

**The Swallows**

Head & Neck Cancer Support Group

Registered Charity Number: 1149794



# AI Voice Technology

**The Swallows**

Head & Neck Cancer Support Group

Registered Charity Number: 1149794

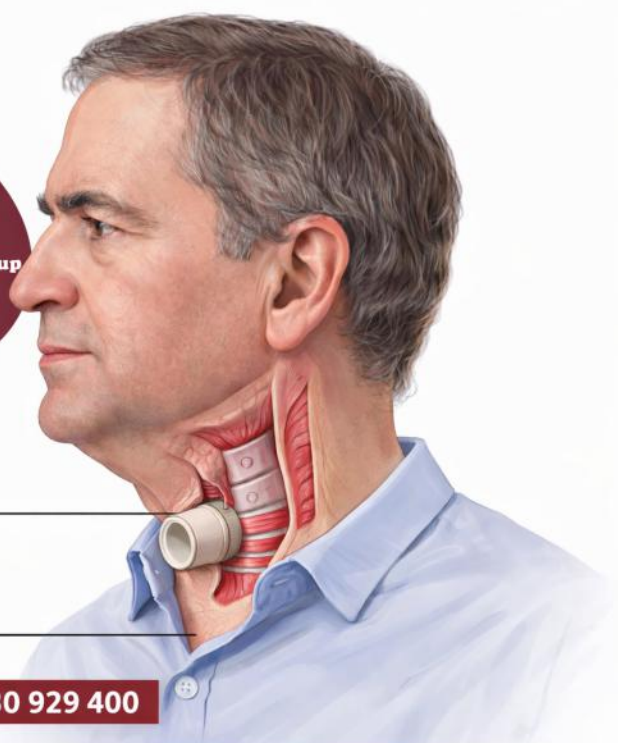


Laryngectomy

Tracheal stoma



24/7 Text Support 07830 929 400







J.H. Snider is a political scientist and public policy analyst who became a laryngectomee in Sept. 2023.

## When technology gives back what cancer took away, a voice that can finally be heard....

The ability to enhance someone's voice in real time is a dream for those of us in the laryngectomy community. The technology to do so has improved dramatically. But is it ready for prime time?

This article follows up on an article I published almost two years ago. (See [Voice-to-Voice Cloning for the Laryngectomee Community: Not Quite There, But Close](#)). At the time I wrote that article, AI-enhanced voice-to-voice synthesis was gaining momentum. Still, real-time voice enhancement was weak due to excessive latency. Voice conversion technology changes an input voice (e.g., my awful, hard-to-understand voice) into an output voice (e.g., a pleasing, easily understandable voice in contrast to mine). But the ability to voice convert in this way is not enough: the conversion also needs to be done in real time (that is, with low "latency") so that a voice conversation sounds normal.

### A Note on Terminology

The voice technology marketplace is changing fast, so the voice conversion terminology used in this article may soon be obsolete. Other terms for AI enhanced customisable voice technology include voice morphing, voice changing, voice cloning, voice modulation, voice conversion, voice dubbing, voice masking, voice swapping, speech-to-speech synthesis, speech transformation, audio persona transfer, vocal style transfer, and generative voice AI.

### Synchronous vs. Asynchronous Voice Conversion

The first type of voice-to-voice conversion to become widely available was asynchronous, such as for audiobooks, podcasts, and advertisers, where the focus is on maximising voice output quality rather than real-time processing. More recently, some companies have begun offering low-latency voice conversion. Most of these products are not directed at the disabled community.

The largest current consumer market for real-time voice conversion appears to be gamers who communicate verbally with other gamers and prefer to use a voice other than their own. Serving the disabled community may be more technically challenging because the input voice for the voice conversion is not as good.

### Altered.AI's RealTime Pro

I recently found a product that meets my usability requirements, including low latency, high-quality voice output, ease of use, and affordability. As a bonus, it can also clone my original voice from up to a 10-minute sample of my pre-laryngectomy voice. Alternatively, I can choose from dozens of other voices (called "skins"), which in some respects make me sound better than I did before my laryngectomy.

The product, RealTime Pro, which describes itself as a "voice changer for voice & video calls," is produced by [Altered.AI](#), a UK-based company run by a former Google AI voice engineer. The product works with video conferencing platforms such as Zoom and Google Meet. It also works with gaming chat services such as Discord, Steam, and Teamspeak. Unfortunately, it does not work with smartphone calling services such as T-Mobile, Verizon, or AT&T, nor for face-to-face communication. For it to work, there must be some telecommunications intermediary between the input and output voice. I also found that if I didn't wear a headset, the receiver, but not me, would hear an annoying echo.

I bought an annual RealTime Pro plan called "Euphonia" for \$72. Euphonia describes itself as alleviating "various forms of dysphonia and voice disfluencies during voice or video calls."

RealTime Pro offers 20 minutes of free use per day. That was ample time for me to test the product with family members and get their feedback on various voices. They thought my British accent was better than my original voice before I had a laryngectomy. But I was only interested in enhancing my original voice based on a recording I made before my laryngectomy. I had no idea what my family was listening to, so I asked them to record my voice. After they played it back, I was amazed by its improved quality. But I also learned that I needed to speak slowly and with good enunciation. Even then, the consonant "r" tended to be mangled, which seemed like a software glitch. Trusting the product is essential because when I'm using it, I cannot hear what the other person hears. If you're on an important call, you want to make sure you are confident in your computer's performance and the software's existing settings. The 20-minute free limit per day was wholly inadequate for me, given the webinars and other long-form communications I regularly engage in. A nightmare for me would be to be in mid-sentence only to have my voice degenerate to my unenhanced voice.



Voice conversion companies are continuously tinkering with their products. What is true today may not be true tomorrow.

The critical point is that real-time voice conversion is now viable, albeit only in limited circumstances.

As for me, I'm thrilled to have been empowered to do many of the things I once took for granted before my laryngectomy





J.H. Snider is a political scientist and public policy analyst who became a laryngectomee in Sept. 2023.

When technology gives back what cancer took away, a voice that can finally be heard....

### Your Experience Might Differ

Just because this product works for me does not mean it will work for you. You would have to try it out to make that determination for yourself. Suppose your voice is tougher to understand. In that case, the technology might not produce sufficiently high-quality voice output for you. You also need a personal computer manufactured in the last year or so that has a fast processor; a low-end computer won't do, but most computers sold today are adequate. Altered.AI lists the recent processors that work fast enough to work well with its product.

It will also work with older processors, but not as well. The relative quality of your voice may also be an essential consideration. I'm a lucky laryngectomee in that I still have a relatively good voice that motivated listeners can understand. But many people are not so motivated. My voice is low and gravelly. I'm often mistaken for an old woman when I call vendors who do not know me. Much worse is when such vendors hang up on me as soon as they hear my voice. But for people who want to listen to what I have to say, especially family members, friends, and colleagues, they can do so with relatively little effort. My general rule is that if I can get someone to stay on the phone for 30 seconds, it will work out. However, TV and radio interviews with reporters have become impossible, as my voice is not good enough for a mass audience (I do public policy work, and reporters sometimes seek me out as a resource). And webinars feel embarrassing and often impractical for me, because I expect I will alienate some audience members, and I don't want to take up time explaining my voice situation.

Altered.AI has empowered me to speak comfortably to many of the same audiences I once did. But I must speak slowly and with good enunciation for it to work well. My words will garble if I talk at the rapid clip that comes most natural to me.

### Caveat Emptor

Voice conversion is a fast-changing field with many emerging competitors and, for some of them, a Wild West sensibility. Don't assume that you can trust a professionally designed website to make accurate product claims, as there seems to be a fake-it-before-you-make-it culture in this field, a common occurrence among high-tech entrepreneurial companies. I'd suggest trying the product before buying it. If the company wants you to spend a lot of money before trying out its product, you might want to consider other options. This market will likely shake out within a few years.

### Real-time Face-To-Face Voice Conversion

One exciting technological breakthrough on the horizon is the ability to use earbuds as an intermediary device for real-time face-to-face voice conversion. Google and Apple are already offering real-time voice language translation services (such as Spanish-to-English and back from English-to-Spanish), and it seems only a matter of time before they reach the voice disability community. I would think that real-time language translation is more technically complex than voice conversion in the same language. But the language translation market might be both a much larger and more dramatic exhibition of AI voice technology.

Hearing aids already have some AI-enabled sound-enhancing features, but they enhance ordinary voices, not laryngectomee voices, and don't require the laryngectomee's audience to have the technology.

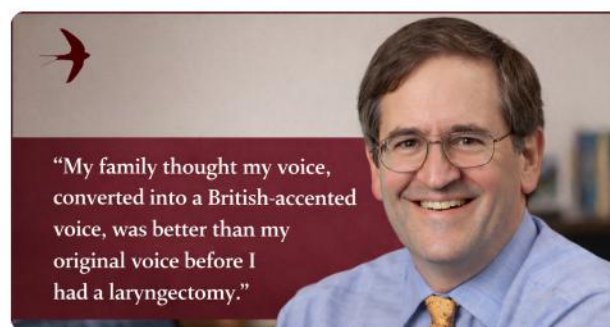
The combination of AI-enabled noise cancellation and voice conversion is magic to me. But the requirement that listeners wear ear devices may limit its use to small, face-to-face audiences, such as family, friends, and colleagues. The barrier posed by that listener requirement may be dramatically reduced if AI-enhanced earbuds become ubiquitous, as they are used for music listening, noise suppression, and other purposes.

### Asynchronous Communication Products

This article is about synchronous, not asynchronous, voice conversion, even though they are closely related products. For asynchronous voice conversion, including text-to-speech and voice cloning, I recommend [ElevenLabs](#), which offers a [free version for laryngectomies](#). While ElevenLabs has a low-latency voice model, it does not currently offer a product that directly competes with Altered.AI.

### Conclusion

Voice conversion companies are continuously tinkering with their products. What is true today may not be true tomorrow. The critical point is that real-time voice conversion is now viable, albeit only in limited circumstances. As for me, I'm thrilled to have been empowered to do many of the things I once took for granted before my laryngectomy.







J.H. Snider is a political scientist and public policy analyst who became a laryngectomee in Sept. 2023.

When technology gives back what cancer took away, a voice that can finally be heard....

To illustrate how the software converts an input to an output voice, I read the following two sentences into a Zoom conversation with my wife:

*“My family thought my voice, converted into a British-accented voice, was better than my original voice before I had a laryngectomy.”*

But I was only interested in converting my voice to a voice based on a recording of my voice before I had my laryngectomy.

My intonation would have been better if I had spoken spontaneously rather than read from a script. But you should still be able to get a sense of how the product works in practice. Note that the lips and the sounds they produce don't sync exactly; the sounds are slightly delayed. My voice begins about four seconds into each clip.



“Addendum: Since this article was submitted to The Swallows, Altered.AI has introduced a cloud based version of RealTime Pro called “Remote Compute” that obviates the need for high powered local computer processors. This product enhancement serves as a reminder that this is a fast-moving marketplace and that what was true a month ago may not be true today.”

