SMATIKA : STIKI Informatika Jurnal
Accredited Sinta 4 Ministry of Higher Education, Science and Technology Republic of Indonesia SK No. 79/E/KPT/2023
E-ISSN: 2580-6939 || P-ISSN: 2087-0256

Design and Development of a Web-Based E-Survey System with Speech-to-Text Feature Using the Waterfall Method

M. Aziz Kurniawan[1*], Yumna Zahran Ramadhan[2], M. Solehuddin[3], Deki Satria[4] Sasmi Hidayatul Yulianing Tyas[5]

[1,2,3,4,5] Telkom University, Faculty of Industrial Engineering, West Minangkabau Street No. 50, RT.1/RW.1, Ps. Manggis, Setiabudi District, South Jakarta City, Special Capital Region of Jakarta 12970, Indonesia

Keyword

Accessibility; E-Survey; Speech-to-Text; SDLC; Web-Based System; WCAG 2.1

****Corresponding Author:***
mazizkurniawan@telkomuniversity.ac.id

Abstract

Web-based electronic survey systems are widely used in higher education due to their efficiency and flexibility in data collection. However, many existing survey platforms still rely on conventional text-based input and provide limited accessibility support for users with disabilities. This study aims to develop a web-based E-Survey system integrated with a speech-to-text feature to improve accessibility and support inclusive participation in digital survey activities. The system was developed using the System Development Life Cycle (SDLC) with the Waterfall method, covering requirements analysis, system design, implementation, and testing. Accessibility considerations were based on the Web Content Accessibility Guidelines (WCAG) 2.1, while speech-to-text functionality was implemented using Automatic Speech Recognition (ASR) technology. System evaluation included black box testing, WCAG-based accessibility assessment, and speech-to-text performance testing using Accuracy and Word Error Rate (WER) metrics. The results indicate that all selected accessibility criteria evaluated in this study were successfully implemented and were consistent with the selected WCAG 2.1 principles assessed through developer-based inspection. The speech-to-text feature achieved an accuracy rate of 85.71% and a Word Error Rate (WER) of 5.41%. These findings demonstrate that the proposed system provides a more accessible and inclusive survey environment by enabling respondents to complete surveys through text or voice input, while highlighting the potential of speech-to-text technology in supporting inclusive participation in higher education.

1. Introduction

The advancement of information technology has driven digital transformation across various academic activities in higher education, including the process of research data collection through web-based electronic surveys. Web-based surveys offer several advantages over conventional survey methods, such as time efficiency, ease of distribution to respondents, and the ability to reach a large number of users quickly and flexibly. In the higher education context, students represent a primary user group who frequently participate

in various academic surveys, including course evaluations and scientific research. Therefore, the development of web-based survey systems must consider usability and user experience aspects to ensure that respondents can provide answers effectively [1],[2],[3].

However, not all web-based systems have been designed with adequate consideration of digital accessibility. Web accessibility refers to the ability of a system to be used by all users, including individuals with disabilities. International standards such as the Web Content Accessibility Guidelines (WCAG) serve as key guidelines in designing inclusive web systems. Previous studies indicate that many educational websites still do not fully comply with these accessibility standards, which can create barriers for users with certain limitations in accessing digital services [4]. This issue has become increasingly critical alongside the growing adoption of digital services in the education sector. The rapid pace of digital transformation requires educational institutions to ensure that developed systems are not only technically functional but also inclusively accessible to all users. Evaluating website accessibility is therefore an essential step in ensuring that digital services can be effectively used by all users without barriers [4],[5].

In addition to implementing accessibility standards, the integration of assistive technologies can further enhance the usability of web-based systems. One rapidly evolving technology is speech recognition, or speech-to-text, which enables users to provide voice input that is automatically converted into text by the system. This technology can assist users who experience difficulties in typing or using traditional input devices, such as individuals with motor impairments or visual disabilities [6],[7],[8]. Beyond improving accessibility, the use of speech-to-text also has the potential to enhance the quality of survey data. Respondents can provide more natural responses through voice compared to typing, particularly for open-ended questions that require more detailed explanations. This approach allows respondents to express their ideas more freely, resulting in richer information. Recent studies show that the integration of speech-to-text technology in smartphone-based surveys can improve user experience and support the data collection process through automatic transcription of respondents' voice inputs [1],[9],[10],[11]. A study involving visually impaired university students found that 80% of participants experienced difficulties accessing digital documents and presentation materials [12], while limited screen reader compatibility and inadequate alternative input support remained major obstacles in online academic systems. Moreover previous studies indicate that many educational websites still do not fully comply with these accessibility standards, which can create barriers for users with certain limitations in accessing digital services [4].

Despite numerous studies discussing web accessibility and speech recognition technology, the implementation of speech-to-text in web-based electronic survey systems specifically designed to support the participation of students with disabilities remains relatively limited. Most existing survey platforms still focus on text-based input methods or predefined answer options without providing voice input alternatives that could enhance system inclusivity. This indicates that the utilization of speech recognition technology in digital surveys still requires further development to support more inclusive user interactions [1],[10],[13]. Although previous studies have discussed web accessibility, speech recognition, and electronic survey systems, limited research has specifically integrated speech-to-text technology into web-based survey platforms to support accessibility and inclusive participation for university students with disabilities and this is the comparison of the limitation of this study.

Table 1. Existing Studies and Limitation of This Study

Existing Studies	Limitation	This Study
Accessibility evaluation only	No assistive input feature	Adds speech-to-text
Speech recognition in mobile apps	Not focused on survey systems	Applied in web survey
General survey platforms	Mostly text-based input	Inclusive voice input
Accessibility studies	Not targeting higher education students with disabilities	Focused on university context

Based on these issues, this study aims to develop a web-based electronic survey system that integrates speech-to-text technology to improve accessibility for students, particularly those with disabilities. The system development process adopts the System Development Life Cycle (SDLC) approach, ensuring that the stages of requirements analysis, system design, implementation, and testing are conducted systematically.

2. Research Method

This study employs the System Development Life Cycle (SDLC) using the Waterfall method to design and develop a web-based electronic survey system equipped with a speech-to-text feature to enhance accessibility for students, including those with disabilities. The Waterfall method was selected because it provides a systematic and sequential approach to software development, allowing each phase to be completed before proceeding to the next stage. This model is suitable for information system development projects with clearly defined requirements and structured development processes. The stage of this method such as requirements analysis, system design, implementation, testing, and evaluation. This approach is widely adopted in information system development as it ensures that each phase is conducted in a planned and controlled manner, resulting in a system that aligns with user requirements [11].

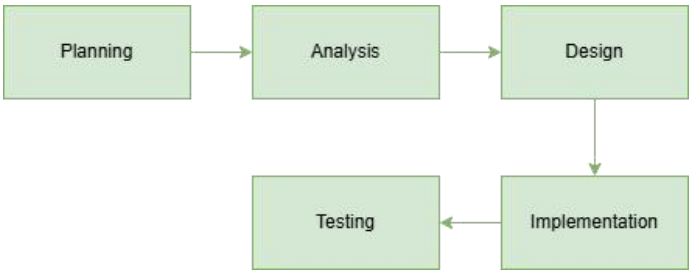

Figure 1. SDLC Method

During the requirements analysis phase, user requirements and system specifications were identified through a literature review and observations of commonly used web-based survey platforms, such as Google Forms and Microsoft Forms. The literature review examined previous studies related to electronic survey systems, speech-to-text technology [1],[3],[5] web accessibility [4],[14] and the implementation of the Web Content Accessibility Guidelines (WCAG) [4],[14]. In addition, observations were conducted to identify survey completion mechanisms, navigation structures, user input methods, and accessibility features available in digital survey platforms.

The requirements analysis process was also informed by empirical findings from recent studies indicating that accessibility remains a significant challenge in higher education digital environments. Liu et al. reported 36,283 accessibility issues across 993 web pages from 250 university websites, with most violations related to alternative text, link identification, and color contrast [15]. Fakrudeen found that many university websites still experience accessibility and usability issues that affect navigation and user [16]. Other studies have also shown that digital educational platforms continue to face accessibility barriers related to content perception, navigation structures, and user interaction mechanisms [17],[18]. These findings suggest that digital systems in higher education still require more inclusive design approaches to accommodate diverse user needs.

Observations of existing survey platforms revealed that most systems primarily rely on visual interaction and keyboard-based input, which may create barriers for users with specific accessibility needs. Therefore, the functional requirements of the proposed system were formulated by incorporating audio-assisted questions to support auditory information delivery and speech-to-text functionality as an alternative input method that enables users to provide responses through voice interaction [1],[2],[9]. Furthermore, the user interface was designed with simple navigation and clear interaction elements in accordance with web accessibility principles [4],[14].

Meanwhile, the non-functional requirements were defined based on the WCAG 2.1 principles of Perceivable, Operable, Understandable, and Robust to ensure that the system can be used more inclusively by a wide range of users [4],[14],[19] The outcomes of this requirements analysis phase served as the foundation for the subsequent system design and development process.

At the system design stage, the system architecture, database structure, and user interface were designed. The system architecture describes the interaction between users (*students* and *administrators*), the web application, the ASR service, and the database system [20]. The database design focused on structuring data related to users, surveys, questions, responses, and survey results. In addition, the user interface was designed by considering accessibility principles based on WCAG 2.1 to ensure that the system could be used effectively by all users, including students with disabilities. Several accessibility-oriented design implementations include readable text and button sizes, simple and consistent navigation, alternative input support through speech, sufficiently large interactive components, responsive layouts compatible with multiple devices.

The next stage is system implementation, in which the survey system is developed as a web-based application that enables respondents to provide responses through both text input and voice input. The voice input is automatically converted into text using Automatic Speech Recognition (ASR) technology, which allows computers to recognize and transform human speech into digital text. The integration of ASR technology in digital systems can enhance accessibility and facilitate user interaction through voice-based input [21].

After the system has been developed, system testing is conducted using functional testing methods to ensure that each feature operates in accordance with the defined requirements. This testing phase aims to verify that all system functions perform correctly and that there are no errors in the input or output processes. The testing phase was conducted using the *Black Box Testing* method to verify whether all system functionalities operated according to the predefined requirements. This testing method focuses on evaluating system inputs and outputs without examining the internal source code. In addition to functional testing, accessibility testing based on WCAG 2.1 principles was also conducted. The evaluation focused on aspects such as navigation consistency, text readability, ease of interaction, and usability of the *speech-to-text* feature. The system was considered successful when all functionalities operated as expected without significant errors. This final stage also aims to assess system performance and ensure that the developed electronic survey application can be used effectively and supports student participation in completing digital surveys [22]. In addition to functional testing, accessibility evaluation based on WCAG 2.1 principles was conducted. The evaluation focused on navigation consistency, text readability, ease of interaction, and accessibility support provided by the speech-to-text feature. Furthermore, the speech-to-text functionality was evaluated using Accuracy and Word Error Rate (WER) metrics to measure the performance of the speech recognition process.

3. Result and Discussions

Study developed a web-based electronic survey system integrated with a speech-to-text feature that allows respondents to complete surveys using either voice or text input. The proposed system architecture for the web-based E-Survey application with a *speech-to-text* feature consists of four main components: *Users*, *Web Application*, *Database*, and *External Services*. The architecture is designed to support accessible survey participation for students while enabling administrators to manage survey results efficiently. This study presents a web-based electronic survey system integrated with a speech-to-text feature that enables respondents to complete survey forms using either voice or text input. The integration of speech recognition in survey systems has been widely explored in prior studies, particularly in the context of improving accessibility, usability, and interaction efficiency [2],[6]. In line with these developments, the proposed system adopts a similar direction by incorporating external speech-to-text services to support more inclusive input modalities. Previous research on Automatic Speech Recognition (ASR) and speech-to-text systems has consistently shown that system performance is highly influenced by vocabulary constraints, speech structure, and environmental conditions during evaluation [3],[21]. Studies further indicate that limited-vocabulary and structured-input scenarios tend to produce higher recognition accuracy compared to open-domain speech,

where variability significantly affects transcription quality. In addition, prior benchmarking studies on speech-to-text systems highlight that performance evaluation should not rely solely on accuracy, but also incorporate error-based metrics such as Word Error Rate (WER) to provide a more comprehensive assessment of transcription quality [10],[23]. This approach is particularly relevant in assistive and educational applications, where controlled conditions are often used to ensure usability while maintaining system reliability. Comparative studies also report that speech-to-text systems generally achieve higher performance under structured and constrained conditions, although their effectiveness decreases in more complex real-world environments [1].

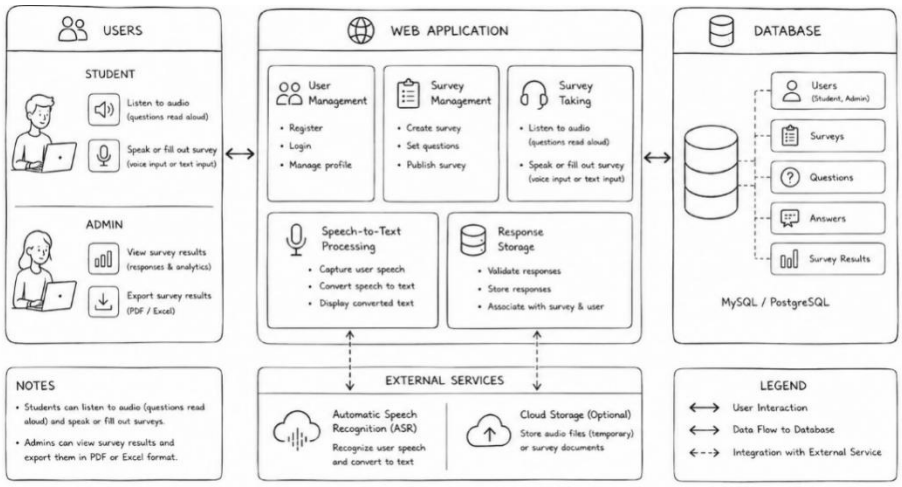

Figure 2. System Architecture

The *Users* component consists of two actors, namely students and administrators. Students act as respondents who can listen to audio-based survey questions and provide responses either by speaking or filling out surveys manually through text input. Meanwhile, administrators are responsible for viewing survey results and exporting survey reports for further analysis and documentation.

The *Web Application* component serves as the core system that manages all application processes. This component includes modules for user management, survey management, survey participation, speech-to-text processing, response storage, and reporting. Through this application, students can interact with the survey system using either voice or text input. The speech-to-text processing module captures user speech and converts it into text automatically using *Automatic Speech Recognition* (ASR) technology before storing the responses into the database.

The *Database* component functions as the central storage system that stores user data, survey data, survey questions, answers, and survey results. The database management system used in the architecture is relational-based, such as *MySQL* or *PostgreSQL*, to ensure structured and consistent data management.

In addition, the system integrates with *External Services*, particularly the ASR service, to support real-time speech-to-text conversion. An optional cloud storage service may also be utilized to temporarily store audio files or survey-related documents. Overall, the proposed architecture supports accessibility, usability, and efficient survey data processing within the web-based E-Survey system.

The system was developed using the Waterfall approach through the stages of analysis, design, implementation, and testing. The results show that the system functioned properly and supported more flexible survey interactions. The speech-to-text feature helped users answer descriptive questions more easily and reduced reliance on manual typing, which may support accessibility for users with motor or visual limitations. In addition, the voice input feature supports WCAG accessibility principles by providing

alternative interaction methods for diverse users. Based on black box testing, all main system functions operated as expected. Nevertheless, speech recognition performance may still be affected by factors such as background noise, internet connection quality, and differences in user pronunciation. Further studies are recommended to conduct broader usability evaluations involving more diverse user groups, including individuals with disabilities. The implementation results indicate that the developed system was able to function properly and support both text-based and voice-based survey interactions. The integration of the speech-to-text feature improved the efficiency of survey completion, particularly for descriptive questions, as respondents were able to provide answers more naturally through voice input compared to conventional typing. From an accessibility perspective, the system provides an alternative interaction method that reduces dependence on manual typing, thereby supporting users with motor or visual limitations in completing survey forms more independently. In addition, the integration of voice input aligns with inclusive interaction principles emphasized in WCAG by providing multiple methods of user input and improving system operability for diverse users. The black box testing results also demonstrate that all main system functionalities operated successfully according to the expected outcomes, indicating that the system is capable of supporting an effective and accessible web-based survey process. However, the performance of the speech-to-text feature may still be influenced by environmental noise, internet stability, and variations in user pronunciation. Therefore, future studies may further evaluate the system using usability testing and larger participant groups, including users with different types of disabilities, to obtain more comprehensive accessibility evaluation results.

The system interface was designed by considering basic WCAG accessibility principles, particularly in supporting alternative interaction methods for users with special needs [14]. Audio-assisted questions support the perceivable aspect, while speech-to-text input supports operable and understandable interaction by enabling users to answer surveys using voice input instead of manual typing. In addition, the interface uses a simple layout and clear navigation buttons to improve usability and accessibility during the survey process.

Table 2. WCAG Principle of the E-Survey System

WCAG Principle	Implementation in the E-Survey System	Description
Perceivable	Audio-assisted questions	The system provides audio playback features that allow users to listen to survey questions, helping users who experience difficulties in reading text visually.
Operable	Speech-to-text input	Users can answer survey questions using voice input instead of manual typing, supporting users with motor limitations or typing difficulties.
Operable	Clear navigation buttons	Navigation buttons such as “Next,” “Previous,” and “Cancel Survey” are provided to create a more structured and consistent interaction flow during survey completion.
Understandable	Simple and centralized interface layout	The interface uses a simple design with minimal complexity to improve ease of use and reduce user confusion during interaction.
Understandable	Integrated text and voice interaction	The system combines text-based and voice-based interaction methods so users can choose the most comfortable way to complete surveys.
Robust	Web-based implementation with ASR integration	The system integrates Automatic Speech Recognition (ASR) technology within a web-based platform to support alternative interaction methods for users with special needs.
Accessibility Support	Inclusive survey interaction	The integration of audio playback and speech-to-text features supports a more inclusive digital survey environment for students with disabilities.

Accessibility evaluation was conducted based on the four principles of WCAG 2.1, namely Perceivable, Operable, Understandable, and Robust. The evaluation focused on the accessibility features implemented in the developed E-Survey system, including audio-assisted questions, speech-to-text input, navigation consistency, interface readability, and browser compatibility.

Table 3. WCAG 2.1 Accessibility Evaluation Results

WCAG 2.1 Principle	Evaluation Criteria	Implementation Result	Status
Perceivable	Audio alternative for survey questions	Survey questions can be played through audio assistance to support users with reading difficulties	Pass
Perceivable	Readable text and interface elements	Text size and interface components are clearly visible and readable on desktop and mobile devices	Pass
Operable	Keyboard-accessible navigation	Navigation buttons (Next, Previous, Cancel Survey) can be accessed consistently during survey completion	Pass
Operable	Alternative input method	Speech-to-text feature allows respondents to answer questions using voice input instead of typing	Pass
Understandable	Consistent navigation structure	Navigation flow remains consistent across all survey pages	Pass
Understandable	Clear instructions and interaction flow	Audio guidance and visual labels help users understand survey completion procedures	Pass
Robust	Browser compatibility	System functions correctly on modern web browsers used during testing	Pass
Robust	Assistive technology support	Speech-to-text integration provides additional accessibility support for users with motor limitations	Pass

The results indicate that all evaluated accessibility criteria were successfully implemented. The system provides alternative methods for both information delivery and user input through audio playback and speech-to-text functionality. Navigation elements remained consistent throughout the survey process, while the interface design supported readability and ease of interaction. Furthermore, the web-based implementation operated correctly across the tested browser environment and provided additional accessibility support for users with motor limitations. These findings demonstrate that the selected accessibility criteria evaluated in this study were successfully implemented and aligned with the selected WCAG 2.1 principles assessed through developer-based inspection and functional accessibility verification. However, this evaluation does not represent full WCAG 2.1 compliance, as it was limited to selected criteria and did not involve a comprehensive accessibility audit.

These results stand in clear contrast to accessibility evaluations reported for general higher-education websites. A large-scale automated assessment of university web pages identified tens of thousands of accessibility violations concentrated in alternative text, link identification, and color contrast, with the majority of pages failing to fully meet WCAG Level A requirements [15]. Similar barriers, including inconsistent navigation and limited usability support, have also been reported across other university website evaluations [16]. The contrast suggests that the accessibility outcomes achieved here are not simply a function of testing leniency, but rather of incorporating accessibility principles directly into the design phase rather than auditing them after the fact. Where prior studies evaluated systems that were not originally designed with WCAG in mind, the present system embedded Perceivable, Operable, Understandable, and Robust considerations from the outset, which likely explains why all evaluated criteria passed without the systemic gaps reported elsewhere. This finding reinforces the broader argument in accessibility research that compliance is more reliably achieved through accessibility-by-design than through retrofitting, although the comparison should be read with caution given the difference in evaluation scope: the cited studies applied automated, large-sample audits across entire institutional websites, while the present evaluation involved developer-based inspection of a single system limited to selected criteria.

During the system design phase, modeling was conducted using Unified Modeling Language (UML) to represent the system structure and interactions. The models employed in this study include the Use Case Diagram and the Component Diagram [24].

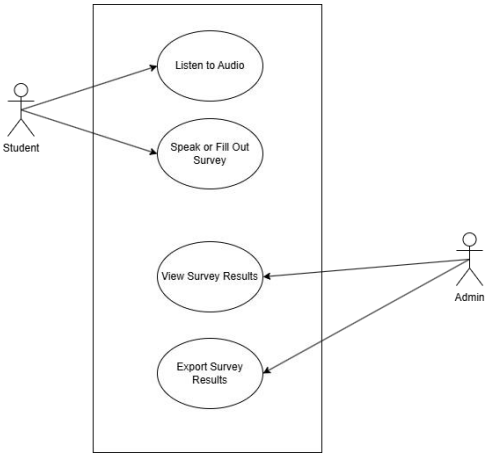

Figure 3. Use Case Diagram Sistem E-Survey

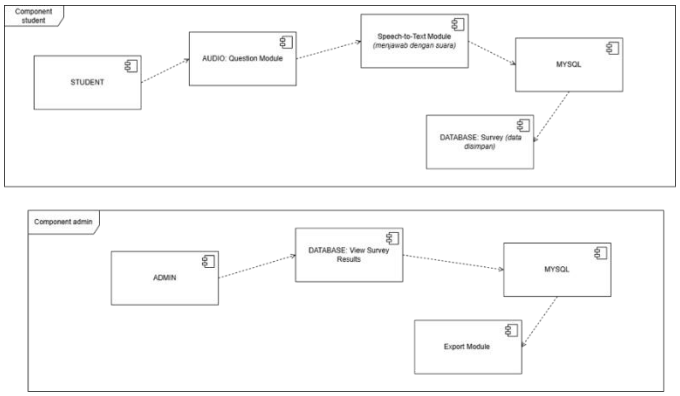

Figure 4. Component diagram Sistem E-Survey

Figure 3 illustrates the interactions between users and the system. The system involves two primary actors, namely the student as the survey respondent and the admin as the system manager. The student is able to listen to survey questions, respond to questions using either text or voice input, and submit their responses to the system. Meanwhile, the admin has access privileges to view survey results and export survey data for analysis purposes. The figure 4 above illustrates the component diagram of the web-based e-survey system integrated with a speech-to-text feature. On the student side, users listen to survey questions through the Audio Question Module and provide responses using voice input. The voice responses are processed by the Speech-to-Text Module, which automatically converts spoken input into text. The converted responses are then stored in the survey database using MySQL. On the admin side, administrators are able to view survey responses that have been stored in the database. In addition, the system provides an Export Module that enables administrators to export survey results into Excel and PDF formats.

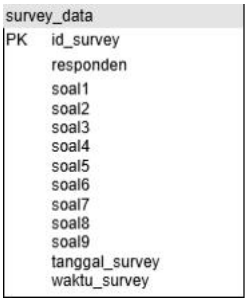

Figure 5. E-Survey System Database

The survey_data table is used to store all survey responses submitted by respondents. This table contains a primary key, namely id_survey, which serves as the unique identifier for each survey record. The responden field stores respondent identity information, while soal1 to soal9 contain answers to each survey question, which may originate from either text input or voice input converted through the speech-to-text feature. In addition, the table includes tanggal_survey and waktu_survey fields that are used to record the date and time when respondents completed the survey.

User Interface (UI) is an important component in system development that functions as the interaction medium between users and the developed system. In this study, entitled Design and Development of a Web-

Based E-Survey System with Speech-to-Text Feature Using the Waterfall Method, the UI was designed to ensure ease of interaction, clarity of information, and user convenience in operating the system. The interface design was developed by considering usability and accessibility aspects, particularly in supporting the speech-to-text feature, so that the system can be effectively used by various types of users, including users with special needs [9].

Figure 6. Survey Home Page

The main page serves as the initial interface accessed by users when opening the E-SKM Disabilitas system. This page displays the application title along with the institutional identity. Users are provided with two main buttons, namely “Survey Results” to view the survey data recap and “Start Survey” to begin completing the survey. The interface is designed with a simple and centralized layout to facilitate easy initial navigation for users.

Figure 7. First Input Page

This page is used to input the respondent’s initial data, such as their name. The system provides an audio playback feature as a question guide, as well as a speech-to-text feature that allows users to respond using voice input. The converted voice input is displayed as text in the input field. In addition, the page includes language selection options and navigation buttons such as “Cancel Survey” and “Next”.

Figure 8. Main Survey Page

On this page, survey questions are displayed for respondents to answer. Each question is accompanied by rating scale options and audio playback features to assist user understanding. Respondents can provide answers using voice input, which is then converted into text by the system. Navigation buttons such as "Previous" and "Next" are also available to allow users to move between questions, making the survey completion process more structured and organized.

Figure 9. Administrator Authentication Page for Accessing Survey Results

This figure shows the administrator authentication page of the web-based E-Survey system. Before accessing survey results, administrators are required to enter a password to ensure that only authorized users can view and manage respondent data. This mechanism was implemented to support data privacy and improve the security of survey response information within the system.

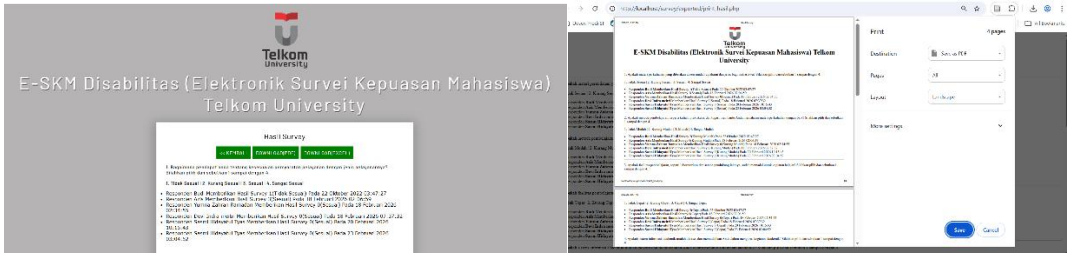

Figure 10. View Survey Page

This page is used by the administrator to view the collected survey results. The data are displayed in the form of a response list from each respondent. In addition, the system provides features to download survey results in PDF and Excel formats, thereby facilitating the documentation process and further data analysis.

The system was implemented as a web-based application running on the Apache web server and MySQL database provided in the XAMPP Control Panel version 3.3.0 package. XAMPP was used as a local development environment to run the web application and manage the system database. System testing was conducted using the black box testing method, a testing technique that focuses on system functionality without examining the internal structure of the program code, where validation is performed based on the conformity between the input and the resulting output [25].

The testing process was carried out on a device with the following specifications: Windows 10 Pro 64-bit operating system (Build 19045), AMD Ryzen 5 3500U with Radeon Vega Mobile Graphics processor (8 CPUs) running at approximately 2.1 GHz, 8 GB RAM, and DirectX 12 support. The device was an Acer laptop used as the testing environment to ensure that the system operated properly under real-world conditions.

Table 4. Black Box Testing of the E-Survey System

Test Scenario	Test Procedure	Expected Result	Status
View Main Page	The user opens the main page of the e-survey system through a browser	The main page of the system is successfully displayed	Valid
Listen to Audio	The user presses the button to listen to the survey question	The system plays the survey question audio	Valid
Speech-to-Text	The user answers the question using voice input	The system captures the voice input and converts it into text	Valid
Save Survey Response	The user submits the survey responses	The system stores the responses in the database	Valid
View Survey Results	The admin opens the survey results page	The system displays survey result data from the database	Valid
Export to PDF	The admin clicks the export PDF button	The system successfully downloads the survey results in PDF format	Valid
Export to XLS	The admin clicks the export XLS button	The system successfully downloads the survey results in XLS format	Valid

This outcome is consistent with findings from other web-based system development efforts that adopted the Waterfall-based SDLC approach. A black box evaluation of an outcome-based education assessment system developed with the same method similarly reported that the system ran properly and produced valid results across all tested cases [25]. The recurrence of this pattern across functionally different systems suggests that the structured, sequential nature of the Waterfall model, in which each phase is completed and verified before the next begins, tends to produce systems with few functional defects by the time they reach testing, provided the requirements analysis phase was conducted thoroughly. In the present study, the seven tested scenarios spanned both core survey functions and the newly introduced speech-to-text input, and the absence of failures across this range indicates that the addition of a more complex, externally dependent feature did not compromise the reliability of the system's existing functions.

The speech-to-text testing was conducted by comparing the transcription results generated by the system with the expected answers. The results were considered correct if the system successfully recognized the spoken words or the intended meaning accurately. The accuracy level was calculated using the following formula:

$$Accuracy = \frac{Number\ of\ Correct\ Data}{Total\ Number\ of\ Data} \times 100\% \qquad (1)$$

Description:
Number of Correct Data = total correctly recognized transcriptions
Total Number of Data = total testing data

The use of accuracy and Word Error Rate (WER) approaches is commonly applied in Automatic Speech Recognition (ASR) or speech-to-text evaluation studies. The evaluation was conducted using 14 speech samples, consisting of 10 proper names and 4 structured phrases. These samples were used to assess the performance of the speech-to-text system under a controlled environment.

Table 5. Speech-to-Text Testing Results

No	Spoken Data	Speech-to-Text Result	Status
1	Arif Rahman Hakim	Arif Rahman Hakim	Correct
2	EEP Kurniawan	EF Kurniawan	Incorrect
3	Devi Meytri	Devi Matrix	Incorrect
4	Ridwan Ariana	Ridwan Ariana	Correct
5	Yusep Sukmana	Yusep Sukmana	Correct

6	Robbihi Awaludin	Robbihi Awaludin	Correct
7	Deki Satria	Deki Satria	Correct
8	Yumna Zahran	Yumna Zahran	Correct
9	Muhammad Solehudin	Muhammad Solehudin	Correct
10	Sasmi Hidayatul	Sasmi Hidayatul	Correct
11	Satu Tidak Sesuai	Satu Tidak Sesuai	Correct
12	Dua Kurang Sesuai	2 Kurang Sesuai	Correct*
13	Tiga Sesuai	Tiga Sesuai	Correct*
14	Empat Sangat Sesuai	4 Sangat Sesuai	Correct*

Note: Differences in numerical and textual representations (e.g., "dua" and "2", "empat" and "4"), as well as capitalization differences, were considered correct because they did not change the semantic meaning of the response and could still be processed correctly by the system.

Table 6. Summary of Speech-to-Text Testing Results

Parameter	Value
Total Test Samples	14
Correct Recognitions	12
Incorrect Recognitions	2
Accuracy	85.71%
Word Error Rate (WER)	5.41%

Word Error Rate (WER) was used as an additional evaluation metric to provide a more detailed assessment of transcription errors.

$$WER = \frac{S + D + I}{N} \qquad (2)$$

Where:

S = Substitutions = 2
D = Deletions = 0
I = Insertions = 0
N = Total reference words = 37

$$WER = \frac{2}{37} = 0.05405 \qquad (3)$$

Despite the promising results, the evaluation is subject to several limitations. First, the dataset is relatively small (14 samples), which limits the statistical significance and generalizability of the findings. Second, the evaluation was conducted in a controlled environment, without variations in background noise, accents, or multiple speakers. Therefore, the results should be interpreted as a preliminary performance assessment rather than a comprehensive benchmark of system accuracy. The speech-to-text system demonstrates satisfactory performance under controlled conditions, achieving an accuracy of 85.71% and a WER of 5.41%. However, further evaluation with a larger and more diverse dataset is required to validate the robustness and generalizability of the system.

This level of performance is broadly consistent with findings reported elsewhere in the ASR literature, where transcription quality varies considerably depending on the system used and the nature of the input. Comparative evaluations of leading speech-to-text services have found that transcript quality can range from roughly 5% to 20% of outputs containing major errors depending on the engine, with more than seventy percent of outputs from the better-performing systems rated as perfect or near-perfect [1]. The error rate

obtained in the present study falls toward the more favorable end of this range, which may be attributable to the relatively constrained and structured nature of the test data, consisting of proper names and short fixed phrases, rather than the open-ended, naturally varied speech typical of survey research conducted in less controlled settings. This aligns with broader benchmarking evidence indicating that ASR performance is highly sensitive to input characteristics, with paid and more recent services generally outperforming open-source alternatives, but with no single tool performing best across all types of audio input [10]. Taken together, these comparisons suggest that the accuracy achieved here reflects favorable testing conditions as much as system capability, reinforcing the study's own caution that the results represent a preliminary assessment rather than a benchmark of performance under real-world, unconstrained survey conditions.

4. Conclusions and Future Works

This study successfully designed and developed a web-based E-Survey system integrated with a speech-to-text feature using the Waterfall-based SDLC approach. The implementation results demonstrate that the system is capable of supporting both text-based and voice-based survey interactions while fulfilling the expected functional requirements. System evaluation was conducted using black box testing, speech-to-text performance testing, and WCAG 2.1-based accessibility evaluation. The results indicate that all main system functionalities operated successfully, while the speech-to-text feature achieved an accuracy rate of 85.71% with a Word Error Rate (WER) of 5.41% under controlled testing conditions. Furthermore, the accessibility evaluation showed that the implemented accessibility features were consistent with the selected WCAG 2.1 principles, supporting more flexible and inclusive digital survey interactions for diverse users. Beyond the technical implementation, the practical contribution of this research lies in providing an accessible digital survey solution that can support inclusive participation in higher education environments. The developed system enables respondents to answer survey questions more naturally and efficiently through voice interaction, which may improve user comfort, reduce interaction barriers, and support broader participation in academic survey activities. In addition, the system can assist educational institutions in implementing more inclusive digital services aligned with accessibility standards and digital transformation initiatives. From a theoretical perspective, this study contributes to the fields of human-computer interaction, assistive technology, and inclusive digital education by demonstrating how speech-to-text technology can be integrated into web-based survey systems to support alternative interaction modalities. The findings extend existing knowledge on accessible digital systems by showing that voice-based input can be incorporated into academic survey environments while maintaining usability, accessibility, and functional reliability. Beyond the development of a specific application, this research provides empirical evidence that combining speech recognition technology with accessibility-oriented design principles can support more inclusive participation in digital data collection processes. These findings may serve as a reference for future studies investigating accessible interaction models and the integration of assistive technologies within educational information systems. However, this study has several limitations. The speech-to-text evaluation was conducted in a controlled environment with limited datasets and participant variations. Factors such as environmental noise, internet stability, accents, and pronunciation differences may influence performance in real-world applications. Furthermore, direct evaluation involving users with disabilities is still required to better assess the system's effectiveness, usability, and accessibility. Therefore, future studies should involve more diverse participant groups and explore multilingual speech recognition, offline ASR capabilities, adaptive accessibility features, and quantitative usability evaluation methods such as SUS or automated WCAG compliance tools to further enhance system inclusiveness and effectiveness.

5. References

[1] J. K. Höhne, T. Lenzner, and J. Claassen, "Automatic Speech-to-Text Transcription: Evidence from a Smartphone Survey with Voice Answers," *Int. J. Soc. Res. Methodol.*, vol. 28, no. 5, pp. 625–632, 2025, doi: 10.1080/13645579.2024.2443633.

[2] G. Almgren Bäck and others, "Speech-to-Text Intervention to Support Text Production Among Students with Writing Difficulties: A Single-Case Study in Nordic Countries," *Disabil. Rehabil. Assist.*

Technol., vol. 19, no. 8, pp. 3110–3129, 2024, doi: 10.1080/17483107.2024.2351488.

[3] J. L. K. E. Fendji, D. C. M. Tala, B. O. Yenke, and M. Atemkeng, "Automatic Speech Recognition Using Limited Vocabulary: A Survey," *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2095039, 2022, doi: 10.1080/08839514.2022.2095039.

[4] J. Ara, C. Sik-Lanyi, A. Kelemen, and T. Guzsvinecz, "An Inclusive Framework for Automated Web Content Accessibility Evaluation," *Univers. Access Inf. Soc.*, vol. 24, no. 2, pp. 1581–1607, 2025, doi: 10.1007/s10209-024-01164-5.

[5] S. Harper and Y. Yesilada, "Web Accessibility and Guidelines," in *Web Accessibility*, London: Springer, 2008, pp. 61–78. doi: 10.1007/978-1-84800-050-6_6.

[6] D. S. Kumar, "Assistive Technology on Audio Visual Speech Recognition for Hearing Impaired Peoples by Using Machine Learning," vol. 20, no. 11, 2022.

[7] L. A. Kumar, D. K. Renuka, S. L. Rose, M. C. Shunmuga Priya, and I. M. Wartana, "Deep Learning Based Assistive Technology on Audio Visual Speech Recognition for Hearing Impaired," *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 24–30, 2022, doi: 10.1016/j.ijcce.2022.01.003.

[8] A. Abdullah, Jumadi, and D. Firdaus, "Implementasi Algoritma Bidirectional Encoder Representations From Transformer Pada Speech To Text Untuk Notulensi Rapat," *SMATIKA J.*, vol. 15, pp. 431–432, 2025, doi: 10.32664/smatika.v15i02.1725.

[9] A. M. Deshmukh and R. Chalmeta, "User Experience and Usability of Voice User Interfaces: A Systematic Literature Review," *Information*, vol. 15, no. 9, p. 579, 2024, doi: 10.3390/info15090579.

[10] A. Ferraro, A. Galli, V. La Gatta, and M. Postiglione, "Benchmarking Open Source and Paid Services for Speech to Text: An Analysis of Quality and Input Variety," *Front. Big Data*, vol. 6, p. 1210559, 2023, doi: 10.3389/fdata.2023.1210559.

[11] H. Suhendar, C. Slamet, and U. Syaripudin, "Analisis Sentimen Hasil Transkripsi Audio Berbahasa Indonesia Menggunakan T5 (Text-to-Text Transfer Transformer)," *SMATIKA J.*, vol. 15, no. 1, pp. 115–125, 2025, doi: 10.32664/smatika.v15i01.1521.

[12] Y. A. Alsamiri and A. B. A. B. A. Alduaylij, "Academic Challenges of Visually Impaired Students at Hail University," *J. Cakrawala Pendidik.*, vol. 44, no. 1, pp. 9–20, 2025, doi: 10.21831/cp.v44i1.77706.

[13] M. Revilla, C. Ochoa, J. K. Höhne, and M. P. Couper, "Transcribing and Coding Voice Answers Obtained in Web Surveys: Comparing Three Leading Automatic Speech Recognition Tools," *J. Surv. Stat. Methodol.*, p. smaf028, 2026, doi: 10.1093/jssam/smaf028.

[14] D. Fithriyaningrum, S. S. Kusumawardhani, and S. Wibirama, "Analisis Aksesibilitas Website berdasarkan Web Content Accessibility Guidelines (WCAG): Ulasan Literatur Sistematis," *J. IPTEK-KOM*, vol. 23, no. 1, pp. 79–92, 2021.

[15] H. Liu, J. Sun, Z. He, H. Huang, X. Liu, and L. Nie, "Web Accessibility of University Websites in China: An Empirical Study," *J. King Saud Univ. Comput. Inf. Sci.*, 2026, doi: 10.1007/s44443-026-00544-6.

[16] M. Fakrudeen, "Evaluation of the Accessibility and Usability of University Websites: A Comparative Study of the Gulf Region," *Univers. Access Inf. Soc.*, vol. 24, no. 2, pp. 1883–1898, 2025, doi: 10.1007/s10209-024-01160-9.

[17] P. Nso-Mangue, C. Cachero-Castro, S. Meliá, and S. Luján-Mora, "Impact of COVID-19 in the Web Accessibility of Higher Education Institutions: A Pending Challenge," *Univers. Access Inf. Soc.*, vol. 24, no. 2, pp. 1439–1460, 2025, doi: 10.1007/s10209-024-01149-4.

[18] P. Ndibalema and W. Kambona, "Impediments to Assistive Technology Accessibility for Students With Disabilities in Higher Education Institutions: A Systematic Review," *ECNU Rev. Educ.*, 2025, doi: 10.1177/20965311251355657.

[19] M. Laamanen, T. Ladonlahti, H. Puupponen, and T. Kärkkäinen, "Does the Law Matter? An Empirical Study on the Accessibility of Finnish Higher Education Institutions' Web Pages," *Univers. Access Inf. Soc.*, vol. 23, no. 1, pp. 475–491, 2024, doi: 10.1007/s10209-022-00931-6.

[20] M. A. Hasibuan and Samsudin, "Analisis dan Rancang Bangun Sistem Informasi Akademik Menggunakan Metode System Development Life Cycle (SDLC)," *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 377–385, 2025, doi: 10.51401/jinteks.v7i1.5589.

[21] H. Ahlawat, N. Aggarwal, and D. Gupta, "Automatic Speech Recognition: A Survey of Deep Learning Techniques and Approaches," *Int. J. Cogn. Comput. Eng.*, vol. 6, pp. 201–237, 2025, doi: 10.1016/j.ijcce.2024.12.007.

[22] A. Perdana, S. Dewi, N. A. Farhana, and D. Febrian, "Comparative Analysis of SDLC and R\&D Methods in System Development: A Case Study of Integrity Zone Management System," *Sinkron*, vol. 9, no. 4, pp. 3197–3209, 2025, doi: 10.33395/sinkron.v9i4.15337.

[23] E. A. Draffan, M. Wald, C. Ding, and Y. Li, "Exploring Practical Metrics to Support Automatic Speech Recognition Evaluations," in *Studies in Health Technology and Informatics*, IOS Press, 2023. doi: 10.3233/SHTI230636.

[24] B. Alturas, "Connection between UML Use Case Diagrams and UML Class Diagrams: A Matrix Proposal," *Int. J. Comput. Appl. Technol.*, vol. 72, no. 3, pp. 161–168, 2023, doi: 10.1504/IJCAT.2023.10058804.

[25] Yulherniwati, M. Gunawan, A. F. Kasmar, R. Hadi, E. Asri, and Humaira, "Black-Box Testing pada Sistem Evaluasi Capaian Pembelajaran Berdasarkan Outcome Based Education," *JITSI J. Ilm. Teknol. Sist. Inf.*, vol. 6, no. 2, pp. 202–207, 2025, doi: 10.62527/jitsi.6.2.279.