Wearable Cleft Palate Speech Interpreter Using Deep Learning and Neural Networks Algorithm

Jaypee P. Canlas[1], Jose Paulo E. Mendoza[1], Josh Nathaniel B. Perez[1], Reynaldo L. De Asis[2], Asil Kastle S. Dela Cruz[2]

[1]Student, College of Engineering and Architecture – Computer Engineering Department, University of the Assumption, City of San Fernando, Pampanga, Philippines

[2]Professor, College of Engineering and Architecture – Computer Engineering Department, University of the Assumption, City of San Fernando, Pampanga, Philippines

Corresponding Author: canlasjaypee123@gmail.com

Abstract— **In a world where communication is important, individuals with cleft palates face difficult challenges in expressing themselves effectively. Traditional communication methods often fall short, hindering their ability to interact confidently in various social and professional settings. Addressing this critical issue head-on, our study embarks on a transformative journey to develop the Wearable Cleft Palate Speech Interpreter. The Researchers developed this device over a ten-month period, from August 2023 to May 2024, using agile methodologies, prototyping methods, and descriptive research, as well as the power of deep learning and neural network algorithms implemented in Python programming to achieve their objectives. Prototype testing, confusion matrix analysis, feedback questionnaires, and extensive internet research formed the foundation of the researchers' comprehensive data collection approach. The study was conducted at the University of the Assumption. The researchers' findings highlight the remarkable efficacy of the Wearable Cleft Palate Speech Interpreter, achieving a 93% accuracy rate in testing and 82% in the actual prototype and 100% precision in speech interpretation. The developed Wearable Cleft Palate Speech Interpreter achieved a grand mean of 3.43 from end-users and 3.67 from professionals on acceptability. The device is considered very acceptable by both professionals and end-users; thus, they concluded that it can be used to serve its intended purpose. This study can be subjected to further development and improvement by future researchers.**

Index Terms— ***Cleft Palate, Speech Interpreter, Deep Learning and Neural Networks Algorithm.***

1. Introduction

Effective communication is a fundamental aspect of human interaction, facilitating the exchange of thoughts, emotions, and ideas (Robinson et al., 2023).

Manuscript revised May 05, 2024; accepted May 07, 2024. Date of publication May 09, 2024.
This paper available online at www.ijprse.com
ISSN (Online): 2582-7898; SJIF: 5.59

However, individuals born with cleft palates, a congenital condition that affects their oral and facial structures, often face significant challenges in this regard. These individuals struggle to articulate speech, impeding their ability to express themselves, share experiences, and fully engage in social interactions. Such communication difficulties can lead to emotional and psychological distress, limiting their participation in various aspects of life (Phalke N, et al., 2023).

The normal palate anatomy comprises the mucosa, hard palate, soft palate, uvula, alveolar ridge, primary palate, secondary palate, and incisive foramen. The hard palate serves to separate the mouth from the nose, playing a crucial role in speech and eating. The soft palate is responsible for speech and swallowing, and the uvula hangs down from its posterior aspect (Odunze, 2022).

Cleft palate classification, as outlined by Millicent Odunze, MD in 2022, serves as a guide for treatment plans. Clefts are categorized as either "complete" or "incomplete." A complete cleft involves both the primary and secondary palates, while an incomplete cleft only affects the secondary palate. Incomplete clefts may present as bifid uvula, submucosal cleft, soft palate cleft, or both soft and hard palate cleft. These variations exhibit varying degrees of severity, impacting speech and feeding in distinct ways.

In the comprehensive systematic review and meta-analysis by Salari et al 2022., which focused on the global prevalence of cleft palate and cleft lip, data from 59 studies involving 21,088,517 individuals were analyzed. The study reported a prevalence of 0.33 per 1000 live births for cleft palate (95% CI: 0.28-0.38). For cleft lip, encompassing 57 studies and 17,907,569 individuals, the prevalence was 0.3 per 1000 live births (95% CI: 0.26-0.34). Additionally, the meta-analysis of 55 studies with 17,894,673 individuals revealed a prevalence of 0.45 per 1000 live births for cleft lip and palate (95% CI: 0.38-0.52).

According to Mölnlycke in 2021, the Philippines exhibits one of the highest incidences of cleft palate and cleft lip globally. Despite the effectiveness of straightforward surgical procedures to address these issues, healthcare access remains limited in the country, particularly for residents living outside the main islands.

Recent advancements in deep learning and neural networks technologies present promising solutions to address the communication challenges encountered by individuals with cleft palates. This project aims to explore and develop an interpreter device as a novel solution to enhance their communication abilities. This interpreter device harnesses advanced deep learning and neural networks techniques, making use of deep learning models for accurate speech interpretation. At its core, the device is powered by a Raspberry Pi 4 model B, a versatile and cost-effective computing platform, functioning as the central processing unit (CPU). The Raspberry Pi 4 model B facilitates real-time data processing, ensuring swift and efficient speech interpretation. This innovation has the potential to significantly improve the quality of life for individuals with cleft palates by enabling clearer and more effective communication, thereby fostering greater inclusivity in various aspects of daily life.

2. Methodology

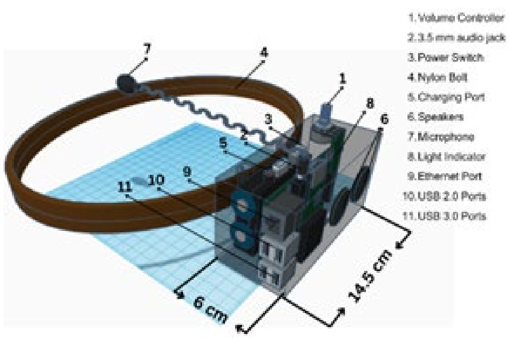

Fig.1. Prototype Design and Layout

The researchers will use a research design that integrates agile methodologies, prototyping methods, and descriptive research methods to develop the "Wearable Cleft Palate Speech Interpreter." Agile methodologies emphasize flexibility, collaboration, and rapid iterations to address evolving user needs and challenges. Prototyping methods allow for hands-on testing and validation of design concepts, while descriptive research methods collect data, including speech patterns of individuals with cleft palates, through recording and questionnaires. This combination creates a dynamic development cycle, ensuring the "Wearable Cleft Palate Speech Interpreter" remains agile, responsive, and user-friendly.

The researchers will gather primary data through rigorous testing and evaluation of hardware and software prototypes, assessing factors such as accuracy, user-friendliness, and responsiveness. A confusion matrix is used to assess the system's classification outcomes, distinguishing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Key performance metrics include accuracy, Misclassification Rate, Precision, Recall, and Specificity.

To ensure reliable operation, an accuracy threshold of at least 80%. Achieving accuracy between 70% and 90% is desirable and realistic, fostering trust in the device's functionality and user satisfaction. The dimensions and weight of a wearable device are crucial factors in determining user comfort. To achieve optimal user-friendliness, questionnaires will be designed and administered to understand user preferences and challenges encountered during interactions with the prototypes.

In the context of the wearable cleft palate speech interpreter, variations in processing time may occur due to factors such as sentence length, vocabulary size, and concurrent tasks being executed. The researchers aim to achieve a maximum delay of 2 seconds to ensure optimal performance.

To ensure comprehensive validation of the hardware and software integration, the researchers will actively seek feedback and validation from professionals with diverse expertise, including Software Engineers and Electronics Engineers.

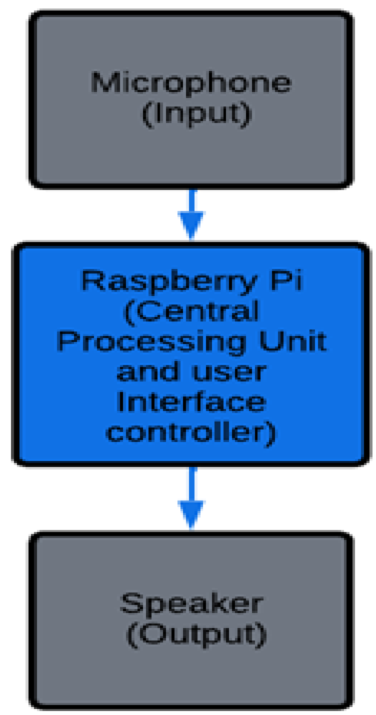

Fig.2. Block Diagram

Raspberry Pi 4 as the central processing unit, a microphone for precise speech input, and a speaker for clear interpreted speech.

The operational process of the wearable cleft palate speech Interpreter. The operation begins with the device being powered on. If there are no Cleft Palate speech detected it will start again. It actively listens for spoken input from individuals with cleft palates. Once spoken, the device's core component, the Raspberry Pi 4, initiates a comparison of the input audio with the information stored in its database.

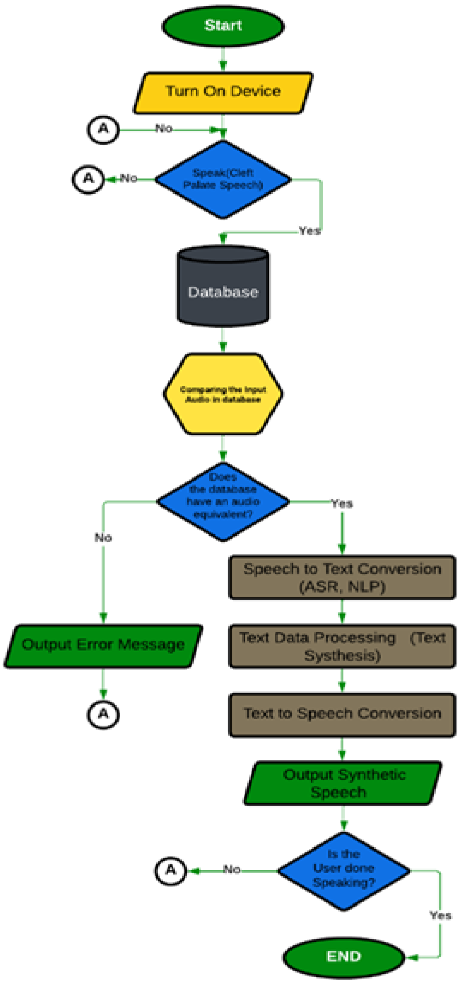

Fig.3. Process Diagram

This comparison utilizes advanced technologies such as Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) to transcribe the spoken words into text. If it does not have an equivalent audio there would be an Error Message. If the audio is in the database the text undergoes a formatting process to prepare it for clear and fluent speech synthesis. The final step in this sequence is the conversion of the transcribed text back into natural speech. This iterative process continues until the device detects that the user has finished speaking, at which point it concludes the interaction.

3. Result And Discussion

A. Confusion Matrix Evaluation

The researchers have selected 100 words that individuals with a cleft palate may find challenging. These words primarily involve the intricate pronunciation of certain letters, including p, b, t, d, k, g, f, v, s, z, sh, ch, m, and n. This list encompasses 50 Filipino words and 50 English words. Each word will undergo four tests: three during the code development phase and one during actual prototype testing to ensure complete evaluation and accuracy.

1. Accuracy = (TP+TN) / (TP+TN+FP+FN)
2. Misclassification Rate (Error Rate) = (TP+TN) / (TP+TN+FP+FN)
3. Precision = TP / (TP+FP)
4. Recall = TP / (TP+FN)
5. Specifity = TN / (TN+FP)

Table.1.
First Testing Analysis

		Actual Values	
		POSITIVE	NEGATIVE
Predictive Values	POSITIVE	77 TP	0 FP
	NEGATIVE	23 FN	0 TN

1. Accuracy = (77+0) / (77+0+0+23) = 77 / 100= 0.77 or 77%
2. Misclassification Rate (Error Rate)=(0+23) / (77+0+0+23) = 23/100 = 0.23 or 23%
3. Precision = 77 / (77+0) = 1.0 or 100%
4. Recall = 77 / (77+23) = 77 / 100 = 0.77 or 77%
5. Specificity = TN / (TN+FP) = 0 / (0+0) = 0 / 0 = 0

These findings underscore the importance of program improvements, particularly in fine-tuning speech recognition for cleft palate-associated speech patterns and reducing misclassifications stemming from English vocabulary. The program utilizes the gtts library for text-to-speech conversion, offering a fundamental means of converting text into audio files.

However, it lacks the advanced voice customization capabilities found in other solutions.

Table.2.
Second Testing Analysis

		Actual Values	
		POSITIVE	NEGATIVE
Predictive Values	POSITIVE	85 TP	0 FP
	NEGATIVE	15 FN	0 TN

1. Accuracy = (85+0) / (85+0+0+15) = 85 / 100= 0.85 or 85%
2. Misclassification Rate (Error Rate) = (0+15) / (85+0+0+15) = 15/100 = 0.15 or 15%
3. Precision = 85 / (85+0) = 1.0 or 100%
4. Recall = 85 / (85+23) = 85 / 100 = 0.85 or 85%
5. Specificity = TN / (TN+FP) = 0 / (0+0) = 0 / 0 = 0

These findings highlight a significant improvement in accuracy, which increased from 77% to 85%. The enhancement is primarily due to the integration of the pyttsx3 library during the second testing, which introduces several features that improve interpretation accuracy. For example, the ability to choose from a wide range of voices creates a more personalized audio experience. In addition, the higher quality of synthesized speech produced by pyttsx3 compared to gtts results in speech that is more natural and understandable.

Table.3.
Final Testing Analysis

Predictive Values \ Actual Values	POSITIVE	NEGATIVE
POSITIVE	93 TP	0 FP
NEGATIVE	7 FN	0 TN

1. Accuracy = (93+0) / (93+0+0+7) = 93 / 100= 0.93 or 93%
2. Misclassification Rate (Error Rate) = (0+7) / (93+0+0+7) = 7/100 = 0.7 or 7%
3. Precision = 93 / (93+0) = 1.0 or 100%
4. Recall = 93 / (93+7) = 93 / 100 = 0.93 or 93%
5. Specificity = TN / (TN+FP) = 0 / (0+0) = 0 / 0 = 0

These findings show a significant improvement in accuracy, from 85% to 93%, with final testing outperforming second testing. Final testing includes the Sphinx recognizer for offline functionality, which ensures consistent accuracy regardless of internet connectivity. It also includes a real-time processing loop to provide immediate feedback and improves data handling for greater interpretation accuracy. Furthermore, final testing provides customizable models and accent support, which improves overall performance and versatility in speech recognition tasks across multiple environments.

Table.4.
Actual Prototype Testing Analysis

Predictive Values \ Actual Values	POSITIVE	NEGATIVE
POSITIVE	82 TP	0 FP
NEGATIVE	18 FN	0 TN

1. Accuracy = (82+0) / (82+0+0+18) = 82 / 100= 0.82 or 82%
2. Misclassification Rate (Error Rate) = (0+18) / (82+0+0+18) = 18/100 = 0.18 or 18%
3. Precision = 82 / (82+0) = 1.0 or 100%
4. Recall = 82 / (82+18) = 82 / 100 = 0.82 or 82%
5. Specificity = TN / (TN+FP) = 0 / (0+0) = 0 / 0 = 0

These findings reveal a lower accuracy compared to the final testing, which achieved a 93% accuracy rate. This discrepancy can be attributed to several factors. Firstly, during the final testing, a laptop was employed, benefitting from advanced features such as noise cancellation and faster processing capabilities. In contrast, the actual prototype utilized a Raspberry Pi 4 Model B for processing. Additionally, differences in the microphone used in the prototype may have contributed to variations in accuracy, despite the fact that both devices execute the same code.

B. *Professional Evaluation*

Table.5.
Professional Evaluation

Completeness	100%
Correctness	100%
Appropriateness	100%
Time Behavior	87%
Resource Utilization	87%
Capacity	100%
Appropriateness recognizability	100%
Operability	87%
User Error Protection	87%
Learnability	100%
User Engagement	87%
Faultlessness	100%
Fault tolerance	87%
Recoverability	87%
Safety	87%

C. *End-Users Evaluation*

Table.6.
End-Users Evaluation

Correctness	80%
Time behavior	75%
Capacity	95%
Operability	100%
Safety	90%
User-Friendly	100%
Overall Satisfaction	85%

Based on feedback from users and professionals, the device was 100% user-friendly, with 90% satisfaction in terms of safety. Professional ratings averaged 87.5%, indicating positive judgments with opportunity for improvement.

In real-time speech interpretation, user feedback showed a 75% rating for Time Behavior, indicating potential delays in providing timely feedback. Professionals also assessed Time Behavior at 87.5%, which was slightly less than perfect, indicating that there was potential for improvement.

4. Conclusion

The Wearable Cleft Palate Speech Interpreter represents a groundbreaking assistive technology tailored for individuals with cleft palates. By harnessing deep learning and neural networks algorithms, it converts speech patterns into text and natural voice output. The researchers adeptly employed these algorithms in Python programming language, using models such as FuzzyWuzzy and speech recognition, supported by code implementation and visual aids like flowcharts. Feedback from both users and professionals underscored the device's high user-friendliness (100%) and satisfactory safety levels (90%). While professional ratings averaged at 87.5%, areas for improvement were identified. Real-time speech interpretation received ratings of 75% from users and 87.5% from professionals, indicating potential areas for refinement. Successful integration of hardware and software components using Raspberry Pi 4 Model B was achieved. Speech interpretation demonstrated commendable accuracy 93% in testing, 82% in the actual prototype and precision 100%, as verified by confusion matrix analysis. While professional feedback reflected flawless performance across key metrics, user feedback on correctness suggested an average rating of 80%, highlighting opportunities for accuracy enhancements. Future researchers should focus on enhancing the accuracy and feedback time of the Wearable Cleft Palate Speech Interpreter.

References

[1]. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architecture, challenges, applications, future directions. Journal of Big Data, 8(1), 1-33.

[2]. Arya, R. S. (2023, December 12). Best Programming Languages for Raspberry Pi.

[3]. Azizan, D. Khan, R. U., Khattak, H., Wong, W. S., AlSalman, H., Mosleh, M. A. A., & Rahman, S. M. M. (2022). Intelligent Malaysian Sign Language Translation System Using Convolutional-Based Attention Module with Residual Network.

[4]. Barkved, K. (March 9, 2022). Accuracy in machine learning: A subjective evaluation.

[5]. Beare, Kenneth. (2023, April 5). The 50 Most Commonly Used Words in the English Language.

[6]. Cambridge University Hospitals NHS Foundation Trust. (Year). How is the palate involved in speech?

[7]. Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., & Mansor, A. (2019). Augmentative and Alternative Communication (AAC) Advances: A Review of Configurations for Individuals with a Speech Disability.

[8]. Ferranco, M. (Year or n.d.). 101 Most Common Filipino/Tagalog Words.

[9]. Frisbee, E. (2021, July 29). Cleft lip and cleft palate. WebMD.

[10]. Gondi, S., & Pratap, V. (2021). Performance Evaluation of Offline Speech Recognition on Edge Devices. Electronics, 10(21), 2697.

[11]. Interaction Design Foundation - IxDF. (2019, October 17). What is Prototyping?. Interaction Design Foundation - IxDF.

[12]. Khanday, S. A. (2023). THE RESEARCH DESIGN. Aligarh Muslim University.

[13]. King's College London. Curtis H. (2023). Wearable augmentative and alternative communication (AAC) devices for people with complex communication needs (CCNs).

[14]. Koppuravuri, S., Pondari, S. S., & Seth, D. (2020). Sign Language to Speech Converter Using Raspberry-Pi. In: A. K. Verma, A. K. Misra (Eds.), Recent Developments in Machine Learning and Intelligent Systems (pp. 35-46). Springer.

[15]. Kulkarni, A., et al. (2020). Confusion Matrix. In Data Democracy.

[16]. Mandal, A. (2023). Symptoms and Diagnosis of Cleft Lip and Palate. News Medical.

[17]. Mölnlycke to Help Establish Cleft Care Center in the Philippines." (2021, March 26).

[18]. Odunze, Millicent. "Different Types of Cleft Palates." Verywell Health, 8 June 2022,

[19]. Papatsimouli, M., Sarigiannidis, P., & Fragulis, G. F. (2023). A Survey of Advances in Real-Time Sign Language Translators: Integration with IoT Technology. Technologies, 11(4), 83.

[20]. Phalke, N., & Goldman, J. J. (2023, July 6). Cleft Palate. National Library of Medicine.

[21]. Ramli, M. N., & Yee, S. K. (2023). Application of Machine Learning in Malaysia Sign Language Translation. Emerging Advances in Integrated Technology,

[22]. Riesener, M., Rebentisch, E., Dölle, C., Schloesser, S., Kuhn, M., Radermacher, J., & Schuh, G. (2019). A model

for dependency-oriented prototyping in the agile development of complex technical systems.

[23]. Robinson, L., Segal, J., Ph.D., & Smith, M.A. What is effective communication? HelpGuide.

[24]. Salari, N., Darvishi, N., Heydari, M., Bokaee, S., Darvishi, F., & Mohammadi, M. (Year of Publication). Global prevalence of cleft palate, cleft lip and cleft palate and lip: A comprehensive systematic review and meta-analysis. 2022 (Issue),

[25]. Taye, M. M. (2023, April 25). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. MDPI, 12(5), 91.

[26]. Tewfik, T. L. (2023, July 26). Cleft lip and palate and mouth and pharynx deformities. Medscape.

[27]. Vyas, T., Gupta, P., Kumar, S., Gupta, R., Gupta, T., & Singh, H. P. (2020). Cleft of lip and palate: A review. Journal of Family Medicine and Primary Care, 9(6), 2621-2625.

[28]. Warburton, D. (2020, February 27). AGILE PROTOTYPING. LinkedIn.