Jurnal Impresi Indonesia (JII)
p-ISSN: 2828-1284 e-ISSN: 2810-062x
website: https: //rivierapublishing.id/JII/ index.php / jii /index

Prototype of Smartglasses English Translator in the LSTM-Based Tourism Sector and Speech Recognition and Speech Synthesis Technology

Jonathan Widodo Wiji Saputra[1], Achmad Hamdan[2]
Universitas Negeri Malang, Indonesia
jonathan.widodo.2309356@students.um.ac.id, achmad.hamdan.ft@um.ac.id

Keywords
Language, Translator, Tourism, Smart glasses, LSTM

Abstract
Language barriers remain a significant challenge in the tourism sector, particularly in Indonesia, where communication between international tourists and local service providers often faces difficulties due to limited English proficiency. Existing translation tools, such as mobile apps and standalone devices, suffer from limitations like reliance on internet connections, one-way communication, and the need for additional devices. This research aims to develop a smart glasses prototype that integrates speech recognition, Long Short-Term Memory (LSTM)-based translation, and speech synthesis technologies to offer real-time, bidirectional communication. The prototype development involved several stages: literature review, design, tool collection, prototyping, and testing using the Tatoeba dataset, which includes 13,570 English-Indonesian sentence pairs. The smart glasses prototype was tested for translation accuracy, speech recognition, and synthesis performance, yielding BLEU scores of 36.5% for Indonesian-English and 39.8% for English-Indonesian, with accuracy rates of 80.8% and 87.8%, respectively. The prototype demonstrated effective real-time translation with speech-to-text and text-to-speech functionalities. These results show that the smart glasses prototype can significantly improve the quality of communication between tourists and local communities, enhancing the competitiveness of Indonesia's tourism industry. The study implies that wearable translation devices could become essential tools in tourism, reducing language barriers and fostering better cross-cultural interactions.

CC BY SA

INTRODUCTION

To overcome barriers in cross-lingual communication, various translation technologies have been developed, one of which is Google Translate, capable of translating text and voice between languages. In addition, research continues to be carried out to develop web-based translation technology and Android applications, including those that use augmented reality concepts to translate text from images into Indonesian (Fitriani Sopayati & Giyoto, 2023; Rakhmyta & Nurmalina, 2023; Widiana & Meilasari, 2024). Nowadays, translator devices such as *interpreter earphones* and *translator smart glasses* are also available as innovative solutions. However, these devices still have limitations, such as reliance on internet connections, one-way communication, and the need for additional devices such as smartphones, which can hinder user mobility. In fact, in the tourism sector, two-way communication has proven to be more often used and is considered more effective compared to one-way communication patterns. Two-way communication patterns allow the parties involved, such as service providers and customers, to interact actively. This allows for constructive feedback, which is very important in improving the quality of service and the tourist experience. Research by

Latifah and Muksin shows that effective communication in the form of mutual feedback increases the success of the communication process, which is relevant for the tourism sector (Latifah & Muksin, 2020).

One of the technologies that enables effective two-way communication is speech recognition and speech synthesis. This technology allows for the conversion of voice to text and vice versa, so interactions can occur more naturally. In addition, the use of smart glasses provides ease of communication without the need to hold the device directly, supporting a more practical hands-free experience (Augusto, 2021; Sergi et al., 2020; Verbeek et al., 2021). When smart glasses are integrated with AR (Augmented Reality) technology, the translated text can be displayed directly without disturbing the view of the surrounding environment. This creates a more immersive experience and improves user understanding. As such, these devices can serve as an alternative to smartphones, which typically require users to look at the screen while reading translated text (Hadiapurwa et al., 2021; Koutromanos & Kazakou, 2023).

Based on these problems, this study aims to develop a prototype equipped with a speech recognition feature to convert the voice of the interlocutor into text that will be translated using the LSTM (Long Short-Term Memory) architecture and displayed on the user's lens, then speech recognition which converts the user's voice into text which will then be translated back using the LSTM architecture and then the output will be converted into voice using the LSTM architecture speech synthesis. This prototype uses a Deep Learning approach using the LSTM architecture. The LSTM architecture was chosen because it can work well in edge computing, such as Raspberry Pi devices, where energy efficiency and processing time are paramount in edge devices (Huang et al., 2018; Miladinovic & Schefer-Wenzl, 2021; "Soft Computing Tools for Intelligent Machines," 2020; Yang et al., 2024; Zhang et al., 2023). Compared to Transformer architectures that have advantages in terms of parallelism and training speed, especially in the context of natural language processing (Shao et al., 2024), however, the implementation of Transformer architecture on edge devices often faces challenges related to compute complexity and higher power consumption, making it less than ideal in some applications that require efficiency (Kim et al., 2023; Park et al., 2023; Peng & Guo, 2022).

Integrating new technologies in the tourism sector has revolutionized how businesses engage with tourists. Communication, a critical element of the tourism experience, often presents challenges due to language barriers between service providers and international visitors. Traditional methods, such as using translation apps on smartphones or relying on human interpreters, can be inefficient and cumbersome, especially in fast-paced or noisy environments. These limitations highlight the need for innovative solutions that offer seamless, real-time translation during interactions. Smart glasses with built-in translation capabilities have emerged as a potential game-changer in this context, providing hands-free, real-time translation for better communication.

In the past decade, significant advancements in machine learning and natural language processing (NLP) have enabled the development of intelligent devices like smart glasses that can assist with language translation. The application of speech recognition and synthesis technologies has also paved the way for more efficient two-way communication, which is more practical and effective than one-way translation tools. These smart glasses can be worn like conventional eyewear, allowing users to converse naturally while the device captures and translates spoken language in real time, enhancing the user experience without requiring them to look away from their surroundings.

In the context of tourism, smart glasses that combine translation technology with augmented reality (AR) can further enrich the communication process. This integration ensures that users can view translated text directly in their field of vision while interacting with tourists or service providers, making it easier to understand and respond to foreign language input without interrupting the flow of conversation. As a result, smart glasses have the potential to significantly improve the accessibility of language services significantly, fostering more inclusive and productive communication in the tourism industry.

This study focuses on the development of a prototype for smart glasses equipped with speech recognition and synthesis capabilities to aid in language translation. The prototype leverages the Long Short-Term Memory (LSTM) architecture, known for its efficiency in processing sequential data such as speech. This research aims to create a tool that allows for real-time, bidirectional communication between tourists and service providers, overcoming language barriers and enhancing the overall tourist experience. Combining cutting-edge translation algorithms with wearable technology, this prototype is designed to make cross-language communication smoother and more efficient in real-time situations.

Furthermore, the development of such technology has the potential to greatly benefit the tourism sector by addressing one of its major challenges—language barriers. The convenience and practicality offered by smart glasses can be a significant competitive advantage for the tourism industry in Indonesia and other countries with high volumes of international visitors. By integrating this innovative tool, businesses can enhance their service delivery and provide a more welcoming environment for tourists, thereby increasing customer satisfaction and overall business success.

Communication in the tourism sector often faces challenges due to language barriers, especially in countries like Indonesia, where the population’s proficiency in English is still relatively low. Tourists and service providers in the tourism industry often struggle to communicate effectively, impacting the overall tourism experience. Existing translation tools, such as smartphone apps and digital interpreters, tend to rely on internet connections, support one-way communication, and often require additional devices, which can be inconvenient for tourists and service providers. These limitations make it clear that there is a pressing need for a more efficient, real-time, and accessible translation tool to overcome language barriers and improve the quality of service in the tourism sector.

The current market offers translation devices like interpreter earphones and some smart glasses, but these products cannot often support two-way communication, which is far more prevalent and effective in tourism settings. Two-way communication allows for better interaction, feedback, and understanding between service providers and customers, making it an ideal mode for enhancing the tourist experience. Despite technological advancements, few solutions integrate speech recognition, speech synthesis, and augmented reality (AR) to provide an immersive and real-time translation experience in a single device.

This research is urgent due to the increasing demand for solutions that address language barriers in the tourism sector. With the growing influx of international tourists to Indonesia, overcoming communication challenges has become essential for maintaining high service standards. The development of a smart glasses prototype that incorporates speech recognition, translation using LSTM (Long Short-Term Memory), and speech synthesis can provide an innovative and practical solution. This device would allow for real-time, two-way communication between tourists and service providers, making it an indispensable tool for improving service quality and fostering better tourism experiences.

In recent years, the development of wearable technology for real-time language translation has gained attention. A study by Koutromanos and Kazakou (2023) reviewed the use of augmented reality (AR) in language translation, highlighting the potential of AR smart glasses in delivering a seamless translation experience. However, their work primarily focused on one-way communication, which limits its application in interactive settings like tourism. Additionally, while many studies have addressed the translation capabilities of various devices, few have integrated speech recognition and synthesis with AR for a fully immersive experience, as in the proposed smart glasses prototype.

Verbeek et al. (2021) explored the advancements in smart glasses technology, focusing on its use in various sectors, including tourism. They emphasized the potential of smart glasses to improve communication, but noted that the technology was still developing in terms of usability and integration with real-time translation systems. The research pointed out that while smart glasses could offer valuable assistance, issues such as processing power, battery life, and device size needed further refinement.

Another important study by Yang et al. (2024) looked into the integration of deep learning models for language translation, particularly using LSTM architectures. While their work demonstrated the effectiveness of LSTM models in improving translation accuracy, it did not address the practical application of this technology in wearable devices for tourism. Their findings underscore the importance of accurate, real-time translation but leave gaps in the application of such models in consumer-friendly, wearable formats like smart glasses.

Although significant research has been conducted in speech recognition, translation technologies, and wearable devices, there is a gap in studies that combine these technologies in a practical, user-friendly device for real-time, bidirectional communication in tourism settings. While existing translation devices support one-way communication, two-way communication remains underexplored in wearable technology, especially when augmented reality is integrated. This research aims to bridge this gap by developing a smart glasses prototype that combines speech recognition, speech synthesis, and LSTM-based translation in a wearable device.

The novelty of this research lies in the integration of several cutting-edge technologies—speech recognition, speech synthesis, and Long Short-Term Memory (LSTM)-based translation models—into a single smart glasses prototype designed for the tourism sector. This device offers real-time, bidirectional communication, enhancing the user experience by enabling tourists and service providers to engage naturally in conversation. The incorporation of augmented reality further sets this prototype apart, as it displays translated text directly in the user's field of vision without obstructing their view, creating an immersive and seamless communication experience.

This research aims to develop a prototype of smart glasses that integrates speech recognition, speech synthesis, and LSTM-based translation to facilitate real-time, two-way communication for tourists and service providers in the tourism sector. The study aims to assess the feasibility and effectiveness of this prototype in overcoming language barriers, improving communication, and enhancing the overall tourist experience.

This research offers several benefits for both the tourism industry and technology development. By providing a practical, real-time solution to language barriers, the smart glasses prototype can significantly improve communication between international tourists and local service providers, thus enhancing customer satisfaction and promoting tourism growth. Additionally, the integration of deep learning and augmented reality into wearable devices opens new avenues for future technological innovations, benefiting industries beyond tourism and contributing to the development of more advanced, user-centric smart technologies.

RESEARCH METHODS

The development of smart glasses prototypes began with the collection of information related to language problems, deep learning, speech recognition, and speech synthesis, as seen in Figure 1.

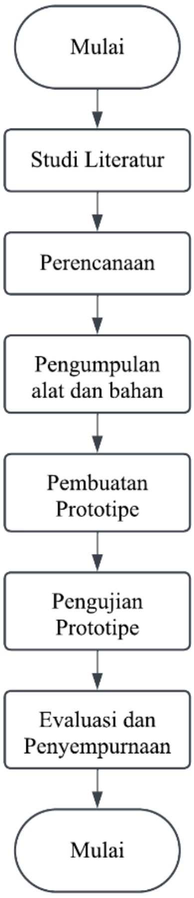

Figure 1. Implementation Method

Next, the prototype design stage is carried out, where, in this case, the workings of the smart glasses prototype technology system are planned, as seen in Figure 2. The 3D design is also designed using Fusion 360 software, as seen in Figure 3.

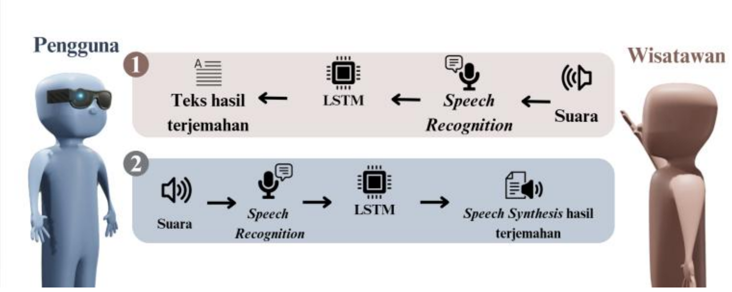

Figure 2. How Prototypes Work

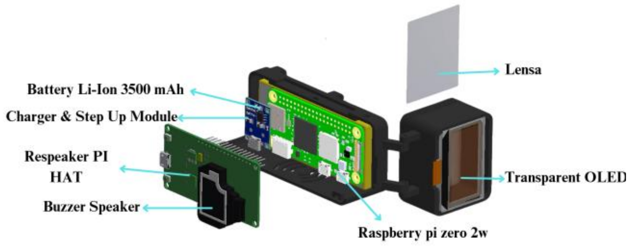

Figure 3. 3D Design of the Prototype Series

After that, tools and materials are collected, including key components such as the Raspberry Pi Zero 2W, Transparent OLED, 3500 mAh lithium battery, Pi HAT 2 Mic ReSpeaker, Buzzer Speaker, and Charger Module. Prototyping involves integrating components and developing software operated by the Raspberry Pi Zero 2W. Translation development is done by training the translation model with the TensorFlow library. Meanwhile, speech recognition and synthesis were developed using the VOSK and PYTTSX3 libraries.

In the translation aspect, the Bilingual Evaluation Understudy (BLEU) test was used to assess the accuracy and quality of the resulting translation. Meanwhile, the test was carried out for speech recognition and speech synthesis by measuring the Word Error Rate (WER) to evaluate the level of word recognition errors in converting voice to text and vice versa. BLEU testing is done using the NLTK Python library, while WER testing is done using the JiWER library (Post, 2018;Roy, 2021).

The evaluation and refinement stage is carried out in line with the testing stage and is based on the results of identifying weaknesses that require improvement. This includes reducing the complexity of translation models by optimizing hyperparameters, improving translation model performance with fine-tuning, improving speech recognition features with noise reduction, and improving the overall system to reduce computational delays.

RESULTS AND DISCUSSION

System Overview

An overview of Smart Glasses technology is shown in Figure 4 below.

Figure 4. Prototype Overview of *Smart Glasses* Translator

With speech recognition and speech synthesis technology, *Smart Glasses* can project real-time translation results in an AR display and output translated sound. By implementing innovative technology, *Smart Glasses* can effectively solve the language barrier problem. The results achieved in the development *of Smart Glasses* include a prototype consisting of a series of hardware and software components. The combination of these components results in an integrated *arrangement of Smart Glasses* hardware. As for the software, the AR interface allows users to view translated information directly in their field of view on a transparent OLED. It also has a translation model with an LSTM architecture that allows text to be converted from one language to another. It is supported by speech recognition and synthetic speech features that can convert voice to text and text to voice.

Smart Glasses have two main working mechanisms. First, when the user listens to the interlocutor say, "What do you know about the history of this place?" The Smart Glasses microphone picks up the sound and converts it into text using speech recognition technology. The text is then translated using the translation model that has been developed, and the translation results are displayed on the lens using AR, as can be seen in Figure 5.

Figure 5. Translation AR Display

Second, when the Smart *Glasses user* speaks to the other person with the sentence "What do you think about this place?", *Smart Glasses* captures the voice with speech recognition, translates it with the model that has been created and the translation results are synthesized back into sound using speech synthesis with an output in the form of a sentence in English, "What do you think about this place?" which the other person can hear through the speaker. This process allows the interlocutor to understand the user's message easily. With these two working mechanisms, *Smart Glasses* can be a seamless cross-language communication solution, even in crowded environments, thanks to the noise reduction feature on *Smart Glasses* speakers. In more detail, the specifications *of Smart Glasses* can be seen in Table 2. Next

Table 2. Hardware Specification *of* Smart Glasses *Prototype*

Heavy	88 gram
Dimension	8.5 cm x 3.5 cm x 3.5 cm
Screen Resolution	OLED 128 x 64 px
CPU	1 GHZ quad-core 64-bit ARM Cortex-A53
Memory	512 MB
Storage	16 GB
Battery	3500 mAh
Charger	Micro USB

This specification is designed to balance high performance and portability. The lightweight and compact dimensions ensure the user's comfort when wearing *Smart Glasses* for a long time.

Test Results

In addition, tests have also been carried out on *Smart Glasses* features such as translation with LSTM and speech recognition and synthesis, resulting in accuracy, BLEU, and WER values in Tables 3 and 4.

Table 3. BLEU Testing and Accuracy

Metric	Model	
	Indonesia - English	English - Indonesia
BLEU	36.5%	39.8%
Accuracy	80.8%	87.8%

Table 4. *Speech Recognition* and *Speech Synthesis Testing*

Feature	Metric	Model	
		Indonesia	English
Speech Recognition	WER	6%	8,4%
	Akurasi	94%	91,6%
Speech Synthesis	WER	12.3%	6.5%
	Accuracy	87.7%	93.5%

Table 3 presents the results of the Smart *Glasses translation quality test* using two main metrics: BLEU (Bilingual Evaluation Understudy) and accuracy. The BLEU metric measures how close an automatic translation is to a human translation, while accuracy measures the percentage of correct translations. Smart Glasses achieved a BLEU score of 36.5% for translation from Indonesian to English. As for the translation from English to Indonesian, the score was slightly higher, at 39.8%. A higher BLEU score indicates that translations from English to Indonesian are closer to human translations. Accuracy testing showed quite good results, with the translation accuracy from Indonesian to English reaching 80.8%, and from English to Indonesian reaching 87.8%. This indicates that the model has a Strong ability to understand and translate languages with a low error rate. Table 4 shows speech recognition and synthesis. Testing the speech recognition feature on *Smart Glasses* showed a Word Error Rate (WER) of 6% for Indonesian and 8.4% for English. The accuracy for speech recognition was 94% for Indonesian and 91.6% for English. These results show that *Smart Glasses* can recognize speech with a low error rate, enabling high-accuracy conversion of speech to text.

The speech synthesis feature on *Smart Glasses* was tested by recording a WER of 12.3% for Indonesian and 6.5% for English. The accuracy for speech synthesis was 87.7% for Indonesian and 93.5% for English. While there is room for improvement in Indonesian speech synthesis, these results show that Smart Glasses effectively synthesizes speech with an acceptable error rate. These tests showed that *Smart Glasses* performed well in translating text and recognizing and synthesizing speech in both languages tested. However, there is room for improvement, especially in the speech synthesis feature for Indonesian, which has a higher WER than English. Overall, these results demonstrate the potential of *Smart Glasses* as a useful device for cross-lingual communication.

CONCLUSION

The test results showed that the *Smart Glasses* prototype performed well in translating English and Indonesian. The *Smart Glasses* translation feature for Indonesian to English reached 80.8%, while the translation from English to Indonesian received an accuracy of 87.8%. Each model also received a BLEU of 36.5% and 39.8%. In the speech recognition feature test, *Smart Glasses* showed a Word Error Rate (WER) of 6% for Indonesian and 8.4% for English, with speech recognition accuracy rates of 94% and 91.6%, respectively. Meanwhile, in the speech synthesis feature, the WER recorded was 12.3% for Indonesian and 6.5% for English, with an accuracy of speech synthesis of 87.7% for Indonesian and 93.5% for English. Overall, these results show that the *Smart Glasses* prototype still has great potential as a useful device for cross-language communication. Developments such as expanding the data set, fine-tuning, and improving the complexity of the model can be undertaken to improve translation accuracy.

REFERENCES

Augusto, J. C. (2021). Smart Cities: State of the Art and Future Challenges. In *Handbook of Smart Cities*. https://doi.org/10.1007/978-3-030-15145-4_95-2

Fitriani Sopayati & Giyoto. (2023). scidac plus Artikel ini menggunakan lisensi Creative Commons Attribution 4.0 International License. In *Berkala Ilmiah Pendidikan* (Vol. 3, Issue 2).

Hadiapurwa, A., Maula, W. A., & Alfitriani, N. (2021). Penggunaan Media Augmented Reality dalam Pembelajaran Mengenal Bentuk Rupa Bumi. *JPP, 38*(1).

Huang, Y., Lu, B., & Zhao, W. (2018). Application of Cloud Computing in Intelligent Machine Tool Control System. *Jixie Gongcheng Xuebao/Journal of Mechanical Engineering*, *54*(8). https://doi.org/10.3901/JME.2018.08.210

Kim, J., Kang, H., & Kang, P. (2023). Time-series anomaly detection with stacked Transformer representations and a 1D convolutional network. *Engineering Applications of Artificial Intelligence*, *120*. https://doi.org/10.1016/j.engappai.2023.105964

Koutromanos, G., & Kazakou, G. (2023). Augmented reality smart glasses use and acceptance: A literature review. *Computers & Education: X Reality*, *2*, 100028. https://doi.org/10.1016/j.cexr.2023.100028

Latifah, W., & Muksin, N. N. (2020). POLA KOMUNIKASI DALAM METODE COACHING PEGAWAI RSUD R. SYAMSUDIN, SH KOTA SUKABUMI. *Desember*, *4*(2), 145–154.

Miladinovic, I., & Schefer-Wenzl, S. (2021). Towards Intelligent Multi-Access Edge Computing Using Machine Learning. *Advances in Intelligent Systems and Computing*, *1192 AISC*. https://doi.org/10.1007/978-3-030-49932-7_104

Park, J., Babaei, M. R., Munoz, S. A., Venkat, A. N., & Hedengren, J. D. (2023). Simultaneous multistep transformer architecture for model predictive control. *Computers and Chemical Engineering*, *178*. https://doi.org/10.1016/j.compchemeng.2023.108396

Peng, Z. Y., & Guo, P. C. (2022). A Data Organization Method for LSTM and Transformer When Predicting Chinese Banking Stock Prices. *Discrete Dynamics in Nature and Society*, *2022*. https://doi.org/10.1155/2022/7119678

Post, M. (2018). *A Call for Clarity in Reporting BLEU Scores*. http://arxiv.org/abs/1804.08771

Rakhmyta, Y. A., & Nurmalina, N. (2023). PENYULUHAN BAHASA INGGRIS BAGI MASYARAKAT DI DESTINASI WISATA SABANG. *JPMA - Jurnal Pengabdian Masyarakat As-Salam*, *3*(2), 41–46. https://doi.org/10.37249/jpma.v3i2.684

Roy, S. (2021). *Semantic-WER: A Unified Metric for evaluating ASR Transcript for End Usability*. http://arxiv.org/abs/2106.02016

Sergi, R., Bellucci, D., & Cannillo, V. (2020). A comprehensive review of bioactive glass coatings: State of the art, challenges and future perspectives. *Coatings*, *10*(8). https://doi.org/10.3390/COATINGS10080757

Shao, J., Qu, P., Wang, S., & Shi, J. (2024). Time series prediction model using LSTM-Transformer neural network for mine water inflow. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-69418-z

Soft Computing Tools for Intelligent Machines. (2020). In *Intelligent Machines*. https://doi.org/10.1201/9781482273939-11

Verbeek, P.-P., Peters, O., Ben Allouch, S., & Zuidhof, N. (2021). Defining Smart Glasses: A Rapid Review of State-of-the-Art Perspectives and Future Challenges From a Social Sciences' Perspective. *Augmented Human Research*, *6*(1). https://doi.org/10.1007/s41133-021-00053-3

Widiana, Y., & Meilasari, P. (2024). Peningkatan Kualitas Pemandu Wisata di Sentani Foresta Madiun Jawa Timur melalui Pelatihan Bahasa Inggris Berbasis Role-Play. *Jurnal Abdi Masyarakat Indonesia*, *4*(6), 1455–1468. https://doi.org/10.54082/jamsi.1351

Yang, C. T., Wei, S. J., Huang, C. K., Liu, J. C., Wang, L. Y., & Kristiani, E. (2024). An Intelligent Thermal Compensation System Using Edge Computing for Machine Tools. *Sensors*, *24*(8). https://doi.org/10.3390/s24082531

Zhang, L., Ma, C., Liu, J., Gui, H., & Wang, S. (2023). Implementation of Precision Machine Tool Thermal Error Compensation in Edge-Cloud-Fog Computing Architecture. *Journal of Manufacturing Science and Engineering*, *145*(7). https://doi.org/10.1115/1.4057011