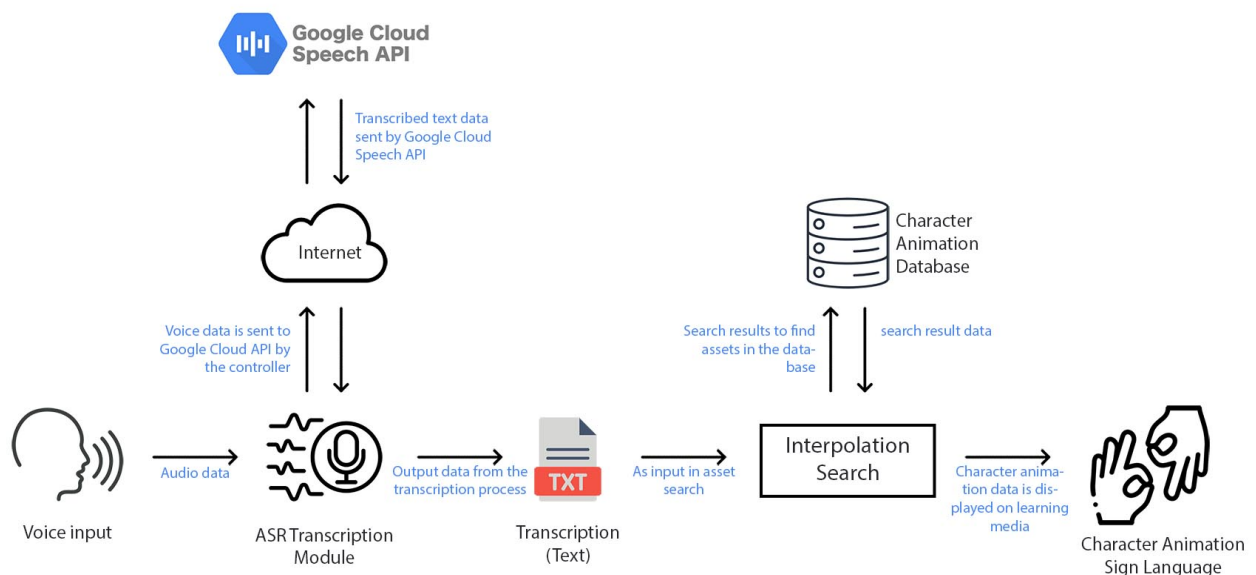


Figure 2. Overview of learning media with real-time ASR transliteration and sign language character animation

The learning media receives input from direct sound, sound from learning videos, and text inputted by the teacher as a user of the learning media. Armed with the ASR transcription module, the learning media translates the sound and text obtained using the Google Cloud API. The result is a transcribed text broken down into vocabulary for further use in searching for sign language character animation assets. The search model uses interpolation search, an algorithm for searching for critical values given in an array at an index ordered by fundamental values [19]. This search is carried out in the sign language animation asset database, which is then displayed by the learning media. From this output, deaf students understand what the teacher means when delivering the material.

In the implementation stage of the ADDIE model, the use of this media aims to ensure that the material delivered can be received by deaf students in an easy-to-understand format without communication barriers. This media is facilitated with ASR Real-Time Transcription and sign language animation, which allows teachers to teach deaf students without the need for extensive sign language skills. Learning scenarios that can show how media based on ASR Real-Time Transcription and sign language animation are used in inclusive learning at SLB. This scenario shows the interaction between teachers and deaf students in the learning process. The teacher prepares material in PWA (Progressive Web Application)-based learning media that supports ASR Real-Time Transcription and sign language animation. The teacher must ensure that the internet connection is working properly to activate the voice transcription feature.

The teacher starts the class by explaining the material with the help of a microphone. The learning media directly transcribes the teacher's voice into text that appears on the student's application screen in the form of text. Sign language animation is also automatically displayed, showing hand movements that represent the words or sentences delivered. Deaf students can use the interactive features in the

application to write down their questions that are not understood during the learning process. The teacher reads the questions from the students that appear in the application and answers them verbally, which are then transcribed back into text and accompanied by sign language animations.

The media ensures that all students can follow the discussion without obstacles, even though not all students can hear directly. This scenario shows how this learning media can bridge the communication gap between teachers and deaf students so that the learning process becomes more inclusive. In this way, deaf students can understand the material directly, and the teacher can provide instructions and explanations that students follow without language barriers.

Sign Language Animation Asset

The animation method used is to design 2D animation based on the vocabulary provided on the website pmpk.kemdikbud.go.id. The government has standardized the Sistem Isyarat Bahasa Indonesia (SIBI) Dictionary through the Minister of Education and Culture Decree No. 0161/U/2994 dated June 30, 1994, concerning the Indonesian Sign Language System Standardization. SIBI is a sign language system used in special education, and it uses excellent and correct Indonesian language rules. The development of information technology has greatly helped the world of education facilitate communication needs for people with disabilities, including people who are deaf or hard of hearing. The development of SIBI in electronic form is one of the demands of the needs in the information technology era [8].

In the Development stage, the asset design for sign language visualization with 2D animation uses a female teacher character with a classroom atmosphere. This female teacher is selected based on a particular teacher for deaf students at SLB Negeri 1 Gianyar. The following is a picture of the teacher asset used in the learning media.

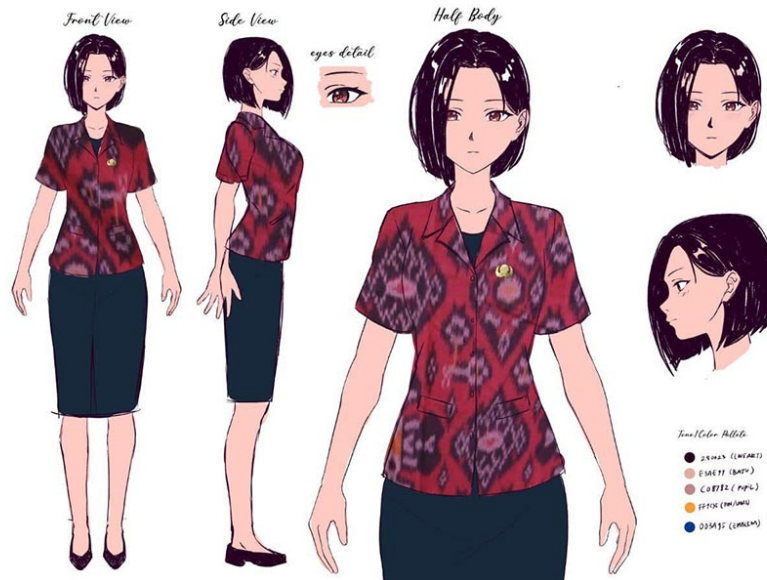


Figure 3. Teacher assets in learning media

Animation technology in learning media for deaf students has an important role because it helps them understand the material more clearly through visualization of movements that are close to native sign language. This animation functions as a bridge between verbal instructions given by the teacher and the understanding received by students, ensuring that complex content can be conveyed easily. To achieve maximum effectiveness, the animation used must be adjusted to the characteristics of the content being developed such as clear and precise movements, animation based on contextual vocabulary, and with an adjusted tempo. The animation created has hand movements and facial expressions, by the Indonesian Sign Language System (SIBI). This is important to maintain proper understanding because sign language relies heavily on the clarity of each hand movement and expression to convey meaning. The content developed will be divided into relevant vocabulary for learning available at

pmpk.kemdikbud.go.id. Each animation is created with a tempo that suits the learning speed of deaf students, allowing them to follow each hand movement without feeling rushed. The duration of the animation must be adjusted so that it is easy for students to follow, especially in explaining more complex concepts.

Learning media are developed according to their design at the design stage. The prototype development process is shown in the following figure. At the analysis stage, needs are identified, data collected, gaps are analyzed, the objectives of learning media are determined, and obstacles and supporting factors in schools are identified. At the design stage, references are collected, and how learning media is built based on findings and objectives is designed. At the development stage, learning media produces an initial prototype to be implemented. Finally, the evaluation stage is carried out, revising the learning media product to produce the final product .

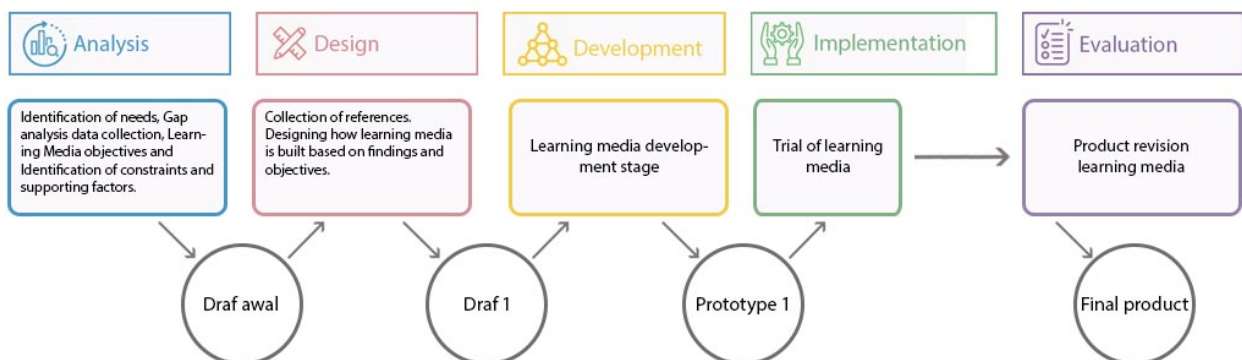


Figure 4. Development of learning media

RESULT AND DISCUSSION

We are developing a particular education application for deaf students at elementary and secondary levels, who generally use sign language as their primary language. Every learning material in Special Schools (SLB) must and can only be delivered by teachers who understand sign language. However, finding experts who meet these qualifications can take time and effort. Another alternative is to present learning videos combined with explanations in sign language. Of course, this requires particular expertise in sign language and terminology and is a significant investment. Videos also need to be updated if there are changes in content and changes in material, adding complexity and cost. Learning media generally use audio to provide instructions and feedback, which is not suitable for deaf children who are unable to hear well. To meet their needs. Sign language character animation offers a more effective solution by presenting instructions visually and flexibly. These animation assets can facilitate understanding new learning materials and provide clear instructions in learning media devices without the need for audio. To manage content in the form of videos in learning media, learning media is created with Progressive Web App (PWA) technology so that the application can optimize the website so that the learning media can provide an experience like using a mobile application.

In the main interaction of users (students) with the system, here is a description of the Use Case of deaf students who can use learning media based on ASR Real-Time Transcription and sign language animations.

Deaf students can immediately start accessing learning materials. Students can see a

list of available materials and choose the topic they want to know. When the teacher gives an explanation or instruction in verbal form, ASR will immediately transcribe the sound into text displayed on the student's screen. This transcription is displayed in real-time so that students can follow the learning flow with their hearing friends. In addition to text, the application provides sign language animations for each important instruction or keyword that is transcribed. Students can see this visualization which helps them understand the concept better. This sign language animation visualization is adjusted according to the learning material so that each instruction becomes easier to understand.

The Use-Case diagram in Figure 5 shows how the system provides easy access for deaf students to access materials, understand teacher instructions directly, and participate in interactive exercises independently. This is expected to increase student involvement in the learning process while strengthening their understanding of the material being taught.

The main interface of the learning media is designed to make it easier for users to manage learning materials and sign language animation assets. Users can quickly access a list of uploaded learning materials on the main dashboard, complete with options to add or edit materials. The material data menu allows users to manage material data by adding titles, descriptions, and video links. On the other hand, sign language animation asset management is done through Asset Data, where users can view, add, or edit animation asset data according to the relevant sign language vocabulary. All of this is summarized in the user interface shown in Figure 6.

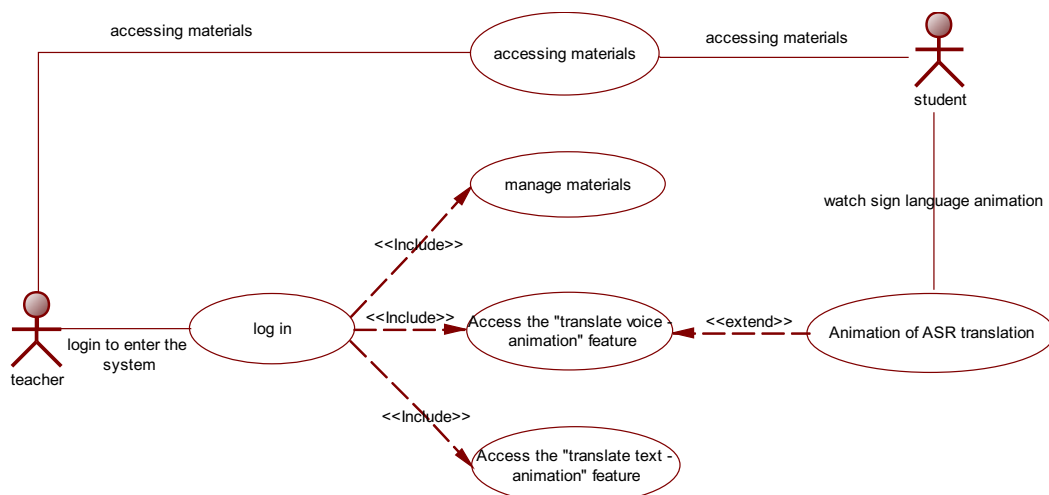


Figure 5. Use case of user interaction with learning media

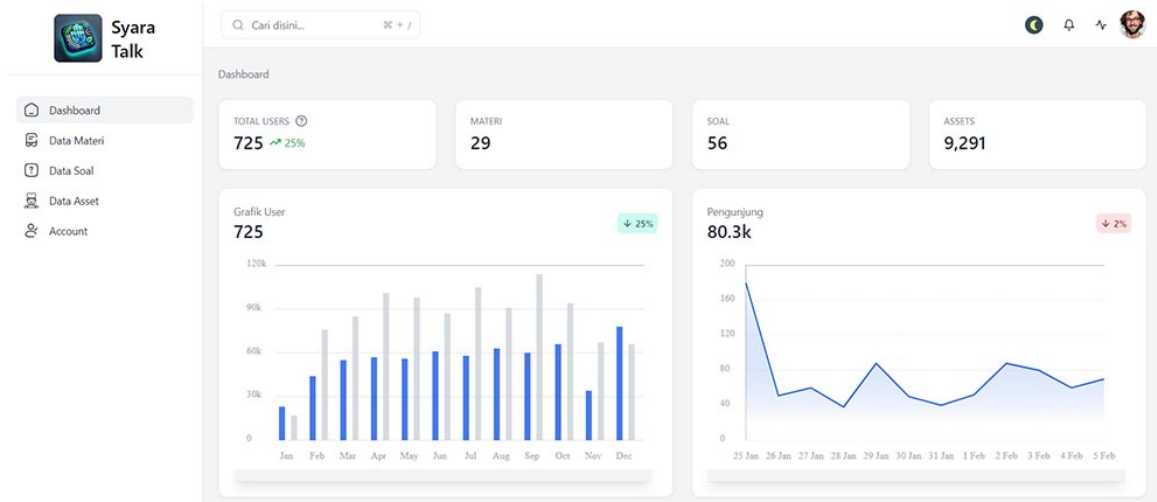


Figure 6. User interface dashboard manage learning content

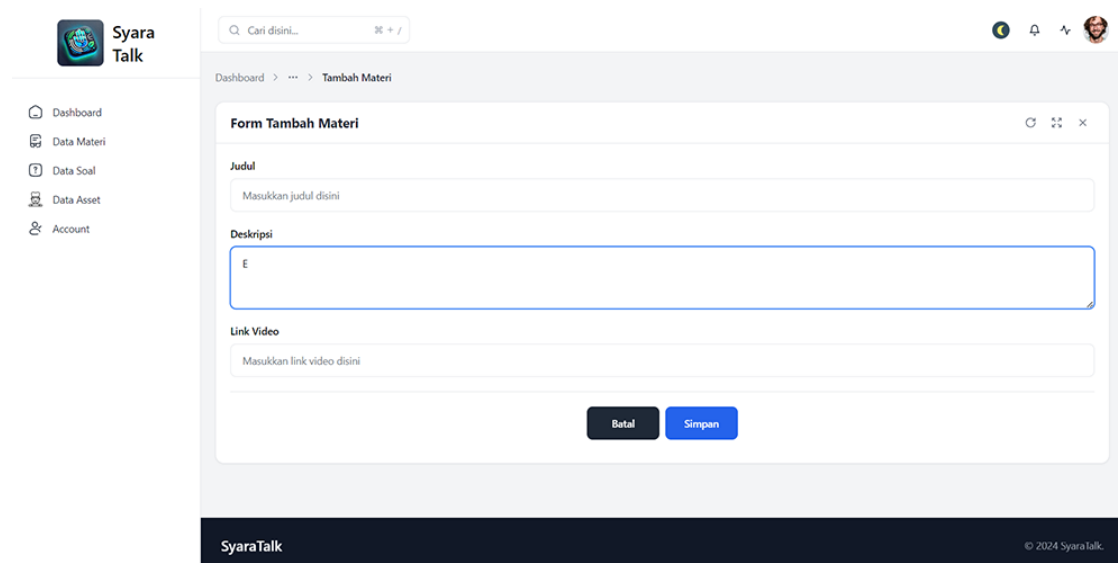


Figure 7. Manage learning materials user interface

Teachers, as users, can manage videos as learning materials in learning media. The user interface for managing learning materials has a feature that allows users to easily add video materials. Users can start the process by selecting the Material Data menu to open an input form. In this form, users can fill in the title of the material, add a description that explains the video content and embed relevant video links. After all the information is entered, users can save the material, which will be displayed in the list of learning materials that can be accessed and managed further.

In Figure 7, the feature designed to simplify the uploading and managing learning material, ensuring that all materials are

available in a format that students can quickly access. With the option to add a title and description, users can provide additional context that helps students understand the video content.

In this learning media, real-time Automatic Speech Recognition (ASR) technology is used to translate the teacher's voice directly into sign language through character animations. ASR technology allows the application to instantly identify and convert the teacher's speech into text. Once the teacher's voice is translated into text, the system will process the text to find relevant vocabulary and search for the appropriate sign language animation from the database.

Students can access this feature through the "Room" menu on the user interface. When students enter the "Room" menu, they will see the option to "Activate Sign Language Translation", which allows them to watch real-time sign language character animations according to the instructions given by the teacher. ASR technology works automatically in the background, ensuring that every instruction or explanation from the teacher is translated and displayed in sign language animations without any significant delays.

Deaf students can select materials and play videos related to the material to be presented on the main page. This feature allows deaf students to learn materials from the videos played with the help of animated characters that translate the audio owned by the video into sign language combined with transcript text as shown in Figure 8. In this way, deaf students can easily follow the lesson material visually, increasing their understanding and engagement in learning.

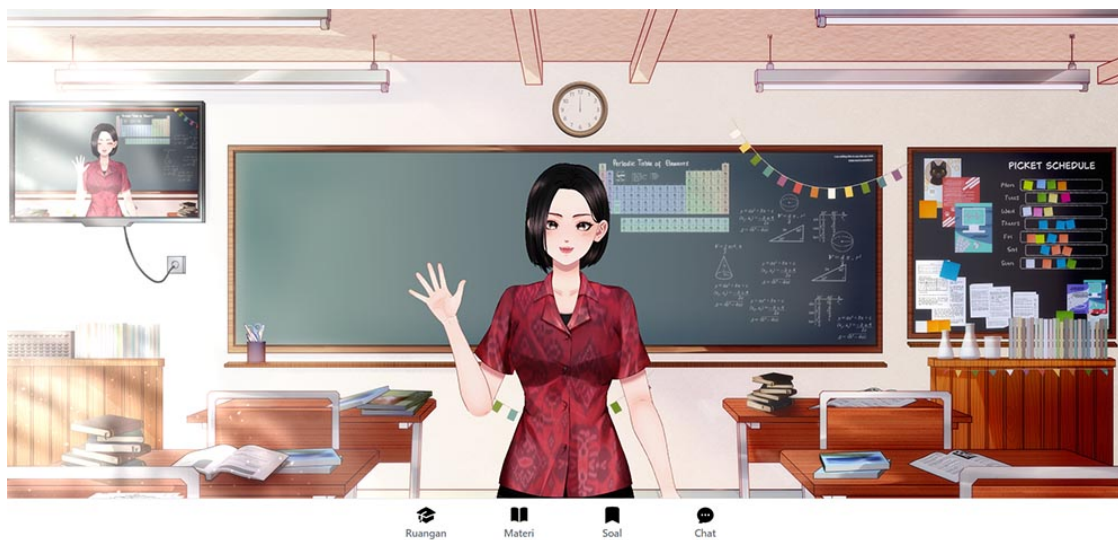


Figure 8. Frame initiation user interface



Figure 9. Learning media with sound input from learning videos

An additional feature in this learning media allows users to input text related to the subject matter, which is then used to search for and display sign language animation assets through an interpolation search algorithm. After users enter text related to the subject matter into the "Text Input" field on the application interface, the system converts the text into relevant vocabulary. An interpolation search algorithm is then used for appropriate sign language animation assets from the database. This algorithm searches for key values in an array based on the inputted text and determines

the most appropriate animation to illustrate the vocabulary.

The search results are displayed on the screen as sign language character animations, which further helps deaf students understand the material more easily. In addition to animations, the inputted text can be displayed as a transcript alongside the animation, providing additional context and ensuring better understanding. This way, the learning media integrates interpolation search technology to provide a more interactive and accessible learning experience.

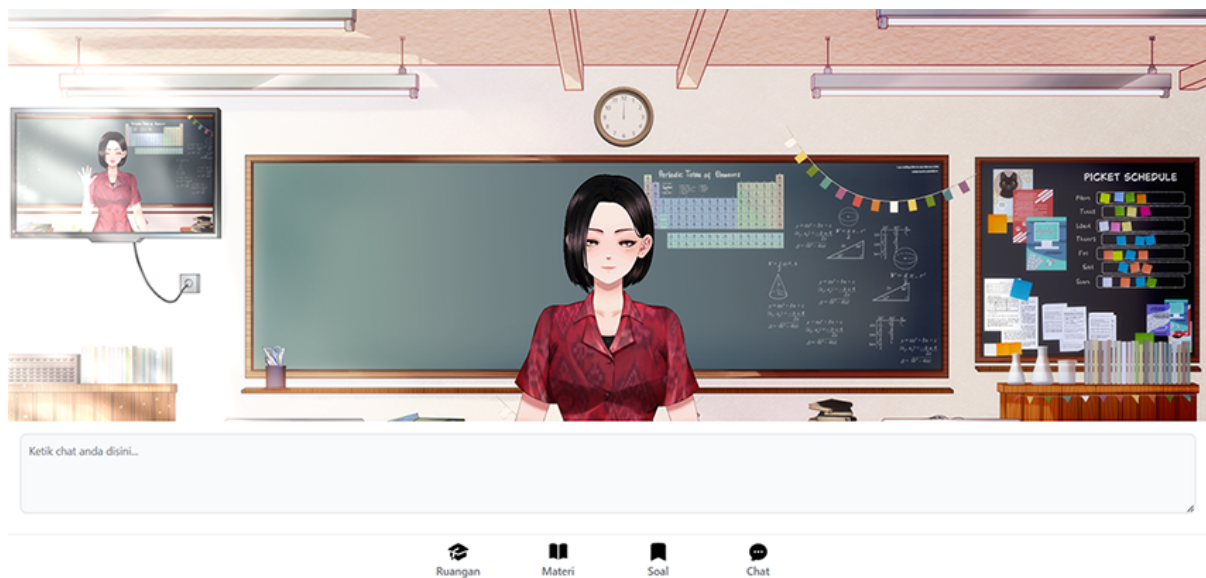


Figure 10. Learning media with voice input from text input



Figure 11. Learning media with transcription output

In Figure 11, the output generated from voice translation, audio learning materials from videos, and text input are all processed using Automated Speech Recognition (ASR) technology. This technology converts all types of input from direct voice, audio video materials, or user-entered text into text that can be further processed. After the text is generated, the system translates it into animated sign language characters displayed on the learning media screen. These sign language animations

CONCLUSION

This study shows that integrating Automated Speech Recognition (ASR) technology into learning media can significantly improve the accessibility of education for deaf students. The application can provide accurate and real-time translations in sign language animations by using ASR to translate the teacher's voice, audio material from videos, and user text input. This allows deaf students to understand the learning material more visually and intuitively, which was previously difficult to achieve using only text or pictograms.

In addition, the feature that allows users to input text and match it with sign language animation assets through an interpolation search algorithm offers a flexible and adaptive solution. This process ensures that the animations aligned with understanding and relevant material, improving the quality of learning. By providing sign language character animations along with text transcripts, the application supports students in understanding the teacher's instructions and provides additional useful context.

This study highlights the great potential of ASR technology and iconic language animations in creating inclusive learning media. Integrating these technologies demonstrates an effective way to overcome communication challenges for deaf students, making learning more accessible and understandable. With this approach, educational apps can better meet the specific needs of deaf students, support their academic success and encourage greater engagement in the educational process.

ACKNOWLEDGMENT

We express our deepest gratitude to SLB Negeri 1 Gianyar for their invaluable support and cooperation throughout this research project. Our deepest gratitude goes to my promoter, Prof. Dr. Ketut Agustini and co-promoters Prof. Dr. Made Tegeh and Dr. Wayan Sukra Warpala for their significant contributions. Their insights, feedback, and dedication have been instrumental

visually illustrate instructions or learning materials, making it easier for deaf students to understand the information. In addition to the animation, a text transcript generated from ASR is also displayed along with the animation. With the combination of sign language animation and text transcript, students can follow the learning materials better, gain a deeper understanding and better access to the information conveyed.

in shaping the development and implementation of our media learning application. We are truly grateful for their commitment and efforts, which have greatly improved the quality and impact of our work. Thank you for your invaluable contributions to advancing inclusive education and improving the learning experience for students at SLB Negeri 1 Gianyar.

REFERENCES

- [1] H. Mansur, A. H. Utama, M. H. M. Yasin, N. P. Sari, K. A. Jamaludin, and F. Pinandhita, "Development of Inclusive Education Learning Design in the Era of Society 5.0," *Soc. Sci.*, p. 496, 2023, doi: 10.4337/9781788974912.S.41.
- [2] S. Agung, A. N. Khoirunisa, and S. Suryaningsih, "Tantangan Guru Sekolah Luar Biasa Pada Pembelajaran Ipa Di Abad 21," *Alotrop*, vol. 6, no. 1, pp. 43–52, 2022, doi: 10.33369/alo.v6i1.21089.
- [3] I. Azizah, "Metode Pengajaran Anak Berkebutuhan Khusus Di Sekolah Luar Biasa (Slb)," *Pena Kreat. J. Pendidik.*, vol. 11, no. 1, pp. 42–47, 2022, doi: 10.29406/jpk.v11i1.1953.
- [4] D. A. Dewantoro, R. F. Pradipta, M. Efendi, A. Huda, and M. H. M. Yasin, "Availability and Utilization Analysis Learning Media at Inclusive Schools in Malang City," *Proc. Int. Conf. Inf. Technol. Educ. (ICITE 2021)*, vol. 609, no. 1, pp. 12–21, 2023, doi: 10.2991/assehr.k.211210.003.
- [5] V. Yunus, A. Zakso, A. T. Priyadi, and A. Hartoyo, "Pendidikan Inklusif Pada Kurikulum Merdeka," *J. Pendidik. DASAR PERKHASA J. Penelit. Pendidik. Dasar*, vol. 9, no. 2, pp. 313–327, 2023, doi: 10.31932/jpdp.v9i2.2270.
- [6] Kementrian Sekretariat Negara RI, "Undang-Undang Republik Indonesia Nomor 8 Tahun 2016 Tentang Penyandang Disabilitas," no. 243, pp. 1–5, 2016.
- [7] N. Made *et al.*, *Problems of Teachers in*

- Inclusive Class Learning in Public Elementary School Bengkulu Buleleng*, vol. 1. Atlantis Press SARL, 2023. doi: 10.2991/978-2-494069-29-9.
- [8] D. Tresnawati, R. Setiawan, N. Alawiah, D. Heryanto, and S. Rahayu, "The development of sign language learning media using multimedia systems," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032059, 2021, doi: 10.1088/1757-899x/1098/3/032059.
- [9] N. Putu, A. Pirdayanti, N. K. Anggriani, K. Jepri, and K. Jaya, "Kolok-Indonesian-English " Through Smart Kolok Application as a Media for Learning and Preserving a Natural Sign Language in Bengkulu Village. Atlantis Press SARL, 2022. doi: 10.2991/978-2-494069-29-9.
- [10] O. E. Ghoul, "Virtual reality for educating Sign Language using signing avatar: The future of creative learning for deaf students," *IEEE Global Engineering Education Conference, EDUCON*, vol. 2022. pp. 1269–1274, 2022. doi: 10.1109/EDUCON52537.2022.9766692.
- [11] D. P. Saridewi, I. W. Santyasa, I. M. Sutajaya, and N. N. Parwati, "application of the mesatua method using cartoon media in Hindu religious learning to develop student independence with disabilities," *Int. J. Health Sci. (Qassim)*, no. 3, pp. 4356–4366, 2022, doi: 10.53730/ijhs.v6ns5.10795.
- [12] L. S. Wilyanti, L. Larlen, and S. Wulandari, "Transkripsi Sastra Lisan Melayu Jambi sebagai Alternatif Bahan Ajar Sastra di Perguruan Tinggi," *J. Ilm. Univ. Batanghari Jambi*, vol. 22, no. 1, p. 247, 2022, doi: 10.33087/jiubj.v22i1.1889.
- [13] J. Zhang *et al.*, "Intelligent speech technologies for transcription, disease diagnosis, and medical equipment interactive control in smart hospitals: A review," *Comput. Biol. Med.*, vol. 153, no. January, p. 106517, 2023, doi: 10.1016/j.compbiomed.2022.106517.
- [14] M. P. Aji and U. G. Aeman, "Aplikasi Speech Recognition untuk Audio Forensik," *J. Media Pratama*, vol. 17, no. 1, pp. 61–73, 2023.
- [15] M. Carrier, "Automated speech recognition in language learning: Potential models, benefits and impact," *Training, Lang. Cult.*, vol. 1, no. 1, pp. 46–61, 2017, doi: 10.29366/2017tlc.1.1.3.
- [16] N. Sutomo, "The Implementation of Automated Speech Recognition (ASR) in ELT Classroom: A Systematic Literature Review from 2012-2023," *VELES (Voices English Lang. Educ. Soc.)*, vol. 7, no. 3, pp. 816–828, 2024, doi: 10.29408/veles.v7i3.23978.
- [17] R. Arakawa, H. Yakura, and M. Goto, "BeParrot: Efficient Interface for Transcribing Unclear Speech via Respeaking," *Int. Conf. Intell. User Interfaces, Proc. IUI*, pp. 832–840, 2022, doi: 10.1145/3490099.3511164.
- [18] X. Li, S. Yang, and H. Guo, "Application of virtual human sign language translation based on speech recognition," *Speech Commun.*, vol. 152, no. March, p. 102951, 2023, doi: 10.1016/j.specom.2023.06.001.
- [19] T. F. Doni, D. M. Midyanti, and S. Bahri, "Implementasi Metode Interpolation Search Pada Kamus Bahasa Indonesia Ke Bahasa Dayak Kalimantan Barat Berbasis Android," *J. Komput. dan Apl.*, vol. 09, no. 03, pp. 436–445, 2021.
- [20] R. M. Branch, *Approach, Instructional Design: The ADDIE*, vol. 53, no. 9. 2009.