MixingBuddy: A Multimodal LLM for Mix Critique and Advice

Pratham Vadhulas, Alexander Lerch

Abstract—this is the abstract. here is some more text. here is some more text.

I. INTRODUCTION

Automatic Mixing, a key subfield of Music Informatics Research (MIR), aims to automate the complex and subjective task of music mixing. This area of study is pivotal to the modern music production and audio engineering market. To date, research in this field has made significant progress, largely by leveraging deep learning. Sophisticated models, such as U-Nets or generative frameworks, have been developed to make mixing systems accurate (predicting parameters that match professional mixes), controllable (allowing for high-level parameters to be set), and diverse (accomadating different genres and styles).

However, a critical limitation of these approaches is their "black box" nature. They can perform the mix, but they cannot explain their reasoning. The recent promise of large language models (LLMs) and multi-modal "agentic" systems introduces a new, necessary paradigm: explainability. We can now envision a tool that reasons about and discusses a mix.

This potential for co-creative, linguistic feedback brings us to our core research question: To what extent can an audio language model, when given a flawed mix, provide correct and useful advice?

Pratham Vadhulas is with the Georgia Institute of Technology, Atlanta, GA, USA (email: pvadhulas3@gatech.edu).

Alexander Lerch is with the Georgia Institute of Technology, Atlanta, GA, USA (email: alexander.lerch@gatech.edu).