Language Translation Glasses

Shivam Raikar
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Jesven Gomes
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Vinayak Halshikar
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Vinayak Naik
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Vaibhavi Naik
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

ABSTRACT

Language barriers cause problems in our global world. People can't understand each other in education, business, travel, and everyday life. Translating languages can be hard, and conversations require translators or devices, making eye contact difficult. This project created real-time translation glasses to break down these barriers. The glasses use advanced technology to translate spoken words instantly. They show the translation as text or play it in your preferred language. These glasses could be very helpful in many situations, allowing people to understand each other during business meetings or casual conversations. The project shows that these glasses can work and have a big impact. They are easy to use and combine technology well. This could make communication better and help people around the world understand and work together.

Keywords

speech-to-text, text-to-text, translation, text-to-display, translated text, Speech recognition.

1. INTRODUCTION

In a world pulsating with the vibrancy of diverse cultures and brimming with the potential for connection, language often remains a stubborn wall, fragmenting interactions and hindering understanding. Imagine the frustration of a tourist lost in the labyrinthine alleys of Tokyo, yearning to ask for directions but rendered tongue-tied by the melody of a foreign tongue. Picture the missed opportunities in bustling marketplaces, where cultural misunderstandings choke off the possibility of fruitful collaboration. These are the everyday realities of a world divided by the invisible fence of language. But what if this wall could crumble, replaced by a transparent bridge of instant translation? What if a subtle pair of glasses could empower you to seamlessly navigate any linguistic landscape, weaving threads of conversation with any soul you encounter, regardless of their native tongue This is the revolutionary promise of real-time language translation glasses, ushering in a new era of global communication where understanding transcends the tyranny of language. These smart glasses are not mere tech toys; they are the culmination of cutting-edge advancements in artificial intelligence, natural language processing, and miniaturized technology. Mini microphones tucked discreetly within the frame capture the spoken word [1], feeding it to a powerful onboard processor equipped with an arsenal of sophisticated algorithms. This linguistic alchemist then dissects the audio stream, identifying the source language with uncanny accuracy. Armed with this knowledge, the AI engine engages in a complex dance of translation, utilizing advanced natural language processing (NLP) models to understand the context, nuance, and cultural subtleties embedded within the spoken word. Finally, the translated text, seamlessly overlays the real world through an augmented reality display, woven into the tapestry of your vision [2]. The implications of this technological leap are nothing short of transformative. Imagine a world where: Business transcends borders: International collaborations flourish as language barriers evaporate, paving the way for joint ventures, knowledge sharing, and cultural exchange. Travel sheds its constraints: The charm of hidden alleyways and the wisdom of remote villages become accessible, enriching journeys like never before. Education embraces inclusivity: Language-diverse classrooms thrive with real- time translations, fostering participation and ensuring no student is left behind. Human connections deepen: Conversations flow freely across cultural divides, bridging the gap between individuals and fostering empathy and understanding. But the impact of these glasses extends far beyond personal convenience. Consider the humanitarian applications: refugees finding their way in unfamiliar lands, medical professionals assisting patients in war-torn regions, or disaster relief teams coordinating with local communities. In such scenarios, language translation becomes more than a tool; it transforms into a lifeline, a beacon of hope in the face of adversity. this nascent technology isn't without its challenges. Refining the AI models to ensure nuanced and context-aware translations remains a continuous pursuit. Optimizing speech recognition for diverse accents and environments requires ongoing research and ensuring seamless integration with augmented reality displays demands technological innovation.

2. RELATED WORK

According to, Maghilnan & M, Rajesh. (2018). This paper presents a system that converts spoken language to text and conducts sentiment analysis on speaker-specific data. It uses tools like Google Speech API, Pyttsx3, TextBlob, and NLTK. While accuracy is not mentioned, the system transforms speech into text and performs speech recognition [1].

Kumar, Vijay and Singh, Hemant and Mohanty, Animesh (2021). This research discusses a real-time speech-to-text and text-to-speech converter with an automatic text summarizer. It utilizes Deep Speech 2 for voice-to-text and AMR parsing for summarization, achieving an accuracy of 99.37% with Lead- 3-AMR. The study identifies Deep Speech 2 as the best model for transcribing spoken audio [2].

According to Wang, C.-S., Huang, W., Chang, Y.- F., Yeh, C.-M., & Xu, Z.-Y. (2020). This paper discusses the development of an assistive device using smart glasses. It captures real-time images, calculates facial features, and displays the person's ID on the right side of the glasses. It utilizes K Nearest Neighbour (KNN) and Support Vector Machines (SVM) for facial recognition, achieving an accuracy rate of 93.3% for KNN and 90.0% for SVM [3].

Wang, Yuluo. (2022). This research explores the combination of Augmented Reality (AR) technology and translation systems. It employs automatic speech recognition, optical character recognition, machine translation, and neural networks. While accuracy is not mentioned, the study highlights the use of real-time speech translation with three components: Automatic Speech Recognition (ASR), speech synthesis, and machine translation [4].

According to P.-S., Lin, Y.-C., Peng, Y.-H., & Chen, M. Y. (2019). This paper introduces PeriText, a method that utilizes peripheral vision for reading text on augmented reality glasses. It employs Rapid Serial Visual Presentation (RSVP) and found that users preferred a 5°eccentricity over 8° in walking scenarios, resulting in better reading efficiency [5].

According to Vyas, R., Joshi, K., Sutar, H., &Nagarhalli, T. P. (2020). An Author presents a real-time machine translation system for English to Indian languages. It uses tab-delimited bilingual sentence pairs as a dataset and the Open NMT model for translation, achieving high accuracy. The study indicates that Rule-Based Machine Translation (RBMT) and Statistical Machine Translation (SMT) offer better accuracy and employs LSTM [6].

According to Nagdewani, Shivangi and Ashika Jain (2020). This review paper provides an overview of methods for speech-to-text and text-to-speech conversion. It discusses Hidden Markov Models (HMM) as effective tools for both processes, although specific accuracy rates are not mentioned [7].

Watanabe, Daiki & Takeuchi, Yoshinori & Matsumoto, Tetsuya & Kudo, Hiroaki & Ohnishi, Noboru. (2018). This research paper focuses on a communication support system using smart glasses for the hearing impaired. It employs noise reduction, automated speech recognition (ASR), and four microphones, achieving an accuracy rate of 90.8% with Gaussian Mixture Model – Hidden Markov Model (GMMHMM). The system successfully recognizes speech and converts it into text [8].

According to Saikiran Gogineni G. Suryanarayana Sravan Kumar Surendran (2020). This paper discusses an effective neural machine translation system for English to Hindi. It employs attention mechanisms, word embedding, and data pre-processing. Although specific accuracy is not mentioned, the research contributes to the knowledge of speech-to-text conversion [9].

According to Chung, Yu-An & Weng, Wei-Hung & Tong, Schrasing & Glass, James. (2019). This paper explores unsupervised speech-to-text translation with an autoencoder. It utilizes a language model and does not require supervision. While accuracy rates vary, the research highlights the unsupervised nature of the system [10].

According to Saini, S., & Sahula, V. (2018). This paper concludes that Neural Machine Translation (NMT) holds promise for English-Hindi translation, showing comparable or superior performance to traditional methods,with potential for further refinement and expansion to other languages[11].

Anto, A., & Nisha, K. K. (2016). The research presents an effective Text-to-Speech synthesis system for English to Malayalam translation, achieving promising accuracy rates and suggesting avenues for further improvement in linguistic accessibility [12].

3. METHODOLOGY

3.1 Design

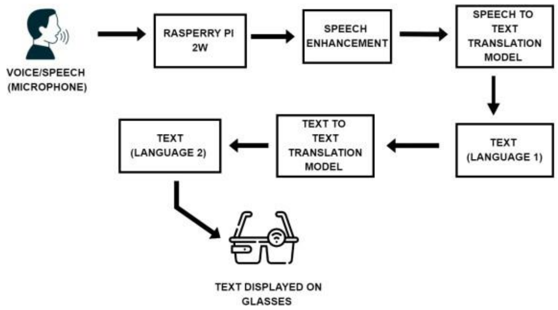

VOICE/SPEECH (MICROPHONE)
RASPERRY PI 2W
SPEECH ENHANCEMENT
SPEECH TO TEXT TRANSLATION MODEL
TEXT (LANGUAGE 1)
TEXT TO TEXT TRANSLATION MODEL
TEXT (LANGUAGE 2)
TEXT DISPLAYED ON GLASSES

Fig 1: Detailed Design

- The language translation glasses start with your voice input where it takes your voice as an input. Imagine you're talking with someone the glasses having microphone captures your words, just like an attentive listener, the microphone converts the user's voice into an electrical signal. Conversations received by the microphone aren't always crystal clear and hence a high-quality microphone is required, especially with background noise which thus increases voice/speech enhancement via raspberry pi 2w.
- Raspberry Pi 2W is a mini-computer inside the glasses that acts like a sound editor which the microphone is connected to it.
- Once your voice is enhanced, the next phase is conversion of voice/speech into text. A speech-to-text translation model is used for the conversion, using its linguistic know-how to decode the patterns of your speech and create a written version of the words in the native language. It acts as a super-fast scribe, which converts your speech to text as you speak.
- Text (language 1) is stored in the memory which is a resultant from the conversion from speech to text.
- The next phase where a model is used for text-to-text translation or language 1 to language 2. The model is trained on vast amounts of text in different languages. It swiftly analysis the meaning of the words and finds the match in the target language. Whether you're speaking English and need Hindi or French, this phase is crucial as it bridges the gap between languages in the system.
- The result is later depicted and stored as language 2.
- The final which is to display the converted text on the glasses. The Smart glasses become a visual guide, projecting the translation onto their lenses, allowing

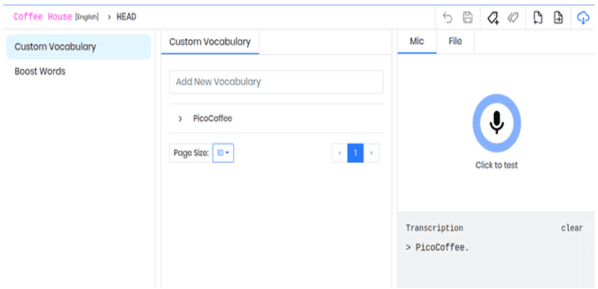
you to read the translated text naturally and effortlessly as you interact with the world around you.

3.2 SPEECH-TO-TEXT

3.2.1 *Cheetah Streaming*

We utilized Cheetah Streaming to capture spoken input with the microphone, converting it into a text file. Speech-to-Text is software that automatically transcribes voice data in real time without network delay or accuracy compromises.

Cheetah Streaming Speech-to-Text processes voice data locally, enabling live transcription on-device, mobile, web browsers, on-premise, or cloud.



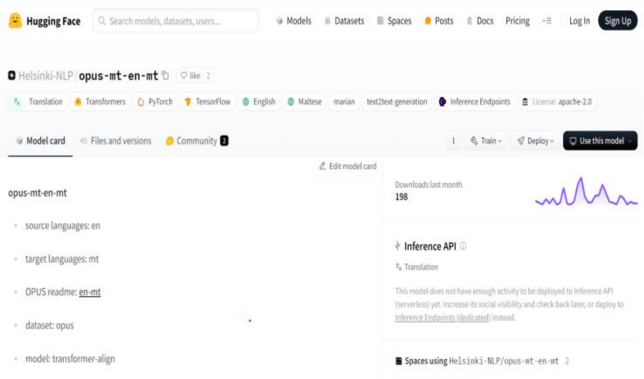
```
> PicoCoffee.
```

Fig 2: Speech to text Speech

3.2.2 *Transformers:*

We used Transformers for train state-of-the-art pretrained models. Using pretrained models can reduce your compute costs, carbon footprint, and save you the time and resources required to train a model from scratch.

Transformers support framework interoperability between PyTorch, TensorFlow, and JAX. This provides the flexibility to use a different framework at each stage of a model's life; train a model in three lines of code in one framework, and load it for inference in another. Models can also be exported to a format like ONNX and Torch Script for deployment in production environments.



opus-mt-en-mt

- source languages: en
- target languages: mt
- OPUS readme: en-mt
- dataset: opus
- model: transformer-align

Downloads last month
198

This model does not have enough activity to be deployed to Inference API (serverless) yet. Increase its social visibility and check back later, or deploy to Inference Endpoints (dedicated) instead.

Spaces using Helsinki-NLP/opus-mt-en-mt 2

Fig 3: Transformers

3.3 PERFORMANCE MEASUREMENT

3.3.1 *BLEU Score:*

BLEU (BiLingual Evaluation Understudy) is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high- quality reference translations. A value of 0 means that the machine-translated output has no overlap with the reference translati.on (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality).

It has been shown that BLEU scores correlate well with human judgment of translation quality. Note that even human translators do not achieve a perfect score of 1.0

3.3.2 *Interpretation Of BLEU Score*

Trying to compare BLEU scores across different corpora and languages is strongly discouraged. Even comparing BLEU scores for the same corpus but with different numbers of reference translations can be highly misleading.

However, as a rough guideline, the following interpretation of BLEU scores (expressed as percentages rather than decimals) might be helpful.

Table 1: Interpretation of BLEU Score

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

The mathematical details

Mathematically, the BLEU score is defined as:

$$\mathrm{BLEU} = \mathrm{BP} * \exp\left(\sum_{(n=1)}^{N} w_n * \log(p_n) \right)$$

With Unigram Precision (p1p_1p1):

$$p_1 = \left(\sum_{(C \in \mathrm{Candidates})} \sum_{(n \in C)} \min(\mathrm{count}(n), \mathrm{max\_count}(n)) \right) / \left(\sum_{(C' \in \mathrm{Candidates})} \sum_{(n' \in C')} \mathrm{count}(n') \right)$$

Brevity Penalty (BP):

$$\mathrm{BP} = \{ 1 \text{ if } c > r \ \exp(1 - r/c) \text{ if } c \leq r$$

- BrevityPenalty

The brevity penalty penalizes generated translations that are too short compared to the closest reference length with an exponential decay. The brevity penalty compensates for the fact that the BLEU score has no recall term.

- N-GramOverlap

The n-gram overlap counts how many unigrams, bigrams, trigrams, and four-grams (i=1,...4) match their n-gram counterpart in the reference translations. This term acts as a precision metric. Unigrams account for adequacy while longer n-grams account for fluency of the translation. To avoid overcounting, the n-gram counts are clipped to the maximal n-gram count.

4. IMPLEMENTATION

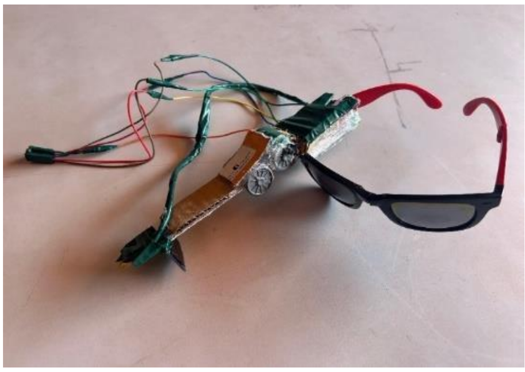


Fig 4: Working Model

The language translation glasses begin by capturing your voice through a high-quality microphone, which then enhances it for clarity using Raspberry Pi 2W, acting as a sound editor. This step is crucial for overcoming background noise. The enhanced voice is converted into text using a speech-to-text model, decoding speech patterns in your native language. The resulting text (language 1) is stored in memory. Then, a text-to-text translation model bridges the gap between languages, swiftly analyzing and translating the text into the desired language (language 2). Finally, the translated text is displayed on the glasses.

5. RESULT

Took Spanish dataset containing 3000 rows of reference(english) and translated (Spanish) data. Calculated Accuracy of the model before training using nltk library to compute BLEU score. The accuracy score of the model is shown in the below table:

Table 2: BLEU score of model before training

langpair	testset	BLEU Score	#words
eng-spa	English-Spanish-dataset	0.5213356	759846

The accuracy score of the model is shown in the below table:

Tabel 3: BLEU score of model after training

langpair	testset	BLEU Score	#words
eng-spa	English-spanish-dataset	0.7485929	759846

6. CONCLUSION

Our project holds immense potential to revolutionize communication, our meticulously crafted real-time language translation glasses represent an epochal leap, aimed at transcending linguistic divides while preserving the essence of genuine connection through sustained eye contact, thus heralding a new era in global interactions. The meticulous orchestration within our design phase encompassed a symphony of intricate connections among pivotal hardware components, by choreographing the Raspberry Pi's functionalities alongside the integration of an omnidirectional microphone. Tests and meticulous fine-tuning culminated in the crowning achievement of a robust voice-to-text conversion model, elegantly interfaced with an OLED display to seamlessly showcase the converted text. It also sets a visionary stage on which our project shall b refinement as we forge an unwavering path toward materializing a transformative solution in the realm of global communication.

7. REFERENCES

[1] Speech to text conversion and sentiment analysis on speaker specific data - S, Maghilnan & M, Rajesh. (2018). Sentiment Analysis on Speaker Specific Speech Data.https://www.researchgate.net/publication/323276680_Se ntiment_Analysis_on_Speaker_Specific_Speech_Data

[2] Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer using Natural Language Generation and Abstract Meaning Representation - Kumar, Vijay and Singh, Hemant and Mohanty, Animesh, Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation and Abstract Meaning Representation (April 3, 2020). International Journal of Engineering and Advanced Technology (IJEAT) , Volume-9 Issue-4, April, 2020, Page no:2361-2365, Available at SSRN: https://ssrn.com/abstract=3925035DOI: 10.35940/ijeat.D7911.049420

[3] Development of an Assistive Device via Smart Glasses - Wang, C.-S., Huang, W., Chang, Y.- F., Yeh, C.-M., & Xu, Z.-Y. (2020). Development of an Assistive Device via Smart Glasses. 2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). doi:10.1109/ecbios50299.2020.9203

[4] Analysis of the Combination of AR Technology and Translation System -Wang, Yuluo. (2022). Analysis of the Combination of AR Technology and Translation System. 10.2991/assehr.k.220105.035.DOI:10.2991/assehr.k.220105.035

[5] PeriText: Utilizing Peripheral Vision for Reading Text on Augmented Reality Smart GlassesKu, P.-S., Lin, Y.-C., Peng, Y.-H., & Chen, M. Y. (2019). PeriText: Utilizing Peripheral Vision for Reading Text on Augmented Reality Smart Glasses. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). doi:10.1109/vr.2019.8798065

[6] Real Time Machine Translation System for English to Indian language -Vyas, R., Joshi, K., Sutar, H., &Nagarhalli, T. P. (2020). Real Time Machine Translation System for English to Indian language. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). doi:10.1109/icaccs48705.2020.9074

[7] a review on methods for speech-to-text and text-to-speech conversion -Nagdewani, Shivangi and Ashika Jain. "A REVIEW ON METHODS FOR SPEECH-TO-TEXT AND TEXT-TO-SPEECH CONVERSION." (2020).https://www.irjet.net/archives/V7/i5/IRJET-V7I5854.pdf

[8] Communication Support System of Smart Glasses for the Hearing Impaired - Watanabe, Daiki & Takeuchi, Yoshinori & Matsumoto, Tetsuya & Kudo, Hiroaki & Ohnishi, Noboru. (2018). Communication Support System of Smart Glasses for the Hearing Impaired. 10.1007/978-

3-319-94277-3_37.

[9] An Effective Neural Machine Translation for English to Hindi Language-Saikiran Gogineni G. Suryanarayana Sravan Kumar Surendran.An Effective Neural Machine Translation for English to Hindi Language.DOI: 10.1109/ICOSEC49089.2020.9215347

[10] Towards Unsupervised Speechto Text Translation-Chung, Yu-An & Weng, Wei-Hung & Tong, Schrasing& Glass, James. (2019). Towards Unsupervised Speech-to-text Translation. 7170-7174. 10.1109/ICASSP.2019.8683550..DOI:10.1109/ICASSP.2019.8683550

[11] Neural Machine Translation for English to Hindi Saini, S., & Sahula, V. (2018). Neural Machine Translation for English to Hindi. 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP). doi:10.1109/infrkm.2018.8464781

[12] Text to Speech Synthesis System for English to Malayalam Translation Anto, A., & Nisha, K. K. (2016). Text to speech synthesis system for English to Malayalam translation. 2016 International Conference on Emerging Technological Trends (ICETT). doi:10.1109/icett.2016.7873642

[13] https://en.wikipedia.org/wiki/OLED

[14] https://www.sid.org/

[15] https://www.oled-info.com/

[16] https://oled.com/

[17] https://spectrum.ieee.org/tag/oled

[18] https://sid.onlinelibrary.wiley.com/journal/19383657

[19] https://spectrum.ieee.org/tag/oled

[20] https://sid.onlinelibrary.wiley.com/journal/19383657

[21] https://www.coursera.org/stanford