



Ethical and Responsible AI Music Making Workshop • 17 July 2024

# **User-Centric Intelligent Context-Aware Assistive Multitrack Music Mixing**

**Soumya Sai Vanka**

Centre for Digital Music, Queen Mary University of London





# Mixing

**Audio mixing is the process of blending multitrack recordings**

- Technical considerations together with creative, artistic or aesthetic decisions

**Achieved with audio effects**

- Gain
- Panning
- Equalization (EQ)
- Dynamic range compression (DRC)
- Artificial reverberation

# More people are creating **audio** content



Music



Podcasts



Short-form content



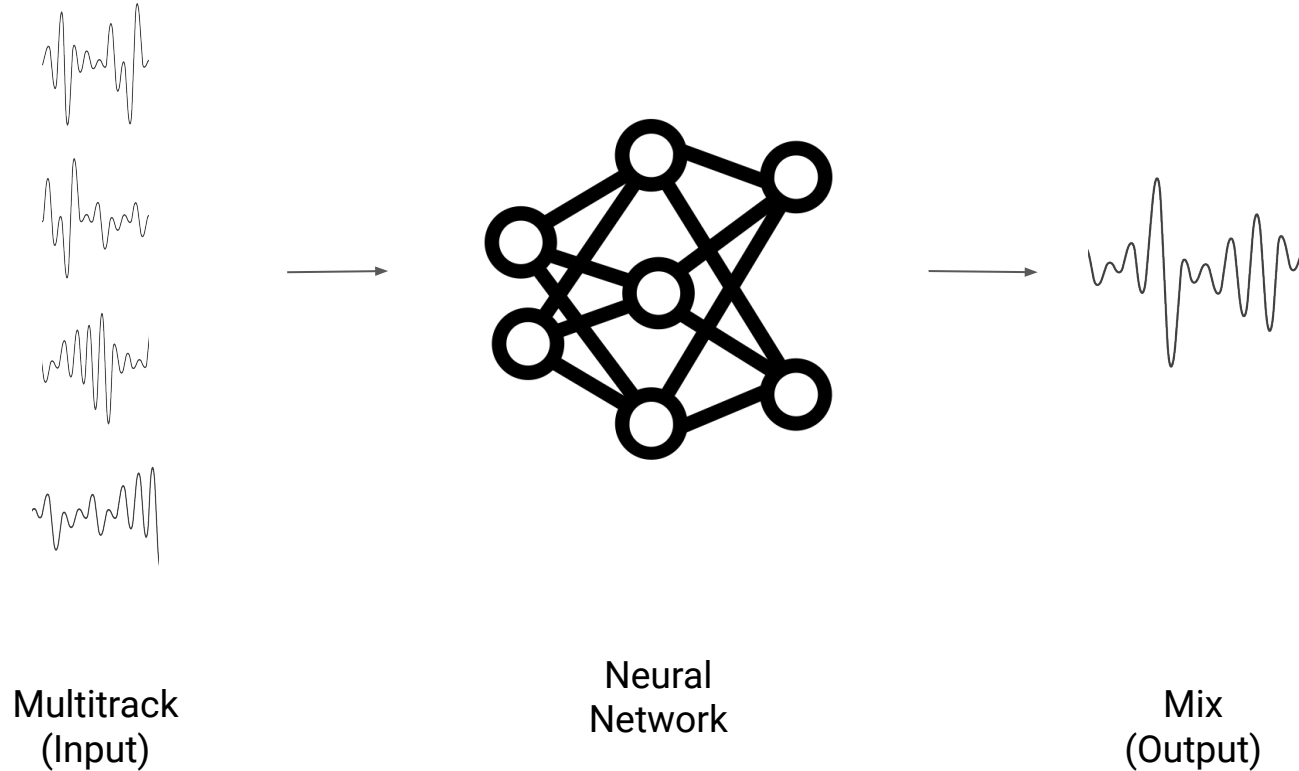
Sound for Video

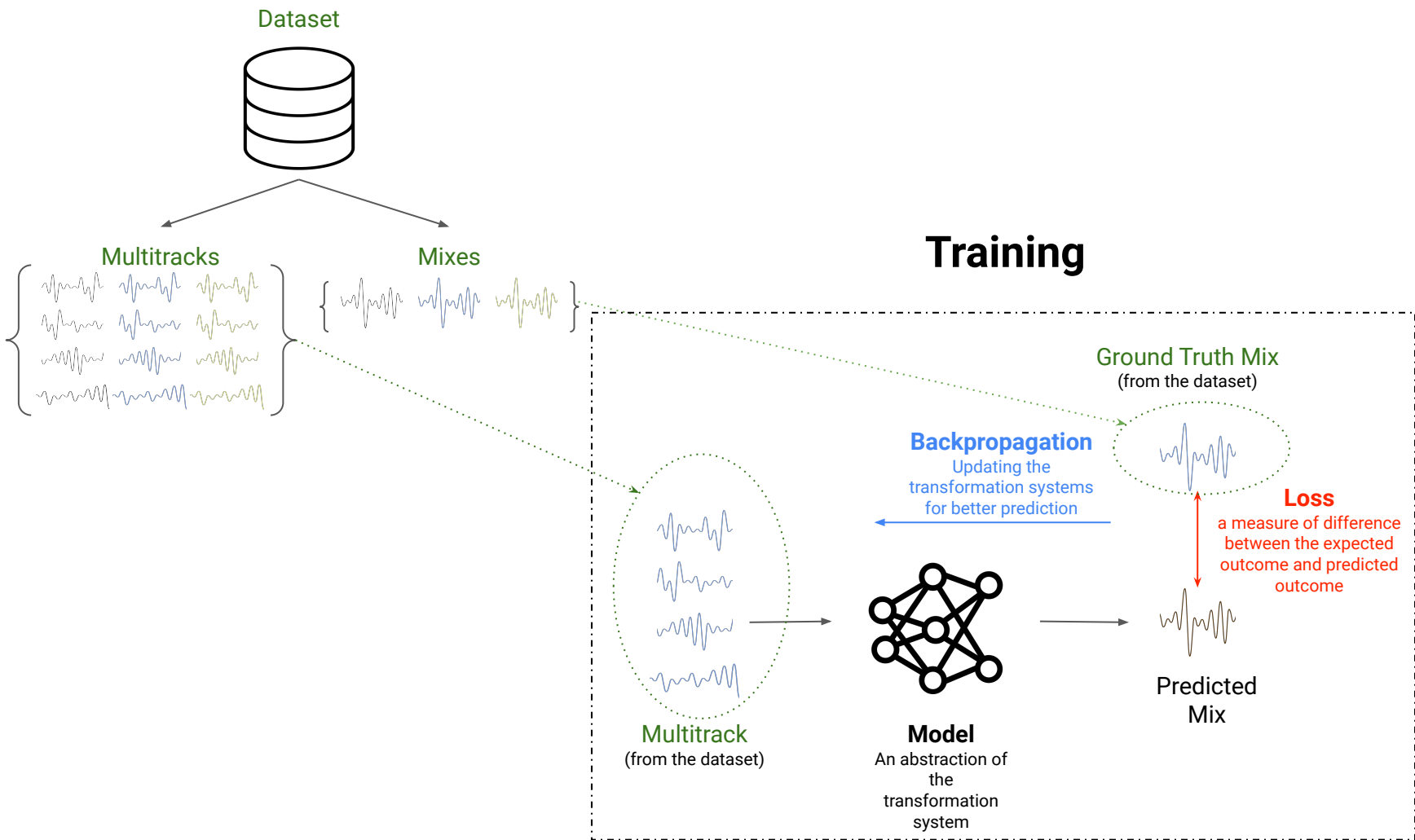
# Demand for high quality audio



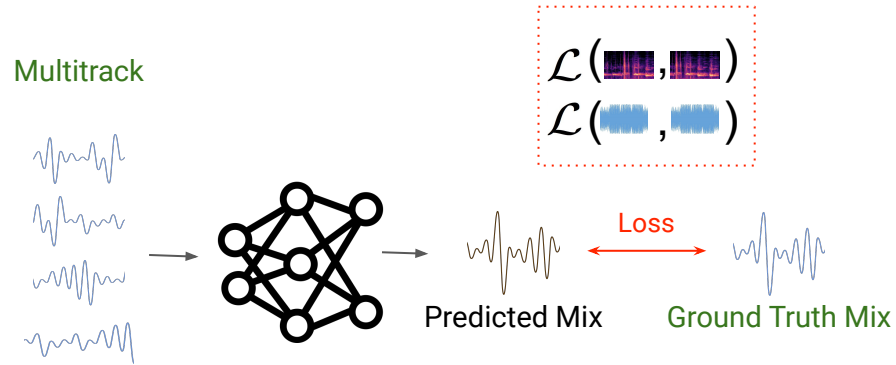
Producing high quality audio requires expertise

# General Idea





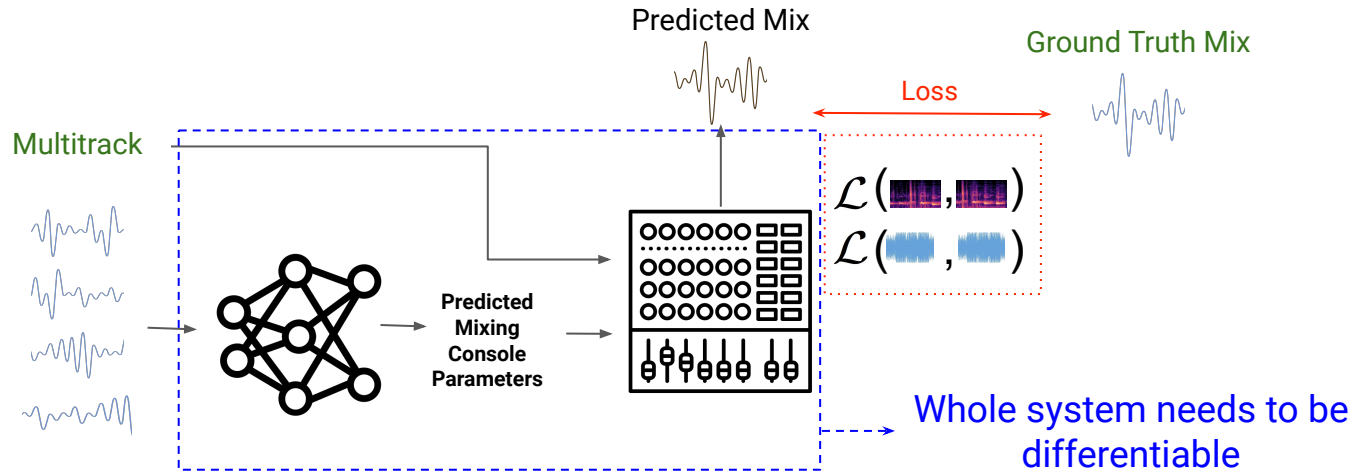
# Model Types



## Direct Transformation

Black box system that lacks interpretability and controllability (context not incorporated)

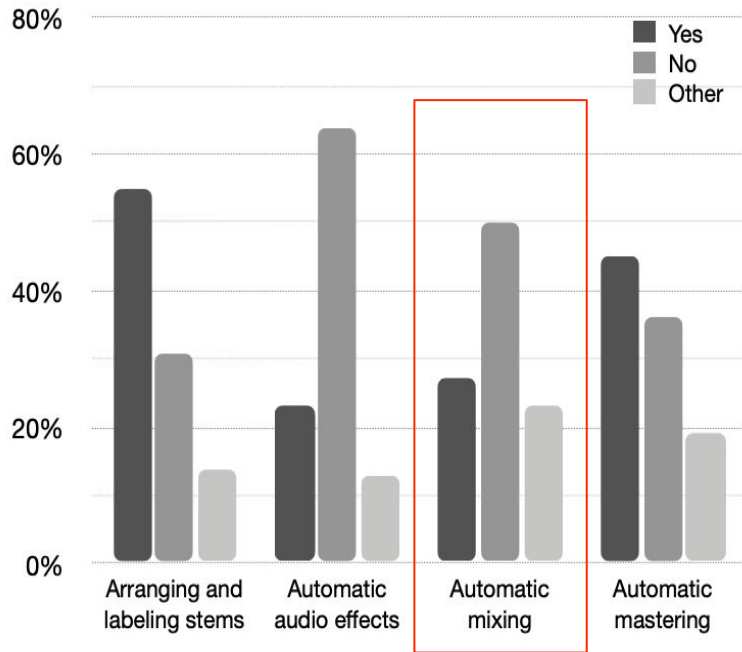
# Model Types



## Parameter Estimation (Audio Loss)

Black box system that allows interpretability and controllability (context not incorporated)





Why such a huge percentage is saying no?



## Audio Engineering Society Convention Paper

Presented at the 154th Convention  
2023 May 13–15, Espoo, Helsinki, Finland

*This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

### Adoption of AI Technology in the Music Mixing Workflow: An Investigation

Soumya Sai Vanka<sup>1</sup>, Maryam Safi<sup>2</sup>, Jean-Baptiste Rolland<sup>2</sup>, and György Fazekas<sup>1</sup>

<sup>1</sup>Queen Mary University of London, London, UK

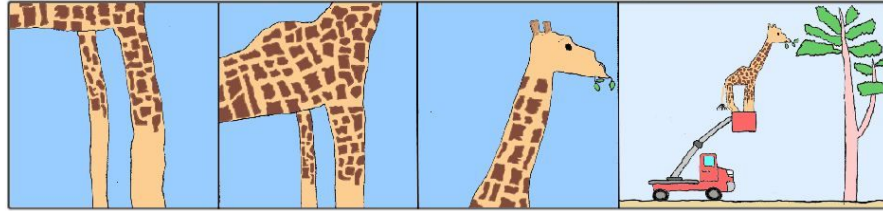
<sup>2</sup>Steinberg Media Technologies GmbH, Hamburg, Germany

Correspondence should be addressed to Soumya Sai Vanka ([s.s.vanka@qmul.ac.uk](mailto:s.s.vanka@qmul.ac.uk))



## OUT OF CONTEXT

PAUL MCGEOWN (pmcgeown@imprint.uwaterloo.ca)

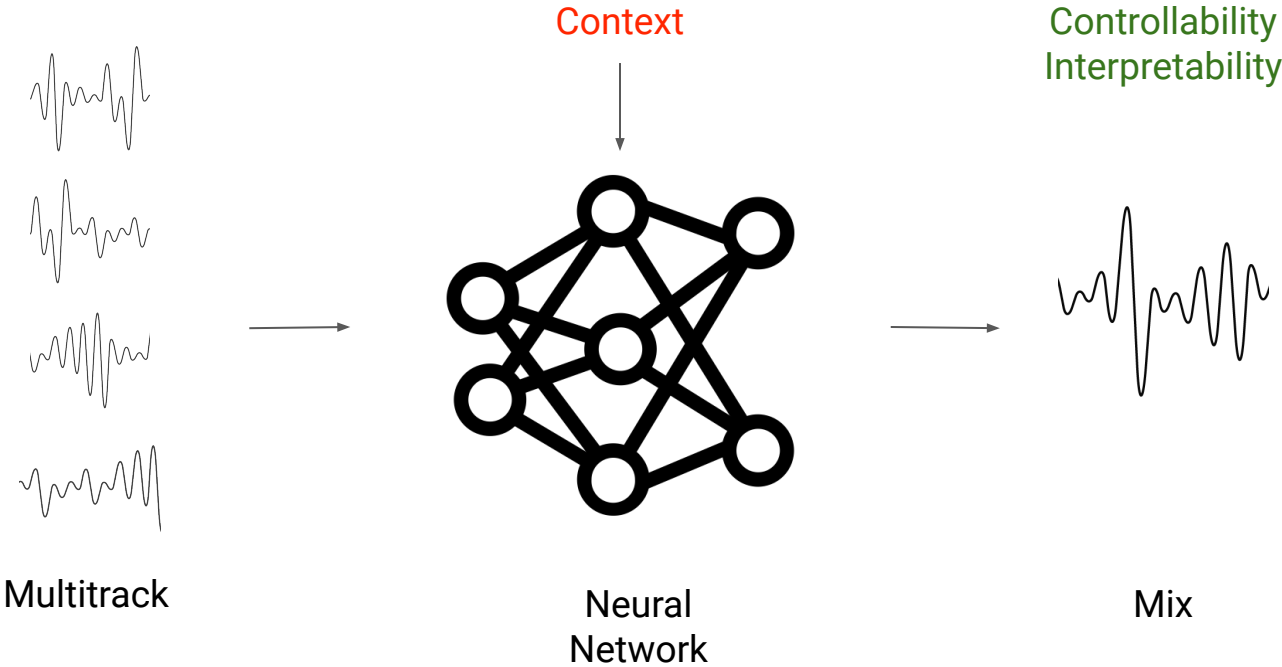


Results are generic and do not understand the context

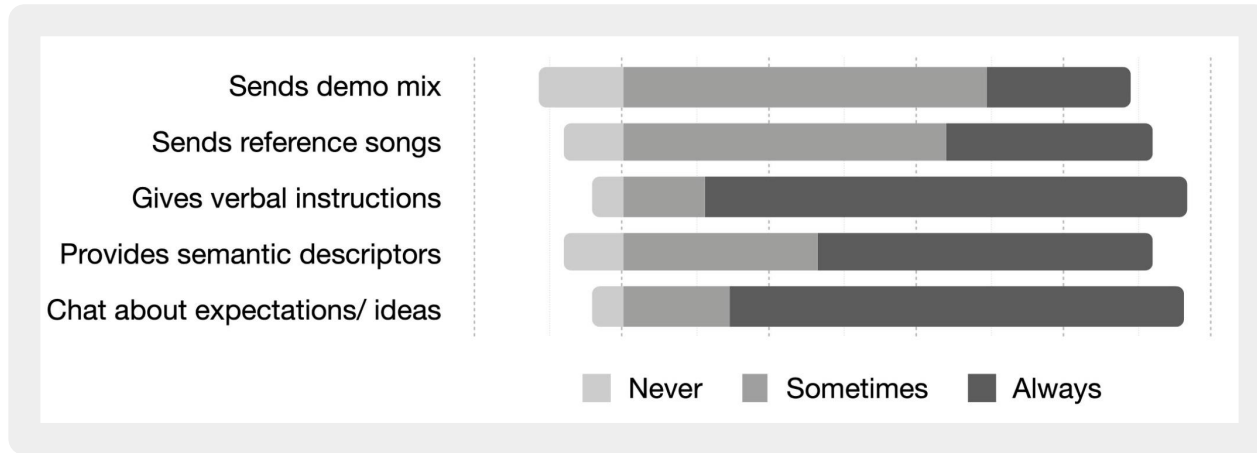


Black box systems: limiting control and interpretability.

# What we want?



## Various media used by artists to communicate their expectations of the mix



How is context communicated?

PAPERS

open access Freely available online

S. S. Vanka, M. Safi, J.-B. Rolland, and G. Fazekas,  
 "The Role of Communication and Reference Songs  
 in the Mixing Process: Insights From Professional Mix Engineers,"  
*J. Audio Eng. Soc.*, vol. 72, no. 1/2, pp. 5–15 (2024 Jan/Feb),  
<https://doi.org/10.17743/jaes.2022.0123>.

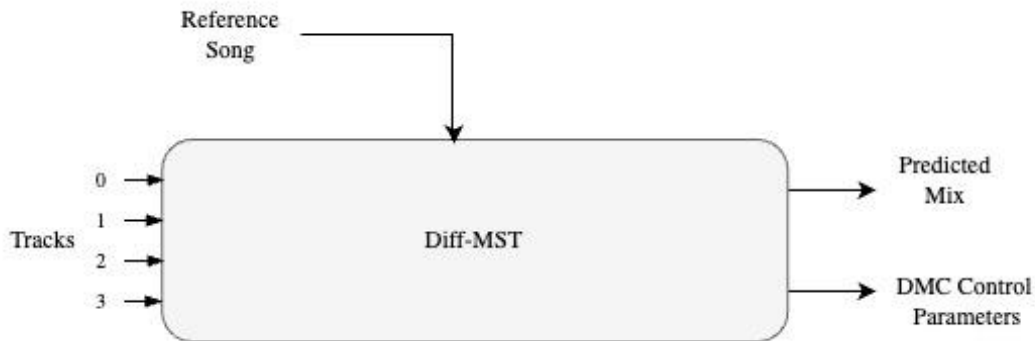
### The Role of Communication and Reference Songs in the Mixing Process: Insights From Professional Mix Engineers

SOUMYA SAI VANKA<sup>1,\*</sup>, AES Student Member, MARYAM SAFI<sup>2</sup>, AES Member,  
 (s.s.vanka@gmail.co.uk) (m.safi@steinberg.de)

JEAN-BAPTISTE ROLLAND<sup>2</sup>, AND GYÖRGY FAZEKAS<sup>1</sup>  
 (jb.rolland@steinberg.de) (georgc.fazekas@gmail.co.uk)

<sup>1</sup>Centre for Digital Music, Queen Mary University of London (QMUL), London, UK  
<sup>2</sup>Steinberg Media Technologies GmbH, Hamburg, Germany

# Diff-MST: Differentiable Mixing Style Transfer



(DMC: Differentiable Mixing Console)

A controllable tool that uses reference song as context to get a starting point in a given direction for creating mix

**DIFF-MST: DIFFERENTIABLE MIXING STYLE TRANSFER**

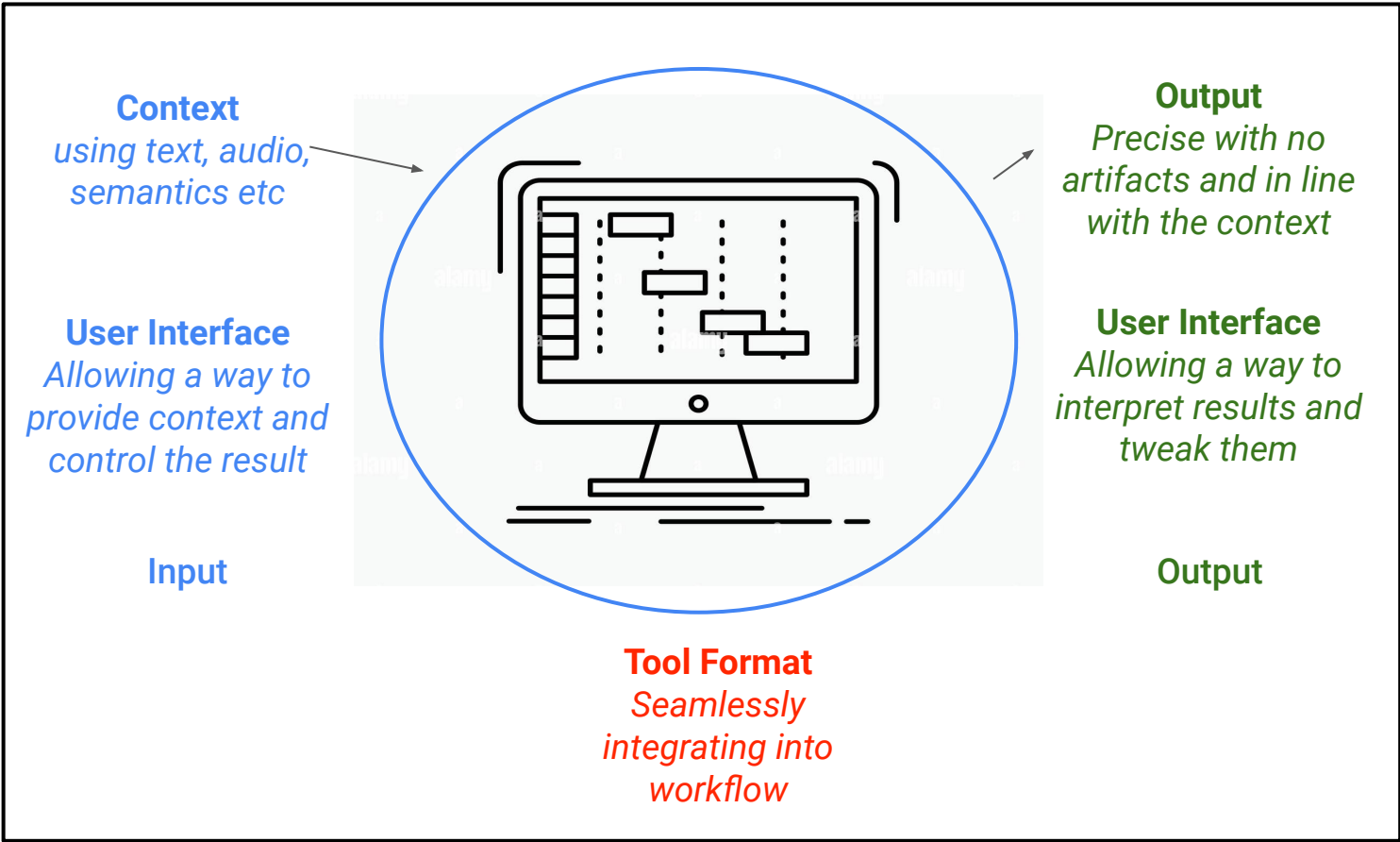
Soumya Sai Vanka<sup>1†</sup> Christian Steinmetz<sup>1†</sup> Jean-Baptiste Rolland<sup>2</sup>  
Joshua Reiss<sup>1</sup> György Fazekas<sup>1</sup>

<sup>1</sup> Centre for Digital Music, Queen Mary University of London, UK  
<sup>2</sup> Steinberg Media Technologies GmbH, Germany  
e.s.vanka@qmul.ac.uk, c.j.steinmetz@qmul.ac.uk

**ABSTRACT**

Mixing style transfer automates the generation of a multi-track mix for a given set of tracks by inferring production attributes from a reference song. However, existing systems for mixing style transfer are limited in that they often operate only on a fixed number of tracks, introduce artifacts, and produce mixes in an end-to-end fashion, without grounding in traditional audio effects, prohibiting interpretability and controllability. To overcome these challenges, we introduce **DIFF-MST**, a framework comprising a differentiable mixing console, a transformer controller, and an audio production style loss function. By inputting raw tracks and a reference song, our model estimates control parameters for audio effects within a differentiable mixing console, producing high-quality mixes and enabling post-hoc adjustments. Moreover, our architecture supports an arbitrary number of input tracks without source la-

11 Jul 2024



**Ideal design for an assistive mixing system**

# What bias are we talking about here?

- There is no penultimate mix for a given set of tracks
  - But training methods may impose the idea - especially with supervised methods
- Limited open source data - very western music dominant - especially pop and rock
  - How does this expand to folk/regional/other music that is not represented?
  - Underrepresented instruments, genres etc
- **How can we use the limited open source multitrack data available to capture a diverse world**
  - Using powerful methods to learn transformation self-supervision or unsupervised learning?
  - Augmentations or smart tricks to learn more from what is available?
  - Explore techniques to learn from existing unorganised data in the world

# Responsible System Design: Thoughts for Discussion

- Giving back to community is key
  - Model's using data; building/learning from the work of giants
  - Ask what community wants.
    - Empirical studies to support and build expert knowledge base -> later used to design systems
  - How do they benefit in the process? Adding value to user's creative workflow
- Allow interaction and control
  - Give power back to the Human
    - This means asking what the user wants
  - Art is a human right
    - Ask users what would help them express themselves
    - Assistive tool most often preferred over automatic tools.



**Thank You!**