Northwestern | SEARLE CENTER
FOR ADVANCING LEARNING AND TEACHING

Generative AI Audio Tools and Applications

Brief Description

An overview of the types of generative AI audio tools available, along with potential uses for these tools in a teaching and learning context.

Introduction

While much of the conversation around using generative AI tools in teaching and learning contexts has been focused on text-based tools like ChatGPT, the pedagogical possibilities of audio generative AI are just starting to emerge. These tools can be used to increase accessibility to course content, allow students to receive feedback in multiple formats, and approach teaching and assessment with creativity.

What are the main categories of generative AI audio tools?

The availability of open-access generative AI audio tools is rapidly growing, but presently there are far fewer options compared to text and image. Most tools fall into one of the categories described below:

- **Speech-to-text:** Convert spoken words, either synchronously or from a recorded file, into text. Many people are familiar with Alexa from Amazon, Siri from Apple, and Google Assistant, which can send hands-free texts or take notes from voice commands. Other options are Otter.ai and Transcribe, which can create timestamped transcripts from live and imported audio.

  - **Potential Use:** Create transcripts for course material. Instructors can record their classes or upload recorded audio or video files and these tools will generate a transcript. While these tools can transcribe live events, the transcripts often need to be reviewed for accuracy. For instance, correcting the spelling of field-specific terminology.

- **Text-to-speech**: Conversely, text-to-speech tools will read aloud text from files, email or text messages, and websites, as well as text entered directly into the app. Aside from Alexa, Siri, and Google Assistant, services like Play.ht can generate natural sounding audio from provided text. Play.ht lets users adjust the speaking style, tone, and speed of the playback. There is also a feature to insert pauses to create a natural cadence for the audio.

  - **Potential Use**: Hear written work read aloud. Students can catch typos and confusing sentences by having these tools read their work back to

them.

- **Potential Use**: Offer creative options for assignments (i.e., podcast, news story, performance). Students can turn text into dynamic interviews, podcast episodes, and more by creating different personas to read the content. The source material could be text written by the student, quotes from literature or research, or text-based interviews.

- **Music and audio creation**: Tools in this category can generate original music from provided prompts. For instance, Beatoven.ai will produce an original track based on inputs like desired length, tempo, musical style (i.e., pop, ambient, techno) and mood (i.e., calm, motivational, triumphant, angry). Similar to Beatoven, Mubert will create royalty-free samples, loops, and tracks for use in podcasts, videos, etc. MusicLM, which has yet to be released to the public, will eventually be able to create music based on text prompts like descriptions of paintings or key words.

  A counterpart to MusicLM is the forthcoming AudioLM. These tools use OpenAI language models to generate realistic speech from text prompts and audio snippets.

  - **Potential Use**: Create songs or audio loops for a podcast episode, presentation, or other creative assignment.

  - **Potential Use**: Generate music from a text prompt relevant to the course (i.e., a quote from a poem or description of a picture). As a creative discussion opener, ask students to predict what the musical sample will sound like based on the prompt. Then have students compare and analyze their predictions against the results.

Conclusion

The potential uses for this technology in teaching and learning contexts will significantly expand once tools like MusicLM, AudioLM, and OpenAI's Whisper become available to the public.

Recommended Reading

Dietz, Lauri and Jennifer Keys. "Exploring the Pedagogical Possibilities of Generative AI." *Searle Center for Advancing Learning and Teaching Guide*. PDF

Dietz, Lauri. "*Heuristic for Designing Learning Experiences with Generative AI.*" *Searle Center for Advancing Learning and Teaching Guide*. PDF

"Introducing Whisper." (2022). https://openai.com/research/whisper.

Nerantzi, Chrissi. (2023). "Creating a Collection of Creative Ideas to Use AI in Education." February 2.

References

Agostinelli, et al. (2023). MusicLM: Generating Music from Text.

Beatoven.ai. https://www.beatoven.ai.

Borsos, et al. (2022). AudioLM: A Language Modeling Approach to Audio Generation

Mubert. https://mubert.com.

Otter.ai. https://otter.ai.

Play.ht. https://play.ht.

Transcribe. https://transcribe.com.

Search Tool Information

[Pedagogy, Course Design & Delivery]

How to Cite this Guide

Pierson, Katie. "Generative AI Audio Tools and Applications." *Searle Center for Advancing Learning and Teaching Guide.* Northwestern University. © 2023. Licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.

We are here for you. As you engage in ongoing reflection about ways to enhance teaching and deepen student learning, The Searle Center offers thought partnership and support. Check our calendar of events or reach out to teach@northwestern.edu to schedule a consultation.