Centre for Mathematical Sciences, Mathematical Statistics | Faculty of Engineering, Lund University
Presented 2026-03-17 | ISSN: 1404-6342 | LUTFMS-3551-2026

MASTER'S THESIS Latent Space Interpolation for Music Transitions
STUDENT Olle Rydén
SUPERVISOR Ted Kronvall
EXAMINER Maria Sandsten

Can neural networks make better song transitions?

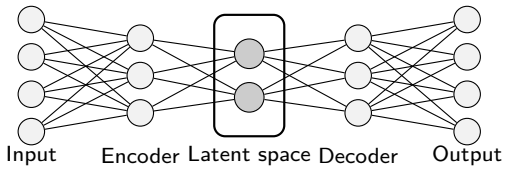
POPULAR SCIENCE SUMMARY **Olle Rydén**

Most music transitions are still made using tools such as beatmatching and crossfades. This work investigates whether a neurally trained audio codec can be used to create transitions directly in a compressed latent space created by neural networks, and whether the result can compete with traditional signal processing.

When a DJ or a music program switches from one song to another, the goal is usually for the change to sound natural. Today, this is most often done using standard signal-processing methods, where the simplest automatic variant is to let one song fade out while the other fades in. This usually works well, but the method only blends two finished audio signals on top of each other.

In my thesis, I test a different approach. Instead of making the transition directly in the audio waveform, both songs are first encoded into a *latent space* using the EnCodec model. There, music is represented as a time series of compact vectors that still carry important information about rhythm, energy, and key. The idea is then to interpolate between two songs in this representation and afterwards decode the result back into audio.

The method begins by selecting suitable segments from two songs. These are tempo-adjusted and aligned so that beats and phrases match. The segments are then encoded into a compact representation, in the so-called latent space. In this latent space, several interpolation methods are tested, including linear interpolation and spherical interpolation, known as SLERP. The result is then decoded back into audio and compared with a standard equal-power crossfade, that is, the type of transition often used in practice.



The evaluation consists of both listening tests and objective measures. In the listening tests, 14 people compared 10 latent space interpolation transitions with ordinary crossfades. The results showed no clear winner: the latent method was preferred about as often as the baseline, and many answers were "no preference".

At the same time, the analyses show that latent interpolation does not behave exactly like a standard mix. The transitions often become somewhat smoother and less detailed, especially in transients and high-frequency content. The diagnostics also suggest that the latent representation captures musical structure, such as rhythmic periodicity and sensitivity to pitch.

The conclusion is that latent interpolation does not replace the classical crossfade, but that it is a promising alternative. In the long term, the technique could be used for smarter music transitions in automated DJ systems, for playlists, and in music tools.