



Potential Diagnostic and Therapeutic Application of AI in Voice Disorders and Restoration

Authors:	*Serena Jiang, ¹ Oreste Gallo ²
	1. Department of Otorhinolaryngology, ASST Valtellina e Alto Lario, Sondrio, Italy 2. Department of Otorhinolaryngology, Careggi University Hospital, Florence, Italy *Correspondence to serena.jiang@asst-val.it
Conflicts of interest:	The authors declare there are no conflicts of interest.
Funding:	The authors declare they have received no funding for this study.
Gen AI:	None.
Peer review:	This article was accepted following double-blind peer review.
Received:	14.03.25
Accepted:	20.04.26
Keywords:	AI, augmentative and alternative communication (AAC), electrolarynx (EL), laryngectomy, voice rehabilitation.
Citation:	EMJ. 2026;11[2]:54-57. https://doi.org/10.33590/emj/J9L112P2

Abstract

Total laryngectomy, a sometimes-necessary surgical procedure, results in the loss of glottal sound production, leading to iatrogenic aphonia that can adversely affect a patient's quality of life and social interactions. Traditional post-laryngectomy speech rehabilitation methods, including oesophageal speech, electrolarynx, and tracheo-oesophageal voice prosthesis, often fail to preserve the patient's spoken identity and effective communication. Recent developments in AI technologies, which can decode complex brain activities into language, offer innovative solutions in the rehabilitative process. Integrating augmentative and alternative communication with traditional aids like electrolarynx can create a personalised communication tool. This paper aims to explore the potential of AI-driven speech rehabilitation methods to transform and improve the lives of laryngectomised patients, highlighting the need for continued research to refine these technologies and address ethical and practical considerations for clinical application.

Key Points

1. Technological innovations in AI-driven communication tools have strong clinical relevance, as they can improve patient experience and functional outcomes following laryngectomy.
2. This article provides a clear explanation of existing communication methods, such as electrolarynx devices and tracheoesophageal puncture.
3. Emerging technologies, including augmentative and alternative communication, voice banking, and electromyography-based systems, show promise, but they must be developed with consideration of practical constraints and ethical issues.

INTRODUCTION

The advent of AI tools in the early 21st century was revolutionary in healthcare. Concerning voice, many studies have focused on neurological patients.¹ For instance, Moses et al.² conducted real-time decoding of words and sentences from the cortical activity of an individual with limb paralysis and anarthria due to a brainstem stroke using high-density electrocorticography. Another team demonstrated that visual cortex neural activities indirectly measured by functional MRI could be decoded into phrases or sentences; their novel language decoding AI model, typically used in machine translation, could convert neural activities into human language using a speech synthesiser.

These ‘silent speech’ technologies (functional MRI, electrocorticography, etc.), while promising, have limitations in voicing prediction and practical application outside the lab.

Augmentative and alternative communication (AAC) refers to a group of methods and technologies designed to facilitate communication by utilising artificially generated voices to support or replace speech or writing for individuals with communication difficulties.³

Real-world application of AAC is reported by various newspapers: in July 2024, Jennifer Wexton, who lost her ability to speak clearly due to progressive supranuclear palsy, regained her voice with an AI program that cloned her speaking voice using old recordings; she simply typed her thoughts and played her voice through an iPad. In 2019, James Vlahos created a chatbox powered by AI that could answer questions about his dad’s life using his father’s voice, leading to the app HereAfterAI (HereAfter, El Cerrito, California, USA). In Seoul, South Korea, DeepBrain AI (also headquartered in Palo Alto, California, USA) creates video-based avatars by capturing extensive video and audio to replicate a person’s face, voice, and mannerisms. Given these premises, it is natural to wonder if AI can also help patients with iatrogenic aphonia in communicating with their own voice.

CURRENT COMMUNICATION METHODS

Total laryngectomy (TL) profoundly alters the physiology of respiration and deglutition as well as the production of voice. It consists of the surgical excision of the larynx due to cancer, accidents, burns, or trauma. Laryngeal cancer accounts for 21.6% of all new head and neck cancer cases in Western countries, and is the second most common head and neck malignancy.⁴ Despite significant advancements in organ-preserving treatments and surgical strategies, preserving laryngeal functions without compromising oncologic outcomes is not always feasible.

A person who has undergone a laryngectomy loses the ability to speak normally while maintaining the ability to breathe through a permanent tracheostomy and eat through the neopharynx. Loss of glottal sound production is the greatest disability for most patients following surgery; a combination of fear of losing their identity, the inability to communicate wishes, and unmet basic needs can impact the therapeutic process.

Nowadays, there are various possibilities for post-laryngectomy speech rehabilitation: erigmophonic voice (oesophageal speech), electrolarynx (EL), and the trachea-oesophageal voice prosthesis, considered the current gold standard of alaryngeal communication in the absence of contraindications. These methods often sacrifice the patient’s spoken identity and hinder communication, reducing patient-reported quality of life and affecting family members and caregivers.

ROLE OF EMERGING TECHNOLOGIES

In 2017, Gilbert et al.⁵ used a permanent magnet articulography system, capturing lip and tongue movements to synthesise speech. During the conversion stage, the sequence of feature vectors extracted from permanent magnet articulography data was processed through a trained recurrent

neural network to predict speech features. The synthesised speech closely resembled the user's own voice, as the training was based on their personal recordings. However, the training process depended on the availability of parallel data for each patient, which may not always be feasible. This could be due to the patient having already lost their voice or insufficient time between diagnosis and laryngectomy to arrange recordings. In such cases, a 'donor' voice from another person, perhaps a relative, could be used as an alternative.

Two years later, Rameau⁶ conducted a pilot experiment using machine learning to analyse surface electromyography signals from the articulatory muscles of the face and neck, enabling the recognition of silent speech in a patient who has undergone laryngectomy. In addition to developing a recognition algorithm, their goal was to create an application that converts recognised speech into a 'user-selected computer-generated voice', allowing for personalised vocal output.⁶

The use of the aforementioned AAC methods in all patients who have undergone laryngectomy could be considered as an emerging benefit: they could potentially increase patients' quality of life and decrease the number of patients who are afraid to accept this type of surgery. This may be one of the best options for communication in the days right after TL, or when the alaryngeal speech is dysfunctional.

AAC works through voice banking and text-to-speech devices.⁴ This process entails recording spoken words and sentences with appropriate intonation and emotional context; these are then stored and organised for future communication. Time plays a crucial role, as TL is often planned shortly after the diagnosis of the malignant disease. To create a high-quality speech corpus, it is necessary to record a substantial number of sentences, ideally in a soundproof studio, although a portable high-quality voice recorder could allow the process to be carried out at home or in the hospital setting; voice recording should be discontinued if the patient's voice quality begins to deteriorate due to

fatigue. For patients whose voice quality is already affected, previous recordings (e.g., voice messages or videos on social media) can be used. It doesn't require lengthy training time, and the expenses are limited to the voice recording device and a smartphone or tablet with an application for voice elaboration and synthesis. However, access to such technologies, levels of digital literacy, and the availability of supporting clinical or technical infrastructure may vary significantly across healthcare settings, potentially affecting the feasibility and scalability of voice banking in different populations.

To have an even more rapid and spontaneous way to communicate, the authors propose the integration of this 21st century technology with a late-1920s invention: the EL. It represents a viable option for those who experience difficulty, restrictions, or complications with other communication tools. EL voice is monotonous, mechanical, and is perceived as unnatural in its overall sound quality; furthermore, it has limited use in tonal languages, such as Mandarin Chinese.⁷ This unpleasant auditory-perceptual outcome may have direct implications for social interactions and communication effectiveness. The speech aid is also limited by its manual control, and this condition affects the spontaneity of the speech. Over the years, some improvements have been made. For example, the setting of the fundamental frequency (F0) and range of pitch modulation can now be user-defined; the tones in Mandarin can be produced effectively by the tone-EL, and the quality of rehabilitated language is better than using a conventional EL.⁷ Researchers have sought to address these limitations using an intra-oral EL or a hands-free EL activated by neck muscle electromyography activity, or by incorporating a mouth-facing camera to control the device's automatic on/off function.

Integrating the EL voice output with AAC methods creates a personalised speech rehabilitation with the possibility of using the patient's own voice. It facilitates essential interactions between the patient, medical team, nursing staff, and speech and

language pathologist, while also helping the patient's caregiver overcome the period of voiceless communication. Although the inability to express emotions nonverbally may be a limitation, the ability to convey wishes, thoughts, and needs through words and sentences in their own voice can serve as a powerful tool for patients against the loss of identity. Therefore, this intervention augments a patient-centred shared decision-making experience.

Evidence on potential negative or unintended psychological impacts, such as emotional responses to hearing a synthesised version of one's own voice or mismatches between expectations and actual experience, remains limited, but further exploration of these aspects would add depth to the discussion.

Although some view AI in competition with human capabilities as harmless fun, others warn of the potential for misuse in misinformation campaigns and fraud. Authenticity and privacy are significant concerns with AI systems, which often rely on large amounts of personal information. AI algorithms and deep learning models, in particular, are like 'black boxes': highly complex and opaque.⁸ Ensuring transparency and accountability in AI is crucial for building trust and safeguarding ethical behaviour. Additionally, the high computational power required for AI can lead to increased electronic waste and environmental impact.

In clinical settings, it is essential to develop and implement AI algorithms ethically and without bias. Clinicians need training to integrate AI tools effectively into their practice while maintaining a patient-centred approach. Addressing extensive data requirements, ethical concerns, and accessibility issues is likewise critical for AI-driven healthcare solutions.

CONCLUSION

AI offers a promising avenue for improving the quality of life and communicative abilities of patients who have undergone a laryngectomy; these innovations could be integrated into existing speech and language therapy pathways by complementing clinician-led interventions with digital tools, and enabling more personalised, continuous support within multidisciplinary care teams. While challenges remain in voicing prediction, data quality, and ethical considerations, the integration of AI in speech rehabilitation represents a step forward in patient-centred care, restoring speech for people who have lost the ability to express themselves with their own voice. Future research should continue to refine these technologies, ensuring that they are accessible, effective, and respectful of patient privacy and autonomy.

References

1. Anderer S, Hswen Y. Digital avatars and personalized voices-how AI is helping to restore speech to patients. *JAMA*. 2024;331(15):1259-61.
2. Moses DA et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N Engl J Med*. 2021;385(3):217-27.
3. Repova B et al. Text-to-speech synthesis as an alternative communication means after total laryngectomy. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2021;165(2):192-7.
4. Lu YA. et al. Seeking medical assistance for dysphonia is associated with an improved survival rate in laryngeal cancer: real-world evidence. *Diagnostics*. 2021;11:255.
5. Gilbert JM et al. Restoring speech following total removal of the larynx by a learned transformation from sensor data to acoustics. *J Acoust Soc Am*. 2017;141(3):EL307.
6. Rameau A. Pilot study for a novel and personalized voice restoration device for patients with laryngectomy. *Head Neck*. 2020;42(5):839-45.
7. Li W et al. Design and preliminary evaluation of electrolarynx with F0 control based on capacitive touch technology. *IEEE Trans Neural Syst Rehabil Eng*. 2018;26(3):629-36.
8. Wong F et al. Discovery of a structural class of antibiotics with explainable deep learning. *Nature*. 2024;626(7997):177-85.