

Breaking Barriers in Learning: The Role of Human-Sounding Text-to-Speech

About the Presenter



Daniel Goerz – Co-Founder, Messenger Pigeon (formerly Note Taking Express)

- 6'3" Caucasian male in his late 30s, with short brown hair and glasses.
- Began career in 2008 supporting students in post-secondary education

Background in disability services and educational accessibility:

- Note taker, Learning Strategist, Exam Proctor
- Alternate Format Specialist (Kurzweil), Assistive Technologist
- Founded Note Taking Express in 2014 to better serve diverse learners
- Leads Messenger Pigeon, a platform built with input from students and accessibility teams
- Advocates for inclusive, human-centred technology
- Believes technology is a tool, not a replacement, for thoughtful, community-driven support

The Challenge and the Opportunity

The Challenge:

College students facing these barriers to learning:

- Reading fatigue and reduced stamina
- Auditory processing issues or fast-paced, accented lectures
- Voice bias in TTS: most are British or robotic
- Lack of representation in learning tools, especially for first-gen students

The Opportunity:

Human-sounding TTS helps students:

- Hear voices that reflect their identity
- Feel a sense of belonging
- Learn at their own pace with greater comprehension and comfort

Sources: The Atlantic (2024), The Chronicle of Higher Ed (2023), EDUCAUSE Review, Gierdowski & Galanek (2019)



Evolution of TTS – From Robot to Human

Robotic TTS (Then):

- Monotone, unnatural rhythm
- Limited emotion or expressiveness
- Often UK-accented or non-relatable

Human-Sounding TTS (Now):

- Natural cadence, pitch variation
- Emotion, warmth, and personality
- Customizable: accent, tone, gender, speaking speed

Let's listen to the difference...



In the upcoming seminar, Dr. Lin will cover CRISPR, AI, and UDL. Students should review the PDF on ATP synthesis and RSVP to the U.S. DOE panel by 3:00 p.m. sharp. Note: “IRB” here means Institutional Review Board—not infrared beacon, like last time.

Dr. Eleanor Vance, after meticulously reviewing the data, found the content of the ancient scrolls remarkably present. She then proceeded to read the document, noting, 'The villagers' simple lives in the desert often yielded unexpected fruit, unlike the urban populace whose primary concern was how to wind their expensive watches.' A recurring anomaly was the notation $E=mc^2$ followed by a cryptic URL: <https://archaeo.log/site-007.html>. Her associate, Siobhán Dubois, concurred, remarking that the archaeological dig provided a unique insight."

The concept of 'duende' in Spanish flamenco music is notoriously difficult to translate directly into English, often described as a heightened state of emotion, authenticity, and mystical connection. While we might approximate it with terms like 'soul' or 'spirit,' these fall short of its profound depth. As many performers would attest, '¡El duende está en el alma de la música y el baile!' This profound essence is what truly captivates audiences and elevates a performance beyond mere technical skill, making it an integral part of the artistic experience.

When learning Korean, one of the most fundamental phrases you'll encounter is 'Hello,' which has different forms depending on the level of politeness required in the situation. The most common and generally polite way to say 'hello' or 'how are you' is 안녕하세요. This versatile greeting can be used in most everyday situations, whether you're addressing an elder, a stranger, or someone of similar age. For a more informal setting among close friends or family, you might use 안녕 (annyeong). Understanding these nuances in politeness, known as 'honorifics,' is fundamental to respectful communication in Korean society, far beyond simply knowing the vocabulary.

As we listen to this **first example**, think about (and then we will review together):

- Is the meaning clear on first listen?
- How does this voice make me feel — engaged or fatigued?
- Are there moments where I felt lost, zoned out, or had to work harder to understand?
- What errors did you catch?

As we listen to this **second example**, think about (and then we will review together):

- Is the meaning clear on first listen?
- How does this voice make me feel — engaged or fatigued?
- Are there moments where I felt lost, zoned out, or had to work harder to understand?
- What errors did you catch?

Now I want to demonstrate to you the real potential of human-sounding text-to-speech when we go beyond just 'neutral' voices. These next few examples feature the exact same content read in different regional accents common across the U.S. As you listen, I invite you to think about two things: first, how easy it is to understand each voice; and second, how hearing a familiar accent might impact a student's sense of comfort or belonging. Let's listen.



Representation in Voice Matters: TTS for Comprehension and Equity

Why accent variation in TTS matters:

- Many students struggle to understand unfamiliar or fast-paced speech
- Research shows that relatable voices improve processing speed and comprehension
- Culturally relevant speech patterns help students feel like they belong
- TTS with voice options gives students equal access to understanding, not just content

Research Highlights:

- Gierdowski & Galanek (2019): Students report TTS improves focus, comprehension, and retention
- EDUCAUSE (2021): Students prefer TTS that “sounds like someone they know or trust”
- Golonka et al. (2017): Accent familiarity improves learning outcomes in audio-based content
- Sato & McDonough (2019): Students process information more easily when it’s delivered in accented speech that matches their linguistic exposure

Conceptual Table: Student Preference for TTS Voices

This table illustrates the impact of voice familiarity on student preference and engagement with Text-to-Speech (TTS) technology, based on findings like those from EDUCAUSE (2021).

Voice Characteristic	Impact on Student Experience	Learning Outcome Correlation
Familiar Accent/Dialect	Increased comfort, sense of belonging, reduced cognitive friction	Enhanced comprehension, improved retention, higher engagement
Trusted/Known Voice Tone	Builds rapport, increases credibility, fosters psychological safety	Greater acceptance of content, sustained attention, positive emotional response
Neutral/Generic Voice	Functional, but may lack emotional resonance, less personal	Basic comprehension, potential for disengagement over time
Unfamiliar/Unnatural Voice	Can cause discomfort, distraction, increased cognitive load	Reduced comprehension, lower retention, potential for avoidance

Key Takeaway: When TTS voices resonate with a student's personal experience or cultural background, it significantly enhances their sense of belonging and reduces the cognitive distance often associated with artificial voices, leading to more effective learning.

How Human-Sounding Voice AI Works — and What Higher Ed Needs to Consider

How It Works:

- Built using deep learning models trained on licensed recordings of real human speech
- Voices come from consenting professional voice actors who are compensated when their voice is used
- AI models learn intonation, rhythm, emotion, and pronunciation
- Speech is generated in real-time and can be customized by accent, tone, pitch, and speed
- Creates more relatable and intelligible audio, particularly for students with learning barriers

Limitations & Ethical Considerations in Higher Ed:

- Does not replace human support
 - TTS, like speech-to-text, is rarely 100% accurate
 - Complex materials (STEM, foreign languages, heavily formatted documents) may still require human alternate formatters
 - Students often need both technology and personal support to fully access content
- Technical limitations:
 - Struggles with formulaic or structured content like equations, tables, or LaTeX
- Ethical concerns:
 - Risk of misuse (e.g., deepfakes, unauthorized cloning)
 - Transparency required: students and faculty should know when AI is used

Use Case: TTS for Accented Faculty

- Some students struggle to follow lectures due to accent or pacing

Especially challenging for:

- Auditory processing disorders
- ELL and multilingual learners
- First-gen or first-year students

TTS Applied to Transcripts:

- Instructor lecture is recorded once
- Transcript is provided to students
- TTS converts it into customizable, private playback
- Students hear it in a voice that improves clarity and comfort

Faculty Benefit:

- TTS offers students voice-based review without having the original recording
- Balances IP protection with student access needs

Key Takeaways

Why It Matters:

- Reading and auditory barriers persist—even in higher ed
- Students deserve tools that support not just access, but understanding
- Voice options foster identity, comfort, and connection

What We Learned:

- Human-sounding TTS improves comprehension and retention
- Familiar voices reduce cognitive load and cultural disconnect
- Accent representation = academic equity

What to Mindful of:

- It's not perfect: still needs human support for accuracy and context
- Must be deployed transparently and ethically
- Used wisely, it can level the playing field for all of learners

Q&A + Thank You

Daniel Goerz

daniel.goerz@habitatlearn.com

References:

- Gierdowski, D. & Galanek, J. D. (2019). The Future of Text-to-Speech in Higher Education. EDUCAUSE.
- <https://er.educause.edu/articles/2019/12/the-future-of-text-to-speech-in-higher-education>
- EDUCAUSE (2021). Student Technology Preferences and Accessibility Reports.
- <https://www.educause.edu/>
- EDUCAUSE Review. Voice and Trust in Digital Learning Tools.
- <https://er.educause.edu/>
- The Atlantic (2024). The Elite College Students Who Can't Read Books.
- <https://www.theatlantic.com/>
- The Chronicle of Higher Education (2023). Is This the End of Reading?
- <https://www.chronicle.com/>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2017). Accent familiarity and learning outcomes in L2 education.
- Available via: <https://www.researchgate.net/>
- Sato, M., & McDonough, K. (2019). The role of accent familiarity in second language comprehension.
- Available via: <https://scholar.google.com/>