A COMPARATIVE STUDY ON SPEECH-TO-TEXT APPLICATIONS: BASIS OF THE DEVELOPMENT OF AN ALTERNATIVE COMMUNICATION APP MODEL

A Research

Presented to the Faculty of Bachelor of Science in Information Technology

Polytechnic University of the Philippines –

Quezon City Campus

In Partial Fulfillment

Of the Requirements for

Fundamentals of Research

By:

Sto. Domingo, Roczpen M.

Sangreo, Sergius Paulo B.

BSIT 3-1

NOVEMBER 2024

Chapter 1

THE PROBLEM AND ITS SETTING

Introduction

The field of speech-to-text (STT) technologies has grown rapidly in recent years, significantly transforming how individuals and organizations approach communication. With the proliferation of digital platforms, the need for efficient and reliable STT solutions has intensified, benefiting diverse groups including those who rely on assistive technologies, professionals in high-stakes environments, and consumers seeking convenience and accessibility (Huang et al., 2022). Today's STT applications are deeply integrated into virtual assistants, transcription services, and accessibility tools, underscoring their role in enhancing digital accessibility and improving productivity across multiple domains (Bahar et al., 2019).

As STT technologies have evolved, a comparative approach to understanding different applications has become increasingly valuable, particularly given the diverse environments in which they operate. For instance, some STT systems are optimized for controlled, quiet settings, while others perform better under noisy or unpredictable conditions. This variability in application design can significantly impact their usability in real-world scenarios, such as medical, legal, and business environments where accuracy and speed are paramount (Chang et

al., 2023). Furthermore, advances in artificial intelligence have introduced a new generation of STT models, with Transformer architecture leading the charge due to their attention mechanisms, which have demonstrated higher accuracy and processing speed compared to traditional recurrent neural networks (Karita et al., 2019). The comparative analysis of these models illuminates the structural choices that improve STT performance and provides insights into the limitations and possibilities inherent to different architectures (Karita et al., 2019).

Recent research has also emphasized the benefits of integrated, multimodal approaches, where STT systems combine audio and visual cues for enhanced recognition accuracy. This approach is particularly beneficial in developing assistive technologies for individuals with hearing impairments, as demonstrated in studies focusing on audio-visual speech recognition systems (Kumar et al., 2022). By capturing both audio and visual data, these multimodal STT systems are better equipped to accommodate diverse user needs and environments, pushing the boundaries of traditional speech recognition applications and setting new standards in accessibility.

Despite these technological advancements, STT applications still face significant challenges in achieving reliable performance across a wide range of use cases. Text quality and pre-processing, for example, remain pivotal to the overall performance of STT systems. Studies show that pre-processing techniques can help improve recognition accuracy, especially when dealing with short, informal texts, such as those commonly encountered on social media platforms (Naseem et al., 2020). Effective pre-processing can help refine the quality of input

data, enhancing outcomes for STT systems across various applications, including those requiring high sensitivity, such as hate speech detection or sentiment analysis (Naseem et al., 2020).

Evaluation and benchmarking of STT systems present another complex challenge, as existing studies often lack standardized metrics for assessing performance across different applications (Chang et al., 2023). This gap hinders a comprehensive understanding of STT technology's capabilities, especially in domains that require high levels of accuracy and real-time responsiveness. Effective evaluation frameworks are essential for a balanced comparison of STT systems, as they can account for nuanced factors such as language diversity, processing latency, and accuracy under various noise conditions. Comparative studies that address these factors offer critical insights into the strengths and weaknesses of different STT applications, guiding both future research and practical deployments in diverse environments (Chang et al., 2023).

Another emerging area of interest in STT research involves the ethical implications of data privacy and security. As STT technologies become increasingly integrated into personal devices, virtual assistants, and enterprise solutions, safeguarding users' privacy and ensuring secure data handling practices are more important than ever. Research has shown that public confidence in STT technologies is closely tied to transparent data practices, where ethical considerations around data collection, storage, and usage are explicitly addressed (Naseem et al., 2020). Without careful attention to these ethical dimensions, the rapid adoption of STT technology could raise significant concerns about privacy,

particularly as applications begin to handle sensitive or personally identifiable information.

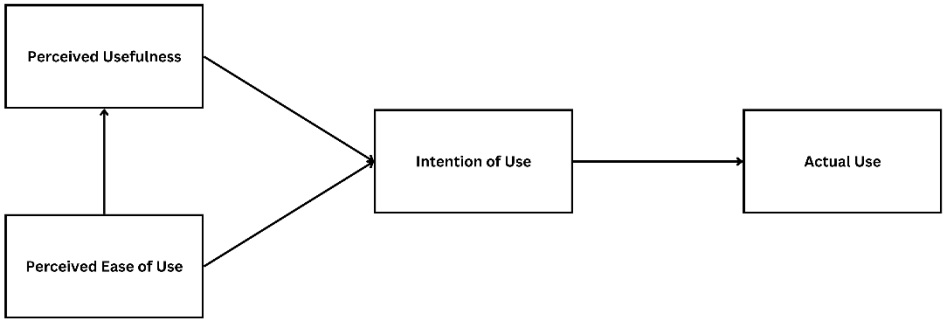
This study aims to fill critical gaps in the literature by providing a detailed comparative analysis of prominent STT applications, focusing on core metrics such as recognition accuracy, processing speed, user experience, and adaptability to varied conditions. Through a synthesis of recent advancements, this research will also identify unresolved challenges and highlight opportunities for future innovation in STT systems. By addressing both technical and ethical aspects, this comparative study seeks to advance the field of STT technology, contributing valuable insights to guide the development of applications that are not only highly functional but also responsive to diverse user needs and aligned with ethical standards in data privacy and security.

Theoretical Framework

The researchers used the following theories as foundations to establish parameters for evaluating models of speech-to-text applications, guiding the development of an alternative communication app model.

Technology Acceptance Model

Figure 1. Technology Acceptance Model



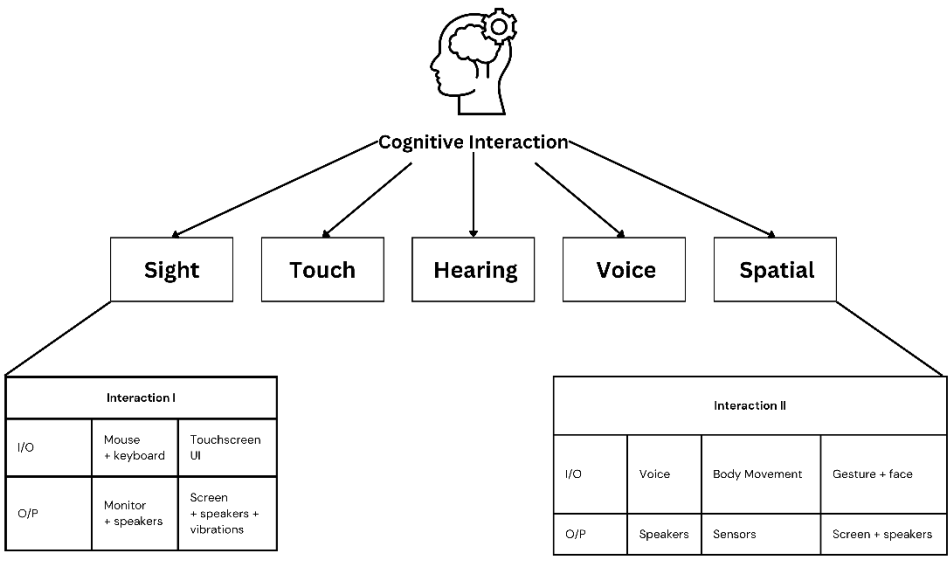
The Technology Acceptance Model (TAM) was developed to understand the factors driving individuals' acceptance of technology, aiming to improve performance, efficiency, and convenience at both organizational and individual levels. Prior to TAM, information systems (IS) research lacked a robust empirical framework to assess users' attitudes and responses to technology, relying instead on subjective performance measures that lacked validity. The Theory of Reasoned Action (TRA) by Ajzen and Fishbein inspired TAM, but TRA's generality limited its application to technology-specific contexts. To address this, Davis (1989) identified two critical constructs—perceived usefulness (the belief that technology improves performance) and perceived ease of use (the belief that technology is easy to use)—as primary determinants of technology acceptance.

TAM posits a three-stage process: external factors influence cognitive responses (perceived usefulness and ease of use), which form an effective response (attitude or intention) that predicts use behavior. This model underscores

that ease of use indirectly boosts technology acceptance by reinforcing perceived usefulness. TAM has been widely validated, helping practitioners design systems that encourage adoption by improving usability and perceived value. Its application has advanced IS research by providing a reliable framework for assessing user motivation across various technologies.

Human-Computer Interaction (HCI) Theory

Figure 2. Human-Computer Interaction Framework



Human-Computer Interaction (HCI) Theory focuses on understanding how people interact with technology and how those interactions can be improved for a better user experience. The theory brings together ideas from computer science, design, and psychology to create more intuitive, effective, and efficient interactions between users and computers. A key concept in HCI is usability, which measures

how easily users can accomplish their tasks with a system. In the case of speech-to-text applications, usability refers to how accurately and efficiently the app transcribes speech and how easy it is to interact with. Shneiderman and Plaisant (2010) explain that usability involves the “effectiveness, efficiency, and satisfaction with which users achieve specific goals” when using a system. For speech-to-text apps, good usability ensures that users can quickly and easily transcribe speech without frustration.

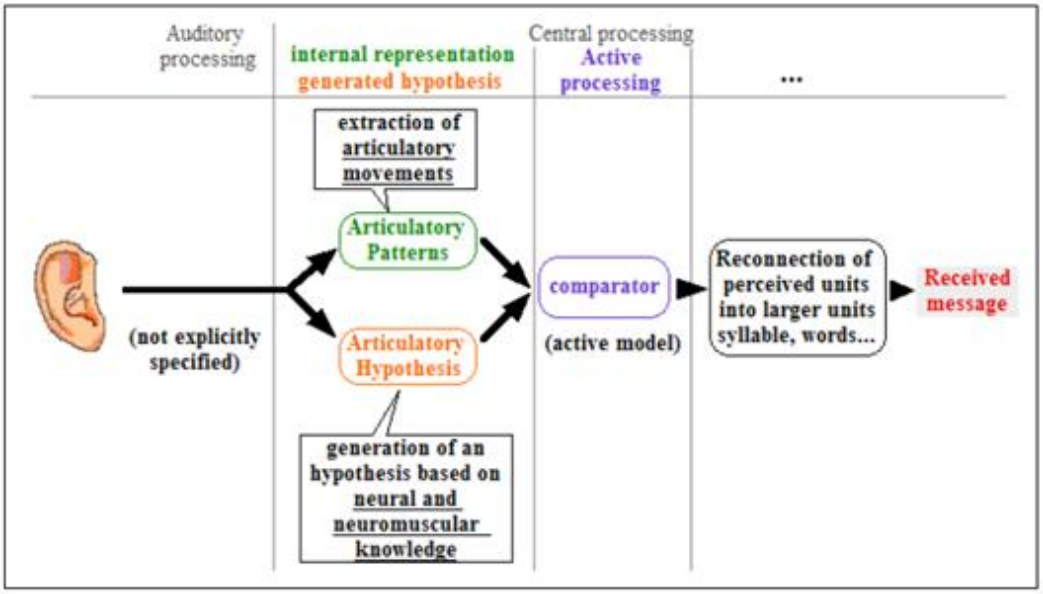
Another important aspect of HCI is User-Centered Design (UCD), which emphasizes designing technology around the needs and abilities of the user. In the case of speech-to-text apps, UCD would involve ensuring the app is accessible to users with different abilities, including those with speech impairments or unique accents. Norman (2013) stresses that design is not just about making things look good, but about making them useful, highlighting the importance of creating technology that genuinely supports the user’s needs. By focusing on users' preferences and challenges, UCD ensures that speech-to-text apps are easy to use and meet diverse user needs. Affordances refer to the features of a system that suggest how it should be used. Norman (1988) described affordances as the relationship between an object’s properties and a user’s ability to interact with it. In speech-to-text apps, affordances could be clear design elements that guide users on how to begin speaking, how to correct errors, or how to navigate through transcriptions. By making the app intuitive, users can engage with it without needing extra instructions or assistance. Interaction Design looks at how users engage with technology and how those interactions can be made smoother. The

goal is to make the system responsive and easy to understand so that users don't get frustrated while using it. Cooper et al. (2014) define interaction design as "the design of behavior between people and products," aiming to create meaningful, seamless interactions. For speech-to-text applications, this means providing immediate feedback, easy navigation, and tools for users to adjust their transcriptions, all of which make the app more user-friendly.

These HCI concepts are essential for improving speech-to-text apps, ensuring they are functional, intuitive, and accessible for users. By applying these principles, the design of these apps can be refined to better support users in their communication needs.

Motor Theory of Speech Perception

Figure 3. Motor Theory Framework

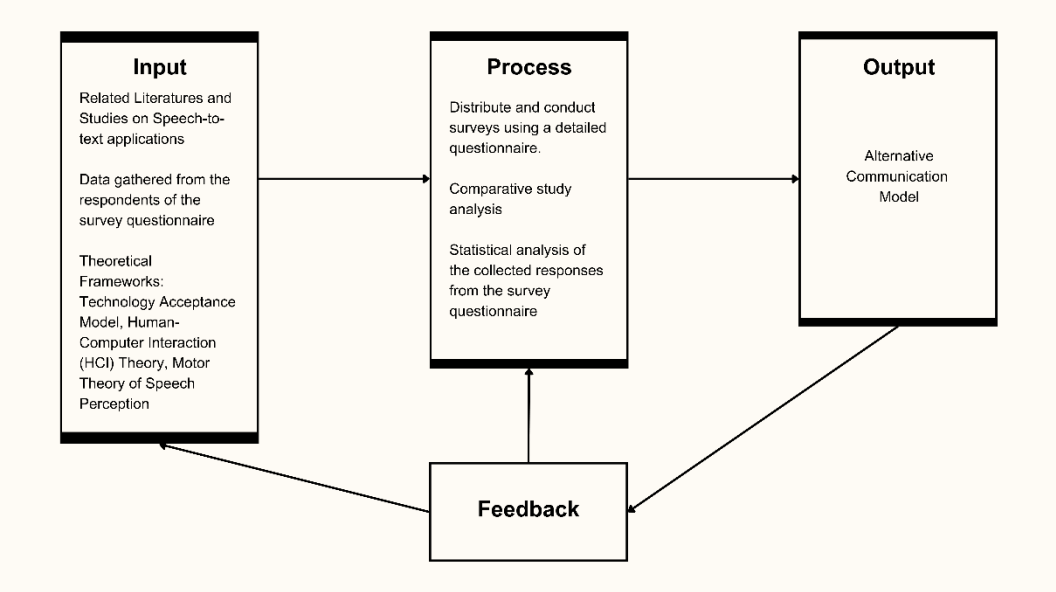


Motor Theory initially proposed in the Haskins Laboratories in the 1950s by Alvin Liberman and Franklin S. Cooper, and developed further by Donald Shankweiler, Michael Studdert-Kennedy, Ignatius Mattingly, Carol Fowler and Douglas Whalen. Considering the lack of acoustic invariance, it is feasible to look for invariance in the articulatory space (i.e., maybe the illustrative units are characterized concerning verbalization). Motor Theory recommends that speech discernment depends on the way things are delivered. In other words, when people see speech, they tap into their comprehension and might interpret how phonemes are verbalized. Units of discernment in this hypothesis incorporate articulatory motions, for example, lip adjusting or squeezing the lips together, which straightforwardly pass phonetic data on to the audience.

Biological specialization for phonetic gestures inhibits listeners from perceiving the signal as a typical sound, yet it allows them to leverage the systematic and special relationship between the signal and sound to interpret the gestures. Initially, the motor commands controlling articulation were deemed as the invariant phonetic features. The updated theory posits that the invariant object of perception is the intended gestures. Sounds are perceived discreetly (categorically) because they are produced with distinct articulators/gestures. The McGurk effect indicates that some features are represented as articulatory.

Conceptual Framework

Figure 4. Conceptual Framework



The research's conceptual framework comprises four key elements. Independent inputs involve literature and studies on speech-to-text applications, theoretical frameworks, and data collection, serving as the foundational components. The literature provides understanding of existing speech-to-text models, while user engagement data illuminates user interactions with these applications. Dependent processes involve comparative study and statistical analysis. The core of the research is comparative study, investigating various speech-to-text applications and their influence on user communication. Statistical analysis thoroughly examines collected data to reveal significant insights and relationships. The dependent output is the creation of an alternative communication app model, the primary research objective. This model, aiming to

improve communication through speech-to-text applications, directly emerges from comparative study and statistical analysis. The framework also includes a feedback loop, wherein insights from the study, analysis, and user engagement data guide the continuous development and refinement of the alternative communication app model.

Statement of the Problem

This study aims to conduct a comparative analysis of speech-to-text applications to establish a foundation for the development of an alternative communication app model. To gain a comprehensive understanding of the field, the researchers seek answers to the following questions:

1. What is the level of awareness of the respondents to the speech text application:
    1.1 Portability
    1.2 Security
    1.3 Functionality
    1.4 Benefits
2. What is the level of agreement of users of alternative communication model of speech-to-text application, considering:
    2.1 Interaction Frequency

2.2 Utilization of Features

2.3 Sustained Involvement

2.4 User Feedback

3. What is the satisfaction level of the respondents using speech text application in terms of:

   3.1 Operability

   3.2 Functionality

   3.3 User-friendliness

   3.4 Effectiveness

4. What is the significant relationship of #2 and #3 for mobile speech text applications?

5. What model can be developed as an alternative communication app?

Hypothesis

H0: There is no significant difference in user communication outcomes between existing speech-to-text applications and the proposed alternative communication app model.

H1: The proposed alternative communication app model demonstrates a significant improvement in user communication outcomes compared to existing speech-to-text applications.

Scope and Limitation

This research focuses on comparing various speech-to-text applications with the aim of developing an alternative communication app model. The research took place at the Polytechnic University of the Philippines, Quezon City Campus. The specific speech-to-text applications that will be used for this study are Speech (Windows), Dictation (macOS), and Gboard (Android). The respondents of this study focused on Polytechnic University of the Philippines, Quezon City Campus, selecting students from 1st year to 4th year. The study will be available to the public upon completion. The selected students and deaf individuals from different programs were the basis of this study. This facilitated the researchers in acquiring further information and valuable ideas for a comparative study on Speech-to-Text applications, which will serve as a basis for developing an alternative communication app model.

Significance of the study

The researchers conducted a comparative study on speech-to-text applications as a basis for the development of an alternative communication app model and aim to be beneficial to the following:

Users - The research empowers users through a comprehensive comparison of speech-to-text applications, offering a foundation for improved communication app models.

Developers - This study aids developers in crafting innovative mobile applications by providing insights into the development of an alternative communication app model.

Future Researchers - The study serves as a foundational resource for future researchers, encouraging exploration of gaps in speech-to-text applications and fostering advancements in communication solutions**.**

Definition of Terms

Technical Definitions

Speech-to-Text (STT) - A technology that converts spoken language into written text. It leverages various algorithms, including deep learning and machine learning, to transcribe speech accurately. Examples include Google Speech-to-Text, IBM Watson Speech to Text, and Microsoft Azure Speech Services.

Natural Language Processing (NLP) - A field of artificial intelligence that focuses on the interaction between computers and humans through natural language. It involves the application of computational techniques to analyze and synthesize natural language and speech.

Artificial Intelligence (AI) - A branch of computer science dealing with the simulation of intelligent behavior in computers. In the context of speech-to-text, AI algorithms are used to improve the accuracy and context understanding of transcriptions.

TRACE Model - A connectionist model of speech perception developed by James McClelland and Jeffrey Elman. It represents speech perception through a network of interconnected units that span various levels of speech units (features, phonemes, words) and involve excitatory and inhibitory interactions.

Comparative Study - A research methodology that involves comparing multiple entities to identify their differences and similarities. In this context, it refers to the evaluation of various speech-to-text applications to determine their strengths and weaknesses.

Statistical Analysis - The process of collecting and analyzing data to identify patterns, relationships, and trends. This method is employed in the study to derive insights from the data collected on speech-to-text application usage.

Direct Realist Theory of Speech Perception - A theory proposed by Carol Fowler, which posits that speech perception occurs through the direct recovery of articulatory gestures, like how visual and olfactory perceptions recover the distal source of perceived events.

Motor Theory and Speech Perception - A theory suggesting that the perception of speech sounds is based on the listener's understanding of the

articulatory movements that produce those sounds. It posits that speech perception is directly linked to the production mechanisms.

Operational Definitions

Alternative Communication App Model - A proposed application framework aimed at enhancing current speech-to-text technologies by addressing gaps in user satisfaction, functionality, and integration. It is designed to cater to specific user needs such as those with speech impairments or those working in multilingual environments.

Real-Time Transcription - The process of converting speech to text instantly as it is spoken, without significant delay. This feature is essential for applications requiring immediate feedback or interaction, such as live captioning services.

Multi-Language Support - The capability of a speech-to-text application to transcribe speech in multiple languages, making the technology accessible to a broader range of users worldwide.

User Interface (UI) - The means by which a user interacts with a computer, software, or application. A user-friendly UI in speech-to-text applications enhances user experience by being intuitive and easy to navigate.

Integration Capabilities - The ability of a speech-to-text application to seamlessly connect and work with other software or systems, such as word processors, communication platforms, and accessibility tools. Motor Theory of

User Engagement Data - Information collected on how users interact with speech-to-text applications, including frequency of use, feature utilization, and feedback. This data is crucial for understanding user behavior and improving application design.

Accessibility - The design of products, devices, services, or environments to be usable by people with disabilities. Speech-to-text applications enhance accessibility by providing transcription services for individuals with hearing or speech impairments.

User Satisfaction - A measure of how well a product or service meets or exceeds the user's expectations. In the context of speech-to-text applications, it includes factors like accuracy, ease of use, and overall effectiveness of the transcription service.

Functionalities - The specific tasks or activities that a software application can perform. For speech-to-text applications, this includes features like real-time transcription, noise handling, multi-language support, and integration with other tools.

Transcription Accuracy - The degree to which the transcribed text matches the spoken input. High accuracy is crucial for the reliability of speech-to-text applications, especially in professional and accessibility contexts.

Context Understanding - The ability of a speech-to-text system to comprehend and accurately transcribe speech based on the context in which it is spoken. This involves recognizing and correctly transcribing homophones, colloquialisms, and context-specific vocabulary.

Chapter 2

REVIEW OF RELATED LITERATURE

In this chapter examines the existing literature on speech-to-text applications and alternative communication app models. The objective of this review is to contextualize and synthesize the current state of research, identifying gaps, trends, and insights crucial for informing the development of an innovative alternative communication application.

Speech-to-Text Technology

The historical development and evolution of speech-to-text technology have been a subject of extensive research and scholarly inquiry, reflecting its profound impact on various domains such as natural language processing, human-computer interaction, and communication technology. This literature review aims to provide a comprehensive overview of the key advancements, challenges, and future directions in this rapidly evolving field. Research by Bhatt et al. (2021) delved into the historical progression of deep learning architectures, particularly focusing on Convolutional Neural Network (CNN) variants in computer vision. While primarily focused on computer vision, the study underscored the overarching influence of deep learning techniques in diverse fields, including speech recognition and natural language processing. Similarly, Borowiec et al. (2022) emphasized the utility of deep learning in ecology and evolution, highlighting its relevance to the broader scientific community. Bi (2020) explored the

advancements in speech recognition and voice generation technologies, emphasizing the evolution of conversation understanding systems. This study provided insights into the challenges of processing multi-turn conversations, particularly in discerning relevant context—a crucial aspect for accurate speech-to-text conversion. Moreover, Nyatsanga et al. (2023) offered a comprehensive review of data-driven co-speech gesture generation, illustrating the evolving landscape of human-machine interaction through the integration of voice and gesture recognition technologies.

The integration of artificial intelligence (AI) and machine learning (ML) techniques has significantly enhanced the capabilities of speech-to-text systems. Luo et al. (2023) demonstrated how triboelectric nanogenerators and machine learning algorithms could enable dual modes of interaction via voice and gesture recognition, opening new possibilities for intuitive human-machine interfaces. Additionally, the study by Aliannejadi et al. (2019) highlighted the potential of harnessing multi-turn conversations for effective answer retrieval, showcasing the importance of AI-driven approaches in enhancing conversational understanding.

While the literature indicates substantial progress in speech-to-text technology, several challenges and opportunities for future research persist. One crucial avenue for exploration is the integration of speech-to-text systems with emerging technologies like edge computing, Internet of Things (IoT), and wearable devices. These integrations could enhance real-time speech recognition capabilities and facilitate seamless communication across various platforms.

Ethical and privacy concerns also warrant attention in the advancement of speech-to-text technology. Issues such as data security, user consent, and algorithmic biases pose significant challenges that must be addressed through interdisciplinary research and regulatory frameworks.

In conclusion, the literature review elucidates the historical trajectory, current advancements, and prospects of speech-to-text technology. By synthesizing insights from diverse research endeavors, this review underscores the interdisciplinary nature of speech-to-text development and the imperative for continued innovation to address emerging societal and ethical considerations.

Features and Functionalities

Speech-to-text applications have become increasingly indispensable in our daily lives, permeating various sectors like healthcare, education, and customer service. They owe their efficiency to cutting-edge artificial intelligence and natural language processing techniques, seamlessly converting spoken words into text. Let's delve into recent research findings to understand their current state and prospects.

Mathews (2019) delves into the realm of explainable artificial intelligence within natural language processing (NLP). This area promises to enhance the transparency and interpretability of speech-to-text applications, a crucial aspect for user trust and comprehension.

In a similar vein, Hakak et al. (2019) shed light on the importance of refining algorithms to boost the accuracy and resilience of speech-to-text systems,

especially in challenging environments with diverse accents. Wagner et al. (2019) evaluated the latest in speech synthesis systems, urging for continued efforts to refine the quality and clarity of synthesized speech. This resonates with the quest for enhancing user experience in speech-to-text applications. Yu et al. (2020) introduced the DurIAN model, advocating for the integration of temporal information to improve transcription accuracy and fluency. Such insights underscore the need for nuanced approaches in accommodating diverse speech patterns. Moreover, Han et al. (2021) proposed leveraging shared semantic spaces to elevate the semantic understanding in transcription, a vital aspect for accurately capturing contextual meaning. Despite these strides, there are notable research gaps to address. Zhang et al. (2022) advocate for exploring multimodal fusion to enhance contextual understanding, particularly in scenarios involving facial expressions. There's a call to enhance the transparency and interpretability of speech-to-text systems, as noted by Mathews (2019). Future research should focus on tailoring explainable AI techniques for these applications to foster user trust and comprehension. The need for advanced algorithms to bolster robustness and accuracy, particularly in challenging acoustic environments, remains pertinent, as highlighted by Hakak et al. (2019). Innovative algorithmic approaches are needed to navigate these hurdles and ensure reliable transcription.

In essence, this review underscores the multifaceted nature of speech-to-text applications, emphasizing transparency, robust algorithms, semantic understanding, and temporal considerations. Addressing these facets will undoubtedly propel the evolution of these applications, enriching user experiences across diverse domains.

Accuracy and Performance Metrics

In recent years, speech recognition technology has experienced remarkable advancements, resulting in notable improvements in accuracy and performance. Let's delve into recent research to gain insights into the current landscape and potential future directions of this technology.

Al-makhadmeh and Tolba (2019) explore the use of natural language processing and ensemble deep learning methods for automatic hate speech detection. Their findings underscore the significance of employing advanced technologies like deep learning to enhance the accuracy of speech recognition systems. This study highlights the potential of such technologies in elevating overall performance metrics.

In a related study, Zhang et al. (2022) introduce WeNet 2.0, an end-to-end speech recognition toolkit demonstrating substantial improvements in recognition performance over its predecessor on various datasets. This showcases the tangible impact of technological advancements on boosting the accuracy and performance metrics of speech recognition systems.

Collectively, research by Al-makhadmeh and Tolba (2019) and Zhang et al. (2022) emphasizes the importance of leveraging advanced technologies and innovative approaches to enhance speech recognition systems. These findings contribute significantly to the existing knowledge base, showcasing the potential for substantial advancements through deep learning and end-to-end speech recognition toolkits.

Despite these advancements, several avenues for future research emerge. One such direction involves exploring novel approaches to tackle challenges

associated with multilingual and accented speech recognition. While current research improves recognition performance, delving deeper into accommodating linguistic and accent variations can foster more inclusive and robust systems. Additionally, developing adaptive models capable of dynamically adjusting to environmental and contextual factors holds promise for future exploration.

In summary, recent literature underscores how advanced technologies have propelled the accuracy and performance metrics of speech recognition systems. The studies by Al-makhadmeh and Tolba (2019) and Zhang et al. (2022) highlight ongoing efforts to refine this technology while pointing towards future directions such as addressing multilingual challenges and developing adaptive models. This research serves as a springboard for further advancements in speech recognition technology.

Applications and Use Cases

Speech-to-text (STT) technology has seen widespread adoption in various fields owing to its versatile nature and practical applications. This comprehensive literature review aims to integrate and synthesize the provided research findings, identify knowledge gaps, and suggest potential future research directions.

One of the primary applications of STT technology is Automatic Speech Recognition (ASR) for language identification and end-to-end ASR for various languages (Semanticscholar). This indicates the diverse linguistic capabilities of STT systems, making them suitable for multilingual and cross-cultural applications. However, racial disparities have been noted in automated speech recognition systems, highlighting the need for research into addressing such biases and

ensuring equitable performance across different demographic groups (Koenecke et al., 2020).

STT technology is also utilized to power chatbots, virtual assistants, and other conversational interfaces (Aljanabi et al., 2023). The seamless integration of STT into conversational systems opens numerous possibilities for human-computer interaction and automated customer support. However, the ethical implications of content moderation using artificial intelligence, particularly concerning hate speech and marginalized voices, warrant further investigation (Oliva et al., 2020).

Moreover, the security and privacy considerations associated with STT technology are significant, given its uses in automatic transcription of conversations and user authentication (Semanticscholar). Research in this area could help in developing robust security measures and privacy-preserving techniques for STT systems. Additionally, the impact of ASR errors on specific demographic groups, such as African Americans, emphasizes the need for culturally sensitive and inclusive STT models (Mengesha et al., 2021).

The healthcare domain has also seen the integration of STT technology in clinical natural language processing systems for extracting clinically important information from electronic health records (Semanticscholar). This application presents opportunities for improving clinical documentation and facilitating data-driven healthcare interventions. However, further research is needed to address the challenges of ensuring accuracy, privacy, and regulatory compliance in healthcare-specific STT applications.

Lastly, advancements in STT technology have led to the development of innovative systems such as WhisperX, a time-accurate speech transcription system for long-form audio (Bain et al., 2023). Similarly, research on noise-resistant multi-modal speech recognition systems and semantic misinterpretation in voice assistant applications opens up new avenues for improving the robustness and accuracy of STT systems (Liu et al., 2021; Zhang et al., 2019).

In conclusion, the literature review provides insights into the diverse applications and use cases of STT technology, highlighting its potential across various domains. However, several knowledge gaps are evident, including the need to address biases and disparities, ethical considerations in content moderation, and the development of specialized STT systems for specific applications such as healthcare. Future research directions could focus on mitigating biases in ASR systems, ensuring cultural and demographic inclusivity, and advancing the security, privacy, and accuracy of STT technology across different domains.

Technological and Ethical Considerations

The use of Natural Language Processing (NLP) technologies, such as ChatGPT, has revolutionized academic writing and research efficiency, but it also raises concerns about the authenticity and credibility of academic work. This literature review aims to integrate and synthesize the findings of various studies to understand the ethical and human rights considerations associated with speech-to-text applications, particularly ChatGPT, in academic and healthcare contexts.

Lund et al. (2023) emphasize the potential of ChatGPT and other NLP technologies to enhance academic writing and research efficiency. However, the study raises concerns about the impact on the authenticity and credibility of academic work, highlighting the need for comprehensive discussions on the potential use, threats, and limitations of these tools. The authors underscore the importance of ethical and academic principles, emphasizing human intelligence and critical thinking at the forefront of the research process.

Similarly, Dergaa et al. (2023) discuss the prospects and potential threats of ChatGPT in academic writing. The study underscores the need to examine the broad social impacts of NLP technology and bring ethical and human rights considerations to every stage of the application life cycle. This includes task formulation, dataset design, model training and evaluation, and application deployment. The authors highlight the importance of integrating ethical and human rights considerations throughout the development and deployment of NLP technologies.

Moreover, Latif et al. (2020) emphasizes the need to examine the social impacts of NLP technology, particularly in the context of healthcare. The study stresses the importance of bringing ethical and human rights considerations to every stage of the application life cycle, from task formulation and dataset design to application deployment. This highlights the relevance of ethical considerations in ensuring that speech-to-text applications in healthcare adhere to ethical and human rights principles.

Furthermore, Kiritchenko et al. (2020) discuss the ethical and human rights perspective of confronting abusive language online. While this study does not

directly address speech-to-text applications, it underscores the broader ethical considerations associated with language processing technologies and the need to address abusive and harmful language in online environments. This highlights the ethical challenges associated with NLP technologies in processing and interpreting human language.

Slavich et al. (2019) also delves into the ethical and privacy considerations related to stress measurement using speech. Although not directly focused on NLP technologies, the study underscores the ethical implications of processing and analyzing speech data. The findings emphasize the importance of ensuring privacy and ethical considerations in the development and deployment of speech-to-text applications, particularly in sensitive contexts such as healthcare.

In summary, the literature review of technological and ethical considerations in speech-to-text applications, particularly ChatGPT and NLP technologies, reveals the crucial need to integrate ethical and human rights considerations throughout the application life cycle. While these technologies offer significant advancements in academic writing, research efficiency, and healthcare, it is essential to address the ethical challenges, privacy concerns, and potential threats associated with their use. Future research should focus on developing ethical guidelines and frameworks for the responsible development and deployment of speech-to-text applications, ensuring that these technologies adhere to ethical and human rights principles. The literature highlights the necessity of comprehensive discussions on the potential use, threats, and limitations of NLP technologies, emphasizing the importance of ethical and human

rights considerations in the development and deployment of speech-to-text applications.

Accessibility and Inclusivity

Ensuring the accessibility and inclusivity of speech-to-text applications is not just a matter of convenience; it's a fundamental aspect of ensuring equitable access to digital technologies for all individuals. However, recent research has highlighted significant challenges and disparities that certain groups, particularly African American users and individuals with disabilities, face when using these applications. Addressing these issues is crucial to creating a digital landscape that is truly inclusive and accessible to all. Koenecke et al. (2020) brought attention to the experiences of African American users with speech-to-text applications, revealing that they often feel othered and must adjust their speech to improve their success with the technology. This highlights a critical issue in the inclusivity of these applications, suggesting a need for design improvements to better accommodate diverse linguistic and cultural backgrounds.

Similarly, Mengesha et al. (2021) emphasized the impact of automated speech recognition errors on African Americans, with many feeling that the devices lack cultural sensitivity. This underscores the importance of considering cultural diversity in the design and implementation of speech-to-text applications to ensure they are inclusive and accessible to all users. The issue of accessibility for individuals with disabilities in the digital realm is also a significant concern. Botelho (2021) highlighted the need for a conscious and systemic effort to ensure that digital technologies, including speech-to-text applications, are accessible to

individuals with disabilities. This requires thoughtful consideration of accessibility features and design principles to accommodate a wide range of needs and capabilities.

In the context of educational technology, Gupta and Sengupta (2021) discussed the potential of webinars as an educational tool in higher education. However, they also emphasized the importance of accessibility and inclusiveness in speech-to-text applications to ensure that educational technologies are accessible to all students, including those with disabilities. Furthermore, Shaheen and Watulak (2019) emphasized the need to consider disability as an equity concern in the field of instructional technology. This highlights the importance of broadening the discussion on accessibility and inclusiveness in educational technology to include speech-to-text applications and other digital tools.

Ngueajio and Washington (2022) conducted a literature review on bias in automatic speech recognition systems and proposed bias mitigation techniques. Their work underscored the importance of addressing biases in speech-to-text applications to make them more inclusive and equitable for all users. Aljedaani et al. (2021) discussed learning sentiment analysis for accessibility user reviews, highlighting the potential for sentiment analysis to capture user feedback on the accessibility and inclusivity of speech-to-text applications. This provides valuable insights into the user experience and areas for improvement.

While there have been strides in understanding the challenges and biases in speech-to-text applications, there are still significant knowledge gaps that need to be addressed comprehensively. Future research should focus on developing more culturally sensitive and inclusive speech-to-text applications, as well as

implementing bias mitigation techniques to ensure equitable access for all users. Additionally, further exploration of user feedback and sentiment analysis will be crucial for informing the design and improvement of speech-to-text applications. By addressing these knowledge gaps, we can work towards creating more accessible and inclusive speech-to-text applications that benefit all users.

Synthesis of Related Literature and Studies

The field of speech-to-text (STT) technology has rapidly evolved, integrating artificial intelligence (AI) and machine learning (ML) to support diverse applications across multiple sectors. From healthcare and education to customer service, STT applications are transforming human-computer interaction and facilitating accessibility for those with various needs. Recent literature highlights both the advancements and challenges within this field, drawing attention to key themes, such as technological progress, ethical concerns, accessibility, inclusivity, and performance optimization. The synthesis of related studies uncovers critical insights and identifies emerging areas for research, laying the groundwork for the development of a more innovative, ethical, and inclusive STT application.

STT technology has undergone significant historical development, driven primarily by advancements in AI and deep learning algorithms. Early contributions by Bhatt et al. (2021) and Borowiec et al. (2021) showcased the foundational role of deep learning, specifically convolutional neural networks (CNNs), which have influenced modern STT systems. These advancements are now integrated into applications that process complex multi-turn conversations and diverse speech

patterns (Bi, 2020; Nyatsanga et al., 2023). A key trend in STT research is the merging of voice and gesture recognition technologies, as highlighted by Luo et al. (2023), enhancing intuitive human-machine interfaces and broadening the potential applications for this technology.

A prominent theme in STT literature is the continuous improvement of accuracy and performance metrics. Studies such as those by Al-Makhadmeh and Tolba (2019) and Zhang et al. (2022) underscore the importance of using ensemble deep learning methods to enhance accuracy, particularly in challenging environments or with accented speech. Furthermore, WeNet 2.0 (Zhang et al., 2022) exemplifies the cutting-edge tools that refine recognition accuracy, setting new benchmarks for STT performance. Despite these advancements, challenges persist in ensuring system adaptability for diverse linguistic backgrounds, highlighting the need for further exploration into multilingual and accented speech recognition to create more inclusive technologies.

Alongside technological developments, ethical and privacy concerns have emerged as essential considerations. Studies by Lund et al. (2023) and Dergaa et al. (2023) stress the importance of ethical frameworks in the design and deployment of STT systems. These works advocate transparency, particularly in applications that involve sensitive contexts such as healthcare and academic writing. Further, Latif et al. (2020) emphasizes the necessity of integrating ethical and human rights considerations throughout the STT application lifecycle to safeguard user privacy and ensure responsible use. Literature consistently signals those ethical challenges, including algorithmic bias and data security, must be

addressed to foster user trust and safeguard sensitive information, especially in areas where speech data may involve personal or confidential material.

Accessibility and inclusivity also feature prominently in STT research. Koenecke et al. (2020) and Mengesha et al. (2021) reveal notable disparities in how STT systems perform across different demographic groups, particularly affecting African American users. This highlights the need for culturally sensitive and inclusive STT applications to avoid marginalizing specific user populations. Additionally, accessibility research by Botelho (2021) underscores the importance of ensuring that STT technology is usable for individuals with disabilities, with accessibility features tailored to various user needs. In educational contexts, Gupta and Sengupta (2021) and Shaheen and Watulak (2019) advocate for more inclusive speech-to-text tools to provide equal learning opportunities for all students. This body of literature calls for further investment in bias mitigation and inclusive design principles to create an STT landscape that is accessible to diverse user groups.

STT technology applications span several domains, including healthcare, customer service, and security. Research in Automatic Speech Recognition (ASR) has demonstrated its utility in language identification and multilingual applications (Semanticscholar). However, the persistence of racial disparities in ASR accuracy (Koenecke et al., 2020) underscores the need for more equitable STT solutions. Healthcare applications, such as clinical documentation systems, show promise in leveraging STT to improve efficiency, though regulatory compliance and accuracy remain key concerns (Semanticscholar). Further, Bain et al. (2023) explore noise-

resistant STT systems like WhisperX, which support long-form audio transcription, broadening the potential use cases for STT applications in real-time and noisy environments.

In summary, the review of related literature reflects the interdisciplinary nature of STT technology and emphasizes the need for continued research in addressing accuracy, inclusivity, ethical considerations, and accessibility. Future advancements in STT technology should aim to integrate AI and ML improvements while adhering to ethical principles that safeguard user privacy and fairness. Moreover, developing culturally sensitive, inclusive models will ensure that STT systems serve a broader range of users effectively, fostering a more equitable digital environment. This synthesis highlights the necessity of a balanced approach that combines technological innovation with ethical and inclusive frameworks, which will be crucial for the responsible advancement of STT technology.

Chapter 3

METHODOLOGY

Research Design

The researchers will employ a quantitative and descriptive research approach in conducting a comparative study on Speech-to-Text applications. The emphasis will be on gathering numerical or statistical data to analyze and compare various aspects of these applications. The quantitative research methodology will involve surveys to collect relevant data. The descriptive research component will focus on providing insights into the comparison of different Speech-to-Text applications.

Source of Data

The target respondents for this study will be students from the Polytechnic University of the Philippines Quezon City Campus. The student population size is estimated to be 1,781 from the following programs: Bachelor of Science in Information Technology, Bachelor of Science in Business Administration Major in Human Resource Management, Bachelor of Science in Business Administration Major in Marketing Management, Bachelor of Science in Business Teacher's Education Major in Home Economics, Bachelor of Science in Business Teacher's Education Major in Information Communication Technology, Bachelor of Science in Business Administration Major in Entrepreneurship, Bachelor of Public Administration Major in Public Financial Management, Bachelor of Public

Administration Major in Fiscal Administration, and Diploma in Office Management Technology.

Using Slovin's formula to determine the sample size, researchers calculated that 327 students would be needed. These students will be equally divided among the 9 programs.

Table 1

Sample Size Per Program

SAMPLE SIZE PROGRAM		
PROGRAM	**SAMPLE**	**PERCENTAGE**
Bachelor of Science in Information Technology	93	28.4%
Bachelor of Science in Business Administration Major in Human Resource Management	73	22.3%
Bachelor of Science in Business Administration Major in Marketing Management	68	20.8%
Bachelor of Technology and Livelihood Education Major in Home Economics	9	2.8%
Bachelor of Technology and Livelihood Education Major in Information and Communication Technology	18	5.5%
Bachelor of Science in Entrepreneurship	53	16.2%
Bachelor of Public Administration Major in Public Financial Management	7	2.1%
Bachelor of Public Administration Major in Fiscal Administration	2	0.6%
Diploma in Office Management Technology	4	1.2%
TOTAL:	**327**	**100%**

The study employs stratified random sampling, which involves dividing the population into different strata based on their program and then randomly selecting respondents from each stratum. This method ensures that the sample accurately

represents the distribution of the entire student population across the various programs. The researchers believe this approach is appropriate for our study because it allows us to capture diverse perspectives from all programs. The criteria for selecting participants are straightforward: respondents must have experience using speech-to-text applications. This includes students who have used these applications for any purpose—academic, personal, or professional. Additionally, the researchers will include deaf individuals to ensure that the study captures the perspectives and experiences of users who may rely heavily on speech-to-text applications for communication.

The data collection will focus on the student population of the Polytechnic University of the Philippines, Quezon City Campus, excluding university faculty and staff. The researchers concentrate on students from the specified programs who have used speech-to-text applications.

Research Instrument

The study's research instrument is a survey questionnaire that is aligned with the study's statement of the problem. The questionnaire consists of four parts, each focusing on critical aspects of speech-to-text applications would be composed of 4 (four) parts comprised the demographics of the respondents, as well as the level of awareness of the respondents to the speech text application (1) portability, (2) security, (3) functionality, (4) and benefits, the level of agreement of users of alternative communication model of speech-to-text application in terms of (1) interaction frequency, (2) utilization of features, (3) sustained involvement, (4) and user feedback, and the satisfaction level of the respondents using speech

text application in terms of (1) operability, (2) functionality, (3) user-friendliness, (4) and effectiveness. The researchers will use an online survey questionnaire through Google Forms which will be distributed to the respondents.

Data Gathering Procedure

The researchers will design an online survey using Google Forms to gather relevant data for their study. Before distributing the survey link, participants will receive a brief orientation about the study's purpose and assurance regarding the confidentiality of their personal information. Here's how the data collection process will unfold:

1. Participants will be asked for their consent, approval of the survey, and given a brief orientation about the study's nature. They will also be assured of the safety of their personal data.
2. The researchers will distribute the survey link via email or social media accounts of consenting participants. These platforms will be used solely for survey distribution.
3. Participants will be given sufficient time to complete the survey.
4. Completed questionnaires will be digitally compiled and stored in Google Sheets. Each response will undergo statistical analysis to calculate percentages and weighted means, facilitating data interpretation and analysis by the researchers.

Ethical Consideration

The researchers made sure that all participants in the study fully understood the topic and purpose of the research. They were informed that their responses in the survey would only be used as data points. The privacy of their personal information was emphasized, ensuring strict confidentiality according to the Data Privacy Act of 2012. No names, emails, or phone numbers would be retained by the researchers for any other purposes. Additionally, the safety and comfort of the respondents were prioritized. They were encouraged to complete the survey in environments where they felt secure, whether it be at home, in a private setting, or on campus. The researchers also took measures to ensure that participants were not exposed to any harm or discomfort during the process, both physically and mentally.

Data Case Analysis

The researchers will analyze the data gathered from respondents using statistical techniques. They'll utilize methods such as calculating percentages, frequencies, and weighted means to interpret the findings.

A. **Percentage and Frequency** were used to calculate the demographic profile from the study's respondents, the formula for the percentage and frequency:

$$P = fn\,x\,100$$

Wherein:

p = percentage

f = frequency

n = total number of respondents

$$F = nN$$

Wherein:

n = the total number of responses

N = the total number of respondents

B. **Weighted Mean** will be used to determine and measure the response to the questions from the survey questionnaires, profiles such as age and identification of the respondents are excluded for this method, the formula for weighted mean:

$$Mean = \sum fn$$

Wherein:

f = number of occurrences from the number of respondents

Mean = population mean or total score

n = the number of sources observed from the respondents

C. **Slovin's Formula** will be used to calculate the sample size from the total population.

$$n = \frac{N}{1 + Ne^2}$$

Wherein:

n = sample size

N = total population

e = margin of error

Table 2

Likert Scale – Level of Awareness

Numeric Value	Scale	Verbal Interpretation	Description
5	4.20 - 5.00	Strongly Aware	The respondents are strongly aware of the statement
4	3.40 - 4.19	Aware	The respondents are aware with most of the statement's claims
3	2.60 - 3.39	Neutral	The respondents are neutral with most of the statement's claims
2	1.80 - 2.56	Unaware	The respondents are unaware with most of the statement's claims
1	1.00 - 1.79	Strongly Unaware	The respondents strongly disagree with most of the statement's claims

Table 2 presents a Likert scale used to measure the level of awareness of respondents regarding a particular statement or set of statements. The scale ranges from 1 to 5, where a numeric value of 5 (corresponding to a scale of 4.20 - 5.00) indicates that respondents are "Strongly Aware" of the statement, suggesting a high level of familiarity and understanding. A value of 4 (scale of 3.40 - 4.19) represents respondents being "Aware," indicating a reasonable level of awareness. A neutral position is denoted by a value of 3 (scale of 2.60 - 3.39), implying that respondents neither agree nor disagree with the awareness

statements. Lower values, such as 2 (scale of 1.80 - 2.56) and 1 (scale of 1.00 - 1.79), correspond to "Unaware" and "Strongly Unaware," respectively, indicating decreasing levels of awareness about the statements (Likert, 1932).

Table 3

Likert Scale – Level of Agreement

Numeric Value	Scale	Verbal Interpretation	Description
5	4.20 - 5.00	Strongly Agree	The respondents strongly agree with the statement
4	3.40 - 4.19	Agree	The respondents agree with most of the statement's claims
3	2.60 - 3.39	Neutral	The respondents are neutral with most of the statement's claims
2	1.80 - 2.56	Disagree	The respondents disagree with most of the statement's claims
1	1.00 - 1.79	Strongly Disagree	The respondents strongly disagree with most of the statement's claims

Table 3 illustrates a Likert scale designed to assess the level of agreement among respondents with specific statements. Like Table 2, this scale also ranges from 1 to 5. A numeric value of 5 (scale of 4.20 - 5.00) represents "Strongly Agree," showing a strong concurrence with the statements. A value of 4 (scale of 3.40 - 4.19) signifies those respondents "Agree," reflecting moderate agreement. A score of 3 (scale of 2.60 - 3.39) indicates a "Neutral" stance, where respondents neither agree nor disagree. Scale scores of 2 (scale of 1.80 - 2.56) and 1 (scale of 1.00 -

1.79) denote "Disagree" and "Strongly Disagree," respectively, illustrating varying degrees of disagreement with the statements (Likert, 1932).

Table 4

Likert Scale – Level of Satisfaction

Numeric Value	Scale	Verbal Interpretation	Description
5	4.20 - 5.00	Strongly Satisfied	The respondents strongly satisfied with the statement
4	3.40 - 4.19	Satisfied	The respondents satisfied with most of the statement's claims
3	2.60 - 3.39	Neutral	The respondents are neutral with most of the statement's claims
2	1.80 - 2.56	Unsatisfied	The respondents unsatisfied with most of the statement's claims
1	1.00 - 1.79	Strongly Unsatisfied	The respondents strongly unsatisfied with most of the statement's claims

Table 4 provides a Likert scale that gauge's respondents' levels of satisfaction with a statement or series of statements. The scale also spans from 1 to 5, where a score of 5 (scale of 4.20 - 5.00) corresponds to "Strongly Satisfied," reflecting a high degree of contentment. A score of 4 (scale of 3.40 - 4.19) indicates "Satisfied," suggesting general contentment with the statements. A score of 3 (scale of 2.60 - 3.39) represents a "Neutral" position, indicating neither satisfaction nor dissatisfaction. The lower end of the scale, with scores of 2 (scale of 1.80 - 2.56) and 1 (scale of 1.00 - 1.79), represents "Unsatisfied" and "Strongly

Unsatisfied," respectively, indicating increasing levels of dissatisfaction with the statements (Likert, 1932).

Chapter IV

RESULTS AND DISCUSSIONS

This chapter will present the quantitative findings derived from the data collection process. It will focus on the tables containing statistical analyses and interpretations, providing a detailed overview of the results obtained.

Demographic Profile of Respondents

The demographic profile of respondents serves as the foundational understanding of the sample population. This section delves into crucial aspects such as the courses of the respondents' offering insights into the diverse backgrounds of the participants in the study.

Table 5

Distribution of Respondents According to Courses

Educational Attainment	Frequency	Percentage
BBTLEDHE	9	2.8%
BTLEDICT	18	5.5%
BPA PFM	7	2.1%
BPAFA	2	0.6%
BSBAHRM	73	22.3%
BSBA-MM	68	20.8%

BSENTREP	53	16.2%
BSIT	93	28.4%
DOMT	4	1.2%
Total	**327**	**100%**

The distribution of courses among the survey participants indicates a balanced selection in the sample. Most (28.4%) were students of Bachelor of Science in Information Technology (BSIT) while nine (2.8%) were Bachelor of Business Technology and Livelihood Education major in Home Economics (BBTLEDHE). Moreover, eighteen (5.5%) were students of Bachelor of Business Technology and Livelihood Education major in Information Communication and Technology (BTLEDICT) while seven (2.1%) were Bachelor of Public Administration Major in Public Financial Management (BPA PFM). Additionally, two of the respondents were students of Bachelor of Public Administration with specialization in Fiscal Administration (BPAFA) while the students of Bachelor of Science in Business Administration major in Human Resource Management (BSBAHRM) accounted for seventy-three (22.3%). Furthermore, sixty-eight (20.8%) were Bachelor of Science in Business Administration major in Marketing Management (BSBA-MM) while fifty-three (16.2%) were from Bachelor of Science in Entrepreneurship (BSENTREP). Lastly, four (1.2%) were from Diploma in Office Management Technology (DOMT).

User's Level of Awareness

Table 6

Distribution of Respondents According to Portability

Statement	*M*	*SD*	Verbal Interpretation
1. I am aware of the speech to text applications compatibility on different platforms (e.g., mobile devices, desktop computers, web browsers)?	4.36	.690	*Strongly Aware*
2. I am aware of its compatibility with several operating systems (iOS, Android, Windows, and MacOS).	4.11	.823	*Aware*
3. The speech-to-text application I use provides customizable accessibility options to meet my specific needs.	3.81	.951	*Aware*
4. I am aware of the various accessibility features offered by speech-to-text applications to assist users with different abilities.	3.92	.910	*Aware*
5. I am aware of the channels or sources of the voice to text program in order to keep up with its current	3.69	.955	*Aware*

features. (Official websites, social media, Newsletters, User forums, Others.)?			
Total	**3.98**	**.866**	***Aware***

Note: *Very Unaware = 1.00 – 1.79, Unaware = 1.80 – 2.59, Neutral = 2.60 – 3.39, Aware = 3.40 – 4.19, Strongly Aware = 4.20 – 5.00*

Analyzing the respondents' portability indicated that most of them understood what was on the list, as depicted by the score of $(M = 3.98, s = .866)$ indicating agreement. Moreover, the statement "I am aware of the speech to text applications compatibility on different platforms (e.g., mobile devices, desktop computers, web browsers)?" got the highest score of $(M = 4.36, s = .690)$ indicating an aware among respondents. On the other hand, the statement "I am aware of the channels or sources of the voice to text program in order to keep up with its current features. (Official website, social media, Newsletters, User forums, Others.)?" given that their average rating is significantly lower $(M = 3.69, s = .995)$ than others.

Table 7

Distribution of Respondents According to Security

Statement	***M***	***SD***	**Verbal Interpretation**

1. I am aware that the spoken text application's security procedures will safeguard saved data.	3.45	1.020	*Aware*
2. I am aware of the encryption techniques used by the speech to text application to safeguard my voice data during transmission and storage.	3.33	1.102	*Neutral*
3. I am aware of the recent changes and advancements in speech-to-text technology.	3.52	1.085	*Aware*
4. I am aware of the app's capacity to select the level of control offered to me.	3.54	1.050	*Aware*
5. The application provides transparent information about the data retention policies, indicating how long my voice recordings and transcriptions are stored.	3.49	1.018	*Aware*
Total	**3.47**	**1.055**	***Aware***

Note: *Very Unaware = 1.00 – 1.79, Unaware = 1.80 – 2.59, Neutral = 2.60 – 3.39, Aware = 3.40 – 4.19, Strongly Aware = 4.20 – 5.00*

Regarding security, most respondents agreed with moderate consensus reflected in a score $(M = 3.47, s = 1.055)$. More importantly, according to the

results, most participants aware that “the app's capacity to select the level of control offered to me.” ($M = 3.54, s = 1.050$), while few of them observed that they were aware of the encryption techniques used by the speech to text application to safeguard my voice data during transmission and storage.($M = 3.33, s = 1.102$).

Table 8

Distribution of Respondents According to Functionality

Statement	*M*	*SD*	Verbal Interpretation
1. I find the interface of the speech-to-text application I use to be user-friendly and intuitive.	3.97	.852	*Aware*
2. The speech-to-text application effectively translates my spoken words into accurate text.	3.78	.846	*Aware*
3. I can easily customize and personalize settings on the speech-to-text application to enhance my user experience.	3.68	.928	*Aware*
4. I have experienced consistent and reliable performance from the speech-to-text application in various situations (e.g., noisy environments).	3.61	.965	*Aware*

5. I feel a sense of accomplishment and worth when using mobile AR applications for literacy-related activities.	3.67	.910	*Aware*
Total	**3.74**	**.900**	***Aware***

Note: *Very Unaware = 1.00 – 1.79, Unaware = 1.80 – 2.59, Neutral = 2.60 – 3.39, Aware = 3.40 – 4.19, Strongly Aware = 4.20 – 5.00*

In reference to functionality, most respondents showed strong concurrence as indicated by the $(M = 3.74, s = .900)$. In particular, the findings show that a good number of participants highly aware whereby there are respondents find the interface of the speech-to-text application I use to be user-friendly and intuitive.$(M = 3.97, s = .852)$, but still some survey respondents have experienced consistent and reliable performance from the speech-to-text application in various situations (e.g., noisy environments).$(M = 3.61, s = .965)$.

Table 8

Distribution of Respondents According to Benefits

Statement	***M***	***SD***	**Verbal Interpretation**
1. The speech-to-text application I use has significantly improved my communication efficiency.	3.87	.865	*Aware*

2. I believe speech-to-text applications contribute positively to inclusive communication for individuals with diverse needs.	4.13	.769	*Aware*
3. Using a speech-to-text application has positively impacted my professional or academic performance.	3.77	.893	*Aware*
4. The benefits I experience from the speech-to-text application extend beyond communication, positively affecting other aspects of my life.	3.78	.891	*Aware*
5. I would recommend speech-to-text applications to others based on the positive benefits I have experienced.	4.05	.758	*Aware*
Total	**3.92**	**.835**	***Aware***

Note: *Very Unaware = 1.00 – 1.79, Unaware = 1.80 – 2.59, Neutral = 2.60 – 3.39, Aware = 3.40 – 4.19, Strongly Aware = 4.20 – 5.00*

The consensus among the respondents regarding benefit ($M = 3.92, s = .835$). Remarkably, large numbers of respondents had a strong believe that speech-to-text applications contribute positively to inclusive communication for individuals with diverse needs as shown by the highest mean score ($M = 4.13, s = .769$). Nevertheless, a group of the participants using a speech-to-text

application has positively impacted my professional or academic performance as signified by the lowest average mark ($M = 3.77, s = .893$).

User's Level of Agreement

Table 9

Distribution of Respondents According to Interaction Frequency

Statement	*M*	*SD*	Verbal Interpretation
1. I find myself engaging with the speech-to-text application regularly for communication purposes.	3.61	.962	*Aware*
2. The speech-to-text application has noticeably increased the frequency of my communication with friends and family.	3.61	.886	*Aware*
3. The speech-to-text application helps me maintain consistent communication with individuals who may have different communication needs.	3.83	.785	*Aware*
4. I feel that the speech-to-text application has positively influenced my overall social engagement.	3.66	.847	*Aware*

5. I find myself relying on the speech-to-text application in both formal and informal communication scenarios.	3.34	1.128	*Aware*
Total	**3.61**	**.922**	***Aware***

Note: *Strongly Disagree = 1.00 – 1.79, Disagree = 1.80 – 2.59, Neutral = 2.60 – 3.39, Agree = 3.40 – 4.19, Strongly Agree = 4.20 – 5.00*

Most respondents' degree of agreement was average in terms of interaction frequency, as shown by the score $(M = 3.61, s = .922)$. Hence, few respondents reported that the speech-to-text application helps me maintain consistent communication with individuals who may have different communication needs $(M = 3.83, s = .785)$. In contrast, a small number stated they find myself engaging with the speech-to-text application regularly for communication purposes$(M = 3.61, s = .962)$ and the speech-to-text application has noticeably increased the frequency of my communication with friends and family $(M = 3.61, s = .886)$.

Table 10

Distribution of Respondents According to Utilization of Features

Statement	***M***	***SD***	**Verbal Interpretation**
1. I am aware of and realtime transcription offered by the speech-to-text application. The application	3.83	.764	*Aware*

contributes positively to my communication efficiently.			
2. I am aware of the voice command recognition offered by the speech-to-text application. Aligns well with my diverse communication requirements.	3.86	.736	*Aware*
3. I am aware of the accuracy of the speech-to-text application in the voice recognition range. The application provides significantly to wide-ranging space discussion.	3.77	.762	*Aware*
4. I am aware of the speech-to-text application's punctuation and formatting. I found the feature excellent.	3.63	.855	*Aware*
5. I make efficient use of the features of the application as it holds multilingual support. Enhancing the overall effectiveness of the communication model.	3.81	.798	*Aware*
Total	**3.78**	**.783**	***Aware***

Note: *Strongly Disagree = 1.00 – 1.79, Disagree = 1.80 – 2.59, Neutral = 2.60 – 3.39, Agree = 3.40 – 4.19, Strongly Agree = 4.20 – 5.00*

Regarding utilization of features, most respondents aware with moderate consensus reflected in a score $(M = 3.78, s = .783)$. More importantly, according to the results, most participants were aware that "the voice command recognition offered by the speech-to-text application. Aligns well with my diverse communication requirements" $(M = 3.86, s = .736)$, while few of them observed were aware of the speech-to-text application's punctuation and formatting. I found the features excellent. $(M = 3.63, s = .855)$.

Table 11

Distribution of Respondents According to Sustained Involvement

Statement	*M*	*SD*	Verbal Interpretation
1. I am constantly inspired to continue using the speech-to-text program in an active manner.	3.70	.796	*Aware*
2. My continued interest and participation are greatly aided by the application.	3.73	.741	*Aware*
3. I picture myself using the application's communication model for a considerable amount of time.	3.64	.839	*Aware*
4. Because it supports various communication scenarios, the	3.68	.726	*Aware*

application's design encourages my constant interaction with the application.			
5. The speech-to-text application effectively retains my attention and commitment over time.	3.57	.788	*Aware*
Total	**3.66**	**.778**	***Aware***

Note: *Strongly Disagree = 1.00 – 1.79, Disagree = 1.80 – 2.59, Neutral = 2.60 – 3.39, Agree = 3.40 – 4.19, Strongly Agree = 4.20 – 5.00*

Most respondents' degree of agreement was average in terms of sustained involvement, as shown by the score ($M = 3.66, s = .778$). Hence, few respondents that my continued interest and participation are greatly aided by the application. ($M = 3.73, s = .741$). In contrast, a small number stated that the speech-to-text application effectively retains my attention and commitment over time ($M = 3.57, s = .788$).

Table 12

Distribution of Respondents According to User Feedback

Statement	***M***	***SD***	**Verbal Interpretation**
1. I feel that my feedback on the speech-to-text application is valued and considered for improvements.	4.02	.755	*Aware*

2. The developers of the application actively seek user opinions and suggestions.	3.87	.781	*Aware*
3. I am satisfied with the responsiveness of the development team to user feedback.	3.82	.764	*Aware*
4. User feedback has been instrumental in enhancing the application's functionality.	4.02	.739	*Aware*
5. I believe that the speech-to-text application is continually evolving based on user input and recommendations	4.09	.693	*Aware*
Total	**3.96**	**.746**	***Aware***

Note: *Strongly Disagree = 1.00 – 1.79, Disagree = 1.80 – 2.59, Neutral = 2.60 – 3.39, Agree = 3.40 – 4.19, Strongly Agree = 4.20 – 5.00*

The consensus among the respondents regarding user feedback ($M = 3.96, s = .746$). Remarkably, large numbers of respondents had a strong believe that the speech-to-text application is continually evolving based on user input and recommendations as shown by the highest mean score ($M = 4.09, s = .693$). Nevertheless, a group of the participants were satisfied with the responsiveness of the development team to user feedback as signified by the lowest average mark ($M = 3.82, s = .764$).

User's Level of Satisfaction

Table 13

Distribution of Respondents According to Operability

Statement	*M*	*SD*	Verbal Interpretation
1. The speech-to-text application is easy to operate.	4.11	.655	*Aware*
2. Navigating through the features of the speech-to-text application is straightforward.	3.97	.664	*Aware*
3. I find it easy to customize settings within the speech-to-text application.	3.84	.766	*Aware*
4. The overall operability of the speech-to-text application meets my expectations.	3.94	.700	*Aware*
5. The response time of the speech-to-text application is satisfactory.	3.91	.674	*Aware*
Total	3.95	.692	*Aware*

Note: *Very Unsatisfied = 1.00 – 1.79, Unsatisfied = 1.80 – 2.59, Neutral = 2.60 – 3.39, Satisfied = 3.40 – 4.19, Strongly Satisfied = 4.20 – 5.00*

In reference operability, most respondents showed strong concurrence as indicated by the $(M = 3.95, s = .692)$. In particular, the findings show that a good number of participants highly aware whereby there are respondents who aware in

the statement “The speech-to-text application is easy to operate” ($M = 4.11, s = .655$), but still some survey respondents find it easy to customize settings within the speech-to-text application ($M = 3.84, s = .766$).

Table 14

Distribution of Respondents According to Functionality

Statement	*M*	*SD*	Verbal Interpretation
1. I can access the speech-to-text application easily from different devices.	3.99	.739	*Aware*
2. The speech-to-text application is accessible whenever I need it.	3.97	.785	*Aware*
3. The application's features are available across various platforms (e.g., mobile, desktop).	4.10	.737	*Aware*
4. The speech-to-text application adapts well to different network conditions.	3.91	.740	*Aware*
5. The application's reachability aligns with my communication needs.	3.99	.705	*Aware*
Total	**3.99**	**.741**	***Aware***

Note: *Very Unsatisfied = 1.00 – 1.79, Unsatisfied = 1.80 – 2.59, Neutral = 2.60 – 3.39, Satisfied = 3.40 – 4.19, Strongly Satisfied = 4.20 – 5.00*

Regarding functionality, most respondents aware with moderate consensus reflected in a score $(M = 3.99, s = .741)$. More importantly, according to the results, most participants were aware that "3. The application's features are available across various platforms (e.g., mobile, desktop)." $(M = 4.10, s = .737)$, while few of them observed were aware that the speech-to-text application adapts well to different network conditions $(M = 3.91, s = .740)$.

Table 15

Distribution of Respondents According to User-friendliness

Statement	*M*	*SD*	Verbal Interpretation
1. The interface of the speech-to-text application is user-friendly.	3.97	.785	*Aware*
2. I find it easy to learn how to use the speech-to-text application.	4.10	.737	*Aware*
3. The application provides clear instructions for its features and functions.	3.91	.740	*Aware*
4. The design of the speech-to-text application enhances user experience.	3.99	.705	*Aware*

5. Overall, I find the speech-to-text application user-friendly.	4.15	.664	*Aware*
Total	**4.02**	**.726**	***Aware***

Note: *Very Unsatisfied = 1.00 – 1.79, Unsatisfied = 1.80 – 2.59, Neutral = 2.60 – 3.39, Satisfied = 3.40 – 4.19, Strongly Satisfied = 4.20 – 5.00*

Analyzing the respondents' user-friendliness indicated that most of them understood what was on the list, as depicted by a mean score of $(M = 4.02, s = .726)$ indicating awareness. Moreover, there were respondents find it easy to learn how to use the speech-to-text application" with highest mean score $(M = 4.10, s = .737)$ indicating strong consent among respondents. On the other hand, the statement "The application provides clear instructions for its features and functions" $(M = 3.91, s = .740)$, got the lowest score.

Table 16

Distribution of Respondents According to Effectiveness

Statement	***M***	***SD***	**Verbal Interpretation**
1. The speech-to-text program accurately turns spoken words into printed words	4.17	.682	*Aware*
2. The application effectively recognizes different accents and speech patterns, leading to	4.05	.719	*Aware*

enhanced user experience and improved accuracy in transcription.			
3. I haven't experienced errors in the transcriptions generated by the application.	4.06	.705	*Aware*
4. The application consistently delivers accurate results in various environments.	4.12	.707	*Aware*
5. I consistently achieve reliable and satisfactory results when using the application	3.89	.789	*Aware*
Total	**4.06**	**.720**	***Aware***

Note: *Very Unsatisfied = 1.00 – 1.79, Unsatisfied = 1.80 – 2.59, Neutral = 2.60 – 3.39, Satisfied = 3.40 – 4.19, Strongly Satisfied = 4.20 – 5.00*

The consensus among the respondents regarding effectiveness $(M = 4.06, s = .720)$. Remarkably, large numbers of respondents says that the speech-to-text program accurately turns spoken words into printed words as shown by the highest mean score $(M = 4.17, s = .682)$. Nevertheless, a group of the participants consistently achieve reliable and satisfactory results when using the application as signified by the lowest average mark $(M = 3.89, s = .789)$.

Table 18

Multiple Regression Analysis with Operability

Predictor	**Estimate**	**SE**	β	**t**	**p**
Intercept	7.329	.901		8.135	<.001
Interaction Frequency	.007	.054	.010	.138	.890
Utilization of Features	.232	.064	.245	3.609	<.001
Sustained Involvement	.159	.063	.190	2.534	.012
User Feedback	.253	.060	.263	4.238	<.001

This research was conducted to determine if interaction frequency, utilization of features, sustained involvement and user feedback predict the operability. It is hypothesized that the four predictors will be positively associated with operability. Results show that the 32% of variance is explained by the four predictors, $(F(3,239) = 37.6, p < .001)$. Specifically, utilization of features $(B = .06, t = 3.609, p < .001)$ and user feedback $(B = .26, t = 4.238, p < .001)$ are positively associated with operability. On the other hand, interaction frequency is not significantly related with the outcome variables $(B = .01, t = .138, p = .890)$ and sustained involvement $(B = .263, t = 4.238, p = .012)$.

Table 19

Multiple Regression Analysis with Functionality

Predictor	**Estimate**	**SE**	β	**t**	**p**
Intercept	5.956	.965		6.172	<.001
Interaction Frequency	.065	.058	.079	1.120	.263
Utilization of Features	.233	.069	.225	3.387	<.001
Sustained Involvement	.128	.067	.140	1.902	.058
User Feedback	.307	.064	.293	4.809	<.001

This research was conducted to determine if interaction frequency, utilization of features, sustained involvement and user feedback predict the functionality. It is hypothesized that the four predictors will be positively associated with functionality. Results show that the 32% of variance is explained by the four predictors, $(F(3.785) = 38.5, p < .001)$. Specifically, utilization of features $(B = .22, t = 3..387, p < .001)$ and user feedback $(B = .29, t = 4.809, p < .001)$ are positively associated with functionality. On the other hand, interaction frequency is not significantly related with the outcome variables $(B = .08, t = 1.12, p = .263)$ and sustained involvement $(B = .140, t = 1.90, p = .058)$.

Table 20

Multiple Regression Analysis with User-friendliness

Predictor	**Estimate**	**SE**	β	**t**	**p**
Intercept	6.081	.867		7.012	<.001
Interaction Frequency	-.109	.052	-.139	-2.100	.037
Utilization of Features	.334	.062	.340	5.412	<.001
Sustained Involvement	.147	.060	.169	2.439	.015
User Feedback	.374	.057	.375	6.520	<.001

This research was conducted to determine if interaction frequency, utilization of features, sustained involvement and user feedback predict the user-friendly. It is hypothesized that the four predictors will be positively associated with operability. Results show that the 32% of variance is explained by the four predictors, $(F(3,891) = 34.7, p < .001)$. Specifically, utilization of features $(B = .06, t = 3.609, p < .001)$ and user feedback $(B = .26, t = 4.238, p < .001)$ are positively associated with user-friendly. On the other hand, interaction frequency is not significantly related with the outcome variables $(B = .01, t = .138, p = .890)$ and sustained involvement $(B = .263, t = 4.238, p = .012)$.

Table 21

Multiple Regression Analysis with Effectiveness

Predictor	**Estimate**	**SE**	β	**t**	**p**
Intercept	2.638	1.023		2.579	.010
Interaction Frequency	.114	.061	.125	1.856	.064
Utilization of Features	.258	.073	.226	3.545	<.001
Sustained Involvement	.243	.071	.240	3.407	<.001
User Feedback	.222	.068	.191	3.279	.001

This research was conducted to determine if interaction frequency, utilization of features, sustained involvement and user feedback predict effectiveness. It is hypothesized that the four predictors will be positively associated with effectiveness. Results show that the 32% of variance is explained by the four predictors, $(F(3,335) = 32.7, p < .001)$. Specifically, utilization of features $(B = .23, t = 3.545, p < .001)$ and sustained involvement $(B = .24, t = 3.407, p < .001)$ are positively associated with effectiveness. On the other hand, interaction frequency is not significantly related to the outcome variables $(B = .13, t = .1856, p = .064)$ and user feedback $(B = .191, t = 3.279, p = .001$

Chapter 5

REVIEW OF RELATED LITERATURE

This chapter presented the summary of findings, conclusions and recommendations which were based on the result of the study conducted.

Summary of Findings

Based on the investigation, the following summary of findings was drawn:

1. The findings of the study revealed that awareness in the use of the particular functionalities of the speech-to-text applications regarding portability and cross-platform capability was highly observed among the respondents, therefore a mean score of 4.36. Awareness on the update and new features and information was less recorded with a score of 3.69. Respondents demonstrated moderate awareness about the security element of data protection (Mean = 3.54), whereas knowledge about encryption methods was lower (Mean = 3.33). In the functional aspects, the user-friendly interfaces were liked and accurate translation of texts; however, performance in noisy environments was less than satisfactory with a mean score of 3.61. More critically, these applications were extremely valued for the assumed benefits-these being their roles in enhancing communication efficiency and inclusivity. The high ratings point to a range of values from 4.13 mean. This echoes insights from Bumann (2023), wherein use increasingly values intuitive and accessible

communication tools, indicating a need for further development where awareness and functionality could be improved.

2. Respondents demonstrated high use of speech-to-text applications as measured in interaction frequency with a mean score of 3.61. They considerably agreed about application impact on maintaining fluent communication with a mean score of 3.83 and acknowledged the proper utilization of features like real-time transcription and voice command recognition with Mean = 3.86. These features were especially appreciated in supporting multiple communication requirements, which is growing as more such technologies become part and parcel of an individual's daily life. The outcomes of this study are in line with Haleem et al. (2022), who emphasized that adaptive features play a vital role in making digital communication. Their importance lies specifically in building inclusivity and bridging differences in accessibility for user requirements.
3. Satisfaction levels of the respondents were very high, especially on operability by a mean of 3.97 and functionality with a mean of 3.78. Applications were perceived to be friendly to users, and their intended purpose as far as functionality was concerned was effective, but some concerns rose based on performance variability in noisy environments, for example. This presents an opportunity for improvements in the robustness of applications under adverse conditions. In this aspect, usability and reliability are the most important factors in user satisfaction, especially where technology plays the bulk of the driving roles, as cited by Qaishammouri et al. (2020). It is essential to continue constant

performance with diverse settings, thereby continuously attempting a full understanding in fulfilling customer expectations.

4. The data were found to have an important positive relationship with the level of agreement among respondents to usage, such as interaction frequency and feature utilization. A higher engagement with application feature utilization was associated with a high level of satisfaction, thus further supporting the development of active feature usage to optimize user experiences. These results align with the study of Kalankesh et al. (2020), who found that active engagement with application features is a key determinant of user satisfaction and loyalty. Results indicate that setting up user familiarity with advanced features through effective onboarding and education can increase overall satisfaction and usage outcomes.
5. Findings indicate the possibility of designing an alternative communication application model that may comprise improved security features, better adaptability to different environmental conditions, and a more user-friendly interface. This will fill in the gap identified in existing applications while building on their strengths to provide a more holistic and user-centric solution. The recommendation made in this study aligns with the findings established by Noguerra (2023), as per which adaptability in real time and user-centric design should become the epitome of effective communication technologies. Such a model, by giving attention to these issues, would significantly help in making online communication access much better, efficient, and secure for most of the user groups.

Conclusions

The following conclusions were drawn based on the findings of the study:

1. The respondents showed a high degree of consciousness about the usability and functionality of speech-to-text applications. Those particularly recognized advantages of portability, user-friendly interfaces, and cross-platform compatibility. This awareness makes clear the role that these applications play in improving the efficiency of communication and creating equality. Such results align with Segun-Falade et al. (2024), who emphasized the impact of the use of mobile and cross-platform technologies on accessibility and productivity. These results underscore how indispensable speech-to-text applications are becoming for both everyday communication and for professional purposes.
2. Although respondents have moderate awareness of the data protection features to avail encryption, they presented ignorance about the advanced security measures that an app may have. It called for education regarding the better transparency of security protocols installed into an app. Sharma et al. (2020) highlights that trust and user confidence inform the adoption of communication tools; effective communication technologies should pay attention to user awareness of privacy and security measures. Addressing these gaps will reduce risks and lead to greater reliance on speech-to-text applications.
3. High levels of satisfaction were reported regarding the operability and functionality of speech-to-text applications. However, users pointed out continued issues, for instance, the decreased performance in noisy

environments. Resolution of these constraints would help increase the usability of these tools, especially in multifaceted situations. Reddy et al. (2023) had deduced that adaptability along with reliability is necessary to maintain user satisfaction. If the environment adaptability feature of voice-to-text applications is enhanced, it can increase the user experience of the utility manifold.

4. Statistical analysis showed a significant positive association between the degree of usage of an app and user satisfaction. Having more frequent interactions with application features, including transcription in real-time, strongly correlated with high satisfaction. Bitrián et al. (2021) reported similar findings; they cited that frequent feature utilization encourages appreciation for the capabilities of the applications. Such associations depict the need for stimulating interaction with the features of an application to realize maximum user satisfaction. Encouraging frequent use might enhance user loyalty and long-term adoption of these resources.
5. Based on the findings, the idea of an alternative model of communication application is highly suggested. The suggested model should implement a number of modern security attributes, better adaptability for various environments, and a more intuitive interface. Such attributes have been also supported by Kim et al. (2020), who mentioned user-centered design as one of the essential aspects to improve the usability of communication technologies. Addressing user feedback is essential to overcoming existing challenges while meeting the diverse needs of users. By incorporating

these improvements, the next generation of speech-to-text applications could set a new standard for communication efficiency and inclusivity.

Recommendations

Based on the findings presented and the conclusions drawn, the researcher recommends the following:

1. Researchers recommend including qualitative approaches such as case studies or structured interviews that will help gather a more in-depth sense of users' perception and specific needs related to speech-to-text applications. These approaches could better elucidate those factors that determine user satisfaction, especially regarding functionality and security. By taking the narrative perspective of users, future research will easily pinpoint unseen problems and improve the design of the applications much better.
2. Future studies should continue in further developing this work through experimentation on better app models that take care of safety and performance at the same time. Developers can concentrate on incorporating noise-adaptive technology along with intuitive security interfaces to overcome those gaps, as highlighted in the findings. This would thus ensure that applications continue to be versatile and practical for a wider variety of environments and users.
3. Improved education features concerning data encryption and protection must be implemented and tested within speech-to-text applications to bridge the identified knowledge gap regarding advanced security

measures. Developer guides, live demonstrations, or even periodic updates may be included in an application to enlighten users on data encryption and protection protocols. This could encourage informed usage with higher amounts of trust and adoption rates.

4. Studies suggest design strategies for applications that will keep users involved, like custom layouts, voice-to-text shortcuts, or integration with popular productivity tools. These kinds of improvements can strengthen the relationship between usage frequency and satisfaction by helping users have more reasons to interact with the application. Such features could be further advertised as part of promotional campaigns.
5. A longitudinal study should be conducted to monitor how users' preference and behaviour evolve as they keep using the updated speech-to-text applications. It would allow researchers to point out the long-term trends, like changes in satisfaction or adaptability in following technological trends. The key findings from such a study can serve as the basis of blueprints for iterative improvements that correspond to the changing needs of the users.