Editorial

Special Issue on Applications of Speech and Language Technologies in Healthcare

Inma Hernáez-Rioja [1,*] [ID], **Jose A. Gonzalez-Lopez [2]** [ID] **and Heidi Christensen [3]**

1 HiTZ Center—Aholab, University of the Basque Country UPV/EHU, 48014 Bilbao, Spain
2 Department Signal Theory, Telematics and Communications, University of Granada, 18015 Granada, Spain; joseangl@ugr.es
3 Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK; heidi.christensen@sheffield.ac.uk
* Correspondence: inma.hernaez@ehu.eus

1. Introduction

In recent years, the exploration and uptake of digital health technologies have advanced rapidly with a real potential impact to revolutionise healthcare delivery and associated industries. Advancements include options for remote assessment and consultation (telemedicine) that enable patients to engage with healthcare professionals and clinicians without travelling into clinics and hospitals. Another development is the investigation of how pervasive technologies, such as wearables and mobile apps, can help track biomarkers, physical activities, and moods. Overall, these advancements in digital health technologies, which are underpinned by artificial intelligence and machine learning, have the potential to improve patient outcomes, reduce costs, and enhance the overall quality of care.

In this context, the use of speech and language processing technologies in healthcare applications is an emerging field with great promise. Applications range from voice-based data entry, assessment and processing of, e.g., electronic health records, to the explorations of how biomarkers that are indicative of various pathologies can be robustly extracted and used for diagnostic support. As evidenced by the spread of topics covered in this Special Issue, speech and language processing technologies are now maturing at a rapid pace, necessitating a focused effort on translational research to make sure solutions are appropriate, meaningful, and accurate for all users in real-world settings.

2. Recent Advances in Application of Speech and Language Technologies in Healthcare

This Special Issue comprises ten papers that cover diverse areas in the field of speech and language technologies. The works are grouped into three main categories: detection of pathologies from speech features (four papers), automatic speech recognition of pathological speech (one paper), and use of speech synthesis and speech conversion technologies to improve communication for people with speech pathologies (five papers). The research featured in this Special Issue analyses a wide range of speech pathologies, including aphasia (two papers), esophageal speech (two papers), hypernasality (one paper), speech from individuals with Parkinson's disease (one paper), speech from individuals attending psychotherapeutic intervention (one paper), and speech from individuals with obstructive pulmonary disease (one paper).

The four papers focused on detecting speech pathologies present innovative approaches. In [1], acoustic and linguistic features that are automatically extracted from patient–practitioner conversations are successfully used to predict practitioners' competency. Ozbolt et al. [2] analysed methodological aspects in detecting speech pathology associated with Parkinson's disease from sustained vowels and provided recommendations to avoid misleading or overoptimistic results. Moreno-Torres et al. [3] focused on detecting

hypernasality using a mobile app that evaluates various types of utterances and concluded that including different utterance types improves the detection of hypernasality. Finally, Farrus et al. [4] analysed speech from patients with chronic obstructive pulmonary disease (COPD) to determine how it is affected by medication and physical effort.

Two papers address esophageal speech, both using speech conversion techniques to improve its quality and intelligibility. Ezzine et al. [5] used a novel sequence-to-sequence model with an auditory attention mechanism, while Raman et al. [6] used synthetic speech as the target of a voice conversion system to improve the quality and intelligibility of the original voice.

Aphasic speech is considered in two of the papers. Torre et al. [7] addressed the difficulty of transcribing very unintelligible speech and the scarcity of annotated data by proposing a new semi-supervised learning method with encouraging results. On the other hand, Cistola et al. [8] explored the use of TTS systems to support individuals with aphasia and reading difficulties and analysed whether the voice quality of synthetic speech affects the reaction times of patients.

Terblanche et al. [9] presented a comprehensive overview of the approaches and strategies used in 58 studies for the generation of child synthetic speech. Despite the rapid development of speech synthesis technology and high-quality adult synthetic speech, their study shows that developing child synthetic speech remains a challenging task.

Finally, Lee [10] proposed the use of harmonic enhancement postprocessing to improve the quality of speech obtained from a silent speech interface. The author used deep neural networks (DNNs) to estimate the spectral coefficients and emphasised the spectral fine structure of speech, while also minimizing the perceptual distance to a target natural speech.

3. Future Applications and Research Directions

The works presented in this Special Issue offer compelling evidence that speech and language technologies can be valuable tools in clinical scenarios for individuals with speech, voice, and language pathologies. These technologies have the potential to serve as diagnostic tools for early detection of speech pathology associated with health conditions, such as Alzheimer's or Parkinson's disease, or as assistive technologies aiming to improve people's lives in their everyday routines (e.g., as a communication tool for people with aphasia). While considerable progress has been made, more research is needed before these technologies can be accepted as reliable diagnostic tools in routine clinical services. This may entail conducting studies with larger sample sizes, thereby providing additional data to train state-of-the-art machine learning techniques. Once these challenges have been addressed, it is expected that these technologies will become commonplace in healthcare settings.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Attas, D.; Power, N.; Smithies, J.; Bee, C.; Aadahl, V.; Kellett, S.; Blackmore, C.; Christensen, H. Automated Detection of the Competency of Delivering Guided Self-Help for Anxiety via Speech and Language Processing. *Appl. Sci.* **2022**, *12*, 8608. [CrossRef]
2. Ozbolt, A.S.; Moro-Velazquez, L.; Lina, I.; Butala, A.A.; Dehak, N. Things to Consider When Automatically Detecting Parkinson's Disease Using the Phonation of Sustained Vowels: Analysis of Methodological Issues. *Appl. Sci.* **2022**, *12*, 991. [CrossRef]
3. Moreno-Torres, I.; Lozano, A.; Nava, E.; Bermúdez-De-Alvear, R. Which Utterance Types Are Most Suitable to Detect Hypernasality Automatically? *Appl. Sci.* **2021**, *11*, 8809. [CrossRef]
4. Farrús, M.; Codina-Filbà, J.; Reixach, E.; Andrés, E.; Sans, M.; Garcia, N.; Vilaseca, J. Speech-Based Support System to Supervise Chronic Obstructive Pulmonary Disease Patient Status. *Appl. Sci.* **2021**, *11*, 7999. [CrossRef]
5. Ezzine, K.; Di Martino, J.; Frikha, M. Intelligibility Improvement of Esophageal Speech Using Sequence-to-Sequence Voice Conversion with Auditory Attention. *Appl. Sci.* **2022**, *12*, 7062. [CrossRef]

6. Raman, S.; Sarasola, X.; Navas, E.; Hernaez, I. Enrichment of Oesophageal Speech: Voice Conversion with Duration–Matched Synthetic Speech as Target. *Appl. Sci.* **2021**, *11*, 5940. [CrossRef]
7. Torre, I.G.; Romero, M.; Álvarez, A. Improving Aphasic Speech Recognition by Using Novel Semi-Supervised Learning Methods on AphasiaBank for English and Spanish. *Appl. Sci.* **2021**, *11*, 8872. [CrossRef]
8. Cistola, G.; Peiró-Lilja, A.; Cámbara, G.; van der Meulen, I.; Farrús, M. Influence of TTS Systems Performance on Reaction Times in People with Aphasia. *Appl. Sci.* **2021**, *11*, 11320. [CrossRef]
9. Terblanche, C.; Harty, M.; Pascoe, M.; Tucker, B.V. A Situational Analysis of Current Speech-Synthesis Systems for Child Voices: A Scoping Review of Qualitative and Quantitative Evidence. *Appl. Sci.* **2022**, *12*, 5623. [CrossRef]
10. Lee, K.-S. Ultrasonic Doppler Based Silent Speech Interface Using Perceptual Distance. *Appl. Sci.* **2022**, *12*, 827. [CrossRef]