On the Development and Practice of AI Technology for Contemporary Popular Music Production

EMMANUEL DERUTY

MAARTEN GRACHTEN (iD)

STEFAN LATTNER (iD)

JAVIER NISTAL

CYRAN AOUAMEUR

Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Although the use of AI technology for music production is still in its infancy, it has the potential to make a lasting impact on the way we produce music. In this paper we focus on the design and use of AI music tools for the production of contemporary Popular Music, in particular genres involving studio technology as part of the creative process. First we discuss how music production practices associated with those genres can differ significantly from traditional views of how a musical work is created, and how this affects AI music technology. We argue that—given the role of symbolic representations in this context, as well as the integration of composition activities with editing and mixing—audio-based AI tools are better suited to support the artist's creative workflow than purely piano-roll/MIDI-based tools. Then we give a report of collaborations with professional artists, in which we look at how various AI tools are used in practice to produce music. We identify usage patterns as well as issues and challenges that arise in practical use of the tools. Based on this we formulate some recommendations and validation criteria for the development of AI technology for contemporary Popular Music.

CORRESPONDING AUTHOR:

Emmanuel Deruty

Sony Computer Science Laboratories (CSL) Paris, 6, rue Amyot, 75005 Paris, France

emmanuel.deruty@sony.com

1 INTRODUCTION

The practice of making music has since long involved instruments and—more generally—technology. As such, advances in technology can have a profoundly transformative effect on music making practice (which we will refer to as music production in this paper). This effect is witnessed for by the roles analog and later digital electronics have played in music, in the form of synthesizers, samplers, sound effect gear, and digital audio workstations.

More recently artificial intelligence (AI) has increasingly found application in music production. Although arguably the use of AI technology for music production is still largely experimental, it does not seem far-fetched that this technology will eventually have a lasting effect on music production.

The versatility of AI technology allows for a wide variety of applications in music, which has only just begun to be explored. Current applications include synthesis of individual musical sounds (Engel et al., 2017; Aouameur et al., 2019), "musical inpainting" (Hadjeres et al., 2017; Bazin et al., 2020), interpolation between musical material (Roberts et al., 2018), or synthesis of complete multi-instrumental audio tracks (Dhariwal et al., 2020).

A variety of AI music production services has emerged in recent years, with different focus areas.[1] These services are often offered on dedicated platforms rather than providing integration with an artist's existing working environment, and as such are not yet a widely used commodity for artists. Understanding, installing, running, and using AI music tools independently is not trivial, as it tends to require technical expertise and familiarity with the basic notions of machine learning. This is also witnessed in the *AI Song Contest* (Huang et al., 2020), a contest in which musicians and AI engineers often team-up in order to jointly produce music with AI tools.

As AI technology is starting to find its way into music production practice, there are several underlying questions that deserve attention. For example, which (sub)tasks of music production should AI tools address to be most useful and relevant for musicians? A further question concerns the possibilities to exert control over the output by the musician, and how the user interface and more generally the affordances of the model should be shaped (Gibson, 1977). Lastly, an important consideration is how AI tools can become part of the music production process, in particular the musician's creative workflow. Answers to these broad questions will necessarily be subjective, and vary depending on individual and style-specific musical practices.

As a music AI research lab specializing in music AI tools for commercial artists,[2] we aim to address the above questions in the context of music production practices of contemporary Popular Music (CPM). Popular Music is a very broad category typically used in contrast to Folk Music and Art Music (Tagg, 1982; Middleton, 1990; Mazzanti, 2019). An important distinguishing feature (among others) of Popular Music in this characterization is the use of recorded sound as the main mode of transmission (storage/distribution) (Tagg, 1982, page 42). This sets it apart from Folk Music (where oral transmission is predominant) and Art Music (where music notation is the main mode of transmission). Frith (2004) lists further features of Popular Music, like the prevalence of commercial/industrial interests, enjoyment/entertainment as the main purpose of the music, and strong ties with mass media like cinema, radio, and television.

The adjective "contemporary" does not seem to have a universal denotation in conjunction with Popular Music beyond "current", or "present-time". The Berklee College of Music[3] mentions a focus on technological innovation and cross-pollination across cultures and genres as characteristics of contemporary music—characteristics that have also been attributed to Popular Music in general (Frith, 2004; Mazzanti, 2019). As such, CPM genres include post-rock, rap/hip-hop, electronica, as well as non-Western contemporary genres such as K-pop, J-pop, Bollywood, and reggaeton.

With this paper we intend to foster the development of AI technology geared towards music production in these genres. To this end we make two main contributions. Firstly, we present a discussion of how music production practices in these genres have diverged in some ways from a more established view of how music is produced (Section 3). We emphasize these differences because, as we will argue, they have implications for the design of AI music tools, but do not seem to be widely acknowledged in literature on AI music generation research.

Secondly, we give a qualitative report of collaborations of our music AI research lab with professional artists/acts producing music in several CPM genres such as urban, ambient, experimental, neo-classical, trance, and mainstream for commercials. In these collaborations the artists experiment with music AI tools developed at our lab in their creative process. We provide a detailed account of the creative process by one of the artists to illustrate the different roles the tools may play in the process, and we present and discuss feedback provided by the artists, linking aspects of their work to issues discussed in Section 3.

In Section 5 we summarize our findings in the form of recommendations for developing music AI tools in the context of contemporary music genres, as well as a list of validation criteria. Conclusions and future work are presented in Section 6. We start with a brief discussion of the possible music production use cases in which AI technology may play a role, and delimit

the class of use cases on which we focus our attention (Section 2).

2 MUSIC PRODUCTION USE CASES

Music production in the general sense as the practice of creating musical works is a very diverse set of activities, and the ways in which AI tools can be used in music production are equally diverse. Without aiming to be exhaustive, we distinguish three broad types of music production contexts for AI tools. Firstly there is music composition in its traditional form, as the creation of a symbolic description of a musical work, be it as MIDI, Western staff notation, or some other symbolic format. Although any type of music may be created through this form of music production, the use of AI tools is somewhat limited to the genres for which substantial data is available in symbolic formats. Examples of AI tools designed for music composition are *Magenta Studio* (Roberts et al., 2019), *DeepBach* (Hadjeres et al., 2017), and the *Piano Inpainting Application* (Hadjeres and Crestel, 2021).

Another type of music production is live performance and improvisation. This includes instrumental performances with or without a pre-determined score or schema to structure the performance. AI tools in this context tend to be focused on responsiveness and real-time interaction. A different type of live music production takes place in clubs and dance venues, where DJs mix electronic dance music, often mixing existing works with musical elements that are added live, and adapting playlists in response to the mood of the audience. See Knotts and Collins (2021) for an overview.

A third type of context is what can be broadly called in-studio composition. Activity in this context centers around the *digital audio workstation* (DAW), a piece of software in which the recording, editing, placement, and processing of recorded material, as well as the creation and synthesis of any symbolic representations is coordinated. AI tools for the first context can be useful in this context too, but a crucial difference from the first context is that the activity of writing music notation is combined with sound recording, editing, and mixing activities, and these activities do not necessarily take place in a fixed order. Knotts and Collins (2021) give a historical account of AI tools in this context.

It is this last context that we focus on when discussing AI for music production in this paper. In the next section we give a brief overview of the development of the in-studio composition practice, which has become a standard form of CPM music production in the past decades (Bell, 2018). We focus on the role of symbolic representations and the different stages of music production. We then discuss some implications for AI technology to support this practice.

3 AI TECHNOLOGY FOR IN-STUDIO COMPOSITION

3.1 MUSIC PRODUCTION AS IN-STUDIO COMPOSITION

Western music production across the 18th and 19th centuries can be roughly schematized as follows: a composer produces a musical work in the form of music notation, after which one or more performers create the sounds the composer has notated in front of an audience. This music production practice was thus a linear process: *composition → performance → consumption,* where the roles of composer and performer/musician were typically separated. The modern sound recording studio elongated this linear chain with a number of steps: *composition → performance → recording → editing → mixing → mastering → consumption.* The recording, mixing, and mastering tasks were carried out by sound-engineers and overseen by record-producers.

The commodification of sound recording and processing technology since the 1980s has given amateur as well as professional artists permanent and low-barrier access to studio technology. Moreover, synthesizers, drum computers, sample libraries, and virtual instruments are nowadays widely used to produce sound, either in addition to or instead of recorded performances. These developments have blurred the distinction between the roles of composer, musician, sound engineer, and record producer (Burgess, 2013), as well as the classification of these roles as either art-oriented or craft-oriented (Lashua and Thompson, 2016). In many contemporary Popular Music genres a song or album is typically the work of a creative collective (Hennion, 1983)—or even a single individual—taking charge of all of these aspects of the music production process.

The notion of in-studio composition has obsoleted the strict linearity of the music production process: a musical work is no longer composed in a symbolic form as a whole before creating a sonic realization in the studio. Instead, compositional activities are often intertwined with sound editing and sound design activities.

Another consequence is that—compared to the aforementioned music practices where the roles of composer and performer are more strictly separated—the need for music notation as a means of transmitting information from composer to performer has diminished. Music notation as "the organization of instructions for the creation of sound" (Jones, 1992) is dispensable when the sound is readily available in the form of recordings to be listened to, distributed, and re-used. As a result, many CPM artists—and in fact some of the professional artists we collaborate with—do not read music notation. They have little use for it in their music production workflow.

This is not to say that symbolic representation formats do not play a vital role in CPM music production—to the contrary, MIDI is widely used in CPM music production.

Rather, we wish to highlight the distinction between the use of symbolic representations such as MIDI to convey the "essence" of the music, and the use of MIDI as a convenient means to organize and trigger sonic events.

In many musical genres (including perhaps some that can be regarded as CPM) the conventions of the genre warrant strong assumptions about the mapping from symbolic representations to sonic events and shapes (in other words, what the music sounds like). For example, the sonic realization of a MIDI file that represents a classical piano performance is relatively predictable, and as such justifies the use of MIDI data as conveying the essence of the music. Moreover, MIDI information is much sparser than audio, and explicitly represents the music in terms of meaningful terms like notes, beats, meter, tempo, and key. This provides a convenient and often very effective starting point for analysis and modelling of musical information.

In other genres however, notably those that substantially involve electronically designed sounds and samples, the mapping between MIDI data and its sonic realization is much looser. *Synth pads* for example are deliberately designed sounds to provide an ambient sonic background, and are ubiquitous in a wide range of music nowadays. Although these sounds are usually triggered in the form of MIDI notes, individual synth pad sounds may be sustained for tens of seconds, and typically have a salient temporal development in terms of timbre, dynamics, tuning, and/or localization that is not conveyed in a piano-roll representation.[4]

It can of course be argued that such genres have their own sonic aesthetics and conventions that constrain the sonic manifestations of the music, and are thus amenable in principle to symbolic representation and analysis. While this is true, the vocabulary needed for this type of analysis is still largely undeveloped, and is the subject of works like those of Moore and Martin (2018) and Moylan (2020).

Although the traditional categories of music analysis are very effective means of understanding music in some musical styles, it is important to realize that they are not always pertinent, or they may not apply cleanly. For example, while melody and harmony are essential notions in Western classical music, jazz, and more traditional forms of rock and pop, they play a much less prominent role in rap and many electronic genres.

Furthermore the pitched quality of the 808 kick drum[5] (named after Roland's early and emblematic TR-808 drum machine)—a widely used percussive element in many CPM genres—requires careful coordination with other pitched elements in the song, just like pitched percussion (e.g. timpani) in orchestral music. In contrast to performed music, where the tuning is done before the performance, pitched percussion in CPM is commonly tuned by way of post-production. Pitched kick drums, especially given their prominent use in some CPM genres, call into question the common view of rhythm

and harmony as two largely orthogonal aspects of the music.

3.2 IMPLICATIONS FOR AI TECHNOLOGY

The above demonstrates how traditional views of music production, specifically the notion of MIDI as conveying the essence of the music, and the separation of composition from other music production activities (recording, editing, and mixing), become problematic in the context of in-studio composition. In research concerning deep learning for music generation such views still seem predominant however. For example, Briot et al. (2020) justify the focus on symbolic representations in their review of deep learning techniques for music generation as follows:

> [W]e believe that the essence of music (as opposed to sound) is in the compositional process, which is exposed via symbolic representations (like musical scores or lead sheets) and is subject to analysis (e.g., harmonic analysis) (Briot et al., 2020, Sec. 4.2)

Another recent survey by Ji et al. (2020) classifies deep learning music generation approaches along the linear view of music production, where information is transmitted from the composer via performer and instrument to the listener. In this perspective composition takes place at the start of the chain on a purely symbolic level, whereas the sonic manifestation of the symbolic score takes place at the end, and is regarded as a task that is largely separate from the composition. Moreover, sound editing, mixing and mastering—crucial constituents of in-studio composition—are not considered at all in the survey.

Nevertheless, awareness of the limitations of such implicit assumptions about the nature of music production has been raised in recent years, and explicit efforts have been made to overcome it, such as the *CompMusic* research project (Serra, 2012), aimed at developing MIR methods for non-Western music traditions such as Hindustani, Carnatic, and Beijing Opera. Gioti (2021) also observes a discrepancy between the note/pitch-centric design of tools such as *MusicVAE* (Roberts et al., 2018) and *Nsynth* (Engel et al., 2017) and contemporary art aesthetics. This diminishes the relevance of the tools for contemporary art music genres like electro-acoustic music, she argues, and calls for increased collaboration between developers and contemporary art music composers to overcome this discrepancy.

It should be emphasized that none of the arguments raised above question the validity and utility of existing AI music tools in themselves. Indeed piano-roll-based modeling approaches like those of Chu et al. (2016); Yang et al. (2017); Payne (2019); Hadjeres et al. (2017); Hadjeres and Crestel (2021); Roberts et al. (2019) address relevant problems, such as generating music with coherent long-range structure, style transfer, style-

specific continuation or in-painting, and often do so with impressive results. Instead, we intend to demonstrate that in-studio composition in recent CPM genres involves a wider variety of use cases that cannot be adequately addressed with piano-roll-based approaches.

How can AI technology be designed to support a broader range of in-studio composition activities and acknowledge the interplay between these activities? Although this question is complex and entails more specific questions like those raised in the introduction, a partial but straight-forward answer may be: design AI tools that are able to deal with all the types of information that CPM artists work with, most notably audio. This applies in particular to the broad category of AI tools that produce outputs conditional on some music inputs. It allows artists to use such AI tools to condition on arbitrary material they have in their DAW project, not just the subset that is encoded as MIDI. A further benefit of audio based AI tools is that they are exposed to the sonic qualities that are not represented by MIDI representations but are musically relevant, such as the pitched quality of the kick drum, or the temporal evolution of a synth pad sound, as discussed before. Moreover, it enables the integration of composition with other stages of music production, in particular mixing. For example, a drum track created by an AI model conditioned on some harmonic tracks can potentially be modulated by changing the mix of the different tracks that go into the model, such as varying the relative gain of the input tracks, the equalization, reverb, or other effects. In a way this allows the artist to use the familiar controls of their DAW to interact with the AI tool beyond the explicit controls of the tool itself.

Of course there are valid reasons for refraining from audio representations when developing AI tools, as mentioned above. In particular it can require significantly more computational resources to store and process audio. A further challenge is capturing long-range dependencies due to the high data rate, and—in generative scenarios—to produce outputs that are coherent both at short and at long time scales. Systems such as *Jukebox* (Dhariwal et al., 2020) show that such challenges are not insurmountable however. Furthermore, conditional generation tasks can also become easier because of the density of audio as the conditioning signal compared to symbolic data.

In conclusion, we believe it is worthwhile to pursue audio-based AI tools, because they can provide CPM artists with a richer spectrum of creative opportunities than AI tools that work exclusively with symbolic notation.

4 USING AI TOOLS IN PRACTICE

In line with the principles of *participatory design* (Muller and Kuhn, 1993), the development of AI tools for CPM is

arguably most effective when it is a joint effort between engineers and musicians. Participatory design may involve end-users at any stage during the development of a product or tool. In our case the collaboration with artists does not usually take place at very early stages of design. One reason for this is that our music AI research lab team includes members with expertise in music production, who give early guidance and feedback on the conceptualization of the tool. Another reason is that the development of AI tools usually takes an engineering-heavy initial phase of experimentation to explore the feasibility of ideas, making smooth interaction with artists more challenging (which is not to say that artist involvement is undesirable at this stage).

In this section we report on experiences from collaborations of our lab with six professional music artists/acts, listed in *Table 1*. These collaborations both allow artists to explore novel ways of producing music, and they allow us make the AI tools more useful and interesting as part of the creative process.

It is important to note that at the time the collaborations with artists started there was no detailed plan to carry out a systematic study of the overall results beyond gathering general feedback regarding the utility and value of the AI prototypes. Although the participating artists obviously have an intrinsic interest in the collaboration (they typically are keen to experiment with new technologies), the collaboration also implies a considerable time investment with uncertain outcomes, which may or may not be acceptable to artists from a professional point of view (the difficulty of engaging creative professionals in studies has also been reported by Csikszentmihalyi (1997); Bennett (2012)). Rather than volunteering as research subjects, the artists agreed to produce works with the prototypes as contractors, and provide documentation of their workflows and experiences. They consent to the use of the material they produce for research and/or promotional purposes.

The results obtained from the collaborations (interviews, songs, workflow diagrams, see Section 4.2) were subjected to a thematic analysis (Braun and Clarke, 2006), similar to the approach taken by Huang et al.

Artist/Act	Genre (s)	Time-span
Niro	Rap/Urban	6 months
Twenty9	Beats/Urban	6 months
Hyper Music	Mainstream for commercial purposes	18 months
Uèle Lamore	Ambient, Experimental, Neo-classical	6 months
Whim Therapy	Rock/Electronic	6 months
Donn Healy	Trance	18 months

Table 1: Collaborating artists/acts.

(2020). The overall question guiding the analysis is: how do artists use the available tools?

The thematic analysis presented in Section 4.3 is not strictly inductive. It was primed in part by the themes identified by Huang et al. (2020) and Clark et al. (2018). Other themes, notably those in Section 4.3.3, emerged from our own use of the tools prior to the collaboration, and were reinforced by the results of the collaborations.

We begin by giving a brief overview of the tools used by the artists (Section 4.1), followed by a general description of the way the collaborations take place (Section 4.2). In Section 4.3 we present and discuss some of the artists's experiences during the collaboration.

4.1 TOOLS USED

The AI tools provided to the artists have been recently developed at our lab, and are generally prototypes in the form of either standalone applications, VST plug-ins for digital audio workstations (DAW), or servers accessible through a web-interface. They cover different aspects of the music production process, ranging from sound design to mixing, equalization, and the generation of melodic and rhythmic material. In line with our conclusions in Section 3.2, the tools work with audio representations as input data. Depending on the tool the output may be audio, MIDI, or both. The tools have been presented in more detail in prior publications, so here we provide only a brief introduction of the tools.

Notono: An interactive tool for generating instrumental one-shots (Bazin et al., 2020). It uses a variational autoencoder (VAE) architecture (Kingma and Welling, 2014) that operates on spectrograms, and is conditioned on instrument labels. You can start from a sound you like and interactively modify it through inpainting of the spectrogram.

Planet Drums, DrumGAN, Impact Drums: Three drum sound synthesizers. *Planet Drums* is based on a VAE architecture that allows the user to explore different drum sounds by traversing a low-dimensional embedding of the latent space (Aouameur et al., 2019). *DrumGAN* and *Impact Drums* are based on generative adversarial networks (Goodfellow et al., 2014). *DrumGan* is conditioned on perceptual features that can be used as controls (Nistal et al., 2020).

DrumNet: A tool for creating drum tracks conditioned on existing audio tracks like guitar, bass, or keyboard tracks (Lattner and Grachten, 2019). The output adapts to the tempo and rhythm of the existing tracks, and users can explore different rhythmic variations by traversing a latent space.

BassNet, LeadNet: Tools for creating bass tracks (BassNet) or lead tracks (LeadNet), conditioned on one or more existing audio tracks (Grachten et al., 2020). The output adapts to the tonality of the existing tracks (if the input is tonal), and users can explore different rhythmic and melodic variations of the output by traversing a latent

space. The model outputs both MIDI and audio, and conveys articulation, dynamics, timbre, and intonation. In terms of model architecture BassNet and LeadNet are identical. They differ in that BassNet was trained on bass guitar tracks, and LeadNet was trained on vocal and lead guitar tracks.

ResonanceEQ, ProfileEQ: Adaptive equalizers for audio mixing and mastering tasks (Grachten et al., 2019). They consist of hand-designed processing pipelines to adjust the spectral characteristics of the sound in an adaptive way, and additional feed-forward convolutional neural networks to estimate optimal control parameters for the equalizer process conditional on the input audio.

4.2 PROCEDURE

At the start of the collaboration we give the artists an overview of AI and machine learning in the context of music, and explain our vision of music AI as tools to enrich the creative workflow in music production. We give them a demonstration of the available tool prototypes in the lab where they can try out the software. When the artists are familiar with the ways the tools work, they use them in their own working environment, experimenting with the tools in their music production process over the course of 6 months or longer (see *Table 1*), next to their regular professional activities. Typically there are follow-up sessions after the first session where the artists talk about their experience, what they like and dislike about the tools, and what changes they would like to see. We modify the tools accordingly, whenever the changes can be realized within a reasonable effort, while proposals that imply more fundamental changes to the tools are used to guide future development. When the artists have finalized their work they send us the outcomes, and a description of their workflow, which typically includes the AI tools, along with several other music production tools they work with. Some of the artists describe their experiences with the tools in interviews.[6]

4.3 RESULTS AND DISCUSSION

In this section we present results of the collaborations in terms of the reports and feedback from the artists. We group the results into themes (such as types of interaction with the tools) that have been identified in part in prior work, such as that by Huang et al. (2020). Before that we give an example of a workflow decomposition (*Figure 1*) in which Luc Leroy and Yann Macé of *Hyper Music* use several AI-driven prototypes in conjunction with mainstream music production technology.[7] Although a detailed analysis is beyond the scope of this paper, the schema exemplifies the non-linearity of the music production process (as discussed in Section 3), involving iteration between compositional activities such as creating harmonic and melodic material, and mixing activities such as equalization. Different equalizations of an audio track may emphasize different acoustical
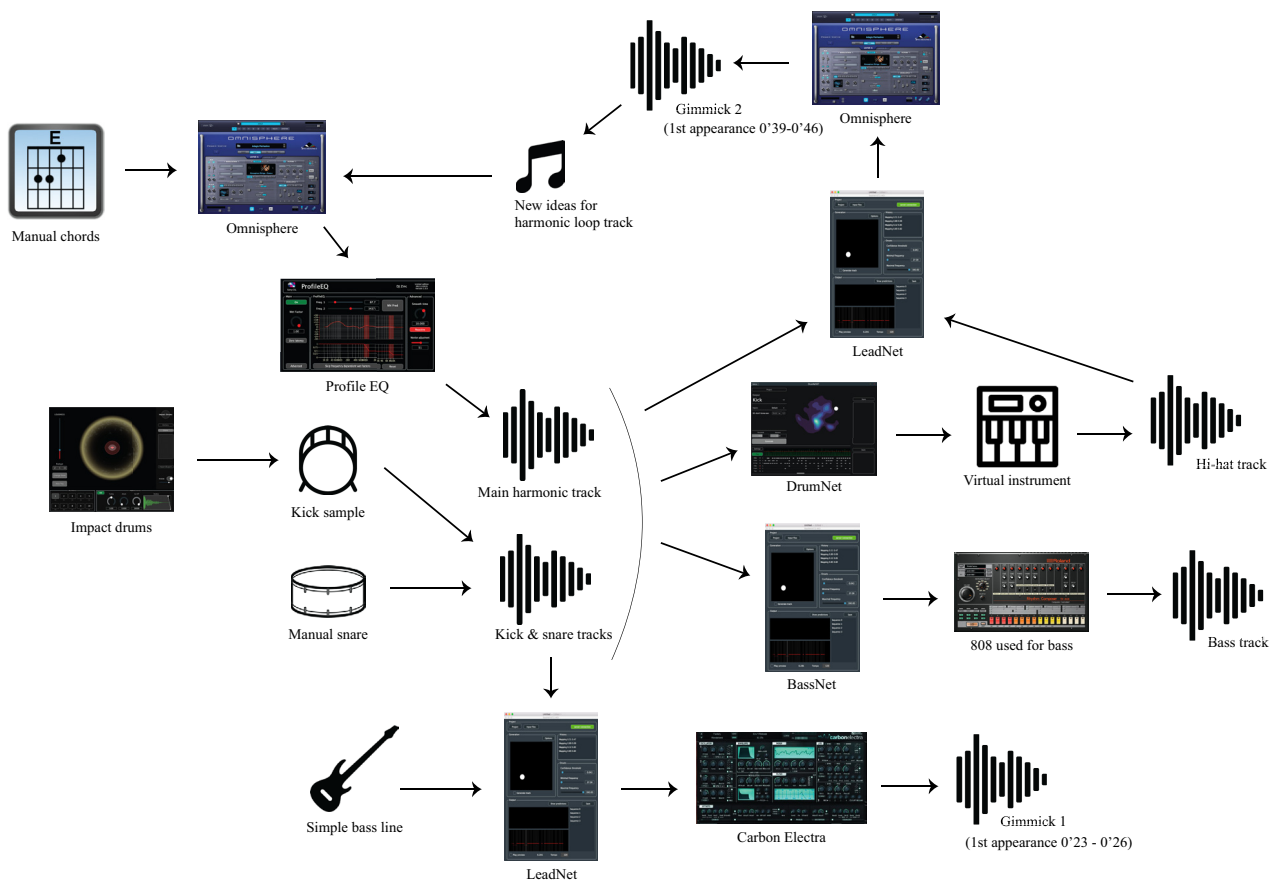
Figure 1: Example workflow diagram by artist duo *Hyper Music*. The diagram shows how various AI tool prototypes are used in the music production process.

elements of the sound, and can thus lead to different rhythmic or melodic variations when used as an input to DrumNet and BassNet/LeadNet.

4.3.1 Push and pull interactions

In the context of creative text writing, Clark et al. (2018) describe two approaches to start an interaction between the user and an AI tool: *push* interactions, where the tool makes spontaneous suggestions (e.g. in an auto-complete fashion), and *pull* interactions, where the user explicitly queries the tool for an output. The tools we focus on here (listed in Section 4.1) are primarily designed for pull interactions: rather than acting autonomously within a session, the tools are operated actively by the artists when they want an output.

One particular case of *pull* interaction is known as *priming* (Huang et al., 2020): the artist designs an input to drive the generation process. The priming input can be used in diferent ways: it can serve as the start of a musical part to be continued by the tool, as the starting template from which variations can be explored, or as a part in a multi-part setting for which the tool generates accompanying parts. Priming amounts to what is referred to as *dense conditioning* (Grachten et al., 2020), where the output of a model is controlled by providing a rich source of information (e.g. an audio or MIDI track) instead of sparser types of information that are provided

by the typical UI elements of a control panel (sliders, buttons, presets, etc.). An example of the *priming* process used in production from *Hyper Music*:

> "Made an 8-bar bounce with kick, snare plus a very simple legato bass part (not used thereafter). Fed this bounce to LeadNet. Tweaked around until I hear something inspiring: it plays a cool part with a 4-note hook that sounds good at the end of the chord cycle."

As mentioned before, BassNet/LeadNet and DrumNet are designed to be primed on audio input. This makes it possible to react even to minor nuances in the input, like expressivity in terms of timing, dynamics, or timbre. Donn Healy's comment underlines this:

> "DrumNet handled this quirky input very well, it followed the expression to a T [very precisely]".

4.3.2 Exploration and higher-level control

As witnessed by Huang et al. (2020) in their study, many musicians adopt the *generate-then-curate* strategy when working with specific AI-driven prototypes; they first generate many samples and then select those they deem valuable for further use. Artist Uèle Lamore adopted this strategy when working with the prototypes:

"The goal was to generate a selection of percussion/drum samples that I could see fit to use in any given setting. [...] generating percussion sounds with DrumGAN and Planet Drums. I'm not interested in generating sounds that sound like a "real" or "classic" kit. I want sounds that are very abstract [...] I now had this selection of sounds available."

Although this strategy is not unique to AI-based approaches, the use of machine learning can potentially make it more efficient and rewarding. When creating a subset of interesting items in a large space of possibilities, there is generally a trade-off between covering a diverse range of interesting items (*recall* in information retrieval terms) and avoiding uninteresting items (*precision*). As such this use case is related to recommendation in music information retrieval. By modeling the distribution of datasets, generative machine learning models are especially suited for this task, potentially reducing the need for cumbersome skipping through sample libraries, or fiddling with numerous controls of a complex synthesizer in order to realize an idea.

An alternative to *generate-then-curate* is *exploration* through *higher-level control*. Rather than producing a batch of possibilities at once, this strategy is interactive, allowing an artist to explore variations of an idea by varying the controls of the tool.

A prominent example of exploration in generative models is the navigation of *latent spaces*, which may or may not have a clear semantic interpretation, depending on how the model was trained (Nistal et al., 2020; Aouameur et al., 2019; Engel et al., 2019).

BassNet/LeadNet and DrumNet also provide control over the output by way of latent space navigation, but here the output is conditioned on one or more input audio tracks. Importantly the latent space in these models does not provide absolute control over the output, which would defeat the purpose of the conditioning input audio, but rather they modify the way the models react to the input. The models are trained in such a way that the latent space is discouraged to encode (and thus provide control over) any qualities of the output that are (largely) constrained by the input. In DrumNet for instance, the tempo and metrical grid are inferred from the input, whereas the latent space encodes control over the remaining degrees of freedom for the drum track, such as its rhythmic patterns and the density of rhythmic events.

4.3.3 AI, the new analog?

Although the machine learning models that power AI tools are trained to perform specific tasks (e.g. to produce a specific type of output given a specific type of input), once an AI prototype is in the hands of artists, the scope and limitations of a model are regularly ignored, or even actively exploited. Here we cover three kinds of examples.

Glitches. Machine learning models often have some degree of systematic bias, which leads outputs to have some persistent characteristics that are not representative of the dataset they were trained on. Although this is considered a fault in terms of machine learning theory, musicians sometimes point out that they like particular artifacts for having a *characteristic identity*. Twenty9 speaks about the Impact Drums and Planet Drums prototype:

"[...] I love [the artifacts'] color, it changes from what I hear in the current available packs that do a lot of recycling. [...] Artistically, this grain is interesting [...], it is the fact of not being able to accentuate it, modify it or even play with it, which is slowing down and which limits the possibilities of sound palettes."

Uèle Lamore speaks similarly about the Notono prototype:

"The biggest weakness of Notono at [this] moment [in development], was its extreme treatment of sound. This resulted in the creation of very "phasy", filtered, samples with a very peculiar acoustic quality. However this was absolutely perfect to represent the *Corruption of the Forest* [song title], an unnatural, evil substance slowly spreading like a disease."

Such a music process is reminiscent of the analog synth's "grain"—a perceptual quality associated with raw, unpolished sound—that is so much sought after by pop musicians.

Of course this does not mean that every imperfection will pass for character. Donn Healy for example is critical of the sound quality of BassNet's audio outputs, which forces him to apply heavy effect processing in order to obtain a usable sound. Nonetheless he prefers the audio outputs over the corresponding MIDI outputs, both because he is used to working with audio from analog synthesizers, and because he likes the intra-note modulation of intonation, timbre and dynamics that is not present in the MIDI output.[8]

The notion of model output confidence can also play a role in the context of glitches. To give an example, Healy found that with an early version of BassNet it was difficult to produce interesting bass lines. A satisfactory resolution of this issue turned out to be simply to tweak the post-processing step of the model that filtered predicted notes based on the model confidence associated with the notes. Allowing more predicted notes to pass—even if the model is uncertain about these notes—produced much more interesting and useful results to Healy's taste.

Out-of-domain input. Just like all machine learning models, the models that are conditioned on input audio

tracks (BassNet/LeadNet, and DrumNet) have been trained on datasets covering a certain range of musical variation. Although the models will still produce outputs for inputs that lie outside that range, the relation between the output and the input is not covered by the training regime. We refer to this scenario as *out-of-domain input*.

Figure 2 shows a transcription of the input and output of BassNet used with out-of-domain input.[9] This version of BassNet was trained on complete multi-tracks of classic rock songs. However, the input to the example consists of the audio of *death metal* solo drums. BassNet (bottom-most track) adjusts its output's spectral envelope to the kick's attacks, and reacts to the tuning of the percussion, in particular the toms and the snare.

Uèle Lamore encounters out-of-domain scenarios as well. She uses a version of BassNet that was trained on rock, pop and EDM multi-tracks with a prominent drum section as input, but her input track has an ambient character. She describes:

> "[…] none of my music on this EP is in 4/4, it's as far as you can get from pop or hip-hop and this track had zero percussion at this point. As a result of this, BassNet did not behave the way you would expect it to. However, I had the pleasant surprise to see the outputs—when transposed upwards by two octaves—were perfect melodies that worked really well in this ambient setting."

Out-of-domain output. We denote "out-of-domain output" when an artist uses the output of a tool for a different purpose than intended. For example, Donn Healy states:

> "I took a new snare pattern that DrumNet suggested and I brought it into a melodic Omnisphere sound, and I spread the notes in a way that they told a musically cohesive story [..] I really enjoyed that."

Also, as illustrated by Uèle Lamore's quote above, some artists took output of BassNet and tranposed it to obtain melodies instead of bass lines. Similarly, we discovered that ResonanceEQ, a tool designed to remove resonances, is usually inverted by artists to *add* resonances to audio.

4.3.4 On machine interference with the creative process

It has been noted before that AI-driven music tools can interfere with musical goals (Huang et al., 2020). We have experienced that even at early stages, prototypes sometimes require compatibility of format (e.g. implementation as DAW plug-ins rather than web applications) and compatibility with the artist's production methods in order to actually be used.

Even then, artists may be reluctant to depart from their creative goals to include inputs from AI tools. Yet
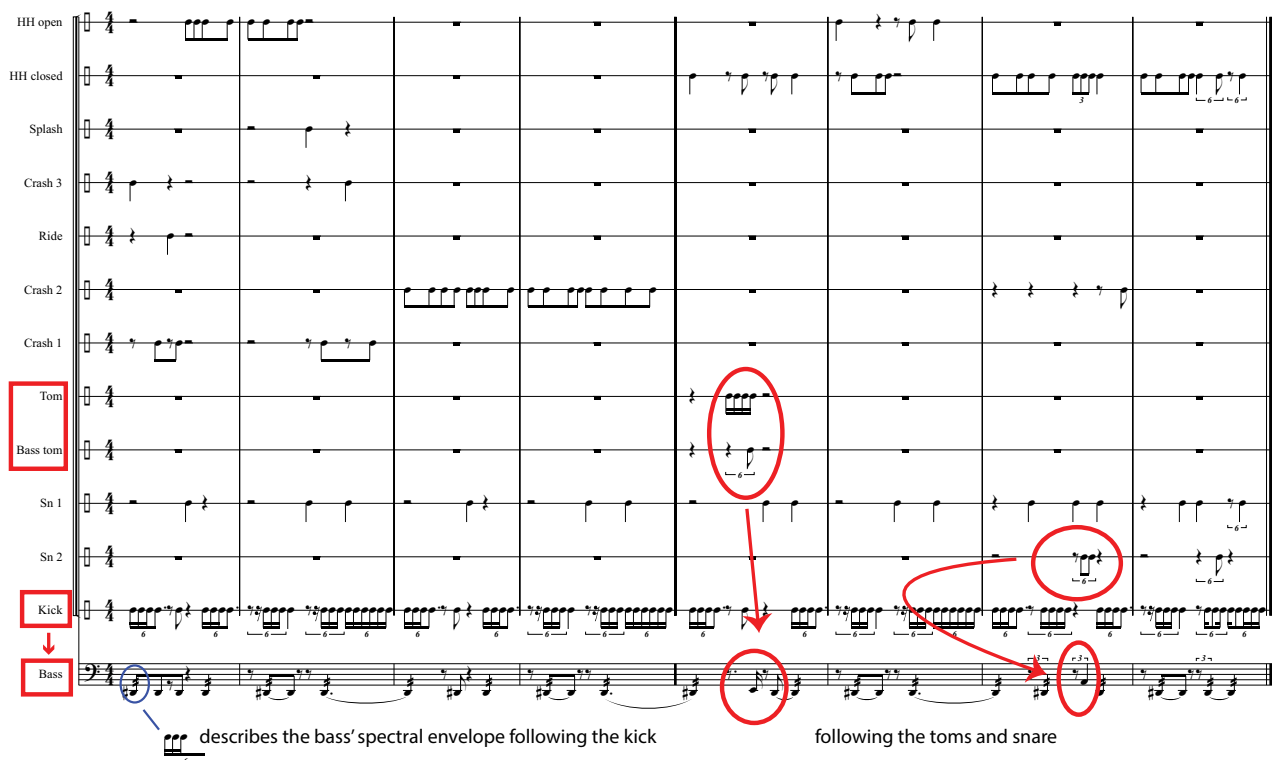

describes the bass' spectral envelope following the kick   following the toms and snare

Figure 2: Example of BassNet's behavior when confronted with out-of-domain input. BassNet (bottom-most track) adjusts its output's spectral envelope to the kick's attacks, and reacts to the tuning of the percussion.

overcoming this barrier can sometimes be rewarding in terms of results. Twenty9 testifies:

> "[...]. Since I was a fan of this loop [...] I went straight to drums. Honestly, in the euphoria, I wanted to jump on my usual sampler and set a rhythm in 5 min. I forced myself to confront DrumNet [...]. To my surprise, [...] I ended up with a pattern that worked well [even though] on my own I would not have placed my kicks like that."

> "[Working with LeadNet], I am confronted with melodies that I would probably never have thought of."

Donn Healy reports working with the AI tools changed his workflow into a more conversational process, and shifted his attitude to being more process-oriented rather than result-oriented[10]

From artist feedback such as the above we witness creativity *emerging* from the machine's interference with the artist's goals, and even from its effect on their workflow. In this respect it seems that AI music technology offers more than just tools for musicians to produce and manipulate sound in order to realize their creative goals more efficiently. In the words of Gioti (2021), it enables co-exploration by "breaking creative habits", and allows one (as an artist) to "reflect on one's own creative practice and aesthetic values".

This is in line with contemporary views of the nature of creativity, which have moved from regarding creativity primarily as an attribute of individuals that can be studied in isolation, to a systems perspective in which creativity stems from interactions between individuals, a group of experts that form a social environment (the *field*), and a set of rules and practices that form a particular cultural context (the *domain*) (Csikszentmihalyi, 1999; Hennessey, 2017). Hennion (1983) and McIntyre (2008). provide a sociological account of the music production process that is similar in spirit. A more elaborate systems perspective on music production is given by Thompson (2019), who highlights how the common view of a creative work such as a song sprouting from the mind of the individual genius is a myth. Almost always, he argues, it is the result of interaction between a collective responsible for different aspects of the music production, like recording engineering, mixing, orchestration/arrangement, performance. Often several individuals collaborate even on a single aspect.

What does this type of creative collaboration look like in music production? A common technique is to start with a small idea (a title, lyric, or motif) that serves as a seed to be developed into a full song (Bennett, 2012). Bennett argues that material produced in this way is subsequently evaluated by the individuals involved, and is either assimilated into the end product (through approval, negotiation, or adaptation), or it is rejected by means of a veto.

Although AI music tools are (at this stage) by no means the equivalent of a human creative collaborator to artists, arguably they can play an active role in the creative process. As we have also seen earlier in this Section, current AI tools typically participate in the process described by Bennett (2012) either by providing starting material (the push/pull generation, and generate-then-curate scenarios), or by exploring variations of existing material, as Notono and PlanetDrums are capable of.

4.4 INTERACTION AND VISUALIZATION

A recurring theme in feedback from the artists concerns the interaction with the tools. Sometimes artists simply suggest convenience functionality to improve usability, such as keeping a record of past interactions with the tool, or the ability to loop over a sound fragment to enable extensive exploration of control parameters. These suggestions are in line with some of the recommendations for design of creativity-support tools by Shneiderman (2007). On other occasions they miss control over some aspect of the output. In terms of AI tool development, this may translate to providing better access to latent dimensions, for example by conditioning on particular perceptual dimensions during training.

Another issue, especially in models with learned latent spaces that do not have a pre-defined interpretation, is understanding model behavior as a function of the latent space. Without any visualization, the only way to navigate variations in output is by trial-and-error, trying out different regions of the latent space. Although extended use of a tool may give the user an intuition of how it will react in different circumstances, a better approach is to signal to the user what behavior they can expect in response to the latent space controls. This can be done for example by a-posteriori mapping of the behavior to perceptual features of the output, and projecting these features on the latent space.

An interesting approach to the problem of *appropriation* of music software by users—the process of getting acquainted with the possibilities and limitations of the tool by exploration—is proposed by Scurto and Bevilacqua (2018). They observe that user interfaces of music software are often intimidating to novice users, and propose an AI framework to support the appropriation by the user through interactive feedback on the utility or value of some function or feature from the user, combined with reinforcement learning by the framework based on this feedback. A similar approach can also be envisioned to help artists explore latent space controls. More generally, this could lead to a two level design where AI at the first level learns (possibly complex, high-dimensional, and use-case agnostic) relationships or patterns from musical corpora based on generic criteria such as information-theoretic principles (van den Oord

et al., 2018; Piantanida and Vega, 2021), and AI at the second level assists the user in tailoring the first level into a personalized and use-case specific AI tool.

5 LESSONS LEARNED FROM AI-BASED MUSICAL RESEARCH

The open-ended, exploratory nature of the creative process in music production (as in other creative fields) makes the development of AI tools to support and enhance this process an exciting, but certainly also a challenging topic. It is at the intersection of several fields of research, most notably machine learning, creativity studies, and user design.

As Gioti (2021) observes, there is a tension between this open-ended, under-determined process and the machine learning paradigm, which relies on well-defined problem descriptions and success criteria that are quantifiable in terms of data and model outputs. For the same reason, collaborating with artists in the development of AI tools for music production is not a straight-forward case of participatory design: user needs can be hard to formulate and may change as a result of interacting with the technology.

In this section we summarize some of the lessons we learned throughout our work in AI-based musical research. They relate closely to the notion of creativity-support in human-computer interaction (Shneiderman, 2007; Cherry and Latulipe, 2014), and may provide orientation for formulating more rigorous success criteria specific to AI-based creativity support for music production.

5.1 RECOMMENDATIONS
Work alongside musicians. Research is only part of the story. A perfectly well-trained model may be irrelevant in music production. Conversely, in music production, it is not always a problem if the model does not work perfectly. Some systematic bias may even be a mark of style (see Section 4.3.3). Going beyond a proof-of-concept and creating usable tools is the only way to assess which qualities make an AI music tool interesting to artists.

Foster chance/serendipity. Create situations with a rich potential for unexpected results. This may mean using different prototypes together or along with third-party tools, modifying models in an unorthodox way to fit some specific purpose, or using models for applications they weren't conceived for. In the most extreme case, AI models do not even need to be trained in order to emit musically valuable output (Steinmetz and Reiss, 2020). As we have argued in Section 3.2, AI tools that work with audio offer better chances of a lucky find in the music production process than MIDI tools, simply because they can be driven by a wider range of inputs available in the

DAW project. In a more practical vein, enabling artists to benefit from serendipity may be as simple as keeping a record of their interactions with the AI tool, and capturing the outputs (as discussed in the previous section).

AI does not need to entail autonomy. In a recent interview,[11] Uèle Lamore states "The computer wants to play everything perfectly, but the music I make isn't perfect. The human will always add something of their own." AI technology for music is sometimes designed in the form of fully autonomous systems creating full songs from only very sparse inputs (e.g. Dhariwal et al., 2020). Although they are important demonstrations of the potential of audio based AI systems, they are not well-adapted to typical in-studio composition use cases, where outputs of a music AI tool are often small parts to be integrated in and coordinated with a larger structure of musical elements.

Adapt to the music at hand. As obvious as this sounds, it is not always straight-forward to assess the limitations that assumptions of AI design and modeling approaches may impose on the success and potential of AI music tools. Specifically, as we have argued in Section 3, learning from scores is only partially relevant for contemporary Popular Music. Beyond the music itself, it is helpful to get acquainted with the workflows of artists. When distributing prototypes, using formats that artists are accustomed to, such as DAW plug-ins eliminates an unnecessary barrier to adoption. That said, there may be valid reasons to use less typical music tool formats. For example we found that web-interfaces and client-server architectures facilitate quick iteration and deployment of designs.

5.2 ON THE VALIDATION OF AI-DRIVEN MUSIC TECHNOLOGY
AI music tools rely on a machine learning model that is designed and trained to perform a specific task, and as such the most obvious way to validate an AI tool may seem to be to measure how well the machine learning performs its task. Although this is clearly an important criterion in initial stages of development of AI technology, as we have seen in Section 4.3.3, artists tend to use AI tools in ways that defy the basic tenets of machine learning, not seldomly to their satisfaction. The tool may be used on data that is unlike the data the model was trained on, the output may be used in other ways than its intended purpose, and whereas models that produce faithful data samples may be uninteresting to artists, they may have a weak spot for artifacts in the outputs, resulting for example from model bias.

This means that we need other ways of measuring the success of AI-driven music technology in addition to standard machine learning evaluation criteria such as prediction error or accuracy. Here we suggest some criteria that have emerged largely from our collaboration with artists.

Workflow integration. Validation may take into account if a tool finds its place in a production workflow. For that a tool needs to be useful, but also should not interrupt the workflow the artist is accustomed to. For example, artists are often reluctant to switch from their DAW to an external stand-alone application or web-interface. This criterion is not absolute however, and relies on specific and clearly defined use cases. For example, in Section 4.3.4 we discuss how a change in workflow may also be a valid outcome of the use of AI music tools.

Facilitation of production. Does the prototype simplify a difficult or time-consuming task? For instance, Yann Macé appreciates latent space navigation in DrumGAN, as it provides much quicker results than spending hours browsing a drum sample library.

Enhanced creativity. Does the prototype stimulate the artist's creativity? Does it provide a good trade-off between quality and novelty (i.e., it avoids frustrating the artist with too many useless outputs or cumbersome usability)? For instance, Twenty9 and Uèle Lamore repeatedly mention that BassNet, LeadNet and DrumNet provide solutions they would have never considered, but turned out to be happy with.

Identifiable results. Did the technology bring recognizable elements to the music? For instance, Twenty9 enjoys the grain of the GAN-based drum generators, and Yann Macé appreciates the characteristic style of DrumNet's hi-hat tracks.

Published content. The commercial compatibility of music content that includes the technology is an indirect form of validation. It should be kept in mind this measure presupposes access to commercial publishing in the first place, and is thus mostly limited to collaborations with professional artists. That said, the integration of AI music tool outputs into the commercial outputs of the artist signals a willingness to endorse the technology in a social context, which is a significant result from the systems perspective on creativity (Csikszentmihalyi, 1999).

Three examples of published content by the collaborating artists that involve the AI tools they worked with:

- A drumkit designed by Twenty9, using Impact Drums, Planet Drums and DrumGAN[12]
- An EP by Uèle Lamore and her label, featuring the AI technology described in Section 4.1[13]
- A worldwide Azzaro advertisement campaign for which Yann Macé and Luc Leroy of HyperMusic wrote the music track, making use of Impact Drums and DrumGAN.[14]

6. CONCLUSIONS AND FUTURE WORK

In this paper we have provided a perspective on the development of AI tools for music production in contemporary Popular Music (CPM) genres, informed both by general considerations of the music production practice in those genres, and by a thematic analysis of reports on real-world AI music tool usage by professional artists.

We have taken a closer look at the in-studio composition process (Section 3) that has become a standard music production practice in many CPM genres, not least because of the increasingly prominent role of electronics and signal processing in music production. There is a discrepancy between what we believe is a widespread view of music production in AI music generation research, and two aspects of the in-studio composition process in particular, namely the role of symbolic representations, and the coupling of composition with sound editing and mixing stages of the process. From these characteristics we conclude that audio based AI tools are better suited to support the creative workflow of the artist than tools that work exclusively with piano-roll/MIDI representations.

In Section 4 we presented a thematic analysis of collaborations of our music AI research lab with professional artists, in which they produced music with various AI tools developed at the lab. Their feedback highlights a variety of aspects to be taken into account in AI tool development, ranging from relatively trivial issues such as the modality of the software (plug-in vs web-interface), to more fundamental issues such as the need for manual control over outputs, and the importance of providing insight in the effect controls have on the output, for example by means of visualizations. The results also show there can be a discrepancy between machine learning success criteria and intended purpose on the one hand, and the utility and value artists find in the AI tools on the other.

We have summarized our findings (Section 5) in the form of recommendations for the development of AI technology for CPM and proposed criteria for validation. We believe these can be a starting point for a more systematic methodology to assess the utility and value of AI music tools in CPM through artist collaborations. More concretely, we plan to use our findings to define a set of specific use-cases within the in-studio composition context, to enable a quantitative assessment of AI technology as creativity support tools using the *creativity support index* (Cherry and Latulipe, 2014).

NOTES

1 Some examples are *Landr* (mastering), *Amper* (media post-production), and *Amadeuscode* (song writing).

2 *https://cslmusicteam.sony.fr/*.

3 *https://www.berklee.edu/news/berklee-now/what-contemporary-music*, accessed January 17 2022.

4 An example of this can be heard in *Reach for the dead* by *Boards of Canada*.

5 *Pitched kick drum sounds* are a cornerstone of hardcore/hardstyle dance genres, for example.

6 Parts of the interviews have been edited into published content: *Uèle Lamore interview*, *Whim Therapy interview*, *Donn Healy interview*, *Donn Healy process description*.

7 The corresponding audio track is available as Additional File 1 (DOI: *https://doi.org/10.5334/tismir.100.s1*)

8 Healy discusses his experiences in the video included as Additional File 2 (DOI: *https://doi.org/10.5334/tismir.100.s2*).

9 The corresponding audio track is available as Additional File 3 (DOI: *https://doi.org/10.5334/tismir.100.s3*).

10 *https://youtu.be/uPvnpZ7mHCM?t=208*.

11 *www.musicradar.com/news/uele-lamore*.

12 *The AI Drum-Kit*.

13 *Heqets Shadow EP*.

14 *Azzaro Advertisement*.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Additional File 1.** Sound example 1. DOI: *https://doi.org/10.5334/tismir.100.s1*
- **Additional File 2.** Donn Healy on sample. DOI: *https://doi.org/10.5334/tismir.100.s2*
- **Additional File 3.** Sound example 2. DOI: *https://doi.org/10.5334/tismir.100.s3*

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Emmanuel Deruty
Sony Computer Science Laboratories (CSL) Paris, 6, rue Amyot, 75005 Paris, France

Maarten Grachten ⬥ *orcid.org/0000-0002-9488-0840*
Contractor for Sony CSL Paris, France

Stefan Lattner ⬥ *orcid.org/0000-0002-3945-7580*
Sony Computer Science Laboratories (CSL) Paris, 6, rue Amyot, 75005 Paris, France

Javier Nistal
Sony Computer Science Laboratories (CSL) Paris, 6, rue Amyot, 75005 Paris, France

Cyran Aouameur
Sony Computer Science Laboratories (CSL) Paris, 6, rue Amyot, 75005 Paris, France

REFERENCES

Aouameur, C., Esling, P., and **Hadjeres, G.** (2019). Neural drum machine: An interactive system for real-time synthesis of drum sounds. In *Proceedings of the Tenth International Conference on Computational Creativity (ICCC)*, Charlotte, North Carolina, USA.

Bazin, T., Hadjeres, G., Esling, P., and **Malt, M.** (2020). Spectrogram inpainting for interactive generation of

instrument sounds. In *Joint Conference on AI Music Creativity*.

Bell, A. P. (2018). *Dawn of the DAW: The Studio as Musical Instrument*. Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780190296605.001.0001*

Bennett, J. (2012). Constraint, collaboration and creativity in popular songwriting teams. In Collins, D., editor, *The Act of Musical Composition: Studies in the Creative Process*, pages 139–69. Ashgate Farnham.

Braun, V., and **Clarke, V.** (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101. DOI: *https://doi.org/10.1191/1478088706qp063oa*

Briot, J.-P., Hadjeres, G., and **Pachet, F.-D.** (2020). *Deep Learning Techniques for Music Generation*. Springer. DOI: *https://doi.org/10.1007/978-3-319-70163-9*

Burgess, R. J. (2013). *The Art of Music Production: The Theory and Practice*. Oxford University Press.

Cherry, E., and **Latulipe, C.** (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):1–25. DOI: *https://doi.org/10.1145/2617588*

Chu, H., Urtasun, R., and **Fidler, S.** (2016). Song from PI: A musically plausible network for pop music generation. *arXiv preprint arXiv:1611.03477*.

Clark, E., Ross, A. S., Tan, C., Ji, Y., and **Smith, N. A.** (2018). Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces (IUI)*, pages 329–340. ACM. DOI: *https://doi.org/10.1145/3172944.3172983*

Csikszentmihalyi, M. (1997). *Creativity: Flow and the Psychology of Discovery and Invention*. Harper-Perennial, New York.

Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In *Handbook of Creativity*. Cambridge University Press, Cambridge, UK. DOI: *https://doi.org/10.1017/CBO9780511807916.018*

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., and **Sutskever, I.** (2020). Jukebox: A generative model for music. *CoRR*, abs/2005.00341.

Engel, J. H., Agrawal, K. K., Chen, S., Gulrajani, I., Donahue, C., and **Roberts, A.** (2019). GANSynth: Adversarial neural audio synthesis. In *7th International Conference on Learning Representations (ICLR)*, New Orleans, USA.

Engel, J. H., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., and **Simonyan, K.** (2017). Neural audio synthesis of musical notes with WaveNet autoencoders. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, Sydney, Australia.

Frith, S. (2004). *Popular music: Critical concepts in media and cultural studies*, volume 1. Psychology Press.

Gibson, J. J. (1977). The theory of affordances. In Shaw, R. and Bransford, J., editors, *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, pages 67–82. Erlbaum, Hillsdale, New Jersey, USA.

Gioti, A.-M. (2021). Artificial intelligence for music composition. In *Handbook of Artificial Intelligence for Music*, pages 53–73. Springer. DOI: *https://doi.org/10.1007/978-3-030-72116-9_3*

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and **Bengio, Y.** (2014). Generative adversarial nets. In *Advances in Neural Information Processing (NIPS)*.

Grachten, M., Deruty, E., and **Tanguy, A.** (2019). Auto-adaptive resonance equalization using dilated residual networks. In *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands.

Grachten, M., Lattner, S., and **Deruty, E.** (2020). Bassnet: A variational gated autoencoder for conditional generation of bass guitar tracks with learned interactive control. *Applied Sciences*, 10(18). DOI: *https://doi.org/10.3390/app10186627*

Hadjeres, G., and **Crestel, L.** (2021). The piano inpainting application. *CoRR*, abs/2107.05944.

Hadjeres, G., Pachet, F., and **Nielsen, F.** (2017). Deep-Bach: A steerable model for Bach chorales generation. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70, pages 1362–1371, Sydney, Australia.

Hennessey, B. A. (2017). Taking a systems view of creativity: On the right path toward understanding. *The Journal of Creative Behavior*, 51(4):341–344. DOI: *https://doi.org/10.1002/jocb.196*

Hennion, A. (1983). The production of success: An anti-musicology of the pop song. *Popular Music*, 3:159–193. DOI: *https://doi.org/10.1017/S0261143000001616*

Huang, C.-Z. A., Koops, H. V., Newton-Rex, E., Dinculescu, M., and **Cai, C.** (2020). AI Song Contest: Human-AI co-creation in songwriting. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, pages 708–716, Montreal, Canada.

Ji, S., Luo, J., and **Yang, X.** (2020). A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions. *ArXiv*, abs/2011.06801.

Jones, S. (1992). *Rock Formation: Music, Technology, and Mass Communication*, volume 3 of *Foundations of Popular Culture*. Sage Publications. DOI: *https://doi.org/10.4135/9781483325491*

Kingma, D. P., and **Welling, M.** (2014). Auto-encoding variational Bayes. In *2nd International Conference on Learning Representations (ICLR)*, Banff, AB, Canada.

Knotts, S., and **Collins, N.** (2021). AI-Lectronica: Music AI in clubs and studio production. In *Handbook of Artificial Intelligence for Music*, pages 849–871. Springer. DOI: *https://doi.org/10.1007/978-3-030-72116-9_30*

Lashua, B., and **Thompson, P.** (2016). Producing music, producing myth? Creativity in recording studios. *International Association for the Study of Popular Music Journal*, 6(2):70–90. DOI: *https://doi.org/10.5429/2079-3871(2016)v6i2.5en*

Lattner, S., and **Grachten, M.** (2019). High-level control of drum track generation using learned patterns of rhythmic

interaction. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA. DOI: *https://doi.org/10.1109/WASPAA.2019.8937261*

Mazzanti, S. (2019). Defining popular music: Towards a "historical melodics". In Vilotijević, M. D. and Medić, I., editors, *Contemporary Popular Music Studies*, pages 17–26. Springer VS, Wiesbaden. DOI: *https://doi.org/10.1007/978-3-658-25253-3_2*

McIntyre, P. (2008). Creativity and cultural production: A study of contemporary Western popular music songwriting. *Creativity Research Journal*, 20(1):40–52. DOI: *https://doi.org/10.1080/10400410701841898*

Middleton, R. (1990). *Studying popular music*. McGraw-Hill Education (UK).

Moore, A. F., and **Martin, R.** (2018). *Rock: The Primary Text: Developing a Musicology of Rock*. Routledge. DOI: *https://doi.org/10.4324/9780429490170*

Moylan, W. (2020). *Recording Analysis: How the Record Shapes the Song*. CRC Press. DOI: *https://doi.org/10.4324/9781315617176*

Muller, M. J., and **Kuhn, S.** (1993). Participatory design. *Communications of the ACM*, 36(6):24–28. DOI: *https://doi.org/10.1145/153571.255960*

Nistal, J., Lattner, S., and **Richard, G.** (2020). Drum-GAN: Synthesis of drum sounds with timbral feature conditioning using generative adversarial networks. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, Montreal, Canada.

Payne, C. (2019). Musenet. *https://openai.com/blog/musenet/*. Retrieved Feb. 2021.

Piantanida, P., and **Vega, L. R.** (2021). Information bottleneck and representation learning. In Rodrigues, M. R. D. and Eldar, Y. C., editors, *Information-Theoretic Methods in Data Science*, chapter 11, pages 330–358. Cambridge University Press. DOI: *https://doi.org/10.1017/9781108616799.012*

Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., Dinculescu, M., Radebaugh, C., Hawthorne, C., and **Eck, D.** (2019). Magenta Studio: Augmenting creativity with deep learning in Ableton Live. In *Proceedings of the International Workshop on Musical Metacreation (MUME)*.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., and **Eck, D.** (2018). A hierarchical latent vector model for learning long-term structure in music. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80, pages 4364–4373, Stockholmsmässan, Stockholm Sweden. PMLR.

Scurto, H., and **Bevilacqua, F.** (2018). Appropriating music computing practices through human-AI collaboration. In *Journées d'Informatique Musicale (JIM 2018)*, Amiens, France.

Serra, X. (2012). Opportunities for a cultural specific approach in the computational description of music. In Serra, X., Rao, P., Murthy, H., and Bozkurt, B., editors, *Proceedings of the 2nd Comp-Music Workshop.* Universitat Pompeu Fabra.

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, 50:20–32. DOI: *https://doi.org/10.1145/1323688.1323689*

Steinmetz, C. J., and **Reiss, J. D.** (2020). Randomized overdrive neural networks. *CoRR*, abs/2010.04237.

Tagg, P. (1982). Analysing popular music: theory, method and practice. *Popular music*, 2:37–67. DOI: *https://doi.org/10.1017/S0261143000001227*

Thompson, P. (2019). *Creativity in the Recording Studio: Alternative Takes.* Springer. DOI: *https://doi.org/10.1007/978-3-030-01650-0*

van den Oord, A., Li, Y., and **Vinyals, O.** (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Yang, L.-C., Chou, S.-Y., and **Yang, Y.-H.** (2017). MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. *arXiv preprint arXiv:1703.10847*.