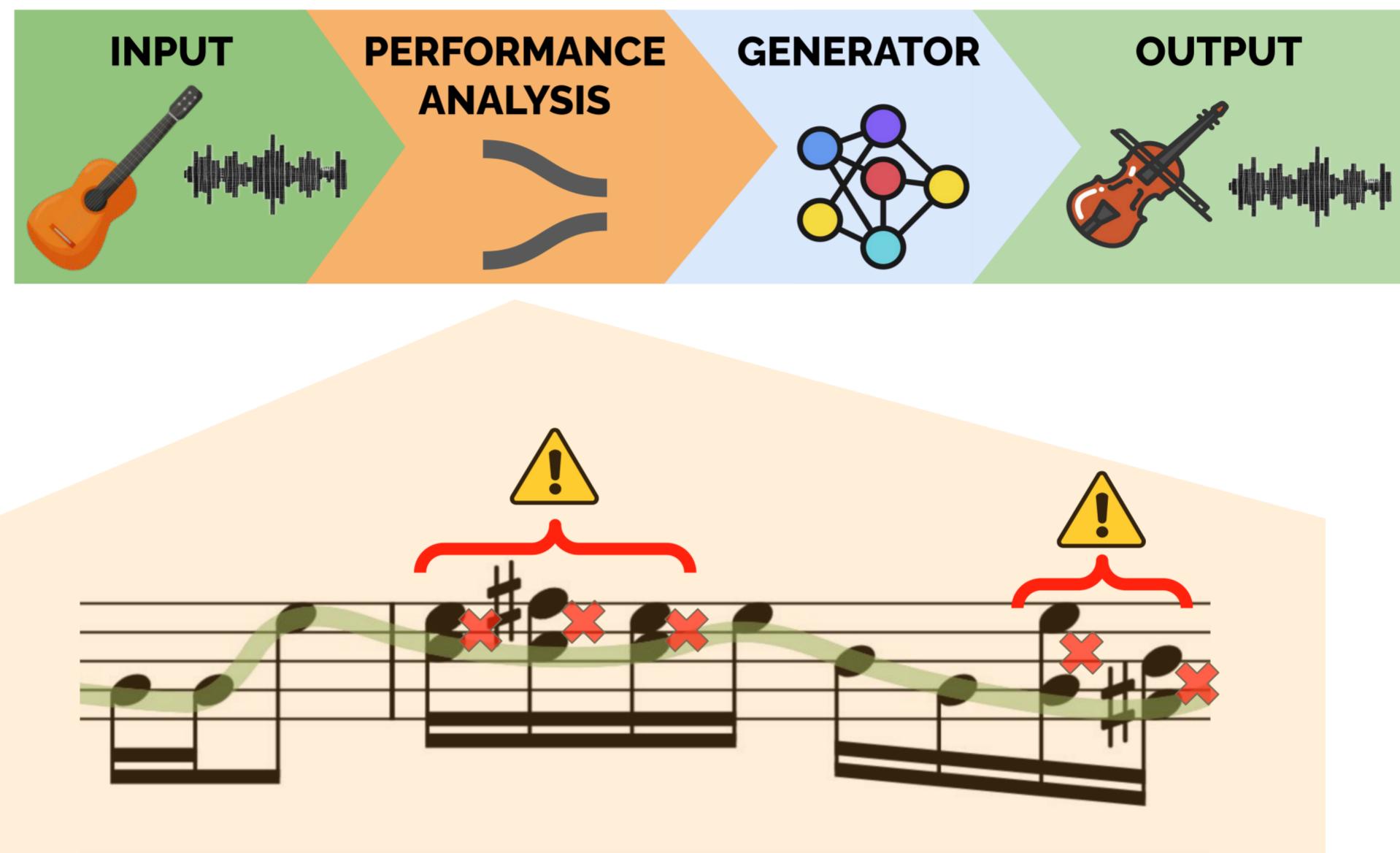


Timbre Transfer for Musical Instrument Interaction

Franco Caspe

2021



Finding

Current real-time Timbre Transfer algorithms impose an interpretative bottleneck to music performance through implicit or explicit rules. The events that don't comply with this may generate sonic outcomes unexpected to the player, breaking their expectations and forcing them to adapt their performance to the system.

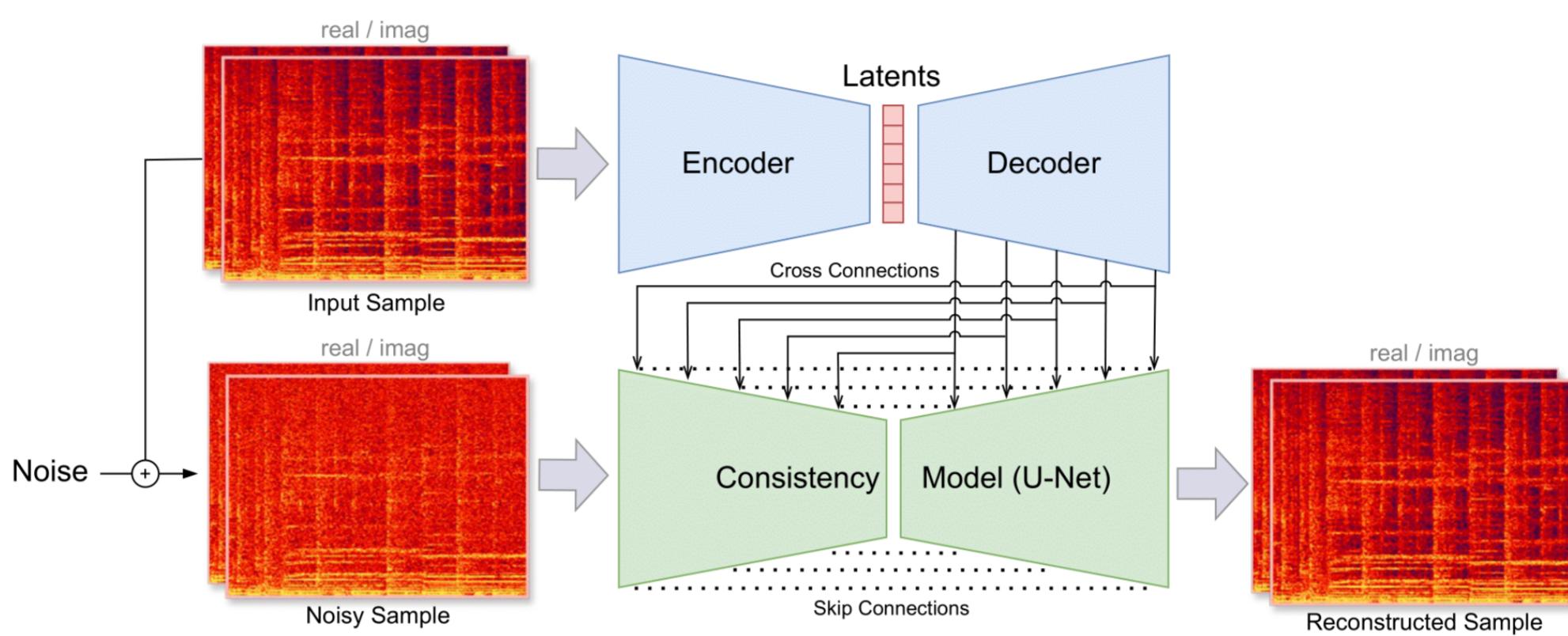
Question

I am interested in pitch & harmony preserving Timbre Transfer. However, imposing a detection bottleneck may lead to failure modes we call ambiguity shifts. How can we relax these assumptions to allow for ambiguous input/output? And what may be an acceptable outcome from that input?

Music2Latent: Consistency Autoencoders for Latent Audio Compression

Marco Pasini

2023



Finding

Music2Latent is a consistency autoencoder that encodes samples into a continuous latent space with a high compression ratio. It is trained fully end-to-end using a single loss function. Its representations perform competitively on downstream tasks, and can be used to train generative models. It provides more accurate reconstructions than comparable models.

Question

We improve on the original consistency models training recipe by introducing continuous noise levels with a novel exponential schedule. We also adopt frequency wise self-attention and an adaptive frequency scaling mechanism to deal with the unbalanced magnitude distribution over different frequency bins. We validate the effectiveness via ablation studies.

Recognition of Jazz Song Names from Audio

Carey Bunks
2022

The image shows a musical score for 'West Coast Blues' by Wes Montgomery. The score consists of five staves of music for a single melodic line. Above the music, the title 'West Coast Blues' and the composer 'Wes Montgomery' are written. Below the music, there is a legend titled 'Harmonic tags can be likened to part-of-speech tagging in NLP'. The legend includes seven categories with corresponding colored squares: Chord tone (red), Semi-tone approach note (blue), Enclosure (grey), Anticipation/Prolongation (purple), Ninth (dark blue), Eleventh (olive green), and Sixth / Thirteenth (yellow). The musical score itself has various chords labeled above the notes, such as Bb7, Ab7, Bb7, Bm7, E7, Eb7, Ebm7, Ab7, Dm7, G7, C#m7, F#7, Cm7, F7, Bb6, Db7, Gb6, and F7.

Finding

Jazz chords can be represented through a process of vector embedding, and chord progressions can be constructed from them as piecewise linear paths in the latent space. I demonstrated this was a useful mechanism for comparing chord progressions and for jazz song contrafact detection.

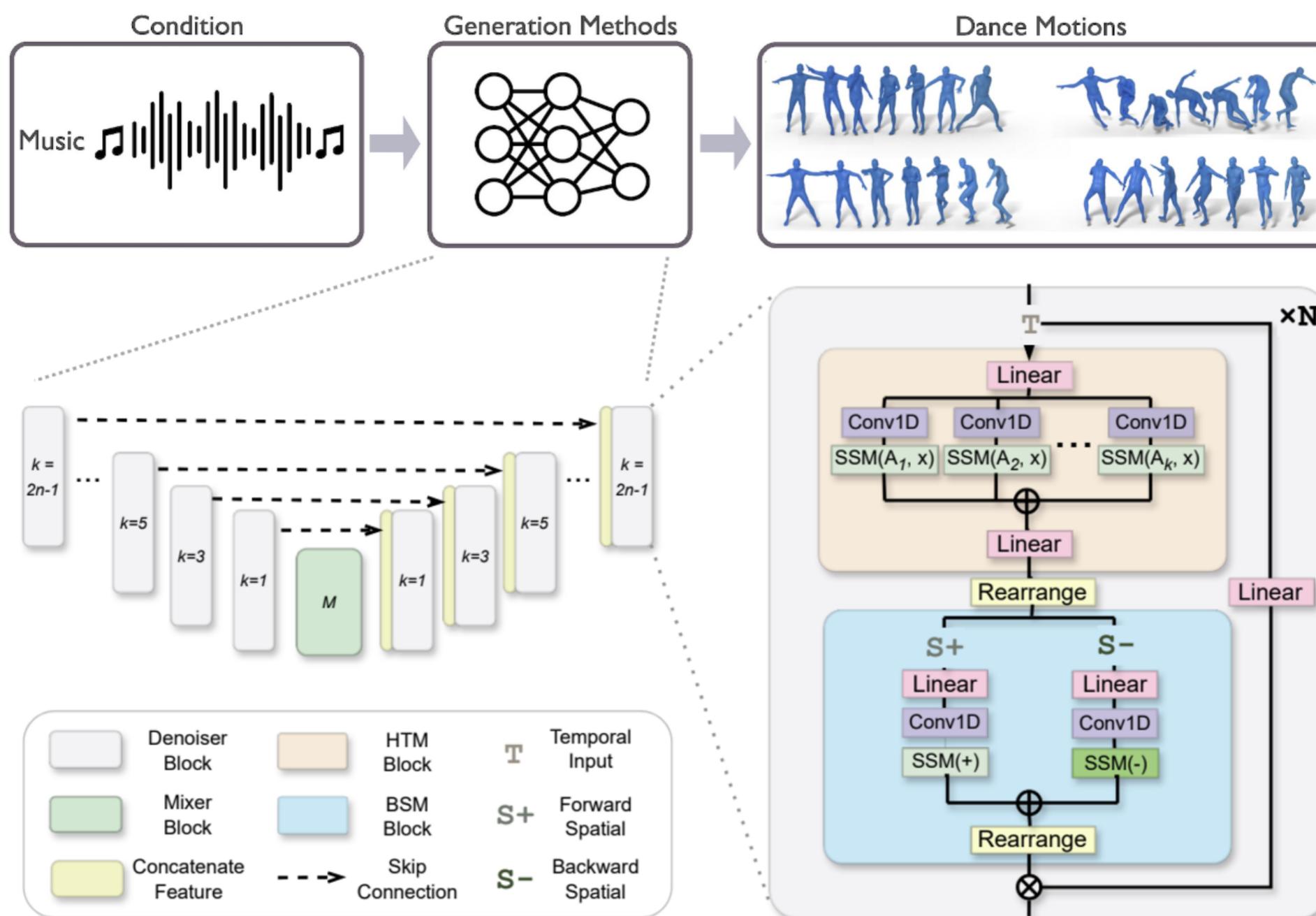
Question

Jazz is a genre that makes free use of improvisation in both melody and harmony, and this makes identifying jazz tunes from audio a difficult problem. I conjecture that symbolic harmony (chord progressions) can be used as representative fingerprints of songs that can be compared to audio data and successfully used for identifying song names.

Music-Driven 3D Dance Generation

Qing Wang

2023



Finding

For music representation, beat information plays a crucial role in the task of generating dance movements from music, but also restrain the generative model from learning more flexible and diverse dance movement. For dance motions, the naturalness hasn't been fully addressed by current works, foot sliding still exists.

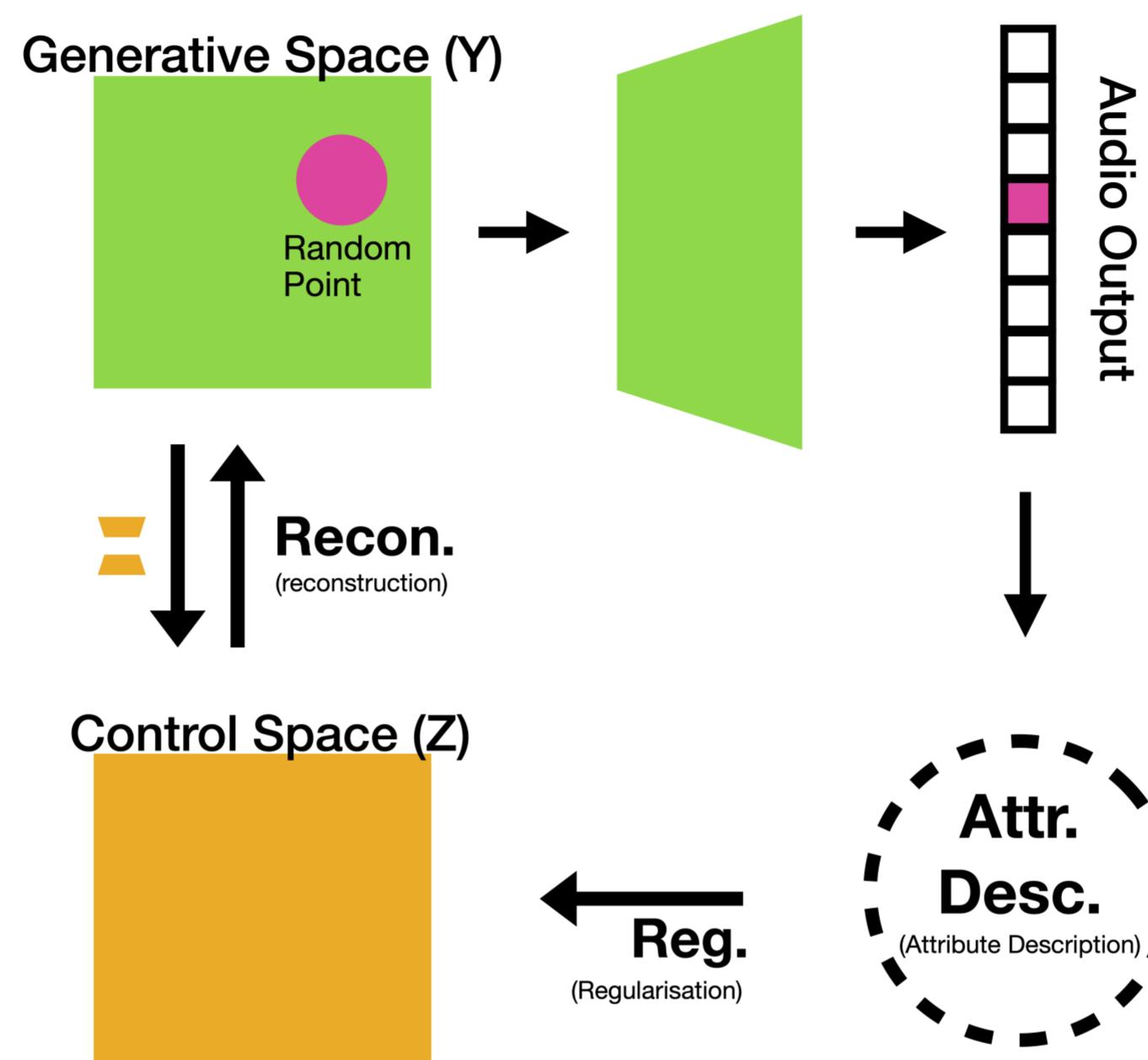
Question

What features of music most significantly influence dance movements? How can we map these musical features to corresponding dance movements in a way that is both artistically expressive and technically accurate?

Latent Spaces for Human-Computer Audio Manipulation

Ashley Noel-Hirst

2022



Finding

We can learn consistent mappings for a range of unseen, pre-trained generative VAEs by creating additional control VAEs regularised on attributes of their output.

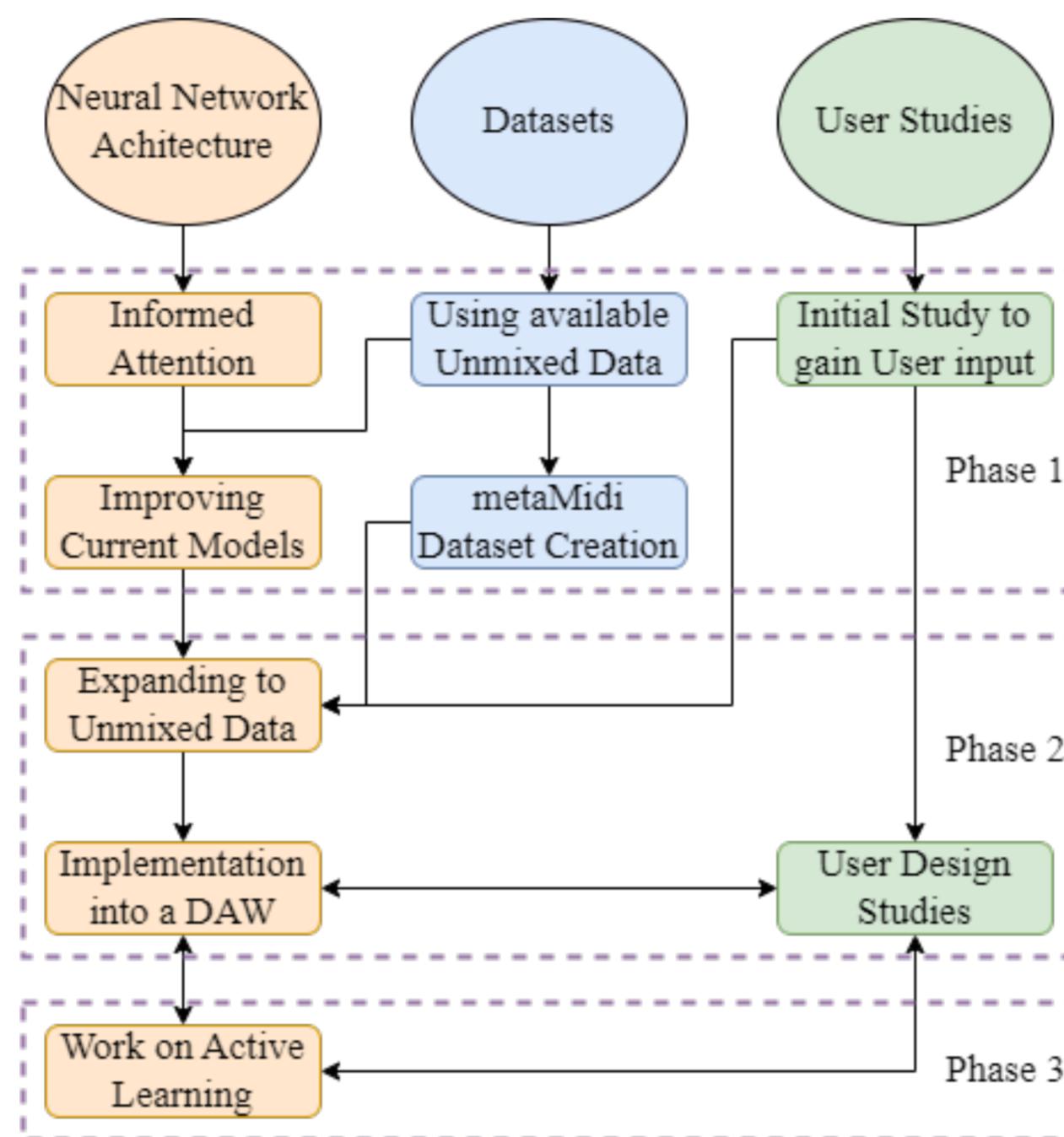
Question

Beyond visual and textual-semantic, what types of latent representations might be helpful for interacting with musical audio?

Intelligent Audio Editing for Multi-track Data

James Bolt

2022



Finding

Can results from previous models be used to inform and improve the results of a new model?

Question

How can multi-track data be utilized to improve MIR tasks.

Self-supervision in Audio Fingerprinting

Aditya Bhattacharjee

2022

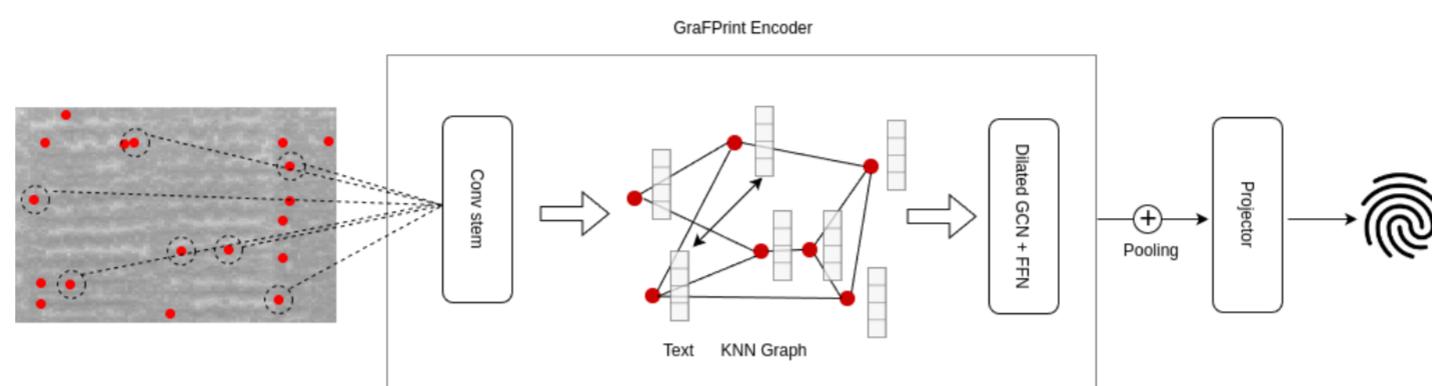


Fig. 1: Schematic design of the GraFPrint Encoder

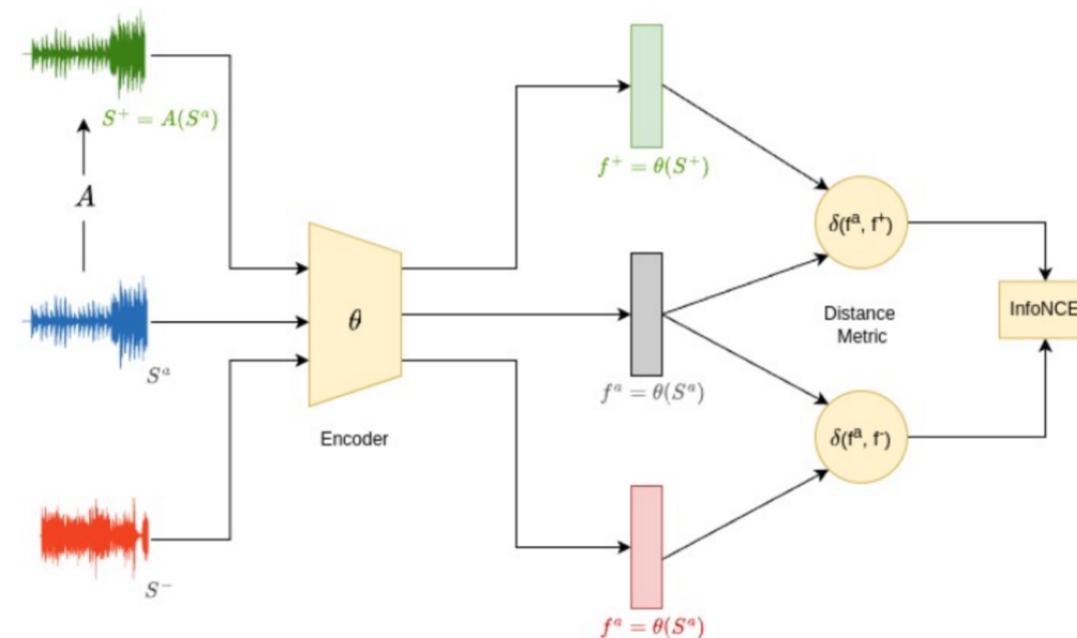


Fig. 2: Data pathways for the self-supervised contrastive training

Finding

Data and network parameter requirements for self-supervised audio fingerprinting are lower than other downstream classification tasks. Lightweight network architectures with informed design choices have been shown to achieve comparable to SOTA performance.

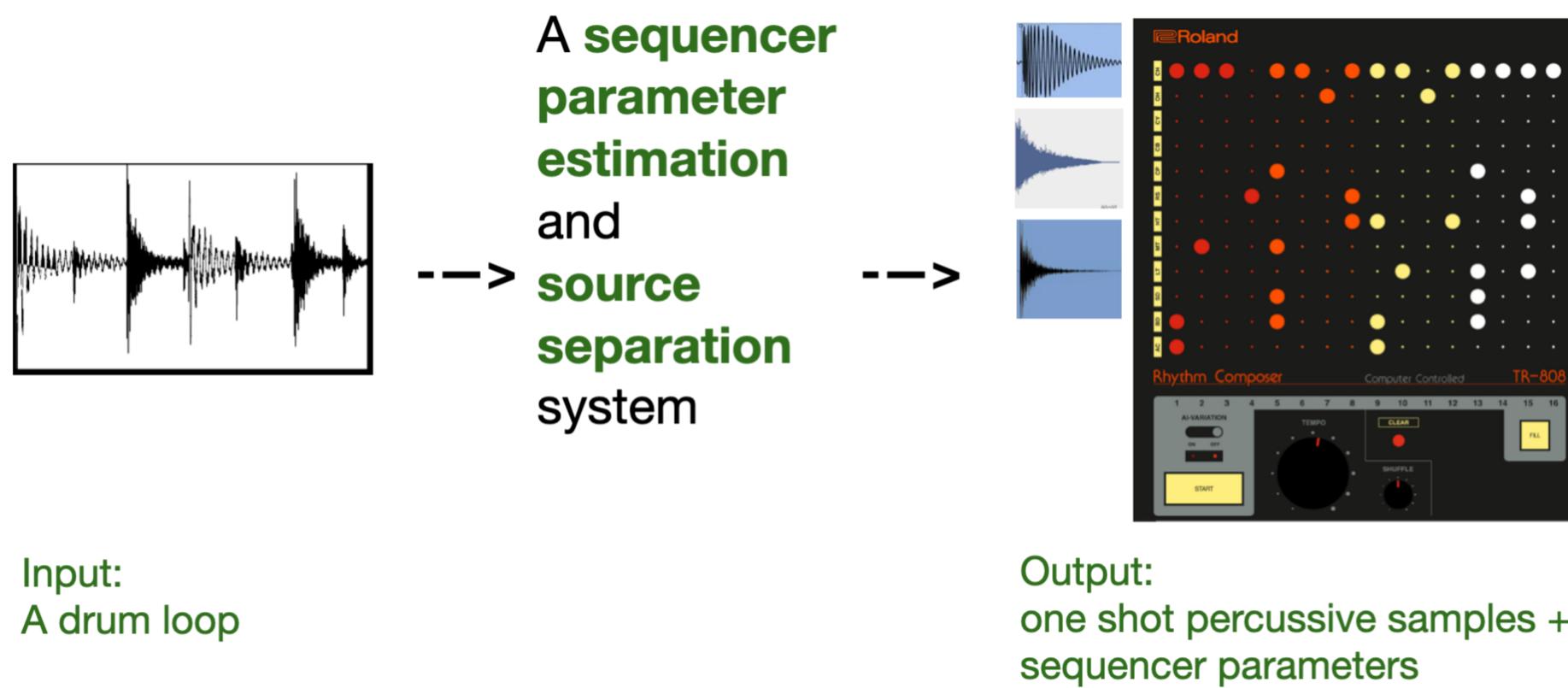
Question

Can self-supervised audio fingerprinting be extended to the task of identification of cases of sampling in musical recordings? Can such a system be objectively evaluated?

AI-based Systems for Facilitating Loop-based Music Composition

Xiaowan Yi

2021



Finding

In Study 1, we worked on compatible loop generation utilising RAVE models Drum2Bass: generating a bass loop in the audio waveform that can 'match' an input drum loop. We trained an extra neural net to learn the mapping from a drum loop's RAVE latent embedding to a bass loop's RAVE latent embedding.

Question

In Study 2, we are working on sequencer parameter estimation for drum loops: mapping a drum loop from the audio waveform domain to sequencer parameters (e.g. global tempo, step activation vectors, etc.). I'm looking for help on how to make the drum sequencer rendering process differentiable to enable a training framework similar to that of DDSP.

Transcribing the Jazz Ensemble

Xavier Riley

2020



A musical score extract in 4/4 time, key of F major. The tempo is 208 BPM. The score consists of two staves. The top staff shows a ground truth transcription with various chords labeled above the staff: Am, D7, Gm, Gm, Gm, Gm, C7, F, Gm, C7, F. The bottom staff shows a transcription by the author (ours), which follows the same harmonic progression but with different note patterns and rhythms.

Figure 1. An extract of “Marmaduke” - ground truth above, transcription (ours) below



A musical score extract in 4/4 time, key of F major. The tempo is 208 BPM. The score consists of two staves. The top staff shows a ground truth transcription with chords labeled above the staff: Fm, C7, Fm, Fm. The bottom staff shows a transcription by the author (ours), which follows the same harmonic progression but with different note patterns and rhythms.

Figure 2. An extract of “The Bird” - ground truth above, transcription (ours) below

Finding

The lack of ground truth data for Automatic Music Transcription can be addressed by aligning digital scores to recordings

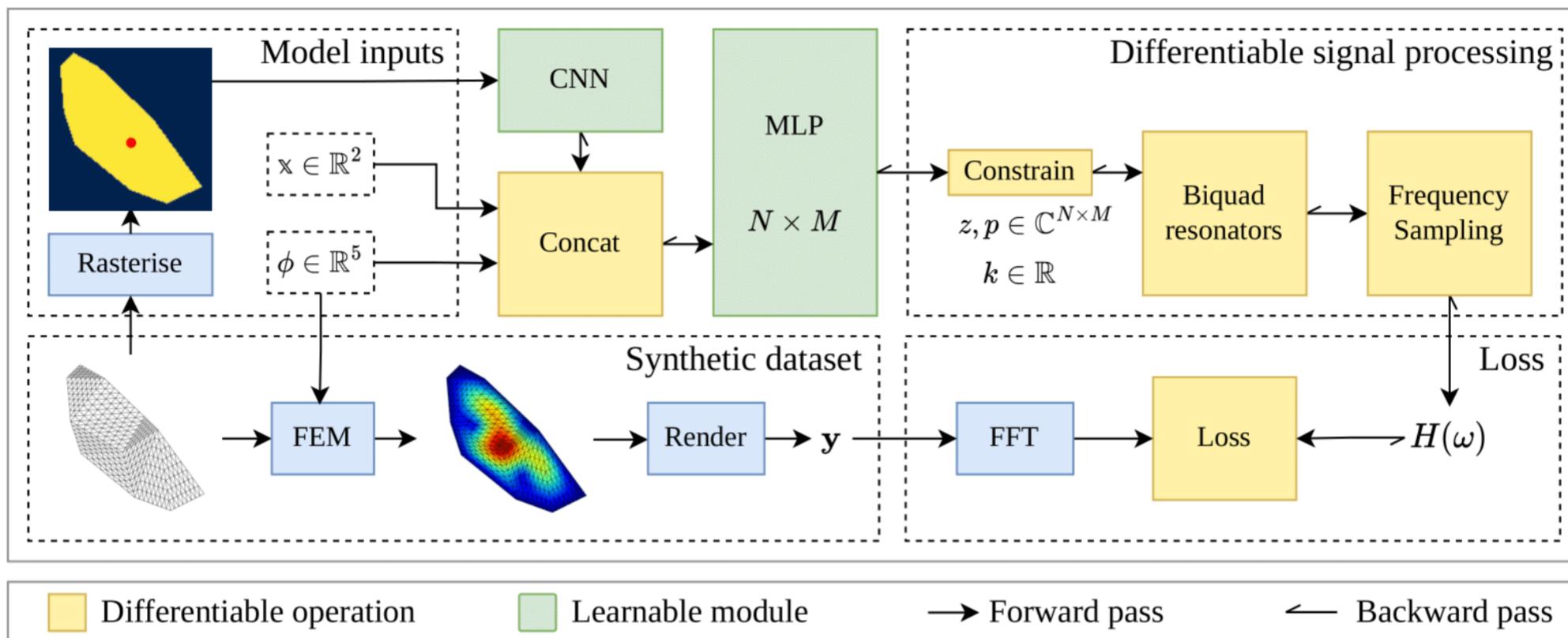
Question

Why are drums, especially cymbals, so hard to transcribe using conventional methods?

Hearing volumetric shapes

Rodrigo Diaz

2021



Finding

It is possible to use neural networks to simulate the vibrational dynamics of membranes of arbitrary shape and strings in real-time.

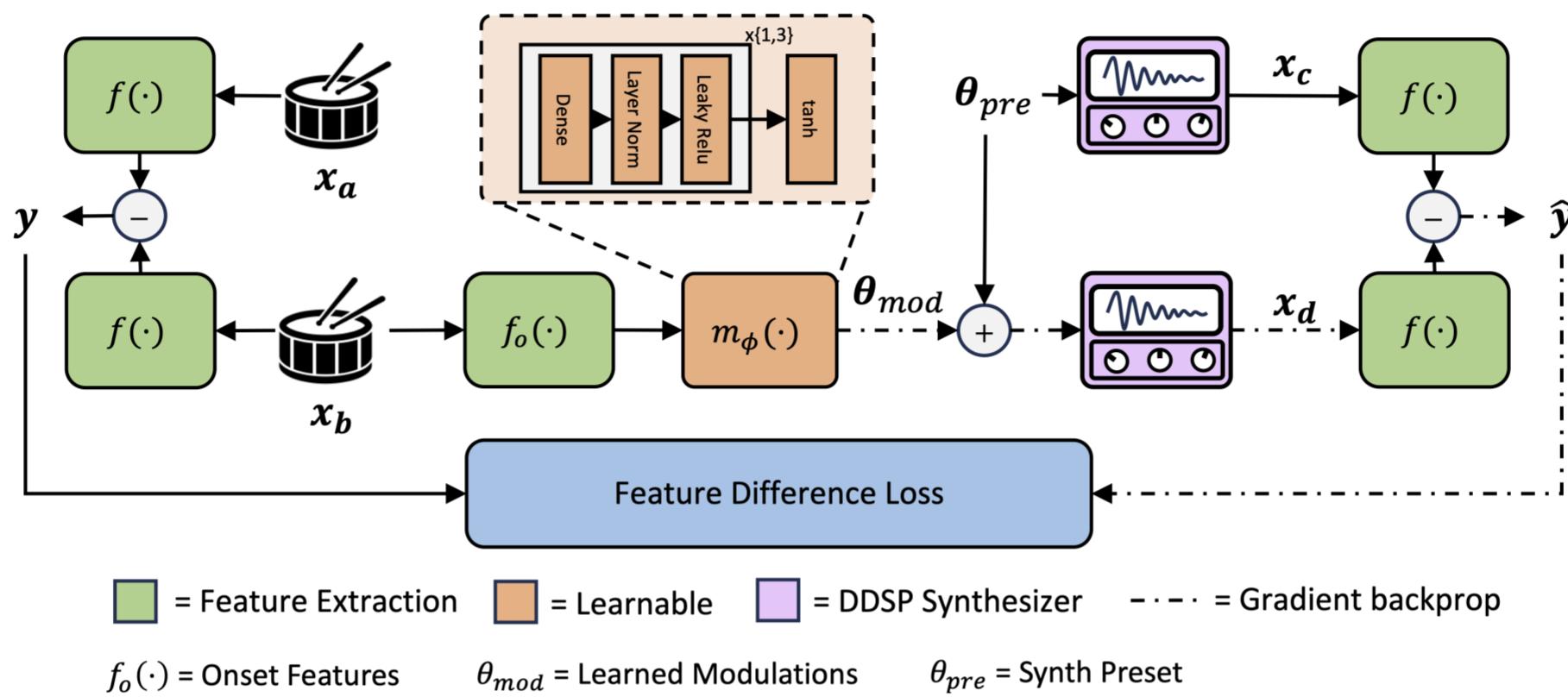
Question

How can we use shape to model sound in an intuitive manner?

Real-time Timbral Mapping for Synthesized Percussive Performance

Jordie Shier

2022



Finding

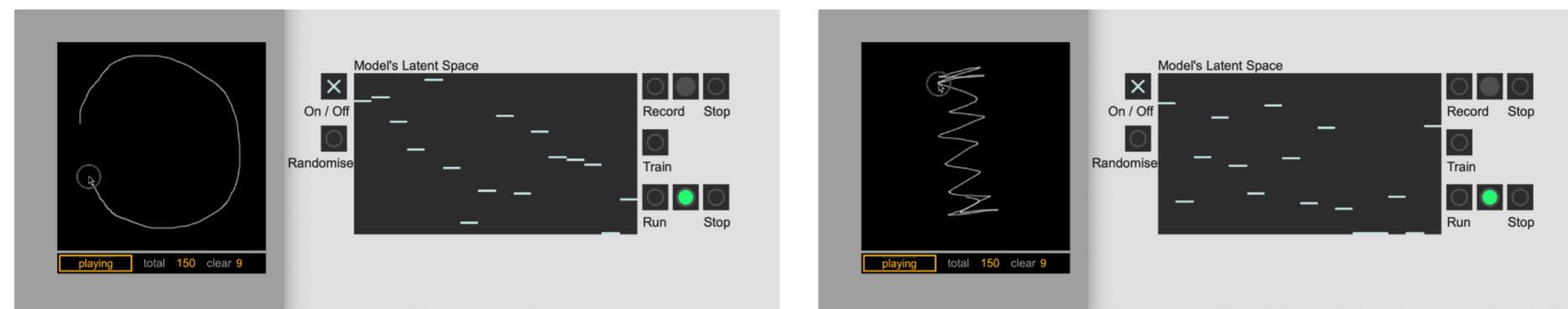
How can we achieve audio-driven synthesis where timbre variation from an input performance is considered as a primary control? Using timbre analogies! Similar to how melodies can be transposed into different keys, we transpose timbral sequences observed in an input onto a synthesiser. This is supported by differentiable DSP.

Question

What audio/timbre features are most relevant for timbre remapping of percussion instruments? How can deep learning and/or learned embeddings support timbre remapping (which is essentially a style transfer task) between acoustic percussion and a synthesiser?

Transparent and Understandable Neural Synthesiser Instruments

Shuoyang Zheng
2023



Finding

A literature review about state-of-the-art methods for explainable neural audio synthesis and latent space explanation, issues and challenges in real-time interaction with neural audio models, and existing evaluation metrics for explainability in the context of generative AI and the arts.

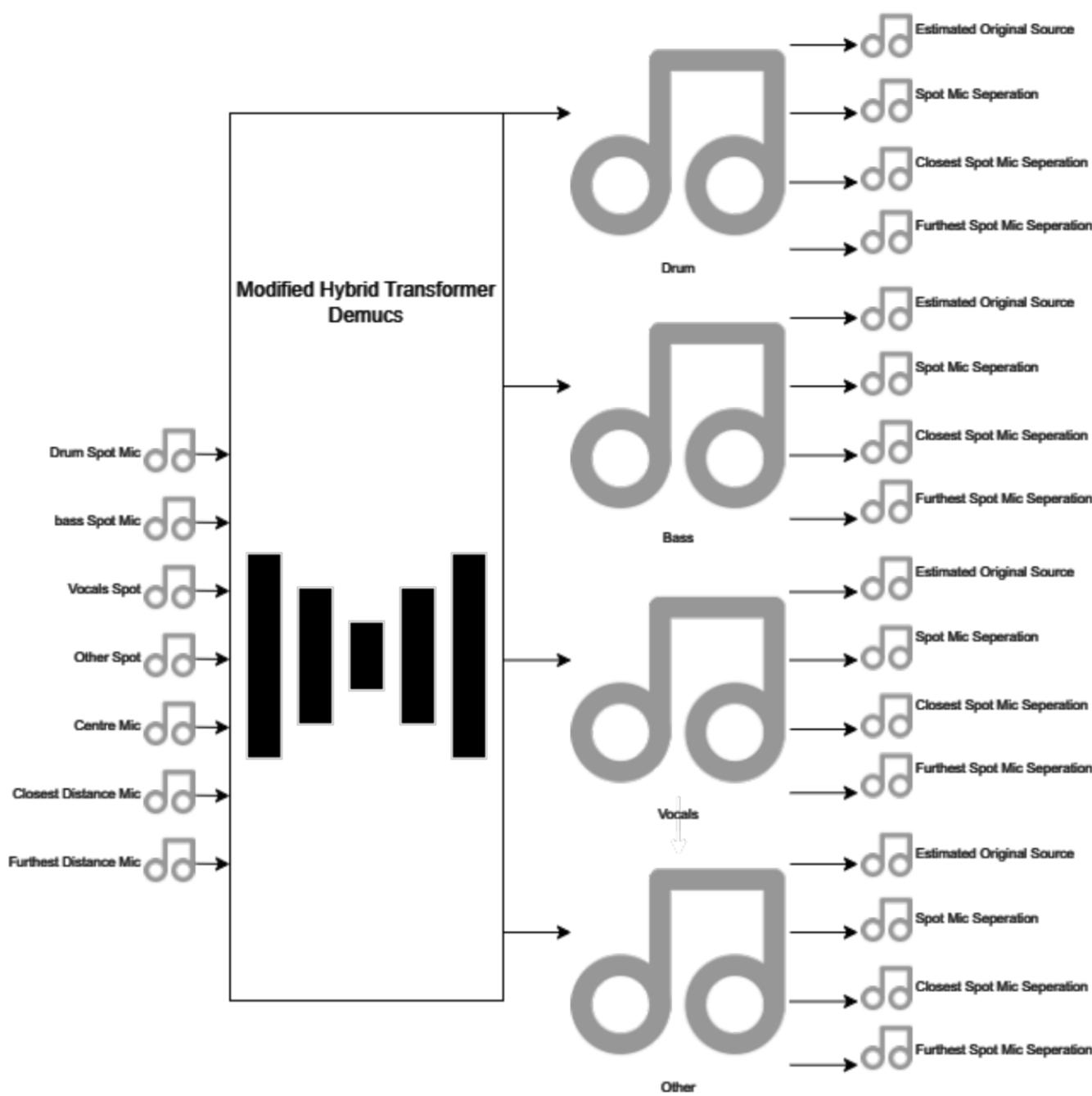
Question

How can techniques in explainable AI be used to create human-understandable features for a neural synthesiser instrument? And how can these explainable controls be tailored into the context of music sound design and production?

Informed source separation for multi-mic production

Harnick Khera

2020



Finding

One key finding from my PhD is that source separation approaches can be effectively adapted to perform multimodal tasks, including both separation and transcription of drum sounds. This demonstrates the potential of these models to leverage shared features and capabilities, enhancing their utility in music information retrieval and broadening their scope.

Question

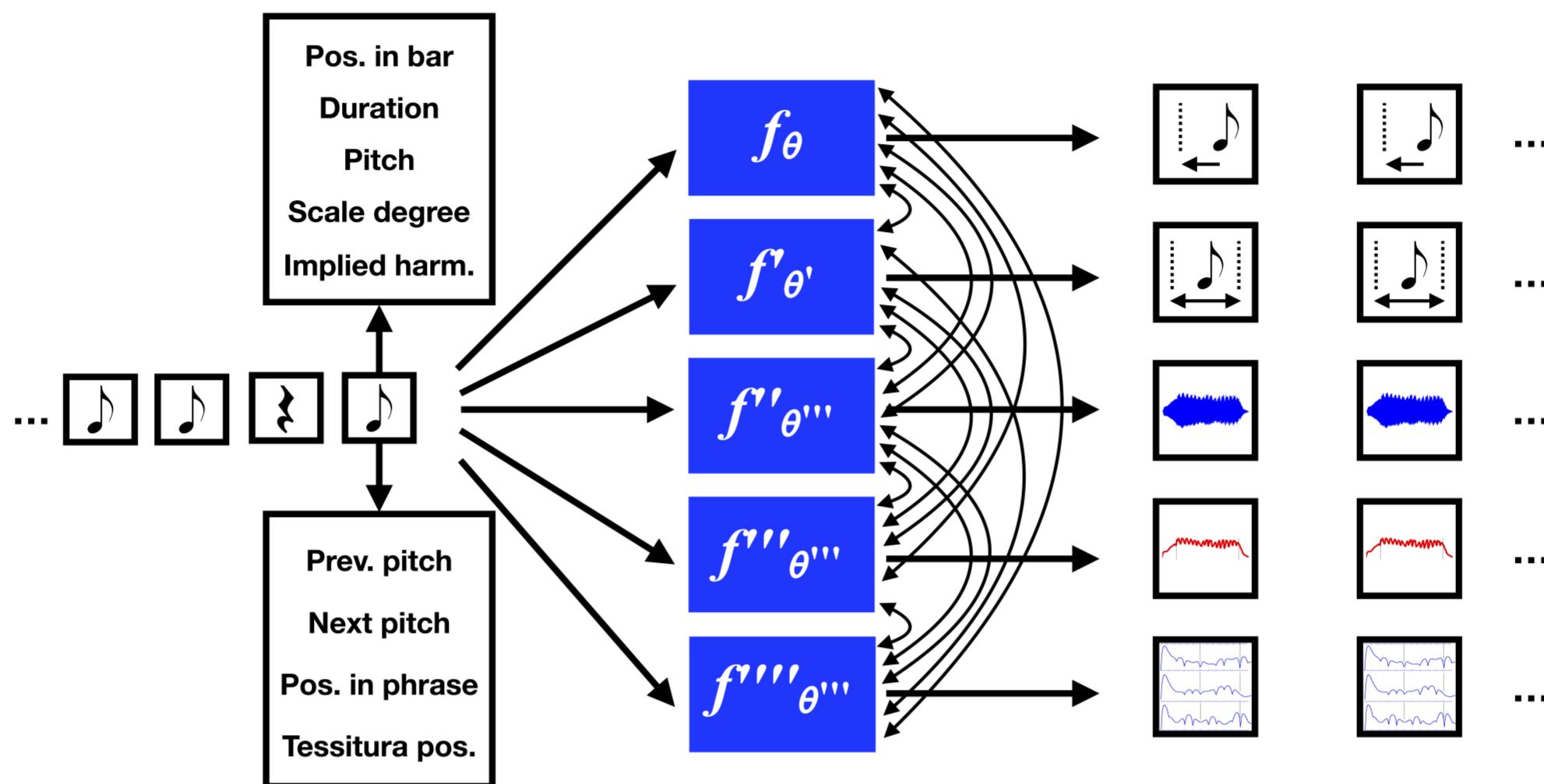
An open question in my research is how using multiple microphones as inputs can enhance source separation capabilities. I am exploring whether spatial information can improve separation accuracy and effectiveness, and subsequently, how this approach might also be leveraged for drum transcription and used for gain control or acoustic contrasting in a mixture.

Supervisor(s): Prof. Mark Sandler, Dr. Johan Pauwels, Dr. Alan Archer-Boyd

Synthesising Jazz Saxophone Performances

Dave Foster

2019



Finding

Performance timing is inconsistent and difficult to accurately pin down - exponentially so when dealing with an ensemble!

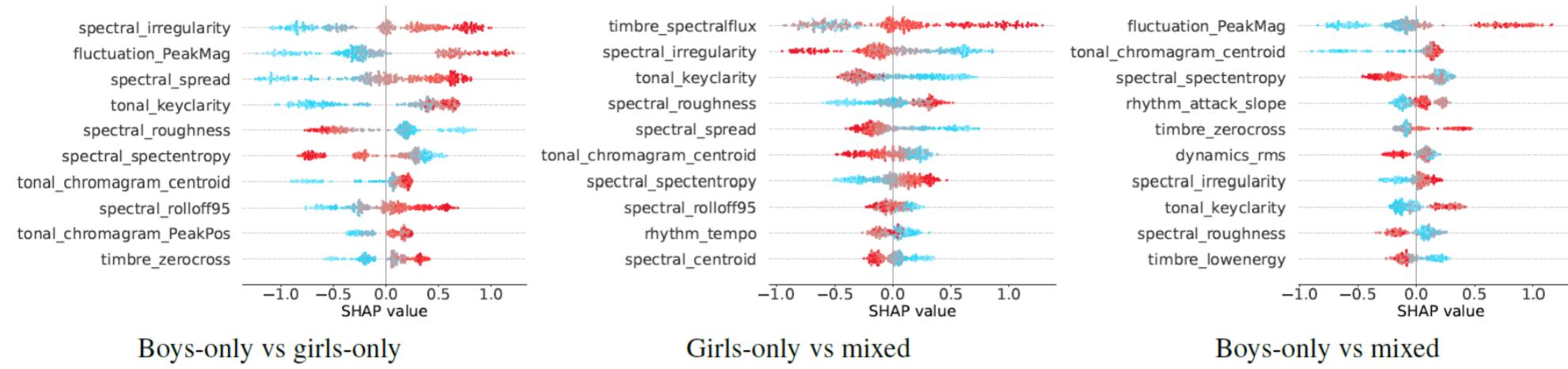
Question

Can we use annotated performance data to generate performance parameters, that can be used to synthesise musical performances which are indistinguishable from that of an expert human?

Gender-Coded Sound: A multimodal analysis of gender encoding strategies in music for advertising

Luca Marinelli

2020



Finding

To what extent is gender-coding deliberately used, when music is secondary to other modalities and serves a clear purpose, such as in advertisement? Our overarching research objective is to provide a theory of the effects that message producers, and their decision-making, have on the selection and composition of music in gendered toy commercials

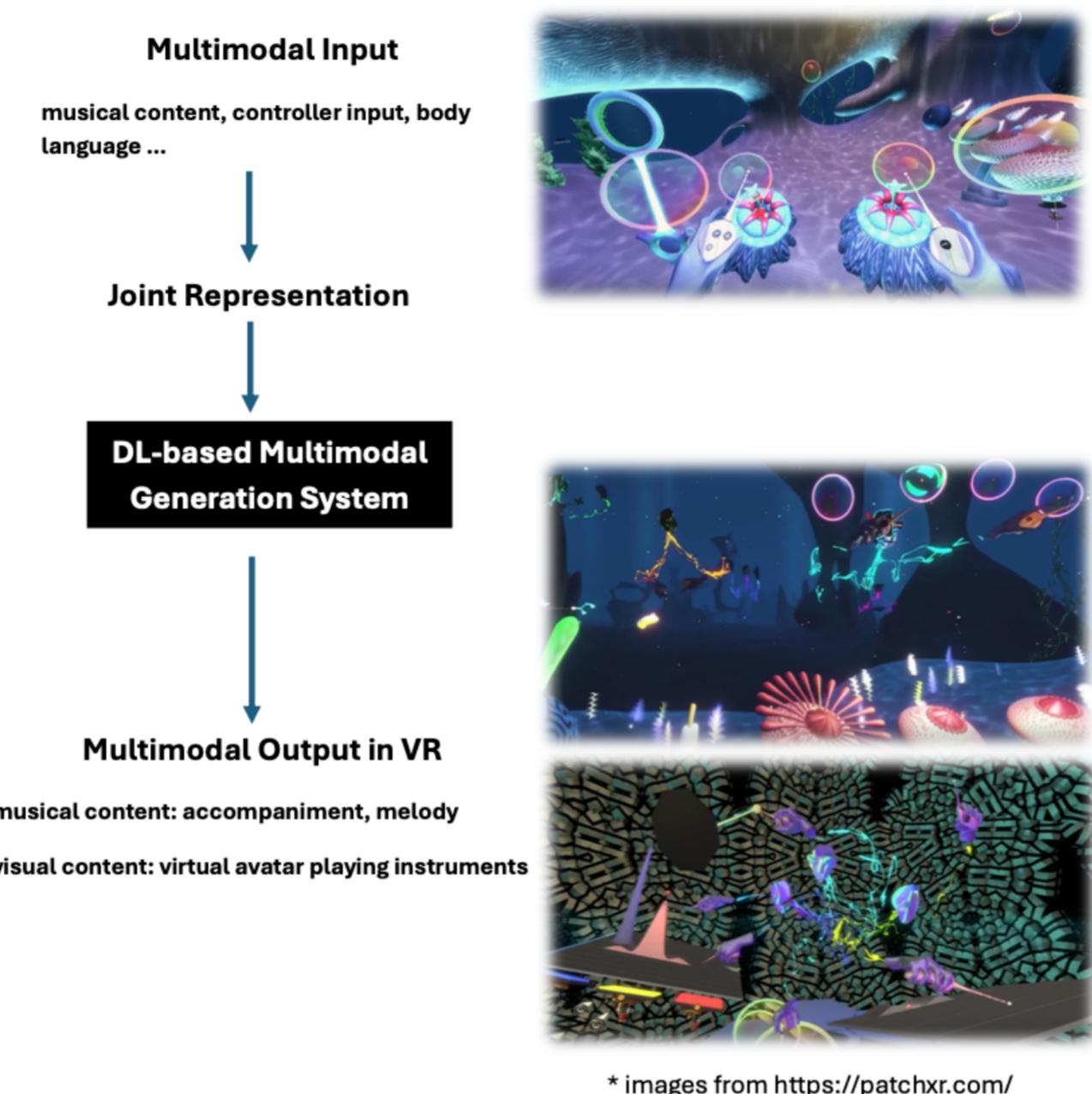
Question

If you're interested, and have experience in retrieval augmented generation with LLMs we could collaborate on a project! I'm relatively up to date with the literature and plan to use DSPy + LlamaIndex.

Multimodal AI for Musical Collaboration in Immersive Environments

Qiaoxi Zhang

2023



Finding

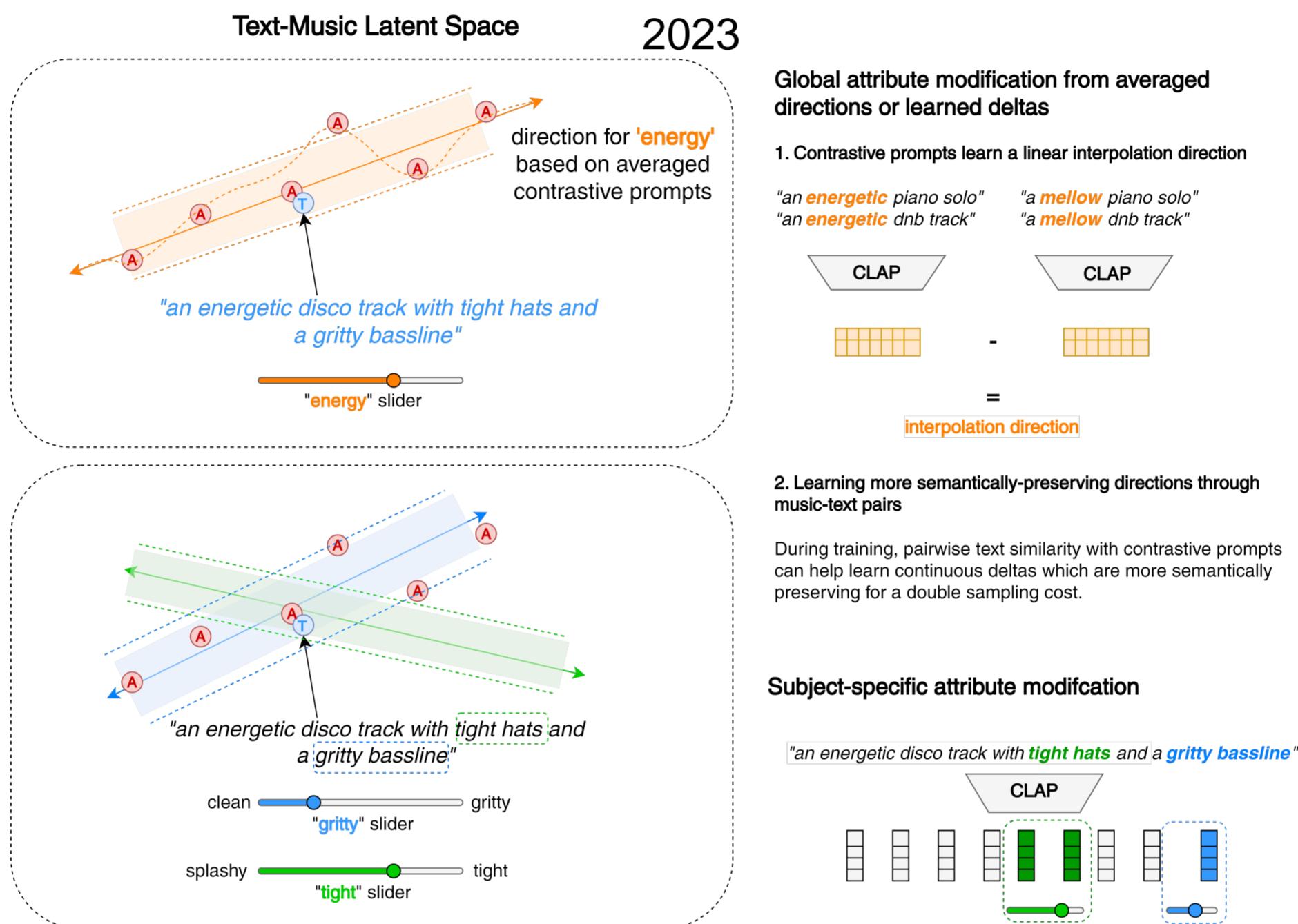
There is a lack of robust methods for integrating and analysing the rich multimodal data collected in VR environments for effective automatic music generation and collaboration.

Question

Which input modalities could benefit the generation of a virtual avatar for music collaboration with humans, and how could a joint representation be designed?

Identifying semantic directions in contrastive text-music spaces for continuous and subject-specific attribute control for music retrieval

Julien Guinot



Finding

Some deterministic audio transformations can be easily associated with musical semantic information (time stretching with tempo, pitch shifting with key). We found that it is possible to learn semantic transformations in the latent space by controlling these factors and learning to reconstruct the modified embedding, which benefits retrieval controllability.

Question

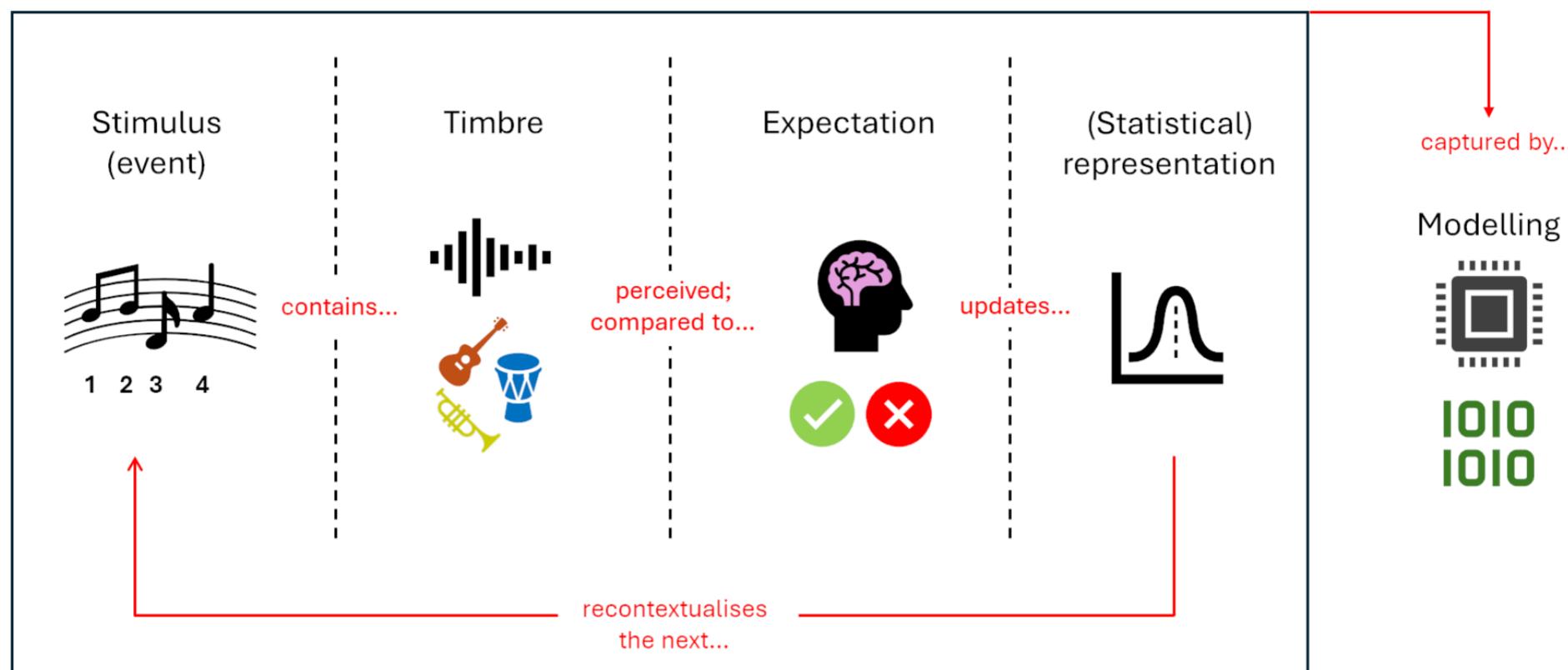
Transformations for more nuanced concepts not directly related to any audio transformation are hard to capture in the audio space. We ask whether discovering semantic 'slider' transformations from contrasting prompts in existing text-music latent spaces is a viable method for modeling transformations relating to these concepts.

Supervisor(s): George Fazekas, Elio Quinton

Computational Modelling of Expectancy-Based Music Cognition from Timbral Structures

Adam Garrow

2022



Finding

A literature review revealed four domains in which timbral patterns significantly inform a listener's musical expectancy:

- Direct timbral prediction, instrument-level timbral attributes for predicting non-timbral material,
- low-level timbral attributes for predicting non-timbral material,
- and timbre-cued style recognition for predicting musical material.

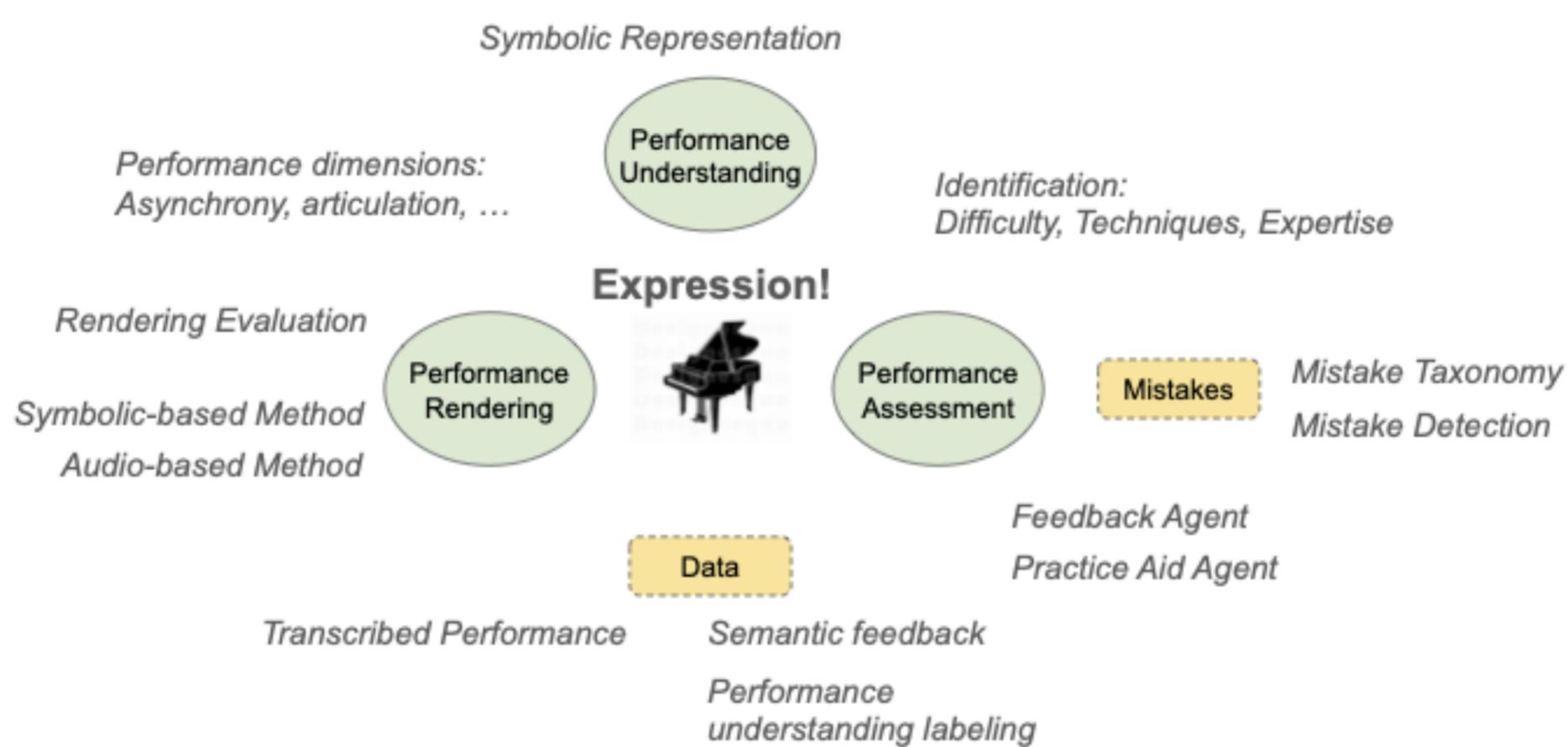
Question

Does the spectral envelope of musical tones interact with the implicit learning of statistical regularities in melodies composed from specific musical scales?

Computational Modeling of Expressive Piano Performance

Huan Zhang

2021



Finding

How much are we able to mimic the human expression with generative models? How well can we assess the student performance and also predict the winner of Chopin piano competition? In the amateur level, how we can neatly clear out the mistakes in performances?

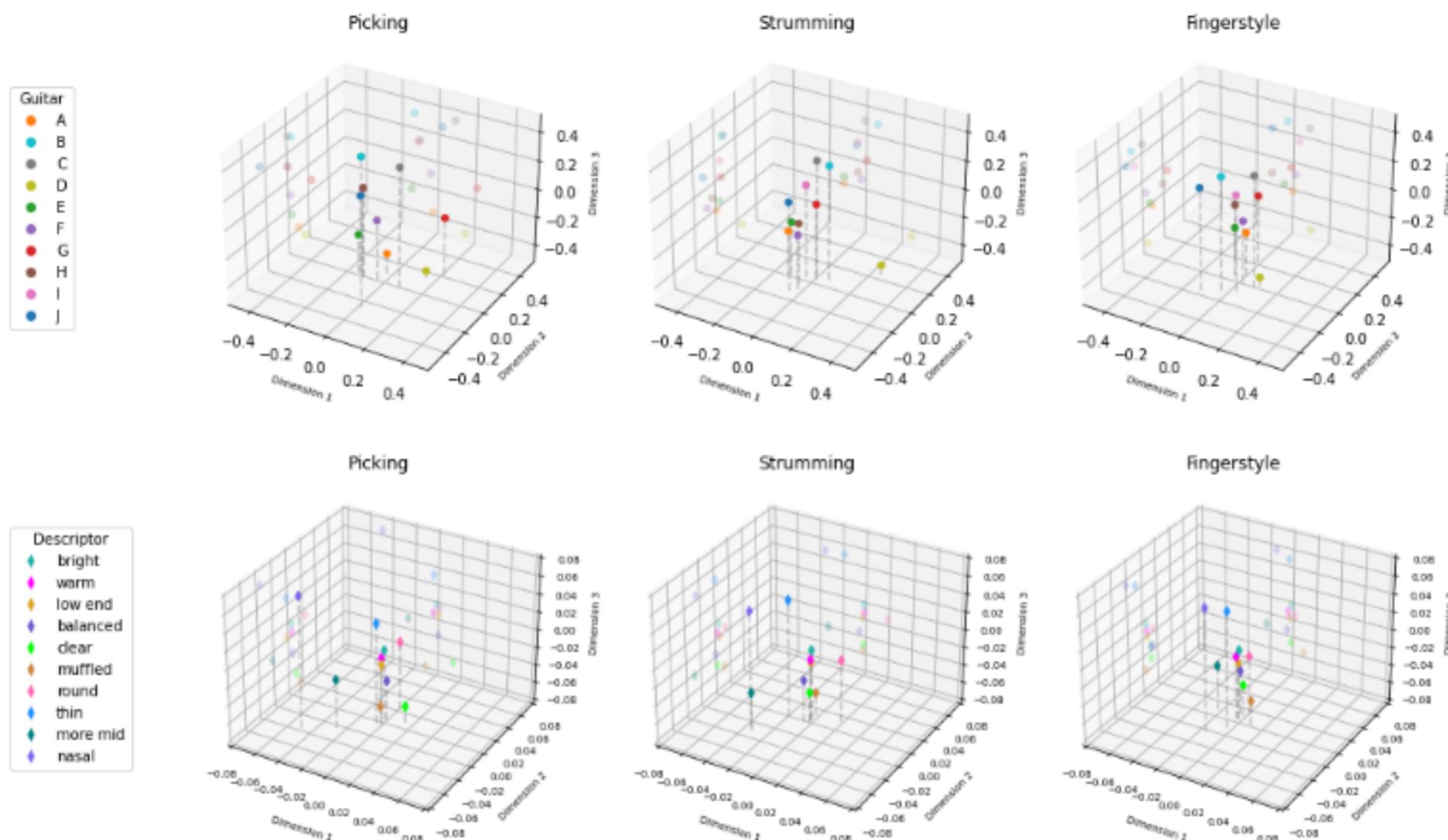
Question

Can we create agents tool to better aids the learning process of instruments?

Timbre Transfer For A Smart Acoustic Guitar

Jack Loth

2021



Finding

Perception of acoustic guitar timbre seems to have a dependence on musical context such as playing style and playing techniques. Listeners tend to prefer the timbre of acoustic guitars which sound 'rich' and 'round', while disliking timbres which sound 'hollow', 'high passed' and 'sharp'.

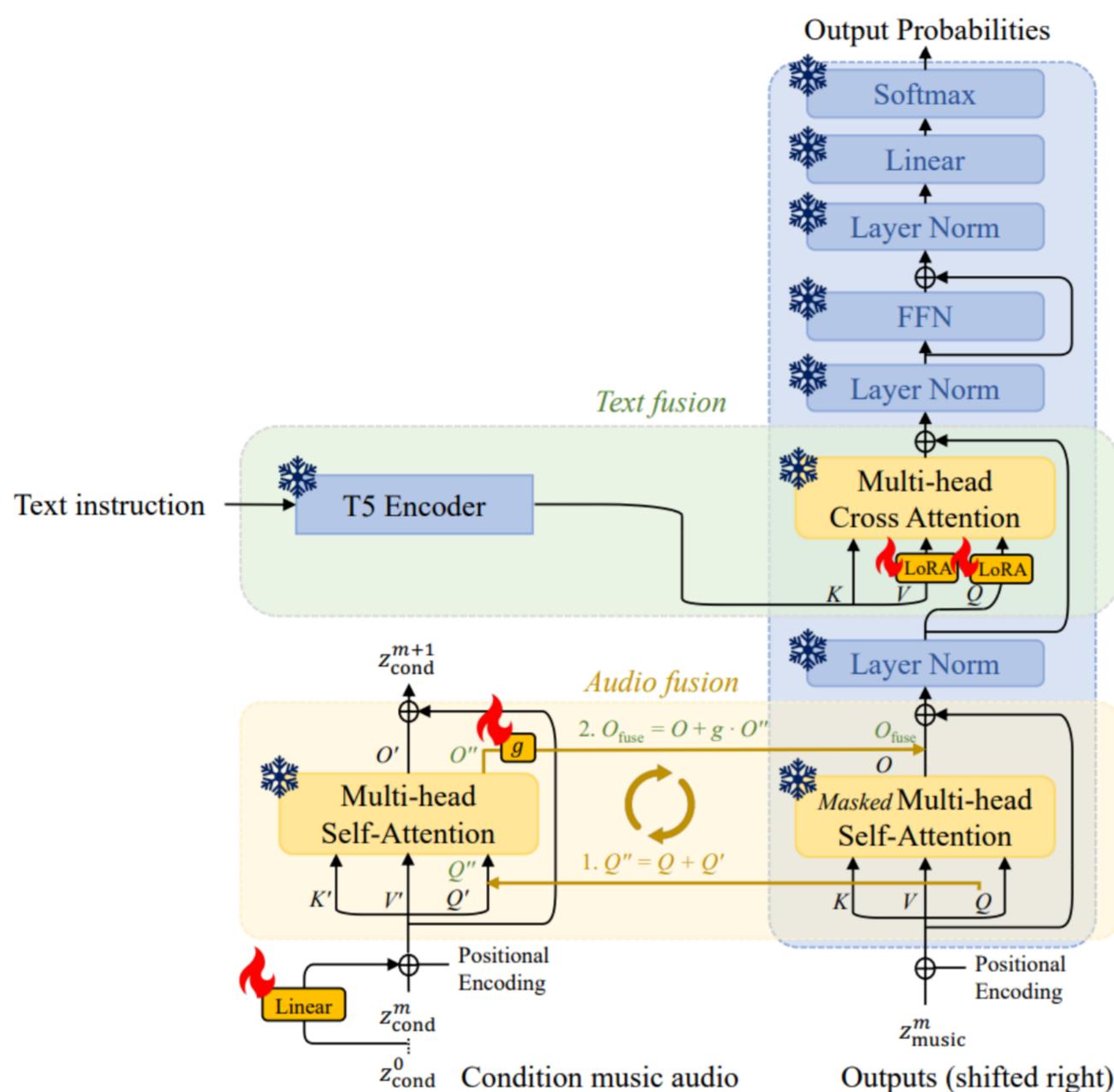
Question

How can we encode acoustic guitar audio into an encoding which specifically contains information on playing techniques as well as notes and velocity?

Instruct-MusicGen: Unlocking Text-to-Music Editing for Music Language Models via Instruction Tuning

Yixiao Zhang

2020



Finding

We already have text-to-music generation model, but what is the next step of text-to-music editing model? We investigate different aspects for text-based music editing methods, including LLM agents, semantic editing for diffusion models, and PEFT methods for music language models.

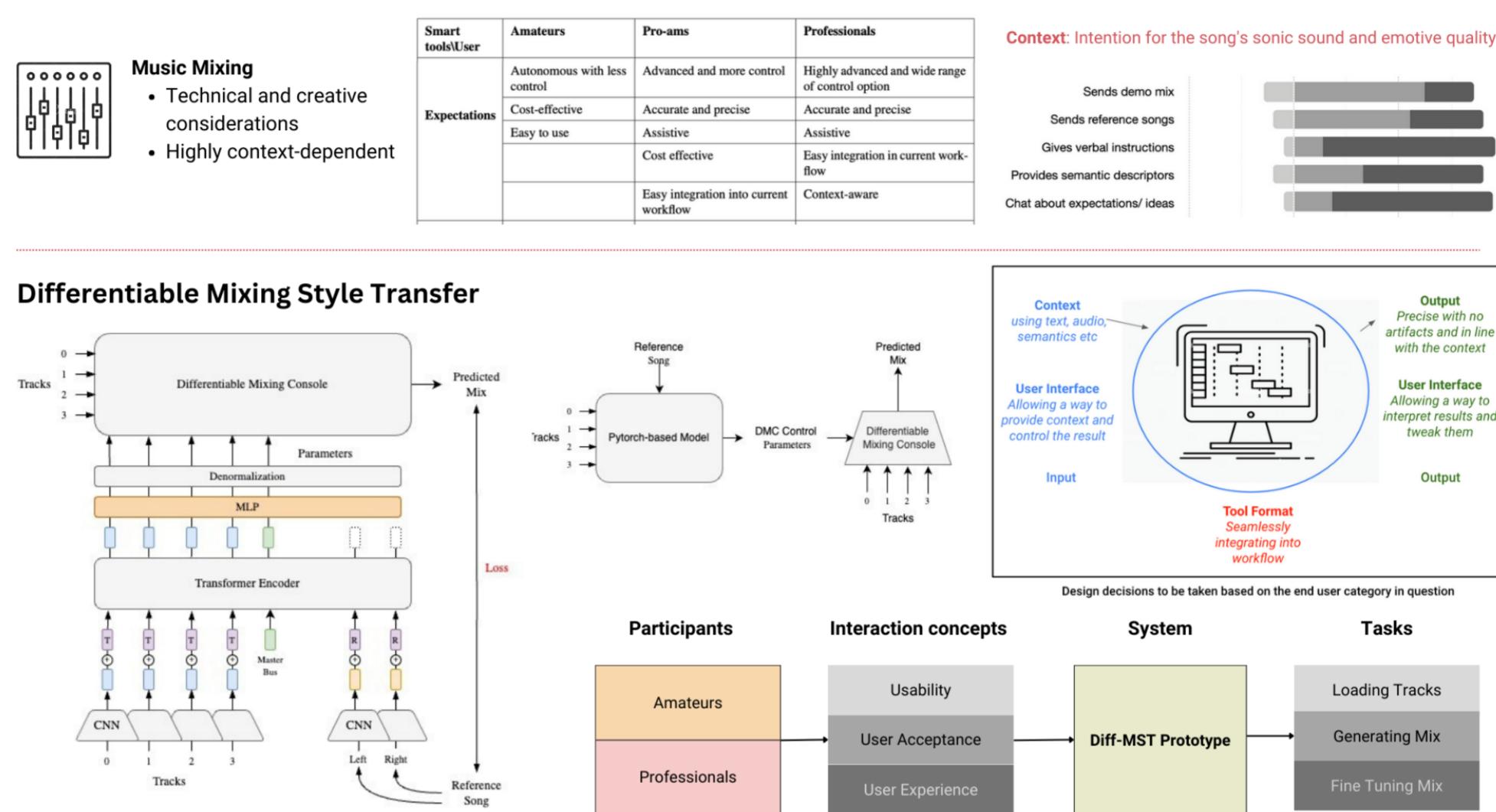
Question

How do we let AI music generation tools facilitates content creators in a better way? What control can be utilised for human-AI music co-creation?

User-Centric Intelligent Music Multitrack Mixing Style Transfer

Soumya Sai Vanka

2021



Finding

We have defined requirements of different user groups from AI mixing tools and identified ways to provide and understand context in real world mixing settings. We have built an interpretable mixing style transfer model incorporating expert knowledge that uses reference song as context to predict mixing console parameters and mix for given tracks.

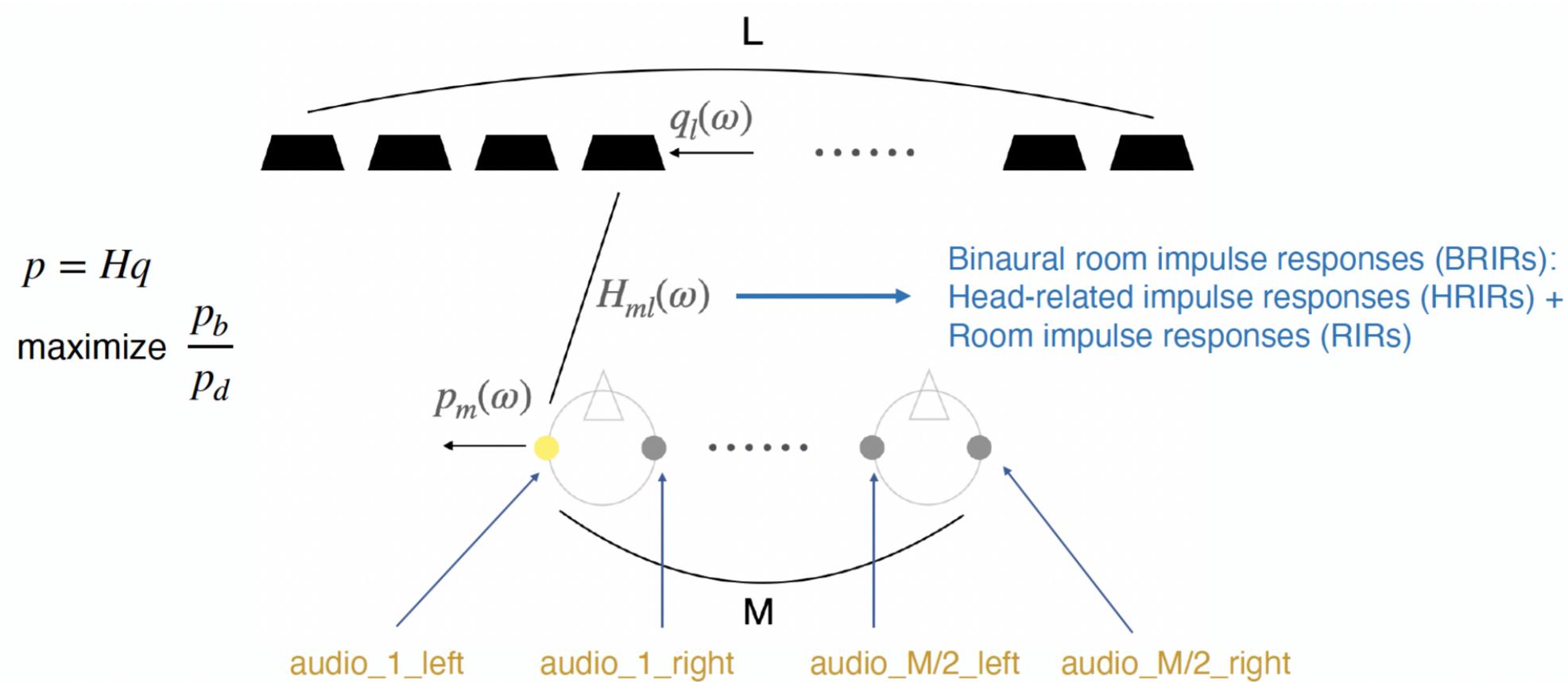
Question

We are looking to find an objective definition for similarity and dissimilarity between two mixes and thereby learn a latent space of all possible mixes. We are also working on a prototype for incorporating Diff-MST model into a DAW and conduct user experience and usability evaluations.

personal and spatial audio reproduction using loudspeaker arrays for home applications

Yazhou Li

2021



Finding

Room acoustics affects personal sound zone reproduction performance, and it is especially important to compensate for the early reflections.

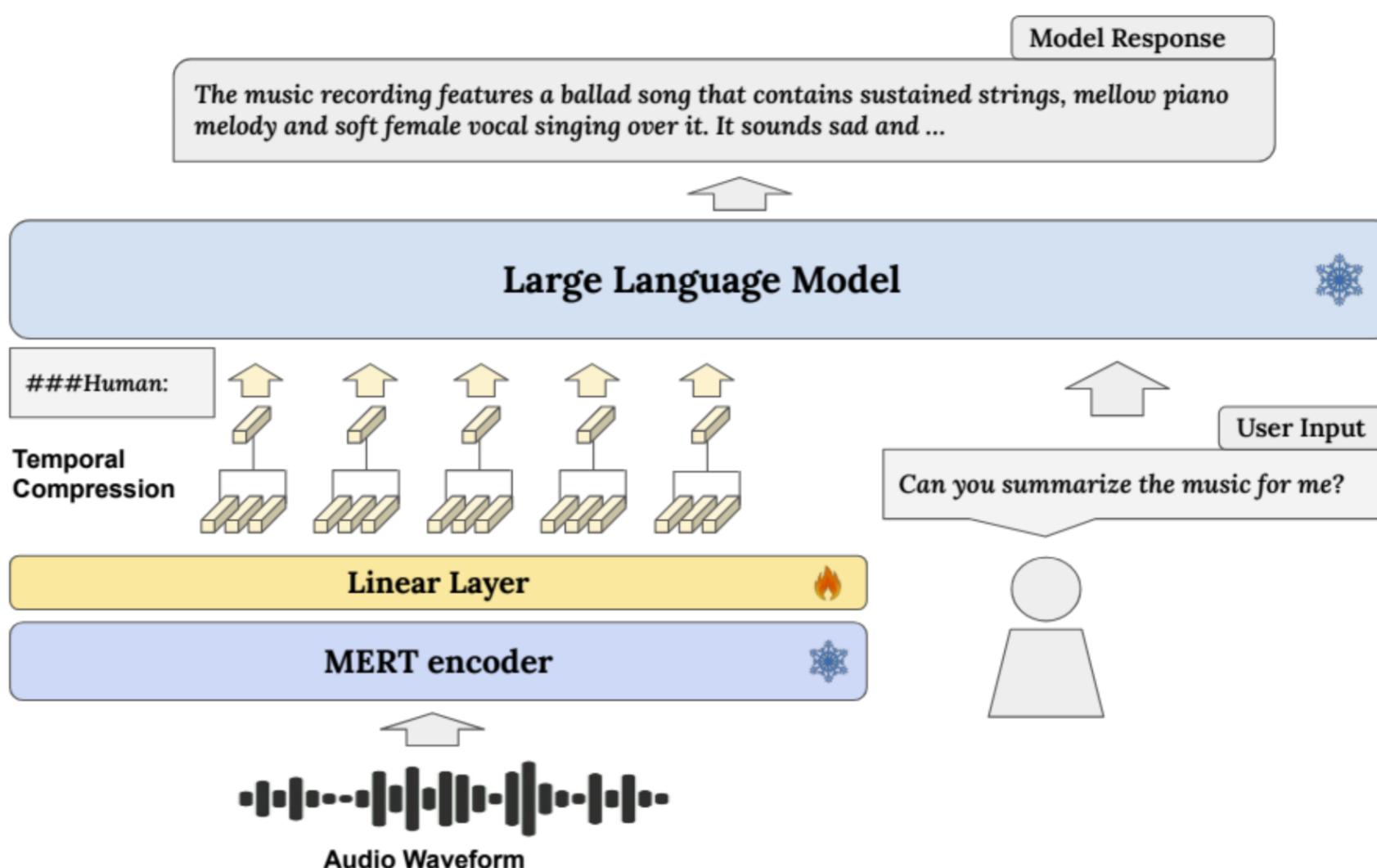
Question

How to estimate the soundfield in a home environment? How well can the estimated RIRs work for sound reproduction problems?

Large Language Model for Music Understanding and Music Generation

Yinghao Ma

2022



Finding

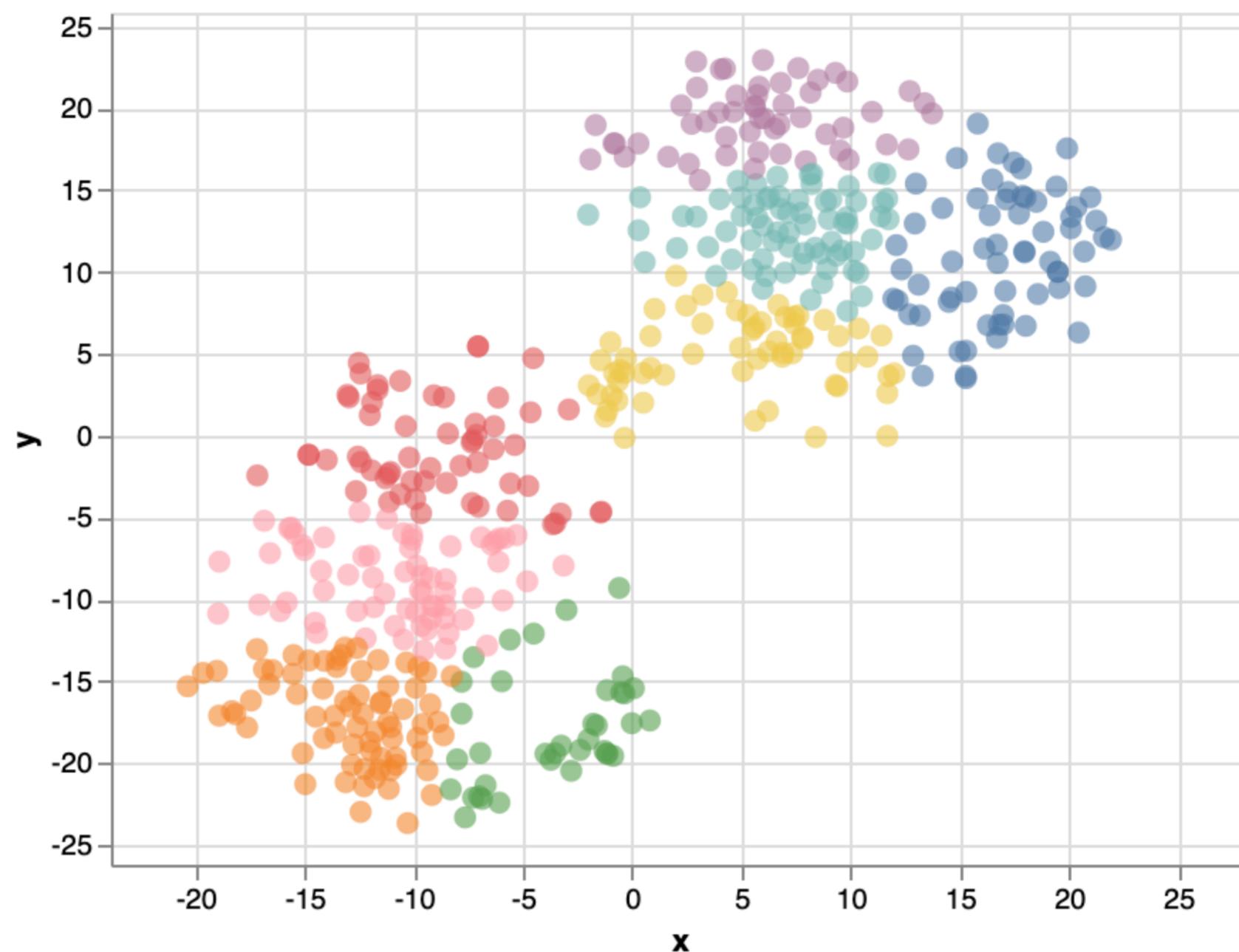
Large Language Models have shown immense potential in multimodal applications. We present MusiLingo, a novel system for music caption generation and music-related query responses. MusiLingo employs a single projection layer to align music representations from the pre-trained music encoder with a frozen LLM, bridging gap between music and textual contexts.

Question

Developing multimodal music-text supervised finetuning (SFT) / instruction finetuning dataset for a better language model finetuning with better generalisation to unseen tasks

Limited-data learning in music

Christos Plachouras
2023



Finding

Is in-domain audio representation learning feasible? Current representation learning models aren't as magical of a solution as they sometimes seem to be; when randomly initialized without pretraining, they often produce representations that perform close to state-of-the-art. With a few minutes of training data, the gap can close further.

Question

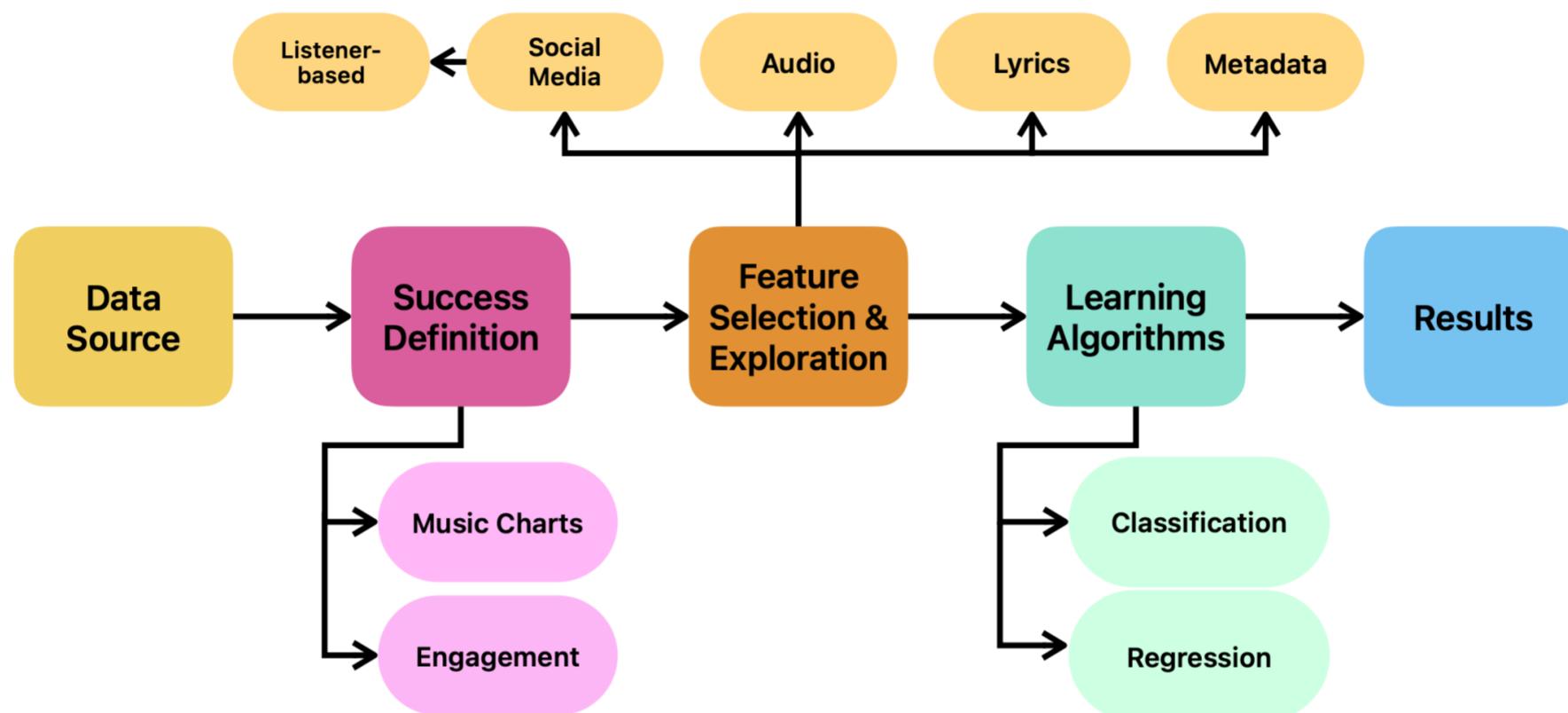
Currently, limited-data scenarios, particularly in a continual learning settings, are tackled with scenario-specific methods. 1. Could we understand what data and task characteristics (e.g. weakly-labeled, many increments etc.) lead to different behaviors (plasticity, forgetting), and 2. is it possible to develop more universal approaches for music?

Predicting hit songs: multimodal and data-driven approach

Kasia Adamska

2021

Hit Song Science: Generic Workflow



Finding

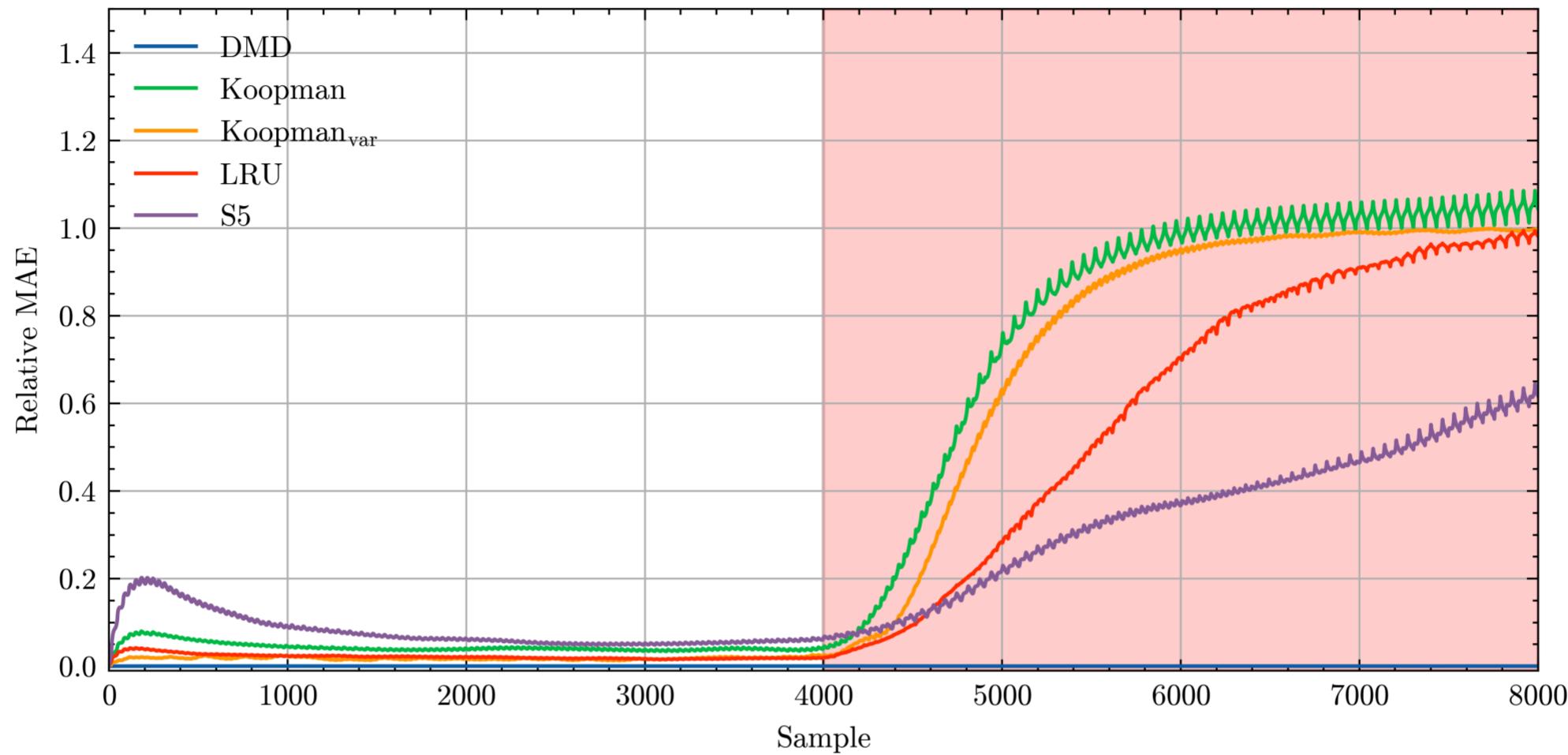
The experiment on predicting chart success of songs based on lyrics found that lyrics affect a song's chart longevity, but not necessarily its peak position. Key features were n-gram repetitiveness and vocabulary richness (type-token ratio). Better classification was achieved by focusing on the dataset's extremes —very popular and very unpopular songs.

Question

I'm searching for refined audio features that best capture how we perceive popular music. I want to move beyond Spotify's audio features, Essentia's music extractor (low-level, rhythm, and tonal features), and Mel-spectrograms. Any suggestions are welcome. Thank you!

Physical modelling with neural operators

Carlos De La Vega Martin
2021



Finding

Autoregressive use of neural operators is not as straightforward as it may seem

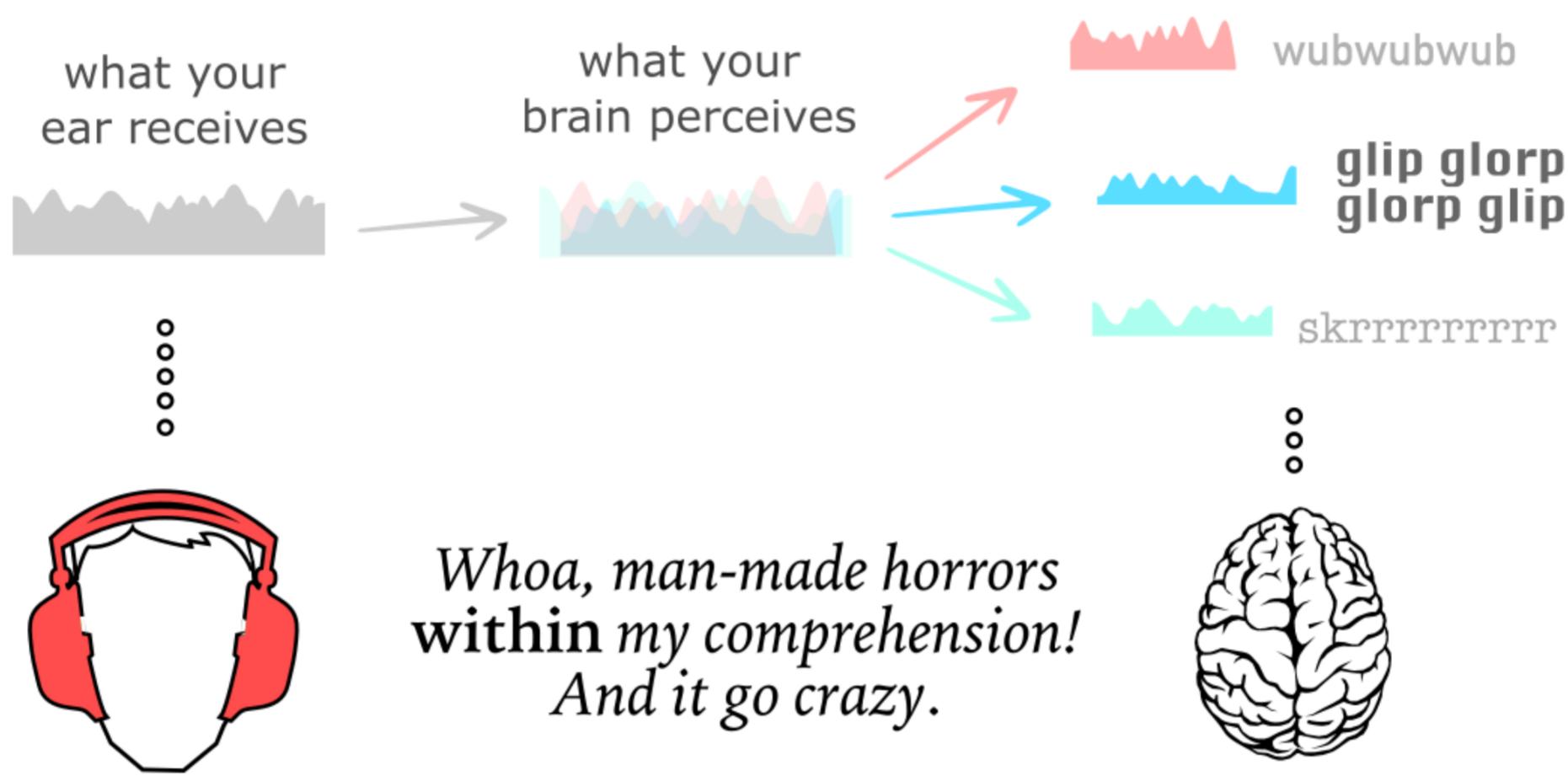
Question

Can we guarantee stability in time extrapolation with neural operators?

Information-theoretic neural networks for online perception of auditory objects

Farida Yusuf

2023



Finding

The auditory neuroscience literature evidences that disentangling sound is a complementary process of bottom-up feature processing and top-down selectivity of sound entities encoded in cortical neurons, i.e. perceptual schema. This framework based on sense data is hardly explored in online source separation, despite advances in ML for pattern recognition.

Question

Imagine an online model processing audio in time and frequency. What grouping and pattern recognition rules recreate our perception of "edge" and "contour" in sound? Can information-theoretic neural networks represent schema-based processing? Ultimately, can a complex model of auditory object perception exceed source-based paradigms of sound separation?

Neuro-Symbolic Automated Music Composition

Keshav Bhandari

2023

YIN-YANG: A NEURO-SYMBOLIC FRAMEWORK FOR DEVELOPING MOTIFS WITH LONG-TERM STRUCTURE AND CONTROLLABILITY



Finding

Sub-task decomposition using a neuro-symbolic framework can lead to generations of melodies with long term structure while infusing domain knowledge into the process. This allows for a semi-interpretable generative approach with the flexibility of human controls and intervention.

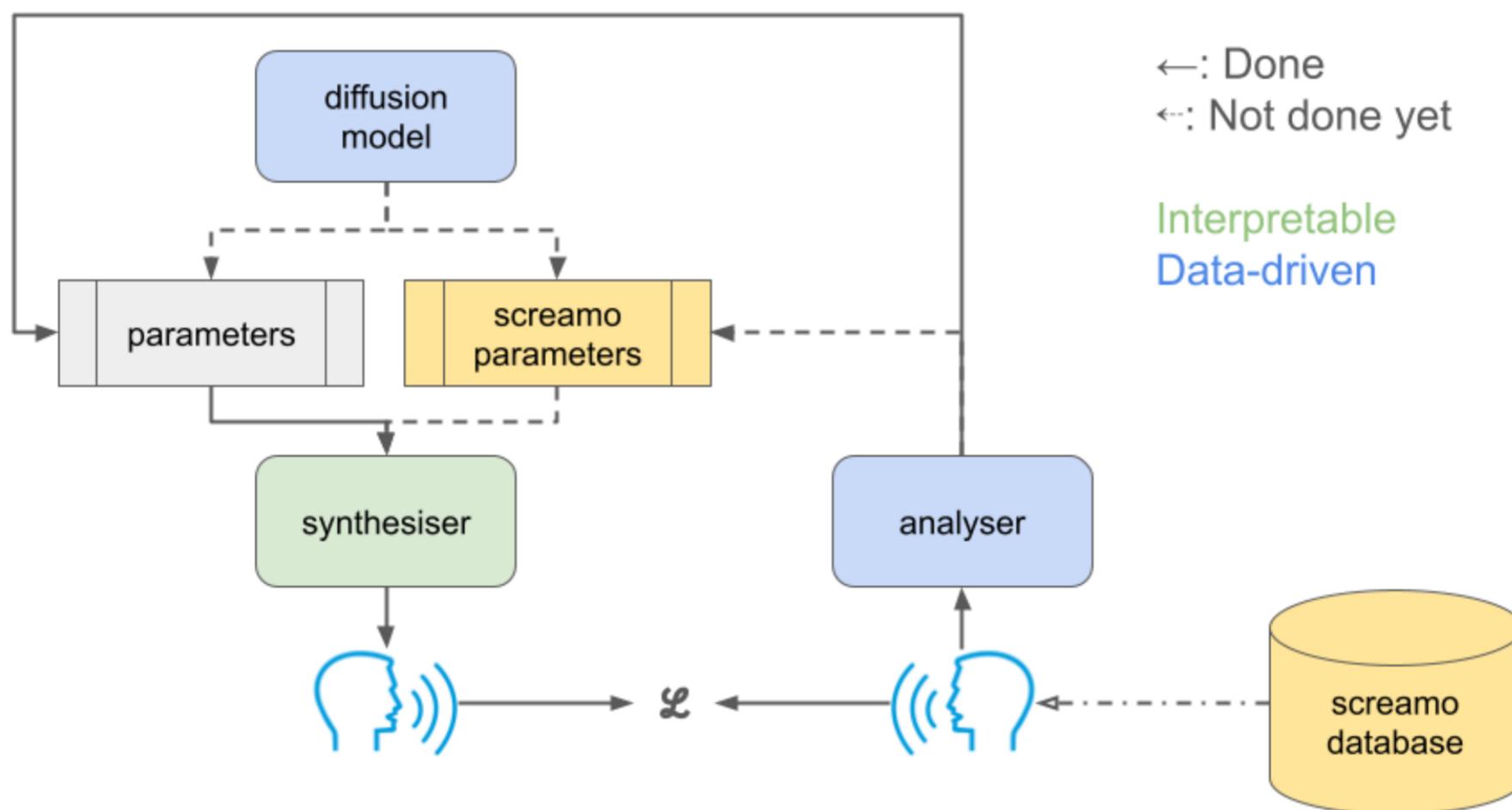
Question

Self-similarity matrix (SSM) is a graphical representation of similar sequences in a data series. In music, SSMs can yield useful information on musical structure and interactions between various elements. Could incorporating SSMs as inputs into the generative framework improve the global structure of the generations?

Analysing and controlling extreme vocal expression using differentiable DSP and generative models

Chin-Yun Yu

2022



2

Finding

- 1) Leveraging DSP components as prior knowledge reduces computational loads and training resources and increases controllability.
- 2) The available techniques for posterior sampling in diffusion models are still not capable of inverse problems that have obscure clues of the original signal (e.g., monotimbral sources).

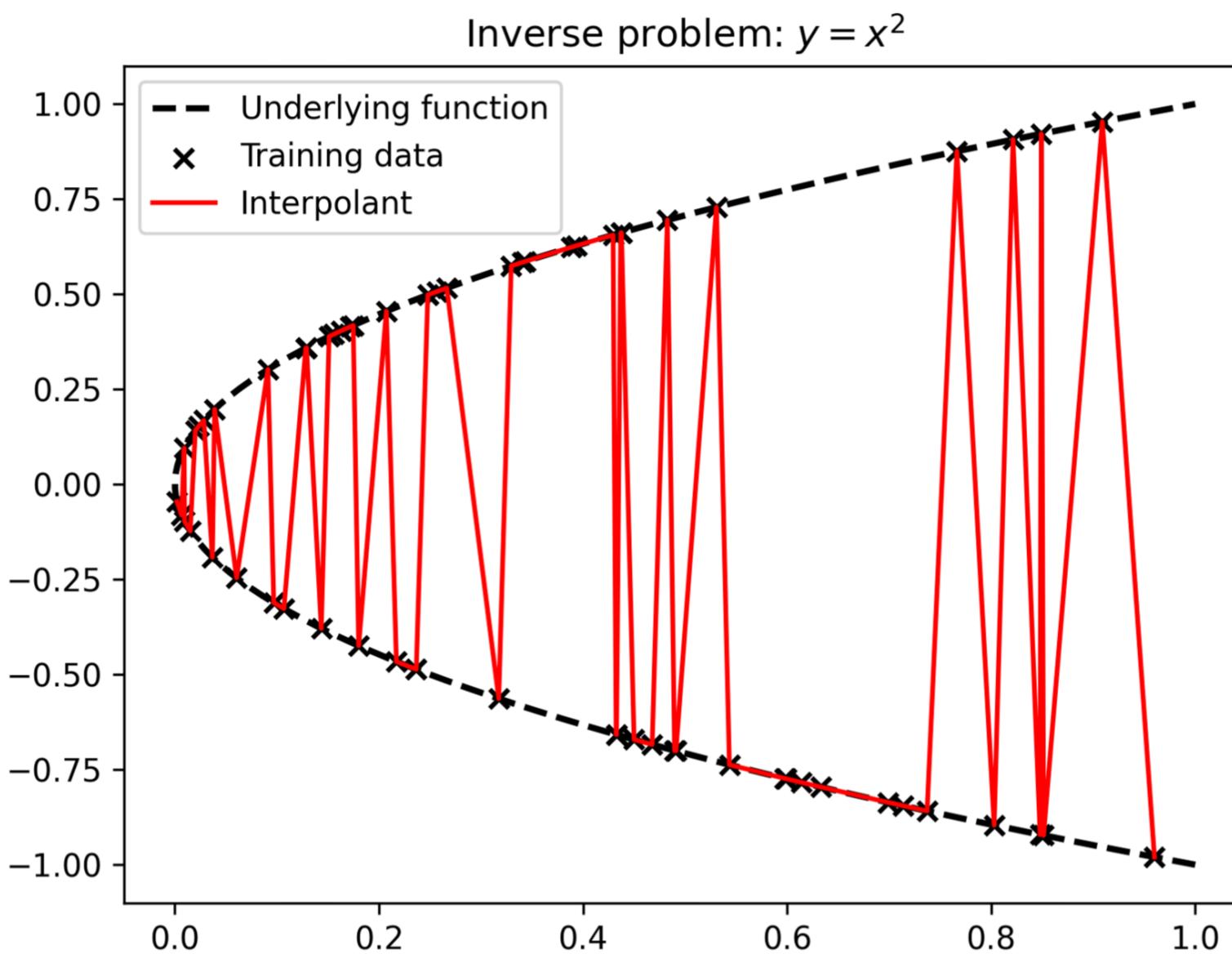
Question

Can we find a set of parameters that 1) describe the screaming vocal in a way that correlates with human perception and 2) can control the synthesiser in an efficient way?

Differentiable DSP and its Underlying Symmetries

Ben Hayes

2020



Finding

In DDSP, if the signal processor is invariant under certain transformations of its parameters, we will observe either (i) failure to converge or (ii) degraded performance. We have found three classes of symmetry that can be solved through carefully designed symmetry breaking: 1. Periodicity/aliasing 2. Permutation 3. Dead parameters

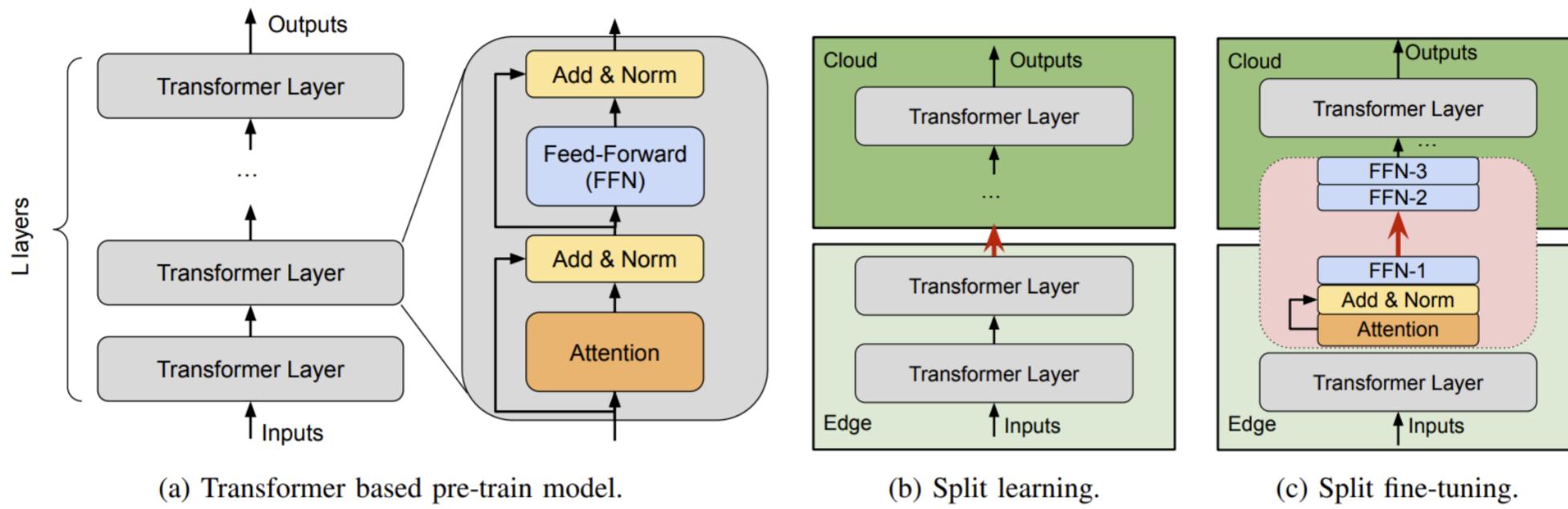
Question

How can we automatically discover and break symmetries in DDSP models?

Optimising Large Audio Models for Efficient Processing and Deployment

Bradley Aldous

2023



Finding

As a preliminary step in my research I trained and profiled two LDMs (audio and image) to identify the bottlenecks in training these models. As expected, the optimiser was a major bottleneck in both models, though more so with the audio model; informing future research in more efficient optimisation strategies.

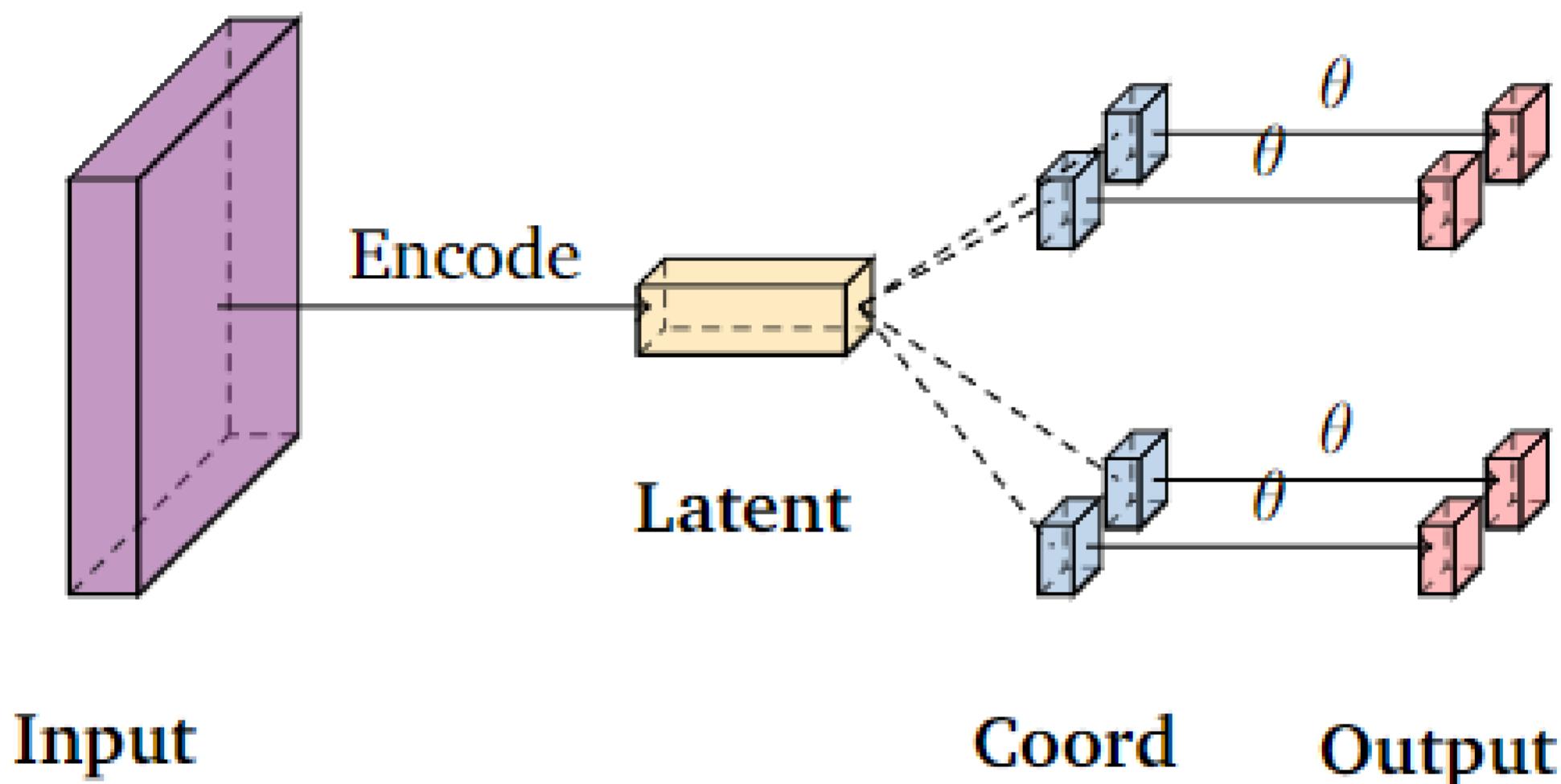
Question

An open question in my research is how to further optimise the training of large audio models, particularly focusing on advanced compression techniques and training strategies. I am exploring methods to reduce computational overhead while maintaining model performance, to eventually allow training of large audio models on resource constrained devices.

Likelihood Oriented Variational AutoEncoder for Representation Learning

Ningzhi Wang

2022



Finding

Coordinate VAE eliminates spatial structure in the latent space and employs manually defined coordinates to capture inter-pixel correlations in the decoder. We observed that this representation method struggled with pitch-related tasks, as pitch information relies on combining adjacent frequency bins, which are missing in the latent representation.

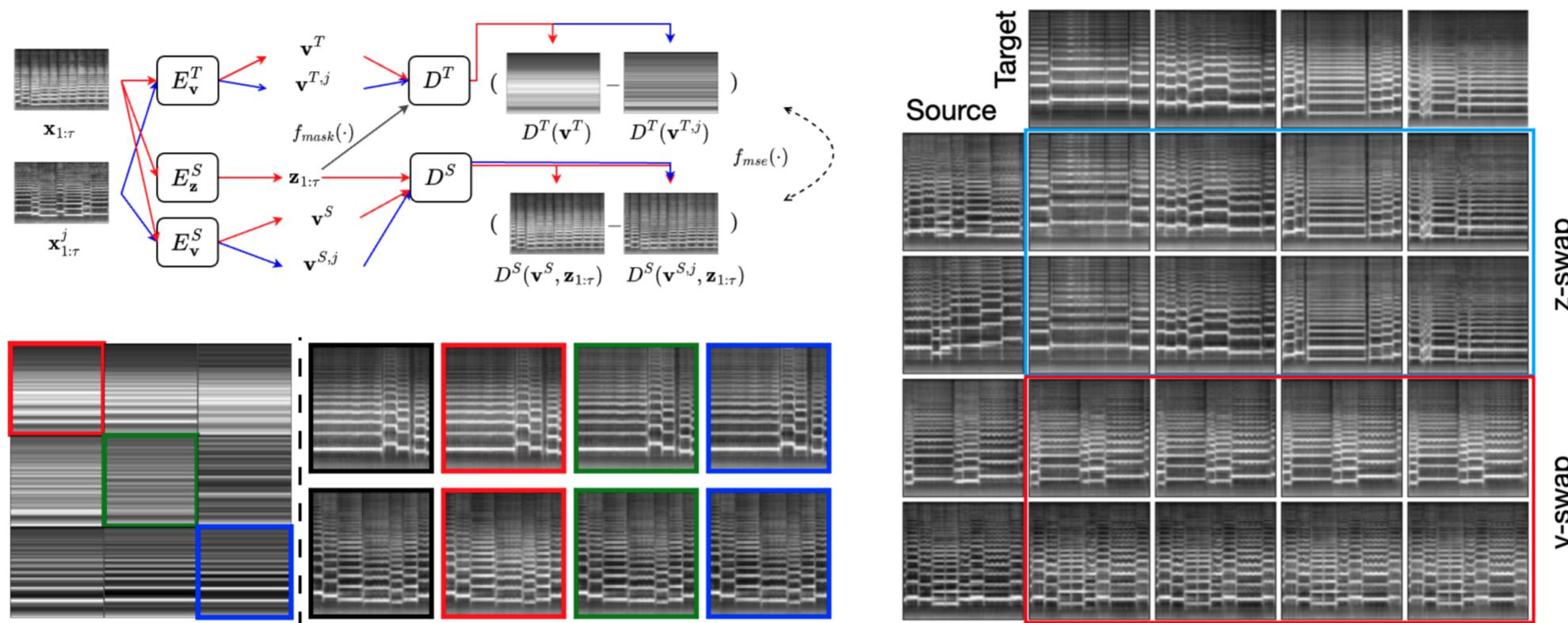
Question

What information should we extract from Mel-Spectrograms, and in what situations will it be useful?

Disentangled Representation Learning for Music Audio

Yin-Jyun Luo

2020



Finding

Exploiting general inductive biases, such as the difference in temporal resolution of factors of variation, facilitates learning semantically meaningful feature representations without supervision.

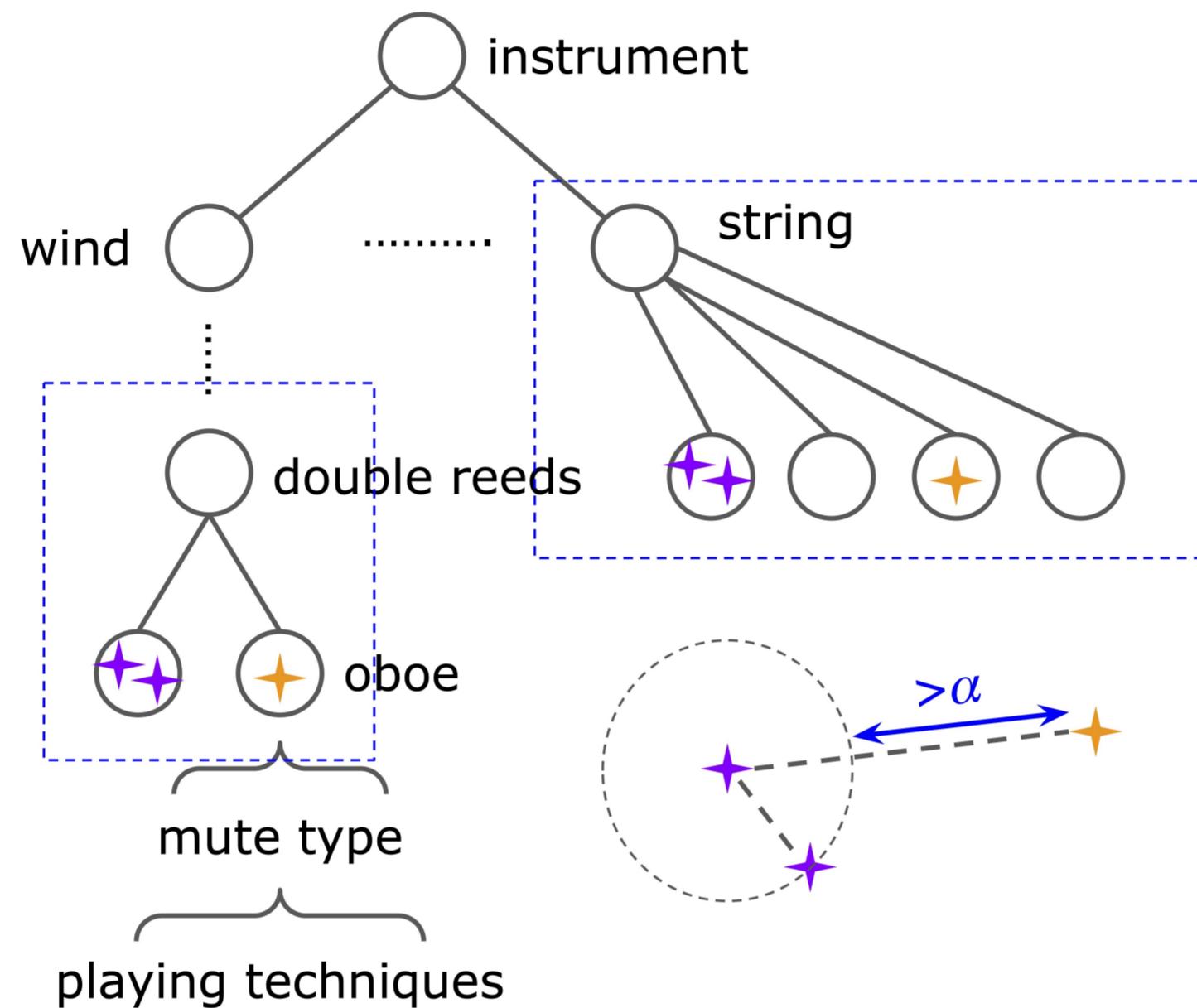
Question

Beyond controllable audio generation, does learning semantically meaningful or disentangled representations benefit classic MIR tasks such as music source separation? And how to apply the technique to more complex data such as music mixture?

Deep Timbre Metrics for Music and Audio AI

Haokun Tian

2023



Finding

Generalised triplet sampling helps to learn an embedding space according to a tree-structured label space.

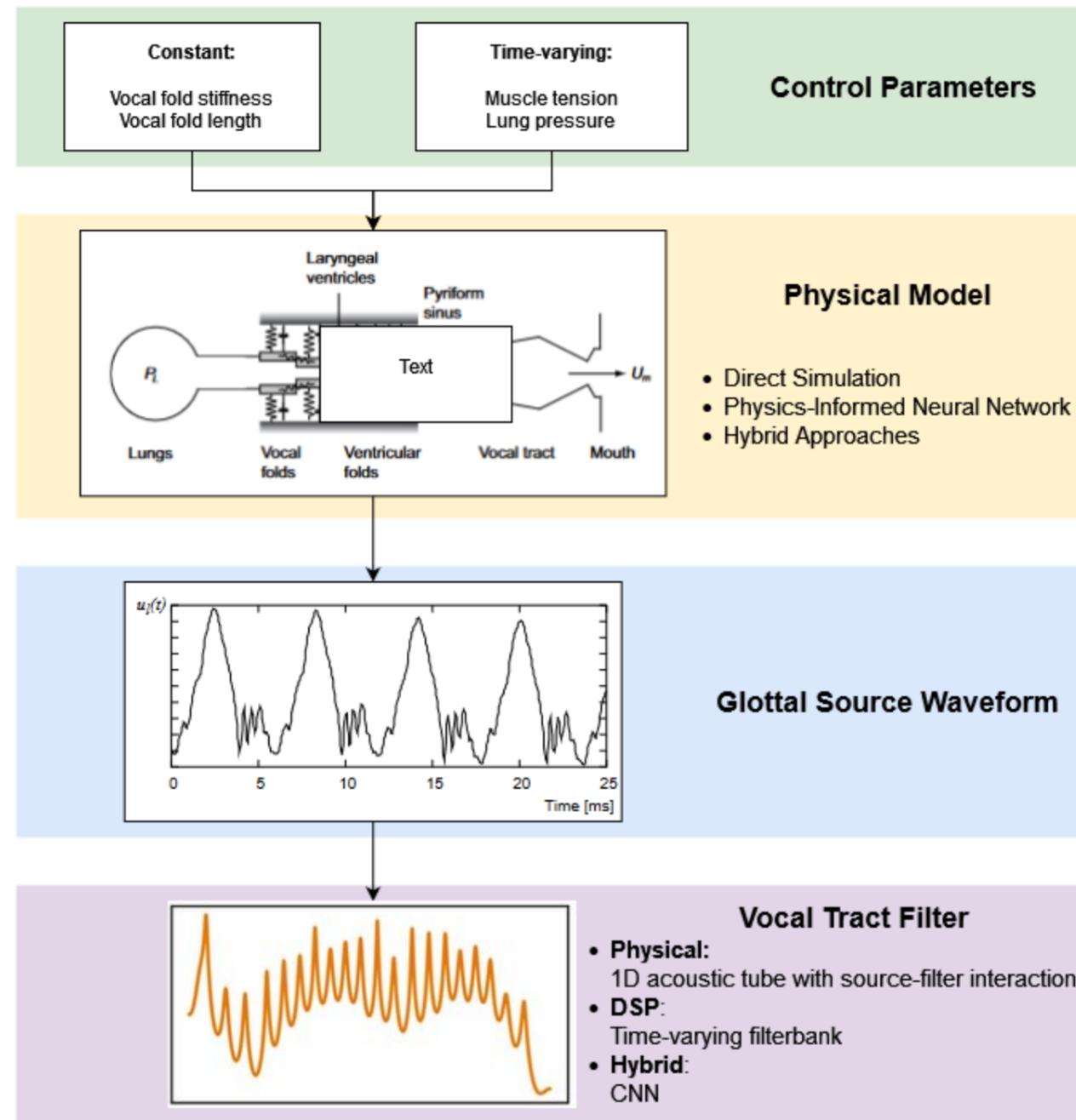
Question

How does it work with sampling methods like semi-hard mining? Does distance in hyperbolic space help learn such hierarchical relationships? How can the learned representations be used as control signals for neural audio synthesis?

Machine learning of physical models for voice synthesis

David Südholt

2022



Finding

More complex and expressive physical models of the vocal folds require non-linear solvers to be run at every glottal cycle to solve the collision of the vocal folds, putting them generally out of reach for real-time synthesis. We have found that we can train neural proxies for the solvers that make the models real-time feasible.

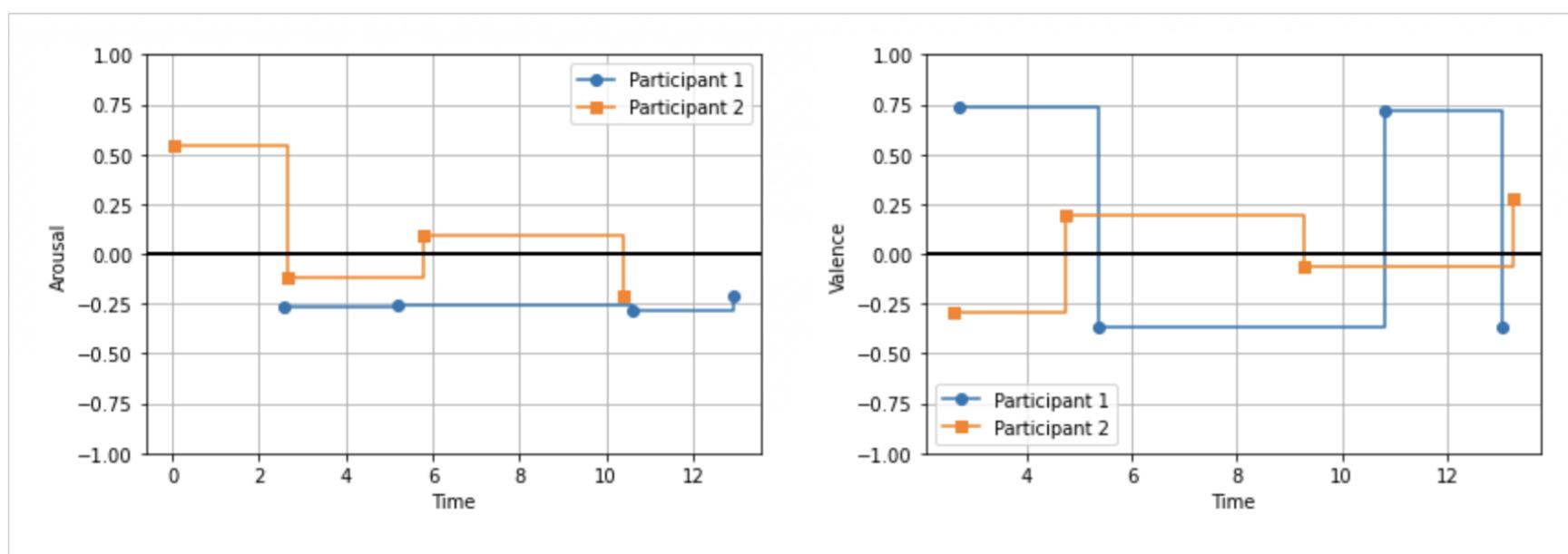
Question

How can we solve the sound matching problem for these physical glottal models? Given a physical model of the vocal folds and a real voice recording, how can we determine the control parameters such that the model will recreate the sound? Can an inductive bias given by a physical glottal model improve

Time-Based Mood Recognition in Film Music

Ruby Crocker

2021



Finding

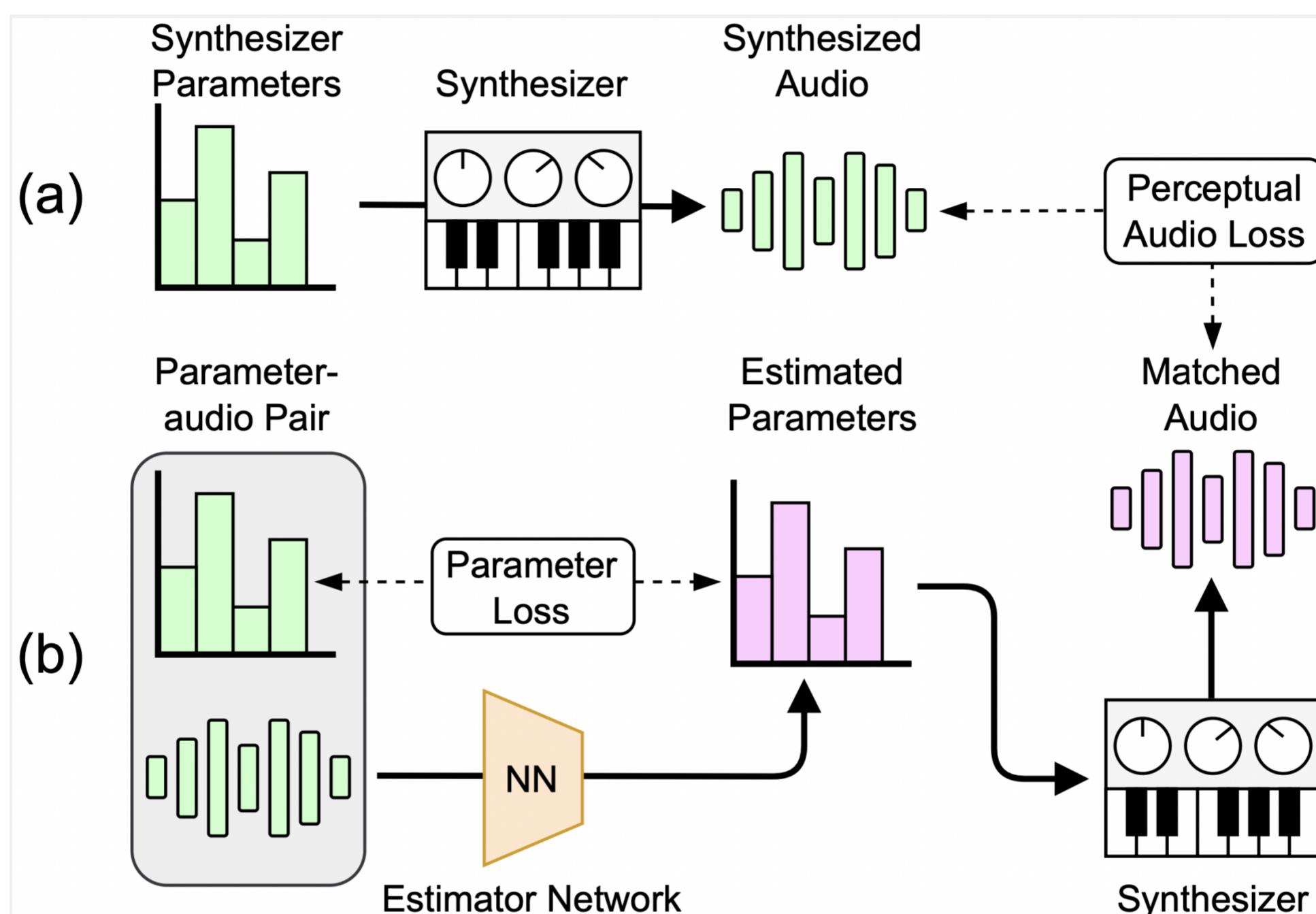
The Film Music Emotion Dataset (FME-24) was created to explore emotion perception in modern and contemporary film music, covering various genres and compositional styles. FME-24 introduces precise temporal control through comprehensive time stamps given by over 170 participants. Timing exhibits the strongest consensus, then arousal ratings and then valence.

Question

Future work involves extracting musical features to compare with arousal-valence values. Statistical analyses and additional annotations of the dataset will validate findings, and help draw definitive conclusions from sparse data. Interviews with composers will provide additional insights. Temporal modelling could help examine emotion changes.

Deep Learning for Time-varying Audio and Parameter Modulations

Christopher Mitcheltree
2022



Finding

Time-varying audio effects and synthesizers can be modeled effectively end-to-end via gradient descent by first extracting parameter modulation signals.

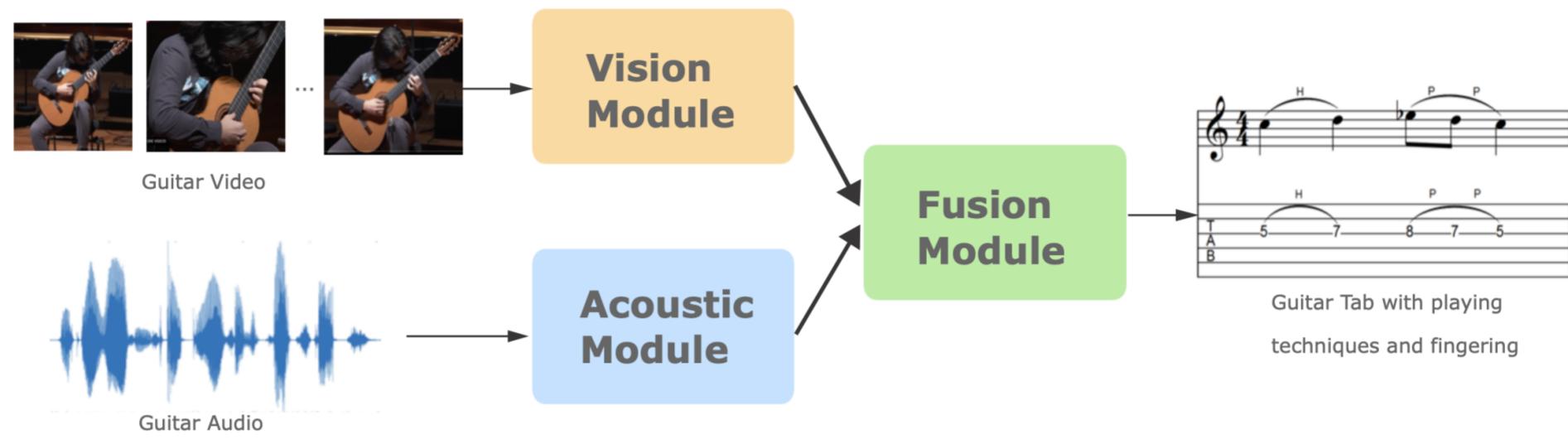
Question

Can a theoretical framework and differentiable loss functions be developed that improve the modeling of time-varying audio systems?

Multimodal Learning for Automatic Music Transcription

Zixun Guo

2023



Finding

There is a significant lack of open-source guitar dataset. Together with my collaborators, we have recently annotated 18 hours of guitar audio from Youtube. Our new dataset, named GAPS, has enabled us to achieve state-of-the-art performance in automatic guitar transcription (audio to midi).

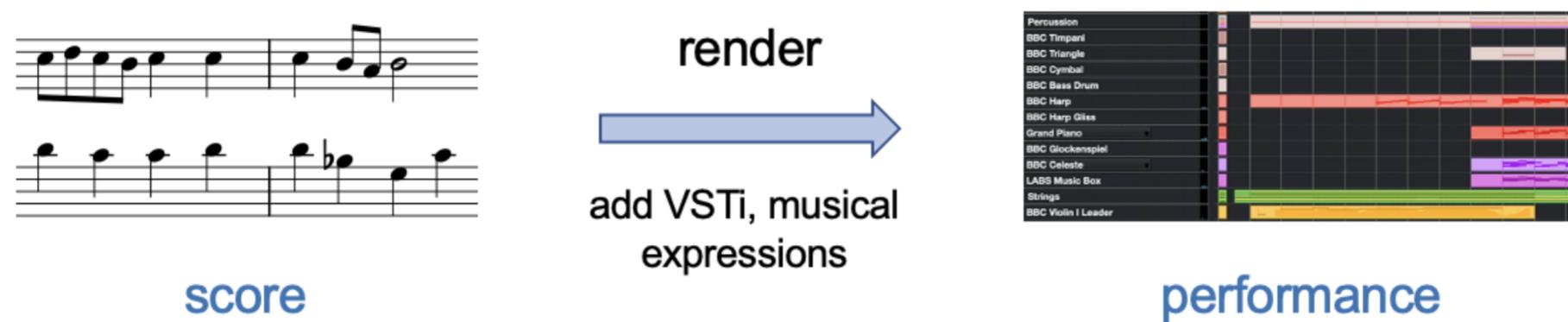
Question

In the future, I aim to fuse the video modality into the transcription network. This mimics how guitarists transcribe guitar audio: first figure out what notes to play from audio, then find out where and how to play these notes on the fretboard. As a result, we will be able to achieve tablature transcription.

Using Deep Learning to Help Render Orchestral Scores to Expressive Orchestral Performances

Yifan Xie

2023



Finding

There is no existing orchestral music dataset that includes both scores and their corresponding expressive performances.

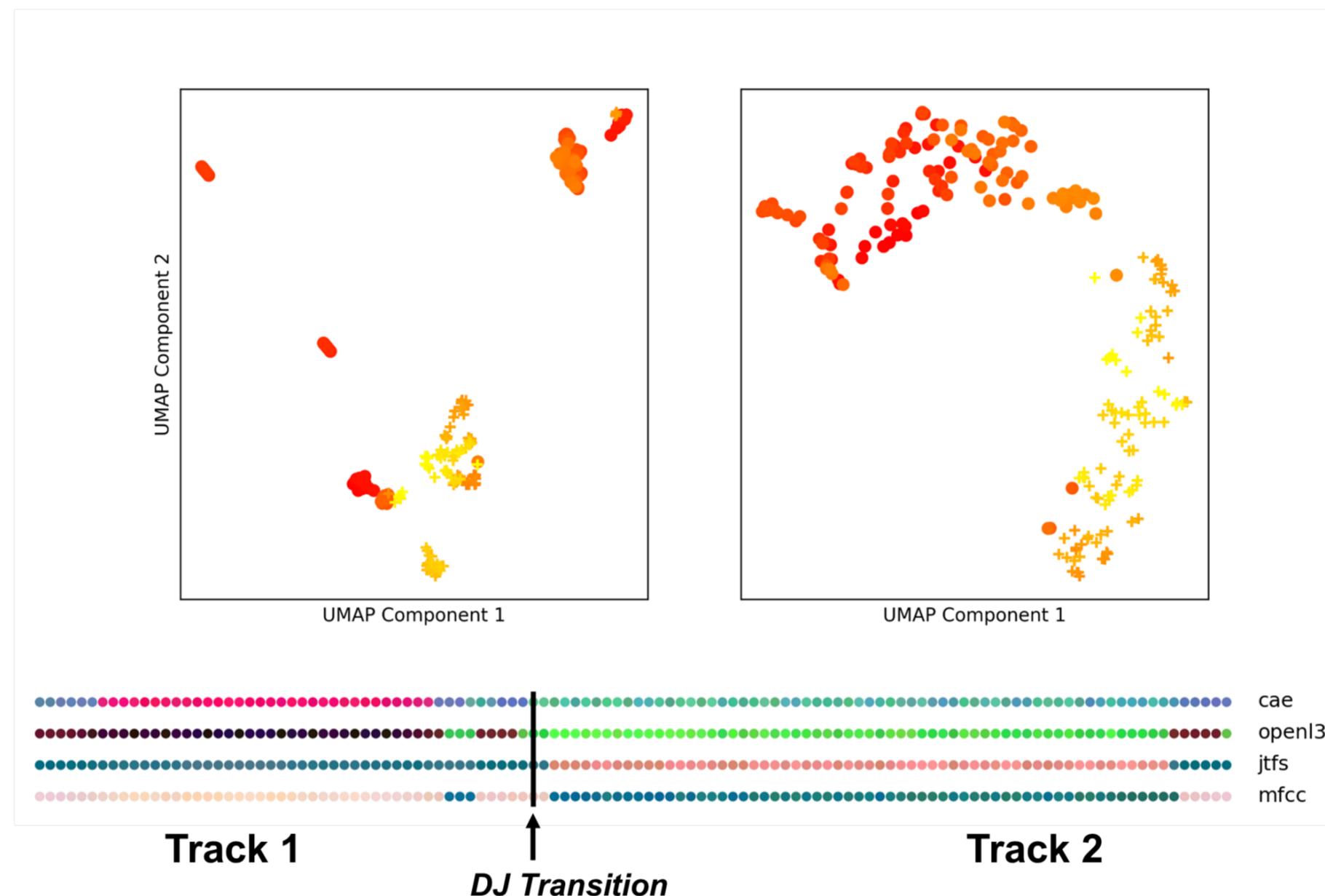
Question

We can find high-quality orchestra music data with individual tracks and aligned scores, ideal for transcribing expressive performances. However, these data are limited for DL. Most available orchestra music is mixed, without separate tracks and aligned scores, and often of poor quality. Is there a way to utilize the large low-quality mixed orchestra music?

Expressive Automatic DJ Mixing of Electronic Dance Music and Digital Audio Workstation Applications

Alexander Williams

2022



Finding

Several DL and signal processing-based audio representations (MFCC, JTFS, OpenL3, CAE) have been found to be valuable for structural analysis of EDM DJ mixes. Observations of recurring structures in the representation space allows us to derive heuristics for estimating track-to-track DJ transition properties and for making track and mix level comparisons.

Question

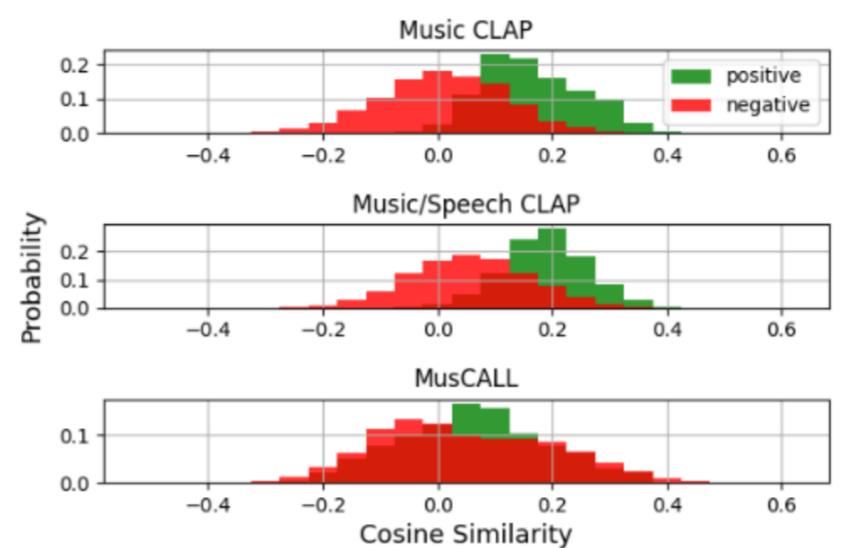
Additional heuristics and metrics are currently being considered to describe subjective DJ mixes in an objective way. What are the important differentiators of DJ mixes and how can these be computationally measured for concise comparison? How are such measures influenced by the actions of the DJ and the mixing controls available to them?

Supervisor(s): Stefan Lattner, Mathieu Barthet

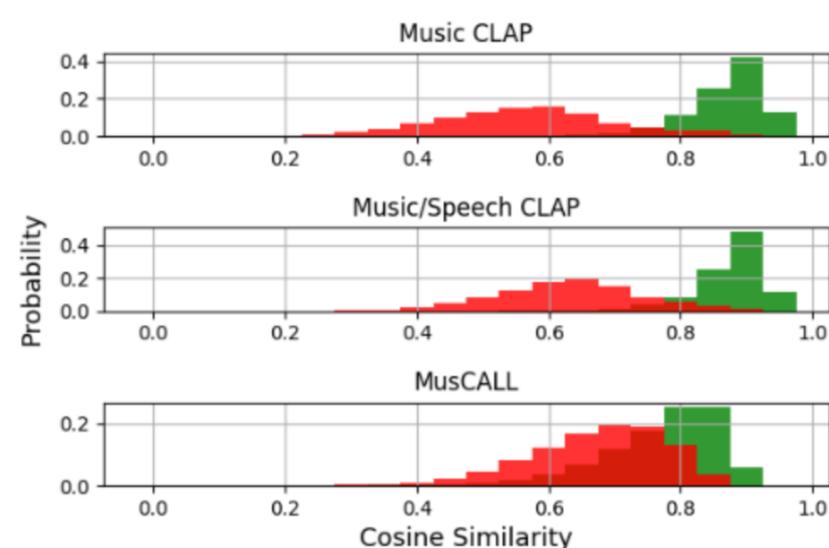
Multimodal User Adaptation

Yannis Vasilakis

2022



(a) Histogram of cosine similarity between TinySOL data and MusCALL prompts in joint audio-text space.



(b) Histogram of cosine similarity between TinySOL audio data and the mean of intra-class embeddings in joint audio-text space.

Finding

Alignment problems: 1) Use of captions as is without any augmentation leads to overfitting 2) BERTs inability to capture music related semantics compared to a taxonomy of concepts We proposed to apply stochastic augmentation (masking, paraphrasing, adding phrases) and musical fine-tuning The Audio encoder provides with informative representations.

Question

How can we effectively utilise external information for tag dependency for better alignment? How can we enforce balance between the modalities? Is intra-modal queries symmetric? How can we adapt the systems to a users concept definition with minimal annotations?

Character-based adaptive generative music for film and video games

Sara Cardinale

2021

Table 2.2: General guide to the association of NRO sequences and affect and/or situational scene elements (Lehman, 2018, 2014).

NRO Sequence	Affect/Situation
P	Duality
L(Maj)	Sorrow, loss (Murphy, 2014)
L(min)	Events of Mythic Significance
R	Authenticity, resolve, seriousness
N	Romantic encounters
S	Boundary spaces (e.g., Life & Death)
Far Fifth	Nature, ominous portent, depth, vulnerability, vastness
H	Intimating catastrophe, death, and transcendence
NR	Fantastical
RL	Wonderment, success
LP(Maj)	Amazement
PL(Maj)	Strong Amazement
LP(Maj/min)	Feeling of harmonic unnaturalness
PL/LP(Maj)	Positive otherness
LP/PL(min)	"Dark Side" (Bribitzer-Stull, 2012), antagonism
T6	Alien and inhuman nature encounters
PRPR(min)	Mortal threats, dangers
PRPR(Maj)	Cosmic spectacles
T6(Maj), PL(Maj), F(Maj)	Magic, awe
NRL	Suspense and mystery
RLRL	Heroism (Lydian)

Finding

How can NRT aid in developing tools for adaptive integration of leitmotifs into media compositions? I created a NRT framework to enable this integration, enhancing the narrative experience. This approach translates music theory into computational concepts emulating composers' processes and involving collaboration to understand their creative methods.

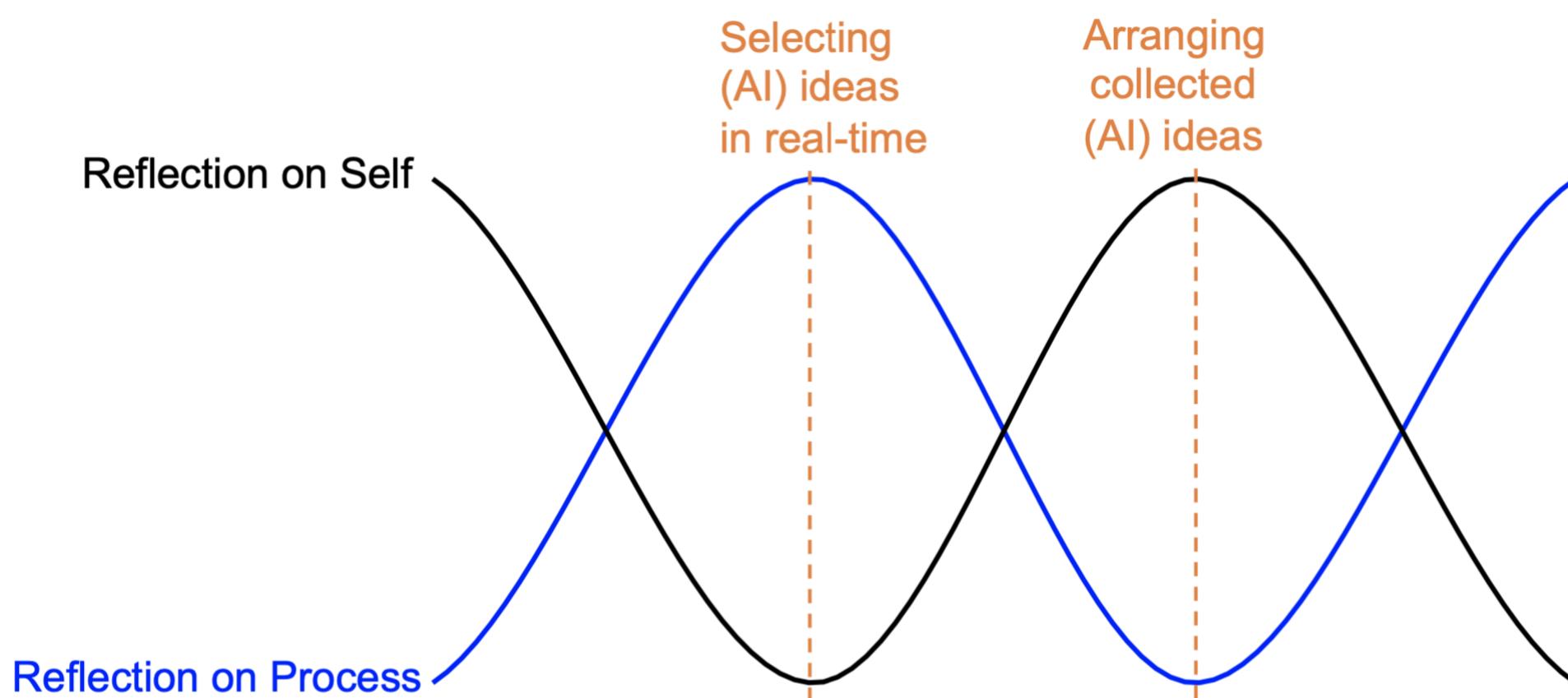
Question

How can AI music generation techniques be enhanced to capture the intricate musical nuances needed for narrative-driven media?

Reflection in AI-based Music Composition

Corey Ford

2020



Finding

Reflection is essential in music making, yet diverse practices with AIGC are underexplored. We examined how artist-researcher composers reflected using AIGC, documenting their reflections hourly through screenshots and first-person accounts. These were analysed with questionnaires, resulting in the speculative model pictured.

Question

An open area could be used to show more empirical proof for this model. I'd also be interested in further exploring the first-person reflective approach used in this paper - perhaps working on tutorials for ways to get more personal perspectives into more traditional HCI user study designs.