A26-134

聽．見字——結合邊緣運算與生成式AI之聽障構音修復與視聽覺整合照護通訊系統

Hear · See Text -Assistive Communication System Combining Edge Computing and Generative AI for Articulation Restoration and Audio-Visual Care



隊伍名稱｜聽見你說 I Hear You

隊　　長｜陳奕瑋 / 逢甲大學自動控制工程學系

隊　　員｜楊子鋐 / 逢甲大學自動控制工程學系

隊　　員｜林承憲 / 逢甲大學人工智慧技術與應用學士學位學程

隊　　員｜洪澄佑 / 逢甲大學自動控制工程學系

指導教授 ADVISOR



秦群立 | 逢甲大學自動控制工程學系

交通大學電機與控制工程博士，現任逢甲大學自動控制工程學系副教授。長期投入醫學資訊與智慧醫療系統研發，具系統實作、學生指導與競賽帶領經驗，曾帶隊獲第16屆國家新創獎。

研究領域

人工智慧、嵌入式系統、醫學影像處理、圖形識別、系統分析與設計。

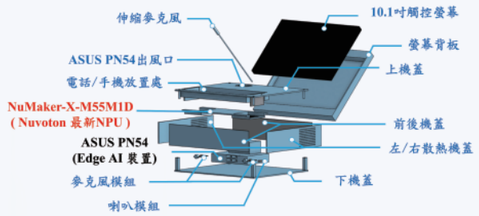
作品摘要

臺灣已正式邁入超高齡社會，高齡者因聽力退化所造成的溝通困難逐漸成為重要議題；同時，年輕族群長時間、高音量使用耳機，也使聽力受損風險逐漸上升，讓聽損不再只是高齡者的問題，而是橫跨不同世代的通訊挑戰。對聽障者與高齡使用者而言，電話溝通的困難並不只是音量不足，更包含語音不清、背景噪音干擾、無法即時理解內容，以及缺乏唇形、表情等視覺線索。現有助聽器、即時字幕或通話轉錄工具多半只能提供聲音放大、單向文字顯示或事後摘要，難以解決使用者「聽不清」與「說不清」的雙向問題；而傳統單一運算架構在面對即時語音辨識、生成式AI與影像互動等高負載任務時，也容易因延遲或反應不及而加劇溝通挫折。

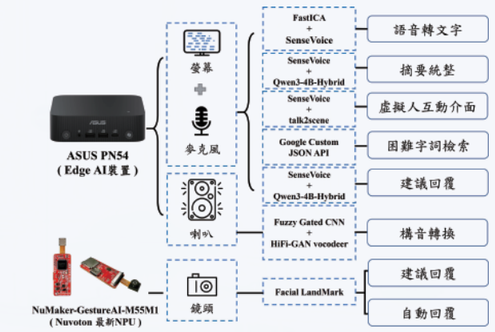
為了解決上述問題，本團隊開發「聽•見字」智慧通訊系統，將傳統電話或手機放入自製3D列印機構後(如圖一所示)，即可透過邊緣運算與生成式AI協同運作，將通話內容即時轉換為文字，並進一步整理重點、查詢困難字詞與產生建議回覆。系統同時導入虛擬人互動介面，透過口型、表情與動作呈現通話內容，補足傳統電話缺乏視覺線索的不足，讓使用者能以更直覺的方式掌握對話重點，系統功能架構圖如圖二所示。此外，本系統搭配模糊門控卷積神經網路進行模糊語音重建，協助構音不清者將語音轉換為較清晰、可理解的表達內容；並透過Nuvoton邊緣嵌入式系統進行即時人臉表情辨識，在使用者緊張、遲疑或無法即時回應時，主動提供AI建議回覆，讓輔具從被動接收資訊，升級為能主動理解、輔助表達與減輕溝通壓力的智慧系統。

本作品已完成實機整合與場域測試，展現良好的通訊輔助成效。針對較難被聽懂的發音內容，語意理解錯誤率平均降低35.0%，有效提升語音清晰度與可理解性。

此外，在86位受測者測試中，有90.7%對系統操作流暢度給予正面評價，顯示本作品具備實際使用可行性。相較於現有通話輔具，本系統不僅協助使用者聽懂通話內容，也能透過視覺化輔助、語音修復與建議回覆，改善傳統電話在理解與表達上的限制。未來可應用於居家照護、日常電話、聽障輔助與高齡通訊場域，協助使用者降低溝通焦慮，提升通訊自主性與社會參與度，讓電話溝通不只是聽得見，更能看得懂、說得清、自在表達。



圖一 自製3D列印機構爆炸圖。



圖二 系統架構圖。

Abstract

Taiwan has officially entered a super-aged society, and communication difficulties caused by age-related hearing decline among older adults have gradually become an important issue. At the same time, prolonged and high-volume headphone use among younger generations has also increased the risk of hearing impairment, making hearing loss no longer an issue limited to older adults, but a communication challenge that spans different generations. For people with hearing impairments and older users, the difficulties of telephone communication are not merely caused by insufficient volume, but also include unclear speech, background noise interference, difficulty in understanding content in real time, and the lack of visual cues such as lip movements and facial expressions. Existing hearing aids, real-time captioning tools, or call transcription systems mostly provide only sound amplification, one-way text display, or post-call summaries, making it difficult to solve the two-way communication problems of users "not being able to hear clearly" and "not being able to speak clearly." In addition, traditional single-computing architectures may suffer from latency or delayed responses when handling high-load tasks such as real-time speech recognition, generative AI, and image-based interaction, which may further increase users' communication frustration.

To address the above problems, our team developed "Hear-SeeText," an intelligent communication system. By placing a traditional telephone or smartphone into a self-designed 3D-printed structure, as shown in Fig. 3, the system enables collaborative operation between edge computing and generative AI to convert call content into text in real time, while further organizing key points, looking up difficult words, and generating suggested responses. The system also integrates a virtual human interaction interface, using lip movements, facial expressions, and body gestures to present call content and compensate for the lack of visual cues in traditional telephone communication. This allows users to grasp the key points of a conversation in a more intuitive way. The interface of system is shown in Fig. 4. In addition, the system adopts a fuzzy gated convolutional neural network for unclear speech reconstruction, helping users with unclear articulation convert their speech into clearer and more understandable expressions. It also uses a Nuvoton edge embedded system for real-time facial expression recognition. When users appear nervous, hesitant, or unable to respond immediately, the system proactively provides AI-generated response suggestions, transforming the assistive device from a passive information receiver into an intelligent system capable of actively understanding, supporting expression, and reducing communication pressure.

This work has completed physical integration and field testing, demonstrating strong effectiveness in communication assistance. For speech content that is more difficult to understand, the semantic understanding error rate was reduced by an average of 35.0%, effectively improving speech clarity and comprehensibility. In addition, among 86 participants, 90.7% gave positive feedback regarding the smoothness of system operation, indicating the practical feasibility of this work. Compared with existing telephone assistive technologies, this system not only helps users understand call content, but also improves the limitations of traditional telephone communication in both comprehension and expression through visual assistance, speech restoration, and suggested responses. In the future, it can be applied to home care, daily telephone communication, hearing assistance, and communication scenarios for older adults, helping users reduce communication anxiety, improve communication autonomy, and enhance social participation. With this system, telephone communication is no longer only about hearing, but also about seeing, understanding, speaking clearly, and expressing oneself with confidence.



Fig. 3 3D-printed structure.



Fig. 4 System interface diagram.