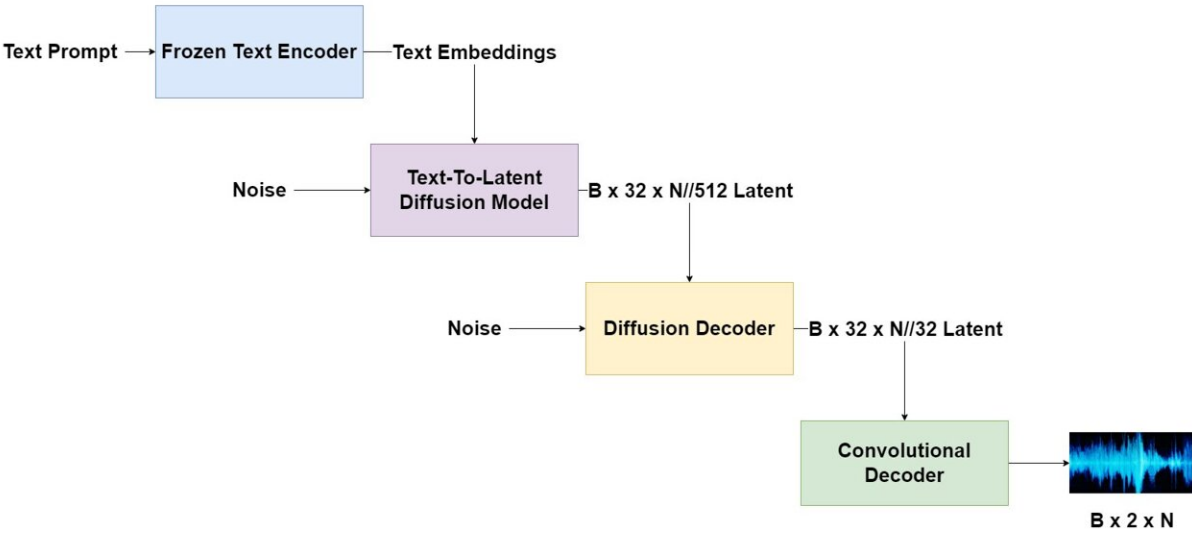


# Exploring text-to-audio models to make music from scratch

December 7 2022



The algorithm transforms a text prompt into audio. Credit: Zach Evans

Type a few words into a text-to-image model, and you'll end up with a weirdly accurate, completely unique picture. While this tool is fun to play with, it also opens up avenues of creative application and exploration and provides workflow-enhancing tools for visual artists and animators. For musicians, sound designers, and other audio professionals, a text-to-audio model would do the same.

As part of the 183rd Meeting of the Acoustical Society of America, Zach Evans, of Stability AI, presented progress toward this end in his talk, "Musical audio samples generated from joint text embeddings."

"Text-to-image models use [deep neural networks](#) to generate original, novel images based on learned semantic correlations with text captions," said Evans. "When trained on a large and varied data set of captioned images, they can be used to create almost any image that can be described, as well as modify images supplied by the user."

A text-to-audio model would be able to do the same, but with music as the end result. Among other applications, it could be used to create [sound effects](#) for video games or samples for music production.

But training these deep learning models is more difficult than their image counterparts.

"One of the main difficulties with training a text-to-audio model is finding a large enough data set of text-aligned audio to train on," said Evans. "Outside of speech data, research data sets available for text-aligned audio tend to be much smaller than those available for text-aligned images."

Evans and his team, including Belmont University's Scott Hawley, have shown early success in generating coherent and relevant music and sound from text. They employed data compression methods to generate the audio with reduced training time and improved output quality.

The researchers plan to expand to larger data sets and release their model as an open-source option for other researchers, developers, and audio professionals to use and improve.

**More information:** Conference: [acousticalsociety.org/asa-meetings/](https://acousticalsociety.org/asa-meetings/)

Provided by Acoustical Society of America

Citation: Exploring text-to-audio models to make music from scratch (2022, December 7)  
retrieved 9 June 2023 from

<https://techxplore.com/news/2022-12-exploring-text-to-audio-music.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.