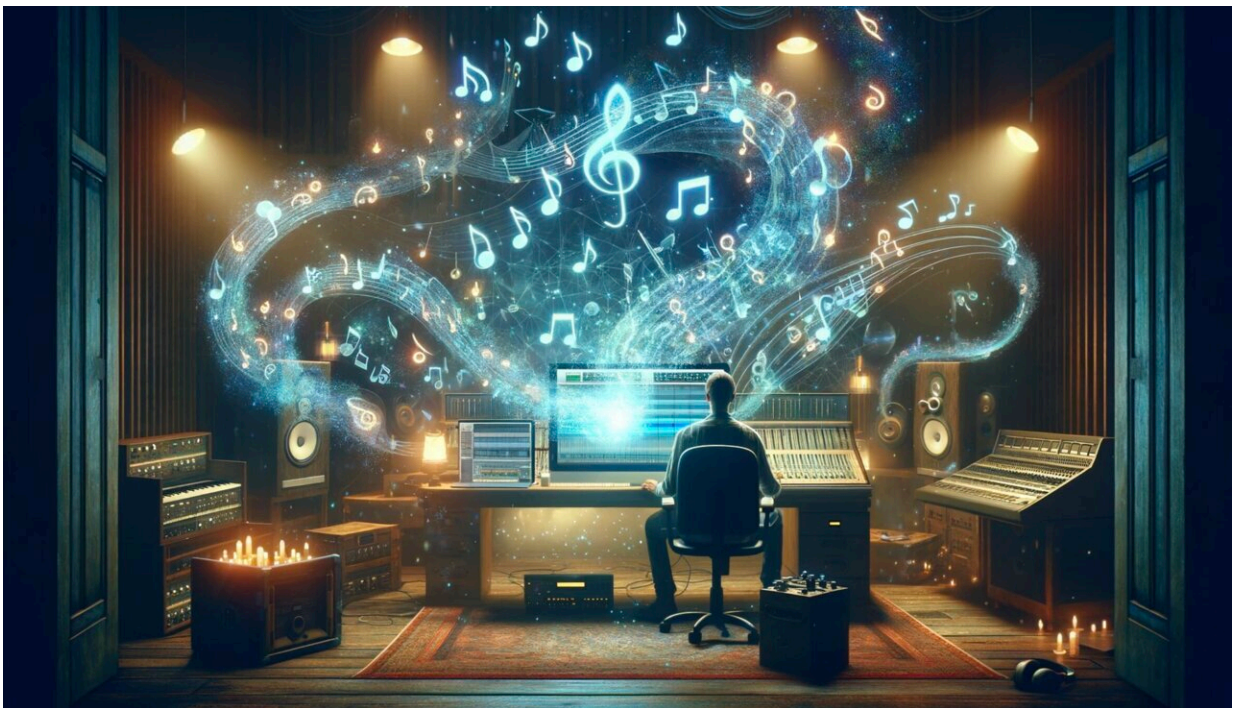


The AI bassist: Sony's vision for a new paradigm in music production

March 6 2024, by Ingrid Fadelli



Credit: Stefan Lattner (DALL-E)

Generative artificial intelligence (AI) tools are becoming increasingly advanced and are now used to produce various personalized content, including images, videos, logos, and audio recordings. Researchers at Sony Computer Science Laboratories (CSL) have recently been working on tools for producers and artists that can assist them in creating new

music.

In a recent paper [posted](#) on the *arXiv* preprint server, researcher Marco Pasini and his colleagues Stefan Lattner and Maarten Grachten at Sony CSL, introduced a new latent diffusion model that can create realistic and effective bass accompaniments for musical tracks. Diffusion models are [deep learning techniques](#) that can learn to generate images, audio or other samples that capture the overall structure underlying a dataset.

"Musical audio generation is currently a popular research topic, with many institutes, companies, and start-ups exploring various use cases," co-author Lattner told Tech Xplore. "At Sony CSL, we aim to assist music artists and producers in their workflow by providing AI-powered tools. However, we have noticed that the most common approach of AI tools generating complete musical pieces from scratch (often controlled only by text input) is not very interesting to artists."

When reviewing previously proposed music generation techniques, the researchers at Sony CSL found that they were not optimal for artists and producers. Specifically, they found that many tools did not allow users to create music aligned with their unique preferences and style.



Credit: Marco Pasini (DALL-E)

"Artists require tools that can adjust to their unique style and can be utilized at any point in their music production process," Lattner said. "Therefore, a generative music tool should be able to analyze and take into account any intermediate creation of the [artist](#) when proposing new

sounds."

In their recent paper, the researchers introduced a new model that can automatically generate bass accompaniments that match the style and tonality of an input music track, irrespective of the elements it contains (i.e., vocals, guitar, drums, etc.). Their proposed tool was designed to generate incisive basslines that complement songs well, thus assisting producers and artists in their [creative process](#).

"Our system can process any type of musical mix that contains one or more sources, such as vocals, guitar, etc.," Lattner explained. "It consists of an audio autoencoder that efficiently encodes the mix into a compressed representation, capturing the essence of the music. This compressed encoding is then used as input to a specially designed architecture based on a state-of-the-art generative technology called 'latent diffusion.' This method generates data in a compressed space, which improves performance and quality."

Lattner and his colleagues trained their latent diffusion model on a dataset of bass guitar encodings containing various music track examples. Over time, the model learned to create a bassline that "plays along" with an input music track.



Credit: Marco Pasini (DALL-E)

"Our system has a unique advantage: it can generate coherent basslines of any length, as opposed to fixed durations," Lattner said. "We also proposed a technique called 'style grounding' that allows users to control the timbre and playing style of the generated bass by providing a

reference audio file."

The researchers evaluated their latent diffusion model in a series of tests and found that it could generate appropriate bass accompaniments to arbitrary song mixes. Notably, the creative bass lines it produced closely matched the tonality and rhythm of an input music mix.

"We presented what we believe is the first conditional latent diffusion model designed specifically for audio-based accompaniment generation tasks," Lattner said. "By training it on paired data of mixes and matching basslines, the model learns the concept of musical coherence."

In the future, the new bassline generation tool created by Pasini and his colleagues could be used by musicians, producers, and composers worldwide, helping them write or improve instrumental parts of their tracks. The researchers now plan to create similar models that produce other instrumental elements, such as drums, piano, guitar, string, and sound effect accompaniments.

"With further development, we envision creative tools where users can customize the bass or other accompaniments that they can seamlessly integrate with their compositions," Lattner added.

"Additional directions for future research involve providing additional, intuitive control mechanisms—in addition to audio references, users could guide the style through free-form text prompts or descriptive stylistic tags. More broadly, we plan to collaborate directly with artists and composers to refine further and validate these AI accompaniment tools to best enhance their creative needs."

More information: Marco Pasini et al, Bass Accompaniment Generation via Latent Diffusion, *arXiv* (2024). [DOI: 10.48550/arxiv.2402.01412](https://doi.org/10.48550/arxiv.2402.01412)

© 2024 Science X Network

Citation: The AI bassist: Sony's vision for a new paradigm in music production (2024, March 6)
retrieved 27 April 2024 from

<https://techxplore.com/news/2024-03-ai-bassist-sony-vision-paradigm.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.