AI as mediator between composers, sound designers, and creative media producers

SEBASTIAN LÖBBERS, Centre for Digital Music, Queen Mary University of London, United Kingdom

MATHIEU BARTHET, Centre for Digital Music, Queen Mary University of London, United Kingdom

GYÖRGY FAZEKAS, Centre for Digital Music, Queen Mary University of London, United Kingdom

Musical professionals who produce material for non-musical stakeholders often face communication challenges in the early ideation stage. Expressing musical ideas can be difficult, especially when domain-specific vocabulary is lacking. This position paper proposes the use of artificial intelligence to facilitate communication between stakeholders and accelerate the consensus-building process. Rather than fully or partially automating the creative process, the aim is to give more time for creativity by reducing time spent on defining the expected outcome. To demonstrate this point, the paper discusses two application scenarios for interactive music systems that are based on the authors' research into gesture-to-sound mapping.

1 INTRODUCTION

Composers and sound designers often work with, or for, people who do not have expert musical knowledge. For example, a composer may work with a filmmaker or choreographer to create an original score serving a narrative, and a sound designer may be hired by a company to produce audio branding suiting their corporate identity. In both situations, the two sides are likely to propose and share ideas in an early exploratory stage to define more precisely what the sounds or music should express. A potential bottleneck lies in the imbalance in musical knowledge between the two parties. Individuals can have strong opinions and preferences for sound and music but when questioned about the underlying reasons, semantic descriptions differ between non-musicians and musicians [9]. Compared to non-experts, musicians tend to rely on technical terms instead of everyday language which can be hard to decode for a composer or sound designer. While technical language (e.g. "a piece in the Lydian mode") can express ideas to domain experts, it is opaque to other stakeholders. In some cases, non-verbal forms of communication like gestures or body movement might be preferred, e.g. when producers describe spatial audio [2]. In practice, these difficulties are overcome by sharing sound and music material that represents intended ideas and can be referenced in the ideation process. However, this introduces two new issues: (1) how is this material found? Stakeholders may use their own knowledge of existing musical compositions

("internal" catalogue of music), but this becomes less viable when searching for specific sounds; (2) despite having a reference in mind, a composer or sound designer may create material that ultimately does not suit the assignment. How can ideas be quickly adapted for prototyping, e.g. fitting composed music to a film or choreography, to reduce "wasted" work?

Recent advancements in artificial intelligence (AI) present new opportunities for music professionals, for example by generating audio from text-prompts [1], creating guitar tablatures in a specific genre [8], visualising sound intelligently [4] or retrieving audio from graphical sketches [3] and gestures [10]. However, AI research often focuses on the optimisation of quantifiable metrics like audio quality or classification accuracy, rather than exploring its role for application in the "real world". Reflecting on the authors' own work that uses AI for gesture-to-sound mappings, two potential applications are imagined to aid human-human music and sound ideation. The paper shortly describes these applications in the next section and discusses them in the context of the *Integrating AI in Human-Human Collaborative Ideation* workshop in Section 3.

2 CASE STUDIES

This section presents two interactive music systems, *Liquid Dance* and *SketchSynth*, and explores their capabilities for collaborative music and sound ideation in two hypothetical case studies that are based on the first author's experience as a professional composer and sound designer.
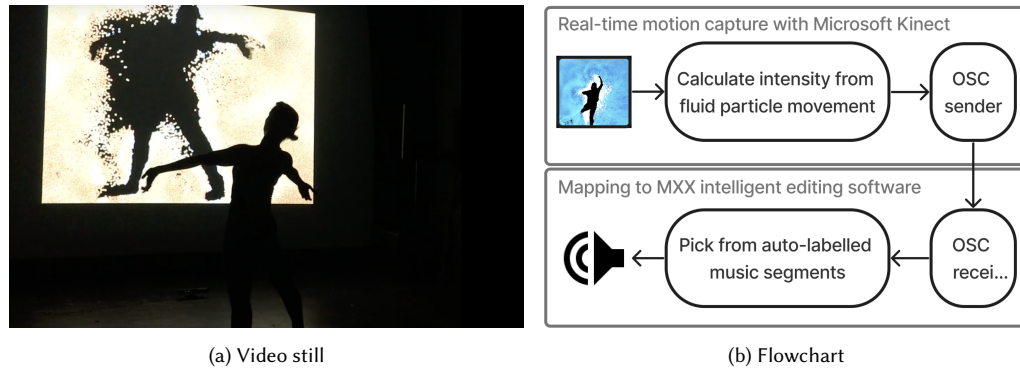
2.1 Liquid Dance

(a) Video still   (b) Flowchart

Fig. 1. *Liquid Dance* overview. Demonstration available at https://youtu.be/fYInCkaOZO8.

As illustrated in Figure 1, *Liquid Dance* uses motion capture and a fluid particle simulation to quantify the intensity of dance movements and automatically assemble musical snippets creating a suitable score in real-time. It was developed in collaboration with a choreographer for the music AI startup MXX[1] to showcase their intelligent music editing software with an interactive performance. At a later stage, an unrelated composition project for dance sparked the idea of using *Liquid Dance* for rapid testing of musical ideas for choreography. This is described in the following scenario:

A choreographer and music composer have chosen several reference music pieces and are now interested in testing their

---

[1] https://www.mxx.ai/

compatibility with a choreographed dance performance. To achieve this, they plan on using a system that will enable them to dance alongside the music pieces, automatically selecting segments that best match the intensity of the dance movements. By using this method, they hope to gain a better understanding of how the selected music will work in tandem with the choreography.
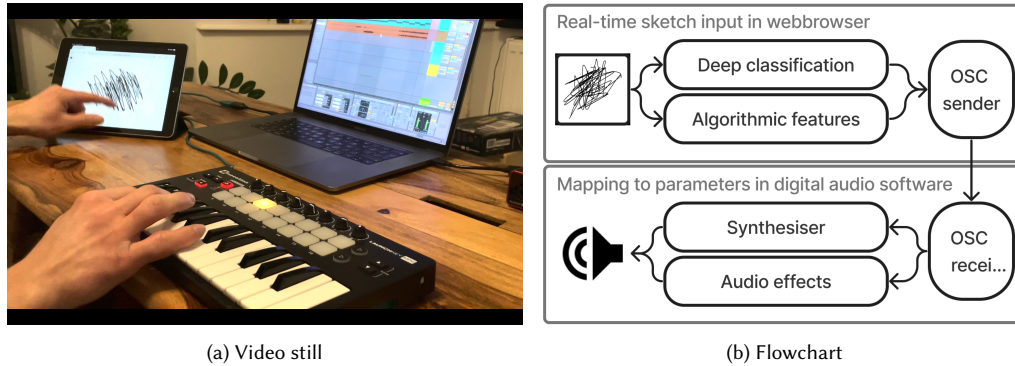
2.2 SketchSynth

(a) Video still   (b) Flowchart

Fig. 2. *SketchSynth* overview. Demonstration available at https://youtu.be/arSFt3iBAUM.

SketchSynth is based on the authors' research into cross-modal perception of musical timbre [5–7]. It employs AI and algorithmic methods to predict sound from a graphical sketch input as illustrated in Figure 2. The following scenario describes a potential application in industrial sound design:

A sound designer and product manager are looking for suitable sounds to incorporate into their product. To facilitate this process, they plan on using SketchSynth, which enables non-experts to express their ideas and explore the sonic space through an intuitive mode of interaction. Both parties can literally draw out ideas until reaching a consensus from which the sound designer can elaborate in greater detail.

3   DISCUSSION AND CONCLUSION

In both case studies AI acts as a translator between non-verbal expression and sound or music creation. While an expert human actor is likely capable of creating suitable content from these expressions, a machine can be trained to perform this task much faster in real-time. This gives stakeholders the opportunity to test initial ideas much faster, focus on their interpersonal relationships and leave more time for the creative process that is arguably done best by a human professional. In this context, AI is seen strictly as a tool that is expected to reliably perform a complex but well-defined task. *Liquid Dance* uses a robust, but very specific setup that might only be useful for a narrow type of projects; *SketchSynth* tries to find universal sketch-to-sound mappings that might be too broad to communicate sound in sufficient detail. As user needs are likely to change between or even during projects, the scope of AI tools might have to change with them. This raises the question of not only how AI is used during human-human ideation, but also how it is trained and set up. One approach is to choose a model that was trained on a diverse dataset to solve a large number of projects. However, this might hand over too much agency leading to unexpected behaviour not suitable for the task at hand. Therefore, a user might want to adjust a system in an initial "briefing" stage to prepare for the

upcoming task. Using the case studies in this paper as an example: for *SketchSynth* this could mean confining the space of possible sounds, e.g. to mechanic sounds for an electronic car project or digital blips for a mobile game; or for *Liquid Dance*, optimising the interface for a specific dance style, e.g. ballet or contemporary, and conditioning the AI-driven music segmentation on a specific genre like classical or electronic music. Ultimately, it remains for the user to define how much agency they want their system to display during the ideation process. In conclusion, recent AI-driven music research can find valuable application in aiding human-human ideation for example as a translator for non-verbal descriptions of sound and music. However, AI tools may have to be adjustable to user needs and project requirements to avoid unpredictable or inadequate behaviour.

4 ACKNOWLEDGMENTS

REFERENCES

[1] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. MusicLM: Generating Music From Text. *arXiv preprint arXiv:2301.11325* (2023).

[2] Thomas Deacon, Patrick Healey, and Mathieu Barthet. 2022. "It's cleaner, definitely": Collaborative Process in Audio Production. *Computer Supported Cooperative Work (CSCW)* (2022), 1–31.

[3] Lars Engeln, Nhat Long Le, Matthew McGinity, and Rainer Groh. 2021. Similarity Analysis of Visual Sketch-based Search for Sounds. In *Proceedings of Audio Mostly 2021*. Association for Computing Machinery, Trento, Italy, 101–108. https://doi.org/10.1145/3478384.3478423

[4] Leandro Garber, T Ciccola, and JC Amusategui. 2021. AudioStellar, an open source corpus-based musical instrument for latent sound structure discovery and sonic experimentation. In *Proceedings of International Computer Music Conference*. Michigan Publishing Services, Santiago, Chile, 62–67. https://hdl.handle.net/2027/fulcrum.t435gg568

[5] Sebastian Löbbers, Mathieu Barthet, and György Fazekas. 2021. Sketching sounds: an exploratory study on sound-shape associations. In *Proceedings of International Computer Music Conference*. Michigan Publishing Services, Santiago, Chile, 299–304. https://hdl.handle.net/2027/fulcrum.t435gg568

[6] Sebastian Löbbers and György Fazekas. 2022. Seeing Sounds, Hearing Shapes: a gamified study to evaluate sound-sketches. In *Proceedings International Computer Music Conference*. Michigan Publishing Services, Limerick, Ireland, 174–179. https://hdl.handle.net/2027/fulcrum.nk322g689

[7] Sebastian Löbbers and György Fazekas. 2023. SketchSynth: cross-modal control of sound synthesis. *EvoMUSART preprint* (2023). https://sebastianlobbers.com/static/76293e97d37a3d69941f2f1b047cec7e/SketchSynth_cross-modal_control_of_sound_synthesis.pdf

[8] Pedro Sarmento, Adarsh Kumar, Yu-Hua Chen, CJ Carr, Zack Zukowski, and Mathieu Barthet. 2023. GTR-CTRL: Instrument and Genre Conditioning for Guitar-Focused Music Generation with Transformers. *arXiv preprint arXiv:2302.05393* (2023).

[9] Simin Yang, Courtney Nicole Reed, Elaine Chew, and Mathieu Barthet. 2021. Examining emotion perception agreement in live music performance. *IEEE transactions on affective computing* (2021).

[10] Michael Zbyszyński, Balandino Di Donato, Federico Ghelli Visi, and Atau Tanaka. 2021. Gesture-Timbre Space: Multidimensional Feature Mapping Using Machine Learning. In *Proceedings of International Symposium on Computer Music Multidisciplinary Research*. Springer, Tokyo, Japan, 600–622. https://doi.org/10.1007/978-3-030-70210-6_39