www.pipelinepub.com
Volume 20, Issue 8

Assistive Voice Technology: the Real-time Solution to Express Yourself

By: Joris Castermans

Worldwide, 300 million people suffer from some speech disability. Losing one's voice has profound and far-reaching effects on a person's life. Communication, a fundamental aspect of human existence, becomes a daily challenge, leading to social isolation and frustration as individuals struggle to convey their thoughts, emotions, and needs to others. Professionally, careers may be jeopardized as many occupations require effective verbal communication. This can lead to financial stress and a loss of independence. Moreover, diminished vocal ability can erode self-esteem, leading to feelings of inadequacy and depression.

There is a wide variety of speech disabilities, with different limitations in speaking easily, in-flow, and intelligibly. Therefore, to better understand and categorize the different types of assistive voice technologies, it is important to first have a better understanding of the concept of speech production.

Communication, talking and expressing oneself, is a fundamental human right. Yet several impairments may hinder this ability, leading to various communication challenges. From Augmented Alternative Communication (AAC) tools to the latest real-time assistive voice technology, advancements in this field facilitate communication for those with speech-related disabilities. The world of assistive technologies is vast and varied, catering to many impairments and challenges. However, amidst this backdrop, the positioning of real-time assistive voice technology remains unique.

The Source-Filter Model: Understanding the Speech Impairment Spectrum

The speech impairment spectrum encompasses various conditions affecting an individual's ability to produce clear and compelling speech. These impairments range from articulation disorders,

where specific sounds are tricky to pronounce, to fluency disorders, like stuttering that disrupts the flow of speech. Voice disorders affect the pitch, volume, or quality of the voice, while motor speech disorders such as dysarthria and apraxia involve difficulties due to muscle control or neurological issues.

A critical framework used to explain how speech is generated is the "Source-Filter Model." This model helps us understand how the vibrations produced by the vocal cord (sound source) are modified and shaped by the form and movements of the vocal tract (filter) to produce clear and intelligible speech.

The Source-Filter Model of speech production is a fundamental framework for understanding how humans produce speech sounds. It conceptualizes speech production in two main components: the "source" and the "filter." Let's delve deeper into how this model explains voice quality and articulation. We'll then explore how this framework can be adapted for individuals using Augmented Alternative Communication (AAC) tools and speech-to-text technologies.

1. *The Source*: The source of speech is primarily the air pushed out from the lungs that causes the vocal cords within the larynx to vibrate. This vibration produces a fundamental frequency, which is perceived as the pitch of the voice. The nature and intensity of the vibration (e.g., loud or soft, high or low pitch) contribute to the voice quality.

2. *The Filter*: The filter comprises the articulatory movements of the vocal tract, including the tongue, lips, and palate. Articulation disorders manifest when these movements are imprecise, as seen in conditions like apraxia, or due to structural anomalies. Dysarthria, characterized by weak muscle control affecting the filtering mechanism, results in slurred or slow speech that can be difficult to understand.

Advancements in Supportive Communication Technologies

For individuals facing challenges in both the source and filter components — such as those with no articulation or compromised voice quality — Augmented Alternative Communication (AAC) tools provide significant support. AAC devices range from simple picture boards to sophisticated electronic systems that generate speech, allowing individuals to communicate effectively despite their impairments.

Supportive communication technologies are crucial in enhancing interaction for individuals with speech impairments. Many forms of these technologies, such as text-to-speech devices, convert typed text into synthetic speech, enabling those who can use manual input or eye-tracking technologies to communicate effectively. Speech-generating devices (SGDs) are tailored with pre-programmed and customizable phrases, facilitating daily communication and social interactions. Voice Output Communication Aids employ digitized pre-recorded human voices and synthesized speech, allowing users to express themselves through touchscreens and buttons.

Using the Source-Filter Model explained earlier, we can categorize assistive speech technology based on the combination of the user's ability to generate enough sound for intelligible speech and their ability to articulate well enough for intelligible speech:

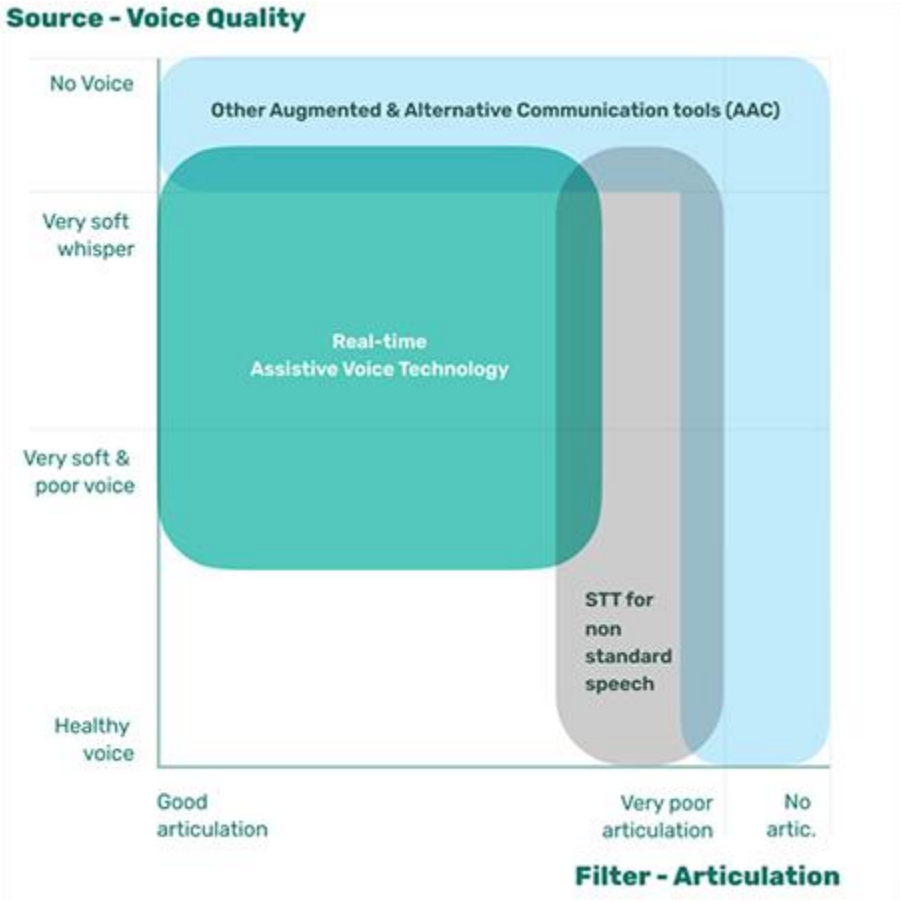

Figure 1. Categorization of assistive speech technology using the Source-Filter Model

Automatic Speech Recognition (ASR) for non-standard speech is for people who are able to generate enough sound but have bad articulation because of disorders that degrade articulation, such as ALS, MS, stroke, and Parkinson's Disease.

Other Augmentative Alternative Communication (AAC) technologies and tools are for people who can't produce any speech sound and/or have articulation that is too difficult to be recognized by ASR. Think of locked-in syndrome or severe cases and combinations of the earlier-mentioned disorders.

Real-time Assistive Voice Technology (RAVT): for people who cannot generate enough sound but (still) have intelligible articulation like laryngeal cancer, vocal cord paralysis, and benign voice disorders like vocal cord polyps and cysts. (See Figure 1 on previous)

Big tech and assistive tech companies predominantly focus on ASR and AAC. Both ASR and other AAC solutions rely on generating text, which is then synthesized by a computer or device. While very helpful for the mentioned target groups, these technologies come with delays, usually around 2 to 3 seconds, which create barriers to natural conversation. As a result, they are not able to provide an adequate solution for many people.

The Unmet Need: Real-time Assistive Voice Technology

As someone who has experienced a stuttering disability myself, I have direct knowledge and understanding of the challenges, emotions, and realities associated with speech disabilities. I know that having this delay is not good enough to solve the problem of being unable to express yourself. Imagine you want to speak over the phone or have a conversation in person with a 2-3 second delay — would you use it?

The predominant focus of these companies for non-standard speech is speech-to-text (STT). Although this STT-TTS approach is beneficial for patients with reduced articulation (ALS, MS, stroke, and Parkinson's' Disease), the high latency creates barriers to having natural conversations with a natural "flow." With this, the current AI speech technology solutions cannot provide an adequate solution for people with voice disorders who have lost their voice but still have good articulation.

In contrast to the current assistive speech tech, assistive voice technology is audio-to-audio based on AI, resulting in a highly scalable real-time conversion due to its language-independent nature.

Real-time assistive voice technology using whisper-to-speech technology fills this gap and improves the quality of life of a currently underserved group of people who have an affected or (almost entirely) lost voice. It is revolutionary because it takes advantage of the fact that deliberately steering the voice towards whispering is helpful for several disorders due to neurological changes in the speech system. For example, people who stutter severely can reduce their stuttering frequency by an average of 85 percent with whispering. Also, people who suffer from Spasmodic Dysphonia or Recurrent Respiratory Papillomatosis speak much more relaxed and fluently when they whisper.

Maximizing Impact in Advancing Assistive Technologies

Collaborating with esteemed governmental institutions like the Netherlands Cancer Institute and engaging in initiatives such as the Dutch Ministry of Social Affairs and Employment's "Technology for Inclusion" pilot can significantly enhance the reach and efficacy of assistive technologies.

My experiences have shown that such partnerships advance technology and ensure it serves those who need it most. Speaking at key conferences and participating in impactful projects demonstrates how technological innovation can directly contribute to societal inclusion, improving the quality of life for people with disabilities. These collaborations are crucial for driving both innovation and social change forward.

I invite others in the field to join us in forging new partnerships and exploring innovative solutions. Together, we can drive technology forward and achieve profound social change, creating a more inclusive world for everyone.