Machine Learning for Musical Expression: A Systematic Literature Review

Théo Jourdan, Baptiste Caramiaux

Machine Learning for Musical Expression: A Systematic Literature Review

Théo Jourdan
Sorbonne Université, CNRS, ISIR
Paris, France
jourdan@isir.upmc.fr

Baptiste Caramiaux
Sorbonne Université, CNRS, ISIR
Paris, France
caramiaux@isir.upmc.fr

ABSTRACT

For several decades NIME community has always been appropriating machine learning (ML) to apply for various tasks such as gesture-sound mapping or sound synthesis for digital musical instruments. Recently, the use of ML methods seems to have increased and the objectives have diversified. Despite its increasing use, few contributions have studied what constitutes the culture of learning technologies for this specific practice. This paper presents an analysis of 69 contributions selected from a systematic review of the NIME conference over the last 10 years. This paper aims at analysing the practices involving ML in terms of the techniques and the task used and the ways to interact this technology. It thus contributes to a deeper understanding of the specific goals and motivation in using ML for musical expression. This study allows us to propose new perspectives in the practice of these techniques.

Author Keywords

Machine Learning, Digital Musical Instrument, Interaction Design, Human-IA Interaction, Literature Review

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); •Applied computing → Sound and music computing;

1. INTRODUCTION

Machine learning (ML) is defined as a set of computational techniques capable of discovering the underlying structure of a data set. These techniques are said to be *trained* on this data set to *learn* the structure, making them capable of making predictions by taking as input data never observed before. In other words, this family of techniques makes it possible to create a program inductively: the computer takes as input a set of data defining the task to be accomplished, and proposes as output the program capable of performing this task (with a certain level of performance). Examples of well-known tasks are: image recognition, induced from a set of image-textual category pairs [49]; machine translation, induced from a set of sentences in one language and its translation into another language [65]; or content generation, induced from a set of typical content to be generated [42]. Due to the remarkable results on complex and real data, this type of technology has attracted interest of creative practitioners and artists working with new technology and new media. In this paper, we propose to provide a survey of the use of ML in the design of New Interfaces for Musicale Expression (NIME).

The NIME community has been involved in the development of ML and its use in a creative context from an early stage with mapping between gesture and sound [20]. Indeed, ML was used to create non-linear relationships between high-level control parameters and low-level sound synthesis parameters, as illustrated in the works of Wessel et al. [89]. This way of building a mapping between signals from different modalities is usually called implicit mapping [91, 36], which has particularly been explored in the creation of interactions between gestures and sounds. Today the use of ML techniques has diversified, especially since the last literature review on ML proposed ten years ago, in 2013, in the NIME community [10]. This computational technique has changed in form and expressive capacity, particularly with the advent of deep learning. A lot of work at NIME has followed and even triggered changes in the use of this technology.

To address this question, we propose in this paper a systematic literature review of the work of the NIME community involving ML over the last ten years. Through the analysis, our questions are: **(1) What musical and technological practices have emerged in the NIME community? (2) Beyond the techniques used, what kinds of interactions are taking place between the user and the ML methods proposed by the NIME community? (3) What are the motivations and expectations of the authors of the publications when using ML?**

We propose an analysis grid allowing us to extract the diversity of the methods used, the actors involved in the design of these systems and their control on them. We report in this article the results of the literature review and discuss their implications.

2. BACKGROUND

Over the past decade the research and tools in Machine Learning have grown exponentially. Shortly after the Deep Learning breakthrough in 2012, showing a qualitative gap between state of the art and what a Deep Neural Network achieves on an image classification task, ML research has exploded, supported by large public and private investments. Among the key steps of this research, we can cite

NIME'23, 31 May–2 June, 2023, Mexico City, Mexico.

the use of Convolutional Neural Network for supervised learning (classification, regression) [50] or content generation through Generative Adversarial Networks [29]. Technological breakthrough, and its widespread integration to many software and services, also does not go without socio-political and socio-cultural consequences such as technological governance [57], accentuation of biases from datasets (e.g. racist or sexist biases) [79], exploitation of cheap labour for repetitive annotation tasks [81], or the enormous impact of such technology on the environment [12].

As the NIME community initiated the integration of ML into instrument design prior to 2012, one wonders what impact technological advances have had on NIME over the past decade. In particular, we ask to what extent the NIME communities have been able to adapt to the new technologies integrated, manipulated, or avoided ML methods in the creation of musical interfaces. In this article, we propose a literature review to understand this diversification, the methods put forward, and also better understand the goals of people applying ML to musical tasks in NIME community.

3. METHOD

In this section, we present the method used for the literature review on the involvement of ML in the design of new interfaces for musical expression. We used the guidelines given by PRISMA for writing and reading systematic reviews and meta-analyses [63]. Our bibliographic research is based on the literature produced in the NIME conference [1]. Figure 1 illustrates the survey methodology, from the identification of the items to the final filtering. We detail below each step of the selection process.

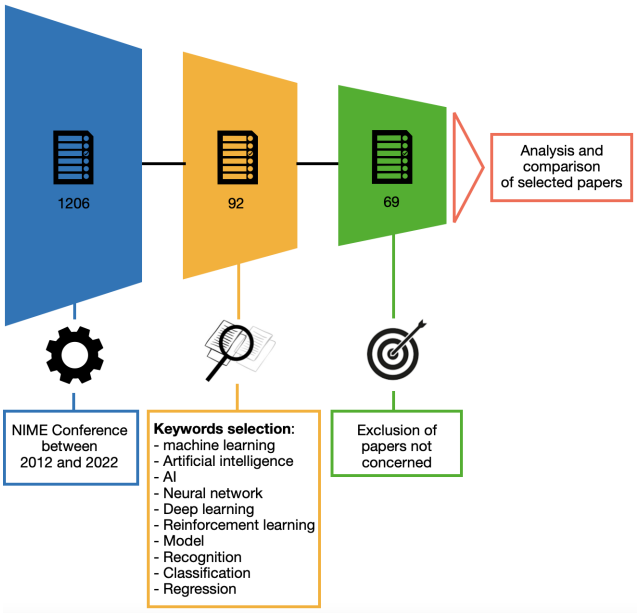

Figure 1: Diagram of the paper selection process broken down into three stages.

3.1 Search Protocol

We propose a review of the literature over the last ten years, between 2012 and 2022. We have searched for papers in the archives of the conference proceedings. We chose not include music proceedings in our search protocol due to a lack of technical and design information in the project descriptions. The archive of the conference proceedings is available online in a GitHub repository[2]. Moreover, the bibliography file for the year 2022 was not available at the time of writing this paper, so we created it from the list of accepted papers indicated on the conference website[3], and we asked for confirmation from the programme committee officers. We have identified a total of **1206** articles.

3.2 Selection of studies

3.2.1 *Filtering by keywords*

The bibliographic search strategy included a collection of keywords defined to identify articles proposing a system in which one of the main elements is an ML model. We used this keyword collection on the metadata of the publications including the title, the keywords of the article and the abstract. The keywords used in the selection of articles were: "*machine learning*", "*artificial intelligence*", "*AI*", "*neural network*", "*deep learning*", "*reinforcement learning*", "*model*", "*recognition*", "*classification*", "*regression*" . These keywords have been selected to consider articles whose main technology is machine learning, through the use of ML model. From this first selection **92** publications were extracted.

3.2.2 *Manual filtering*

We performed a final step of reading the content of the selected articles and removed any research articles that did not include a system involving a ML model, such as articles reporting reviews or interviews. A total of **69** articles were finally selected with this filtering process: [16, 93, 30, 51, 28, 74, 56, 15, 66, 60, 98, 76, 55, 18, 68, 31, 54, 84, 6, 7, 92, 100, 83, 67, 8, 95, 90, 22, 43, 87, 33, 94, 38, 64, 46, 70, 32, 85, 88, 73, 99, 35, 24, 40, 5, 75, 62, 27, 53, 61, 80, 48, 78, 59, 4, 82, 14, 26, 72, 52, 44, 23, 47, 2, 19, 34, 86, 97, 77].

3.3 Data extraction

For each publication, nine features were extracted using a custom-made data extraction worksheet. These features were chosen according to three research questions presented in Introduction (Section 1), framing our analysis.

1. **The techniques used**
   (a) **Type of learning**: we differentiate two types of learning with (i) *Shallow learning* which concerns all models using manually calculated descriptors (e.g. Decision tree [39], Support vector machine [17], k nearest neighbours method [13])); (ii) *Deep learning* which concerns all models which automatically learn the most efficient descriptors for the defined task, this generally concerns the different types of neural networks.
   (b) **Type of model**: gives details of the ML algorithm used (e.g. Recurrent neural network [71], Variational Autoencoder [45]).
   (c) **Type of task**: defines the goal for the model with either (i) *classification* where the data is predicted in discrete class labels; (ii) *regression* where the model predicts a continuous quantity; or (iii)

[1] https://www.nime.org/

[2] https://github.com/NIME-conference/NIME-bibliography

[3] https://nime2022.org/

generation where the model can generate new instances of data.

2. **The possibility of interaction between the author/user and the ML system**

   (a) **Modifiable parameters**: defines whether or not a user can modify certain parameters of the system to influence its operation (e.g. learning rate of the ML model)

   (b) **Trainable by user**: defines whether users have the possibility to train the proposed model with their own data.

   (c) **User intervention**: defines at what point in the process of designing and conceiving the contribution, the users intervened. We distinguish three moments: (i) *Beginning*, meaning that the users were solicited from the beginning of the process, whether it was the design or the first prototypes; (ii) *Middle*, means that the users have tested and evaluated the first prototypes designed with quantitative and/or qualitative feedback, or if they have participated in the training of the learning model; (iii) *End*, means that the users are only content to use the given system at the end of its design.

   (d) **Evaluation**: defines how the system is evaluated. This may be quantitatively through performance indicators or qualitatively through analysis of the user experience.

3. **The goals of the authors in using ML techniques**

   (a) **Terminology**: gives the vocabulary used by the authors to define the contribution presented in the publication (e.g. agent, instrument, system)

   (b) **Usage expectations**: defines for what purposes the authors develop the systems.

Finally, we were able to map each article in this grid of analysis, creating fequency of occurences of each category (type of learning, type of model, type of task, task details, terminology, intended user, modifiable parameters, author training, user intervention, evaluation) in our pool of selected articles. In the following section, we report the results of the analysis.

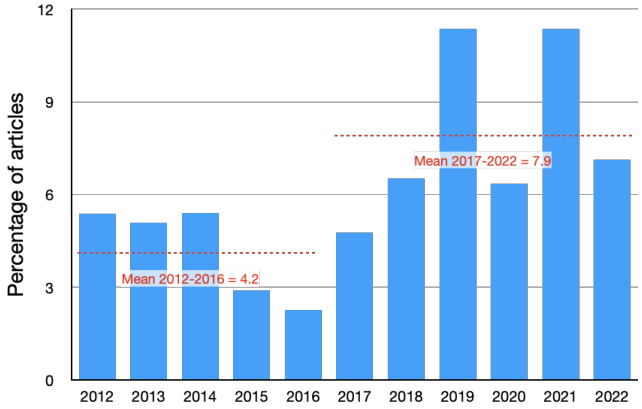

Figure 2: Percentage of articles per year offering a ML model in their system.

4. RESULTS

In this section, we present the results of the analysis. First, we found that the number of publications using ML has increased significantly since 2012. Figure 2 shows the proportion of these papers among the total number of accepted publications at NIME each year, from 2012 to 2022. Another finding is that the use of ML models remains a minority in the published articles, as they concern between 2% and 12% of the accepted papers. We notice that on average the percentage of articles is twice as high between 2017 and 2022 compared to the 2012-2016 period. The remaining results are structured in three parts: learning models and tasks, interactions between the user and the ML methods, and user motivations analyzed through the applications and terminologies (additional details are reported in Appendix A).

4.1 Learning models and tasks

Here we describe the types of ML models used in the selected articles of the literature review, as well as the specific tasks performed with these models and their purpose in a musical context.

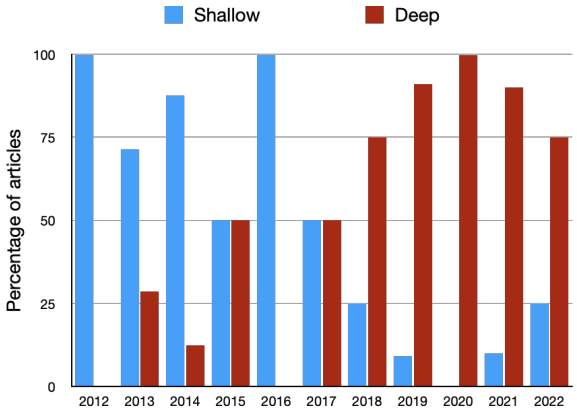

Figure 3: Evolution of the distribution of papers according to the type of models.

4.1.1 *Types of models*

Figure 3 depicts the evolution of the percentage of articles using *Shallow learning* and *Deep learning* models between 2012 and 2022. Our analysis shows that the number of articles using ML models has increased significantly over the years, with a shift from a predominant use of *Shallow learning* models between 2012 and 2016, to a predominant use of *Deep Learning* models since 2017. The first works highlighting *Deep learning* methods appeared in 2013 [47, 23], one year after the seminal article [50] that heralded the onslaught of artificial neural networks on the field of machine learning and artificial intelligence. It is also worth noting that several papers use both *Shallow learning* and *Deep learning* models [47, 67, 8, 64].

4.1.2 *Task types*

Figure 4 shows the evolution of the distribution of tasks performed by the proposed models, divided into three categories: *classification*, *regression* and *generation*. The first two tasks are supervised tasks while the last one is unsupervised. Several articles can combine several tasks at the same time by integrating several models into the proposed system.

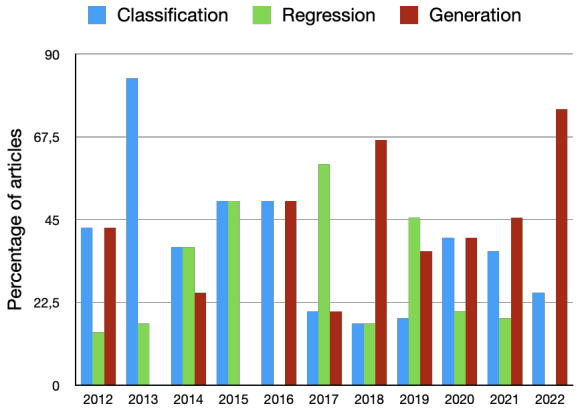

Figure 4: Evolution of the distribution of papers according to the type of tasks

We found that classification and regression models are mainly used for two specific musical tasks. Firstly, they have been used to build **mappings**, which consist of learning to associate input data to output data. For example, it associates gestures , or a set of control parameters from an ad-hoc interface, to sound synthesis parameters [62, 76, 84]. These mapping methods have historically been based on *Shallow* models, such as probabilistic methods based on Gaussian Mixture Models [22]. One of the reasons is that mapping design is preferably done iteratively by musicians or performers, requiring fast training and short iteration cycles between model training to testing.

Deep learning models have opened up new possibilities in mapping applications by allowing the consideration of parameter spaces of greater complexity. For example, these methods have enabled video analysis to detect gestures [59], or multi-modal analysis with virtual reality (VR) including gesture, sound and images [51]. Increasing the complexity of mapping with deep learning allowed the development of more "*open-ended mapping processes*" [59] in which the creator looks for new musically expressive mappings, without having a specific vision of what it might be, but opinions and intuitions about whether something works or not.

Secondly, supervised ML methods have been used for **audio analysis**, which includes all sound processing tasks such as annotating sounds to recognise a played instrument [80, 48], predicting the mood of a sound environment (e.g. pleasant, boring, chaotic) based on soundscape affect collected on users [78], sorting sounds according to the user's tastes [24], or speech recognition [54].

With the significant increase in the use of deep learning methods between 2017 and 2022, we notice the development of contributions involving generation tasks (see Figure 4), which have been commonly used for **sound synthesis**. Synthesis concerns tasks for generating sound signals with, for example, Variational Auto-Encoders (VAEs) [88, 27, 87] or Generative Adversarial Networks (GANs) [74]. This also applies to melodic pattern generation tasks in MIDI format, which, prior to the use of deep learning models, used hidden Markov chains (HMMs) (e.g. with drum pattern generation[40]). Methods based on HMMs were still limited, especially on the size of the generated sequence. The generation of longer sequences was made possible with deep learning via the use of recurrent neural networks (RNNs) [5, 66, 60] and then an extension of RNNs with networks known as *Long Short Time Memory* (LSTM) [16], which are particularly well suited to the analysis of time series, to learn and memorise past events.

4.2 Designing interactions with ML

In this section we focus on the analysis of how users interact with the proposed systems from different perspectives: whether it is their ability to act on the system, their inclusion in the design process, or the way they evaluate the systems.

4.2.1 Users' control on the system

We analysed the extent to which the user had control over the learning models embedded in the musical system. We define control as the ability of the user to act and influence the ML model behavior. We observed that this is characterised by changing either the training data or the model parameters such as the learning rate or the architecture of the model. The Table 1 represents the percentage of items in which a user has the ability to modify the model he or she is using, either on the model's training data or its parameters.

There is a general trend, with 54% of the publications, where the proposed approaches do not allow users to act on the training data or on the model parameters, in other words, most papers propose a "black box" system for users. We believe that this result may come from the fact that it is difficult to have interpretable parameter changes for a user on a deep learning model, and that training is often done on large dataset where the user is less often left with the possibility to personalise it. That said, these articles also include a minority of publications (16% for model parameters and 19% for model training data) that do not explicitly state whether such interactions with training or model parameters are possible. This lack of information is partly due to articles presenting a system built and used by the authors themselves, which do not always document certain parts of the system design.

Users' control	**Authorized**		
	Not specified	**No**	**Yes**
Parameters	16% (11)	54% (37)	30% (21)
Training data	19% (13)	54% (37)	27% (19)

Table 1: Percentage distribution of articles according to the possibility left to the users to modify the model by acting on its parameters or its training data

4.2.2 Inclusion of the user in the design process

We analysed the extent to which the user was included in the process of design, conception or evaluation of the interactive musical system or instrument based on ML methods. This analysis is reported in Table 2. We found that 43% of the publications do not include users in the system design process or giving no information to suggest that they do.

None	**Begin**	**Middle**	**End**
43% (29)	15% (10)	13% (9)	29% (20)

Table 2: Percentage distribution of articles defining at what point in the process of designing and conceiving the contribution, the users intervened

When users are included in the development of the proposed system, it is mostly at the end of the process (with 29% of the publications). In this context, one or more people are invited to evaluate the proposed system and then

to give feedback on it, either through questionnaires or through discussion with the authors (more details on the evaluation methods are presented in Section 4.3.3).

Including users in the middle of the design process was found in only 13% of the publications reviewed. This usually results in the evaluation of prototypes at intermediate stages of system development so that the feedback given can be put into practice in the next design iteration. In this way, the evolution of the system can be reported in the research documentation. These papers therefore use a user-centred design (UCD) as their methodology. More precisely, we can relate this approach with *technology probe* methods [37], where the objective is to put parts of the system to be designed in the hands of users in order to collect feedback on them and to better understand interaction phenomena in an ecological context.

Finally, 15% of the publications include a user or a group of users in the whole design process. Two categories can be distinguished. (1) When the authors are distinct from the users. For example, Jordà et al. [40] following an UCD approach, conducted preliminary studies in the form of interviews in the target community of practice in order to collect suggestions for the design of the system but also to learn more about the musical practice of this community and their potential interest in the system. (2) When the authors are the users. Evaluation is then intrinsic to the design process, but not necessarily made explicit in the publications. The development of the author's practice is often poorly documented, i.e. these articles do not explicitly inform about the design steps that led to the presented system. In general, only the last prototype is presented in the article. This practice is related to research-creation [11], a research approach combining creative and academic research practices.

4.2.3 Evaluation

We have analysed the evaluation methods used in the articles, categorising them as qualitative evaluation, quantitative evaluation, or no evaluation. The results of this analysis can be seen in Table 3, where we aggregated the results over two 5-year periods. We note that a significant part of the contributions do not carry out any evaluation or at least no explicit evaluation, as often some systems are developed within the practice of the author who therefore does not necessarily describe the development process based on what he/she has assessed as limitations or advantages of the system.

Years	Qualitative	Quantitative	None
2012-2016	33% (9)	52% (14)	15% (4)
2017-2022	36% (16)	22% (10)	42% (19)

Table 3: Distribution of the percentage of articles according to the type of evaluation carried out between 2012 and 2016 and between 2017 and 2022

On the one hand, 52% of the papers published between 2012 and 2016 use quantitative evaluation, while they are 22% between 2017 and 2022. This type of evaluation generally allows the performance of the proposed ML model to be monitored, thanks to a training and test data set, or via a predefined metric such as the *accuracy* [88, 94, 35]. This type of evaluation is less represented in the last five years. On the other hand, 33% of the papers published between 2012 and 2016 use qualitative evaluation and 36% between 2017 and 2022. This type of evaluation includes sessions of use of the system or artistic performances by a person or group of people. Users give feedback either through interviews [64], questionnaires [46] or through feedback more or less informally [33]. The number of articles using this method is balanced over the two 5-year periods reported in the Table 3.

To summarise, the NIME community develops systems that rarely allow the user to modify and influence the behaviour of the ML model, but also to modify and influence the design process. In addition, the publications produce fewer evaluations of the proposed systems, especially at the quantitative level.

4.3 Applications and terminology used

This section analyse how the authors of the publications perceive and represent the systems they are developing through terminology analysis, and for what purpose these systems are used based on the application analysis.

4.3.1 The different applications

We have analysed the applications involved in the articles of the literature review using the taxonomy proposed by Scurto [69] and based on the interaction paradigms as presented by Beaudouin-Lafon for Human-Computer Interaction (HCI) [3]. This taxonomy involves three families of applications:

1. **Music information retrieval**: ML is seen as a tool. The model is used to automate tasks related to musical data. For example, annotating sounds according to the emotions it provokes [78] or according to the type of instrument used [48, 80]. **19**% of the selected publications belong to this category.

2. **Artificial creativity**: ML is seen as a partner. Here the model is used to automatically generate new sound sequences. For example, the generation of realistic drum sequences [61], MIDI sequences according to the styles, genres or composers on which the model is trained [53], or even MIDI sequences associated with ways of playing them via rhythms to specify timing and expressive dynamics [27]. **16**% of the selected publications belong to this category.

3. **Human-machine improvisation**: ML is seen as a medium. Here the model is used to generate sounds but adapts in real time to the user's musical data in order to create a dynamic process of interaction. The system can thus be seen as musically expressive in order to be able to improvise with the human. Scurto [69] differs from the original definition given by Beaudouin-Lafon [3] and talks about *reflexive medium*. For example, through real-time motion-sound mapping [62, 75], or by creating an agent capable of synthesising sounds or generating melodies according to the way the musician plays his own instrument [16, 5]. **65**% of the selected publications belong to this category.

From this analysis, we see that the majority of the literature considers ML as a medium, focusing on creating processes where the system adapts and reacts to the sounds, gestures, and other ways of interacting of the user. This type of application thus opens up the possibility of building interactive music systems that support embodied forms of human expression. This is specific to the NIME community, which aims to create systems that make performative musical practice possible and opens new prospects to build

interactive music systems that supports embodied forms of human expression.

4.3.2 Terminology used and associated objectives

We analysed the terminology used to refer to ML methods in the selected articles and grouped them into two categories associated with the way technology can be perceived: Technology as **technical object** and Technology as **musical object**.

Technology as technical object. It refers to system that must achieve objective tasks. In this first category, we group together articles which use the following terminologies: "*system*" [87, 59, 28, 56, 76, 18, 31, 33, 6, 7, 70, 90, 24, 35, 14, 82, 4, 23, 34], "*method*" [96, 83, 19], "*technique*" [88], "*algorithm*" [48, 38], "*instrument*" [51, 74, 75, 32, 60, 54, 73, 26, 2], ou "*model*"[93, 61, 47, 44]. Around 77% of the articles use these different terminologies. They are accompanied by objectives in terms of performance and quality assessment in the accomplishment of the task aimed by the tool. For example, these goals may be to provide a generator of rhythmic sequences that are as realistic as possible [61], or to produce a system that efficiently sorts a bank of sounds according to the user's tastes [24]. These terminologies are used in music information retrieval applications and also sometimes in the case of applications related to artificial creativity (see Section 4.3.1).

Technology as musical object. It refers to system with which users can dialogue. The intention here is to create a musical relationship between the user and the technology, whether for the purpose of improvisation, live performance, or exploration. In this category, we group together articles that use the following terminologies to designate ML methods: "*agent*" [16, 94, 77, 5, 40], "*AI*" [27, 53, 80] or "*companion*" [94]. Around 23% of the articles use these different terminologies. In this case, they imply the will to ascribe some agency to the machine in the musical process more than a tool as described by the human-machine improvisation applications in Section 4.3.1. This may refer to methods of sound mapping with gestural inputs or predefined parameters [32], co-creation [27] or collaboration [16].

To summarise, the vast majority of publications (around 77%) use tool-related terminologies to talk about ML technologies, regardless of the type of application that are diverse in these cases. The different terminology used are very much related to how the systems developed are perceived by the authors which may impact how the system interaction is designed. But these aspects are rarely explicitly described in the articles and may deserve to be better investigate.

However, for the few publications that use terminologies associated with a technology as an musical object, the type of application is generally associated with human-machine improvisation process. For example, Benetatos et al. [5] talk about a computer agent that improvises with a performer in real time, Erdem et al. [16] talk about a shared exploration between a human performer and an artificial agent for an interactive performance, or Gillick et al. [27] use an *AI* for co-creation.

5. DISCUSSION

In this section, we propose to discuss in more depth three aspects that we find interesting in view of the literature review: the practice required to have expressive models, the factors that limit interactivity and the limitations of the review that call for further studies.

5.1 Expressive models call for practice

We have shown in Section 4.1 the transition from a limited use of deep learning models between 2012 and 2016, to a predominant use between 2017 and 2022. This transition does not seem to be the effect of the emergence of new musical tasks but is linked to the increasing complexity of these tasks. For example, in motion-sound mapping applications, deep learning has made it possible to move from the exploration of a limited number of manually selected descriptors [22] to the possibility of continuous exploration with a large number of descriptors automatically extracted by the model, as for example in gesture recognition from video streams [59].

However, their integration into a long-term musical practice remains an open question because there is a compromise between being able to do increasingly complex tasks and the limits this raises in terms of the practical complexity of use, such as the need for large dataset, the difficulty to understand the behavior of the model, the biases that can be implicitly learned in the training, among others. Getting to grips with these systems takes time, in the same way that learning a musical instrument takes time. This temporality, maybe specific to NIME, may also contribute to the fact the community uses only a limited number of deep learning techniques (e.g. VAE, RNN, MLP) compared to published ML models. Some authors use personal architecture of GAN [74] or Transformer [61, 53] but they are rare and necessitate knowledge to be able to personalize and play with the architecture of the models and adapt them to specific needs. We assume that the community primarily values models for which tools are available and robust, enabling the development of practice, which does not yet seem to be the case for a large part of deep learning models.

5.2 The limits of interactivity

Although a majority of the publications studied in this review give the user relatively few possibilities to act on the training data and the model parameters, 27% of the publications propose systems where this is allowed. These articles highlight the value of exposing users, without a technical background, to the underlying learning mechanisms in order to infer greater creative possibilities. These forms of interaction with ML has been formalised under the name Interactive Machine Learning (IML) in HCI, and ML-mediated musical expression was part of the genesis of this endeavour [1]. IML involves the user in the process of selecting, creating and labelling examples and then parameterising the model. Several contributions have aimed at developing interfaces that allow easy development of interactive ML models in music making, such as Wekinator [21], which is a tool still used by the community. However, the increasing complexity of deep learning models creates disincentives for their use in an interactive context for discovery, exploration, and shaping.

Indeed, paradoxically, the use of deep learning has enabled more complex interactions to be created but has, at the same time, reduced user agency over the models. These become designed to produce automated processing and the user has rarely access to how these models are formed. For example, this implies that the training phase, which takes place offline, remains opaque to the user, which can sometimes limit her understanding of the system's behaviour. There are technical options to partially overcome these limitations, for example by using *transfer learning*, where part of the model is pre-trained and another part is refined according to specific needs. This solution remains a compromise: one part of the model remains a black box, with its biases and flaws, while another part will be adapted to the

user's wishes. Further study of such trade-offs (including their ethical and political implications) would then be necessary.

5.3 Limitations

In general, this review suffers from a double penalty: the choice of features used for the analysis, which are not exhaustive, and the lack of information given in the articles studied on a large number of extracted and analysed features. For example, with regard to the techniques used, the majority of articles give information on the type of model used but generally without giving further details and sometimes even giving only the name of a toolkit used. By considering the features about the techniques used and the design of the interaction (see the items 1. and 2. in Section 3.3), around 65% of the articles do not explicitly provide information on at least one feature analyzed in the review. This lack of documentation has already been raised in NIME community by Calegario et al. [9]. The authors argue that the presence of a documentation such as detailed text description, source code or models is a necessary process to make the system "live", evolve and reuse by the community. We believe that this documentation effort should be accompanied by good practice in ML, and previous work on the construction and documentation of datasets [25] and models [58] is a good starting point for the community to build on. Finally, by considering only paper proceedings, the analysis deprives itself of an important part of NIME's practices, whether in musical performances, sound installations, or pedagogical workshops. Some of the knowledge of the NIME community does not come from scientific articles but from practices. This would require going beyond the framework of systematic review to open the analysis to other archives. To go beyond the articles and better understand the practices around ML in the community, we conducted a thematic analysis based on a series of interviews with seven researchers and artists from the community [41].

6. ETHICS STATEMENT

Within the literature review only publicly available resources were used. All the publications that have been used for the reported work were cited.

7. CONCLUSION

In this literature review we have looked at how previous works involving ML in instrument making and musical interaction. This literature helped us to understand the types of models and tasks used, the involvement of 'users' in the design process of the musical instrument, and finally the interaction styles used. Given the increasing importance of ML models in the systems developed, we believe that there is an interest in the NIME community, and the wider HCI community, to question the design of interactions with ML and in particular deep learning models so that they can be more inclusive of users. This inclusion can take place in design and evaluation methodologies, as well as in interaction techniques that improve their agency over systems, making them more transparent and explainable. This literature review provides research avenues in the current theme of integrating ML technologies in an interactive and creative context.

8. ACKNOWLEDGMENTS

This research was supported by the ARCOL project (ANR-19-CE33-0001) – Interactive Reinforcement Co-Learning, from the French National Research Agency. We want to acknowledge and thank everyone involved in each stage of the research. We want to express our sincere gratitude to the anonymous reviewers for their constructive comments.

9. REFERENCES

[1] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, 35(4):105–120, 2014.

[2] M. Astrinaki, N. D'Alessandro, and T. Dutoit. Mage -a platform for tangible speech synthesis. In *NIME*, 2012.

[3] M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces*, pages 15–22, 2004.

[4] M. Ben-Asher and C. Leider. Toward an emotionally intelligent piano: Real-time emotion detection and performer feedback via kinesthetic sensing in piano performance. NIME, 2013.

[5] C. Benetatos, J. VanderStel, and Z. Duan. BachDuet: A Deep Learning System for Human-Machine Counterpoint Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020.

[6] D. Bennett, P. Bennett, and A. Roudaut. Neurythmic: A rhythm creation tool based on central pattern generators. NIME, 2018.

[7] P. Beyls. Motivated learning in human-machine improvisation. NIME, 2018.

[8] J. Bullock and A. Momeni. Ml.Lib: Robust, cross-platform, open-source machine learning for max and pure data. NIME, 2015.

[9] F. Calegario, J. Tragtenberg, C. Frisson, E. Meneses, J. Malloch, V. Cusson, and M. M. Wanderley. Documentation and Replicability in the NIME Community. In *NIME*, 2021.

[10] B. Caramiaux and A. Tanaka. Machine learning of musical gestures: Principles and review. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 513–518, 2013.

[11] O. Chapman and K. Sawchuk. Creation-as-research: Critical making in complex environments. *RACAR: Canadian Art Review*, 40, 2015.

[12] K. Crawford. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press, 2021.

[13] P. Cunningham and S. J. Delany. k-nearest neighbour classifiers - a tutorial. *ACM Computing Surveys*, 54(6):1–25, 2022.

[14] N. Derbinsky and G. Essl. Exploring reinforcement learning for mobile percussive collaboration. NIME, 2012.

[15] M. O. DeSmith, A. Piepenbrink, and A. Kapur. SQUISHBOI: A multidimensional controller for complex musical interactions using machine learning. NIME, 2020.

[16] C. Erdem, B. Wallace, and A. R. Jensenius. CAVI: A coadaptive audiovisual Instrument–Composition. In *NIME*, 2022.

[17] T. Evgeniou and M. Pontil. Support vector machines: Theory and applications. volume 2049,

pages 249–257, 01 2001.

[18] A. Faitas, S. E. Baumann, T. R. Naess, J. Torresen, and C. P. Martin. Generating convincing harmony parts with simple long short-term memory networks. NIME, 2019.

[19] S. Fasciani and L. Wyse. A self-organizing gesture map for a voice-controlled instrument interface. 2013.

[20] S. S. Fels and G. E. Hinton. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE transactions on Neural Networks*, 4(1):2–8, 1993.

[21] R. Fiebrink, D. Trueman, P. R. Cook, et al. A meta-instrument for interactive, on-the-fly machine learning. *NIME*, 2009.

[22] J. Françoise, N. Schnell, R. Borghesi, and F. Bevilacqua. Probabilistic models for designing motion and sound relationships. NIME, 2014.

[23] O. Fried and R. Fiebrink. Cross-modal sound mapping using deep learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 531–534, 2013.

[24] O. Fried, Z. Jin, R. Oda, and A. Finkelstein. AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 281–286, 2014.

[25] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. M. Wallach, H. D. III, and K. Crawford. Datasheets for datasets. *CoRR*, 2018.

[26] N. Gillian and J. A. Paradiso. Digito: A Fine-Grain gesturally controlled virtual musical instrument. NIME, 2012.

[27] J. Gillick and D. Bamman. What to Play and How to Play it: Guiding Generative Music Models with Multiple Demonstrations. In *NIME*, 2021.

[28] A.-M. Gioti. A compositional exploration of computational aesthetic evaluation and AI bias. 2021.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[30] M. Graf and M. Barthet. Mixed reality musical interface: Exploring ergonomics and adaptive hand pose recognition for gestural control. PubPub, 2022.

[31] J. Gregorio and Y. Kim. Augmenting parametric synthesis with learned timbral controllers. NIME, 2019.

[32] J. Gregorio and Y. E. Kim. Evaluation of Timbre-Based Control of a Parametric Synthesizer. In *NIME 2021*, 2021.

[33] L. Hantrakul. GestureRNN: A neural gesture system for the Roli Lightpad Block. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 132–137, 2018.

[34] F. Hashimoto and M. Miura. Operating sound parameters using markov model and bayesian filters in automated music performance. In *14th International Conference on New Interfaces for Musical Expression*, 2014.

[35] J. Hochenbaum and A. Kapur. Drum Stroke Computing: Multimodal Signal Processing for Drum Stroke Identification and Performance Metrics. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2012.

[36] A. Hunt, M. M. Wanderley, and R. Kirk. Towards a model for instrumental mapping in expert musical interaction. In *ICMC*, 2000.

[37] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, et al. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24, 2003.

[38] S. C. Ianigro and O. Bown. Exploring Continuous Time Recurrent Neural Networks through Novelty Search. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 108–113, 2018.

[39] B. Jijo and A. Mohsin Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2:20–28, 01 2021.

[40] S. Jordà, D. Gómez-Marín, Ángel Faraldo, and P. Herrera. Drumming with style: From user needs to a working prototype. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 365–370, 2016.

[41] T. Jourdan and B. Caramiaux. Culture and politics of machine learning in nime: A preliminary qualitative inquiry. *NIME*, 2023.

[42] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, 2018.

[43] C. Kiefer. Musical instrument mapping design with echo state networks. NIME, 2014.

[44] T. Kim and S. Weinzierl. Modelling Gestures in Music Performance with Statistical Latent-State Models. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 427–430. NIME, 2013.

[45] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2013.

[46] N. Klügel, T. Becker, and G. Groh. Designing Sound Collaboratively Perceptually Motivated Audio Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 327–330, 2014.

[47] N. Klügel and G. Groh. Towards mapping timbre to emotional affect. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 525–530. NIME, 2013.

[48] A. Kobayashi, R. Anzai, and N. Tokui. ExSampling: a system for the real-time ensemble performance of field-recorded environmental sounds. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *NeurIPS*, 2012.

[51] M. Lee. Entangled: A multi-modal, multi-user interactive instrument in virtual 3D space using the smartphone for gesture control. PubPub, 2021.

[52] B. Levy, G. Bloch, and G. Assayag. Omaxist dialectics: Capturing, visualizing and expanding improvisations. NIME, 2012.

[53] J. A. T. Lupker. Score-Transformer: A Deep Learning Aid for Music Composition. In *NIME*, 2021.

[54] M. J. Macionis and A. Kapur. Sansa: A modified sansula for extended compositional techniques using machine learning. NIME, 2018.

[55] C. P. Martin and J. Torresen. An interactive musical prediction system with mixture density recurrent neural networks. NIME, 2019.

[56] L. McCallum and M. S. Grierson. Supporting interactive machine learning approaches to building musical instruments in the browser. 2020.

[57] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru. Model cards for model reporting. *CoRR*, 2018.

[58] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.

[59] T. Murray-Browne and P. Tigas. Latent Mappings: Generating Open-Ended Expressive Mappings Using Variational Autoencoders. In *NIME*, 2021.

[60] T. R. Naess and C. P. Martin. A physical intelligent instrument using recurrent neural networks. NIME, 2019.

[61] T. Nuttall, B. Haki, and S. Jorda. Transformer Neural Networks for Automated Rhythm Generation. In *NIME*, 2021.

[62] E. Nyström. Intra-Actions: Experiments with Velocity and Position in Continuous Controllers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020.

[63] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, and D. Moher. Updating guidance for reporting systematic reviews: development of the prisma 2020 statement. *Journal of Clinical Epidemiology*, 134:103–112, 2021.

[64] S. T. Parke-Wolfe, H. Scurto, and R. Fiebrink. Sound Control: Supporting Custom Musical Interface Design for Children with Disabilities. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 192–197, 2019.

[65] M. Popel, M. Tomkova, J. Tomek, L. Kaiser, J. Uszkoreit, O. Bojar, and Z. Žabokrtský. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 2020.

[66] R. Proctor and C. P. Martin. A laptop ensemble performance system using recurrent neural networks. NIME, 2020.

[67] J. C. Schacher, C. Miyama, and D. Bisig. Gestural electronic music using machine learning as generative device. NIME, 2015.

[68] M. Schedel, J. Ho, and M. Blessing. Women's labor: Creating NIMEs from domestic tools. NIME, 2019.

[69] H. Scurto. *Designing With Machine Learning for Interactive Music Dispositifs*. PhD thesis, Sorbonne Université, 2019.

[70] H. Scurto, F. Bevilacqua, and J. Françoise. Shaping and Exploring Interactive Motion-Sound Mappings Using Online Clustering Techniques. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 410–415, 2017.

[71] A. Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *CoRR*, 2018.

[72] B. D. Smith and G. E. Garnett. Unsupervised play: Machine learning toolkit for max. NIME, 2012.

[73] J. Snyder and D. Ryan. The Birl: An Electronic Wind Instrument Based on an Artificial Neural Network Parameter Mapping Structure. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 585–588, 2014.

[74] K. Tahiroğlu, M. Kastemaa, and O. Koli. AI-terity 2.0: An autonomous NIME featuring GANSpaceSynth deep learning model. 2021.

[75] K. Tahiroğlu, M. Kastemaa, and O. Koli. Al-terity: Non-Rigid Musical Instrument with Artificial Intelligence Applied to Real-Time Audio Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020.

[76] A. Tanaka, B. Di Donato, M. Zbyszynski, and G. Roks. Designing gestures for continuous sonic interaction. NIME, 2019.

[77] N. J. W. Thelle and P. Pasquier. Spire Muse: A Virtual Musical Partner for Creative Brainstorming. In *NIME*, 2021.

[78] M. Thorogood and P. Pasquier. Impress: A Machine Learning Approach to Soundscape Affect Classification for a Music Performance Environment. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2013.

[79] T. Tommasi, N. Patricia, B. Caputo, and T. Tuytelaars. A deeper look at dataset bias. *CoRR*, abs/1505.01257, 2015.

[80] A. Tsiros and A. Palladini. Towards a Human-Centric Design Framework for AI Assisted Music Production. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020.

[81] P. Tubaro, A. A. Casilli, and M. Coville. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1):2053951720919776, 2020.

[82] D. Van Nort, J. Braasch, and P. Oliveros. Mapping to musical actions in the filter system. NIME, 2012.

[83] F. Visi, B. Caramiaux, M. Mcloughlin, and E. Miranda. A knowledge-based, data-driven method for action-sound mapping. NIME, 2017.

[84] F. Visi and L. Dahl. Real-time motion capture analysis and music interaction with the modosc descriptor library. NIME, 2018.

[85] R. Vogl and P. Knees. An Intelligent Drum Machine for Electronic Dance Music Production and Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 251–256, 2017.

[86] C. Volioti, S. Manitsaris, and A. Manitsaris. x2Gesture: how machines could learn expressive gesture variations of expert musicians. In *New Interfaces for Musical Expression NIME2016*, 2016.

[87] N. Warren and A. Çamcı. Latent drummer: A new abstraction for modular sequencers. 2022.

[88] A. Weber, L. N. Alegre, J. Torresen, and B. C. da Silva. Parameterized Melody Generation with

Autoencoders and Temporally-Consistent Noise. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 174–179, 2019.

[89] D. Wessel, C. Drame, and M. Wright. Removing the time axis from spectral model analysis-based additive synthesis: Neural networks versus memory-based machine learning. In *ICMC*, 1998.

[90] D. J. V. Wikström. Musical composition by regressional mapping of physiological responses to acoustic features. NIME, 2014.

[91] T. Winkler. Making motion musical: Gesture mapping strategies for interactive computer music. In *ICMC*, page 26, 1995.

[92] J. Wu, M. Rau, Y. Zhang, Y. Zhou, and M. Wright. Towards robust tracking with an unreliable motion sensor using machine learning. NIME, 2017.

[93] L. Wyse and P. T. Ravikumar. Syntex: parametric audio texture datasets for conditional training of instrumental interfaces. 2022.

[94] A. Xambó, G. Roma, S. Roig, and E. Solaz. Live Coding with the Cloud and a Virtual Agent. In *NIME*, 2021.

[95] G. Xia and R. Dannenberg. Duet interaction: Learning musicianship for automatic accompaniment. NIME, 2015.

[96] Q. Yang, A. Scuito, J. Zimmerman, J. Forlizzi, and A. Steinfeld. Investigating how experienced ux designers effectively work with machine learning. In *Proceedings of the 2018 designing interactive systems conference*, pages 585–596, 2018.

[97] R. Yang, T. Chen, Y. Zhang, and G. Xia. Inspecting and interacting with meaningful music representations using vae, 2019.

[98] V. Yaremchuk, C. B. Medeiros, and M. Wanderley. Small dynamic neural networks for gesture classification with the rulers (a digital musical instrument). NIME, 2019.

[99] A. V. Zandt-Escobar, B. Caramiaux, and A. Tanaka. PiaF: A Tool for Augmented Piano Performance Using Gesture Variation Following. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 167–170, 2014.

[100] M. Zbyszyński, M. Grierson, and M. Yee-King. Rapid prototyping of new instruments with CodeCircle. NIME, 2017.

APPENDIX

A. SUMMARY OF THE RESULTS

Table 4: Summary of results by year, model type and task type. Abbreviations used: LSTM (Long Short Time Memory), MDN (Mixture Density Network), VAE (Variational Auto-Encoder), HMM (Hidden Markov Model), KNN (K-Nearest Neighbors), CNN (Convolutionnal Neural Network), VGG (Visual Geometry Group), MLP (Multi Layer Perceptron), GAN (Generative Adversarial Network), RNN (Recurrent Neural Network), HHMM (Hierarchical Hidden Markov Model), CPG (Central Pattern Generator), RL (Reinforcement Learning), SVM (Support Vector Machine), DT (Decision Tree), GVF (Gesture Variation Follower), GRT (Gesture Recognition Toolkit), LR (Linear Regression), MLR (Multiple Linear Regression), LDA (Linear Discriminant Analysis), GMM (Gaussian Mixture Model), ESN (Echo State Network), SMO (Sequential Minimal Optimization), GA (Genetic Algorithm)

Author	Year	Shallow or Deep Learning	Type of model	Type of task
Erdem et al. [16]	2022	Deep	LSTM, MDN	Generation
Warren et al. [87]	2022	Deep	VAE, HMM	Generation
Wyse et al. [93]	2022	Deep	Not specified	Generation
Graf et al. [30]	2022	Shallow	Wekinator, KNN	Classification
Nuttall et al. [61]	2021	Deep	Transformer	Generation
Lee et al. [51]	2021	Deep	CNN (VGG-Style)	Classification
Xambó et al. [94]	2021	Deep	MLP	Classification
Lupker et al. [53]	2021	Deep	Transformer	Generation
Gillick et al. [27]	2021	Deep	VAE, LSTM	Generation
Murray-Browne et al. [59]	2021	Deep	VAE	Generation
Gioti et al. [28]	2021	Deep	MLP	Classification,Regression
Tahiroğlu et al. [74]	2021	Deep	GAN	Generation
Gregorio et al. [32]	2021	Deep	AE + Regressor	Regression
Thelle et al. [77]	2021	Shallow	Not specified	Classification
Benetatos et al. [5]	2020	Deep	RNN	Generation
McCallum et al. [56]	2020	Deep	Not specified	Classification-Regression
Kobayashi et al. [48]	2020	Deep	CNN (MobileNetV2)	Classification
Tahiroğlu et al. [75]	2020	Deep	GAN (GANSynth)	Generation
DeSmith et al. [15]	2020	Deep	Wekinator (NN)	Regression
Nyström et al. [62]	2020	Deep	NeuralNet(SuperCollider)	Classification
Tsiros et al. [80]	2020	Not specified	Not specified	Classification, Generation
Proctor et al. [66]	2020	Deep	RNN	Generation
Næss et al. [60]	2019	Deep	RNN	Generation
Yaremchuk et al. [98]	2019	Deep	RNN,MLP	Classification
Weber et al. [88]	2019	Deep	VAE	Generation
Tanaka et al. [76]	2019	Deep	HHMM,MLP (Wekinator)	Regression
Parke-Wolfe et al. [64]	2019	Shallow,Deep	KNN,MLP	Classification,Regression
Martin et al. [55]	2019	Deep	LSTM	Regression
Faitas et al. [18]	2019	Deep	LSTM	Generation
Schedel et al. [68]	2019	Deep	Wekinator	Regression
Gregorio et al. [31]	2019	Deep	LSTM - Autoencoder	Regression
Yang et al. [97]	2019	Deep	VAE	Generation
Macionis et al. [54]	2018	Shallow	Wekinator	Classification
Ianigro et al. [38]	2018	Deep	Novelty Search + RNN	Generation
Hantrakul et al. [33]	2018	Deep	LSTM, Wekinator	Generation
Visi et al. [84]	2018	Not specified	Wekinator	Regression
Bennett et al. [6]	2018	Not specified	Central Pattern Generator (CPG)	Generation
Beyls et al. [7]	2018	Deep	RL (Q-learning)	Generation
Wu et al. [92]	2017	Shallow	KNN, SVM, DT, MLP	Classification
Zbyszyński et al. [100]	2017	Not specified	Not specified	Regression
Visi et al. [83]	2017	Deep	GVF	Regression
Vogl et al. [85]	2017	Deep	Not specified	Generation
Scurto et al. [70]	2017	Shallow	GMM	Regression
Jordà et al. [40]	2016	Shallow	HMM	Generation
Volioti et al. [86]	2016	Shallow	HMM	Classification
Schacher et al. [67]	2015	Shallow, Deep	GRT, GVF	Classification
Bullock et al. [8]	2015	Shallow, Deep	GRT, GVF	Regression, Classification
Xia et al. [95]	2015	Shallow	LR	Regression
Wikstr´om et al. [90]	2014	Shallow	MLR	Regression
Fried et al. [24]	2014	Shallow	LDA, Kernlized Sorting, K-means	Classification
Kl´ugel et al. [46]	2014	Shallow	Generative Topographic Map, KNN, Kmeans	Generation
Françoise et al. [22]	2014	Shallow	GMM, HMM	Regression
Zandt-Escobar et al. [99]	2014	Shallow	HMM	Classification, Recognition
Kiefer et al. [43]	2014	Shallow	ESN, RNN	Classification
Snyder et al. [73]	2014	Deep	Wekinator, MLP	Classification
Hashimoto et al. [34]	2014	Shallow	Markov model	Generation
Fried et al. [23]	2013	Deep	Autoencoder	Classification
Thorogood et al. [78]	2013	Shallow	MLR	Classification, Regression
Kl´ugel et al. [47]	2013	Shallow, Deep	Naives Bayes, Bayes Net, Random Forest, SVM	Classification
Kim et al. [44]	2013	Shallow	HMM	Regression
Ben-Asher et al. [4]	2013	Shallow	Bayes Classifier	Classification
Fasciani et al. [19]	2013	Shallow	LDA	Classification
Van Nort et al. [82]	2012	Shallow	GA	Regression
Derbinsky et al. [14]	2012	Shallow	RL	Generation
Gillian et al. [26]	2012	Not specified	Not specified	Classification
Smith et al. [72]	2012	Shallow	Self-Organising maps, Adaptive resonnance theory, MLP	Classification
Hochenbaum et al. [35]	2012	Shallow	MLP, DT, Naive Bayes, SVM, SMO, LR	Classification
Levy et al. [52]	2012	Not specified	Not specified	Generation
Astrinaki et al. [2]	2012	Shallow	HMM	Generation

Table 5: Summary of the results according to the criteria of task detail, terminology and training by the authors

Author	Task details	Terminology	Author training
Erdem et al. [16]	Parameter-Gesture Mapping	CAVI, musical agent	Complete
Warren et al. [87]	MIDI synthesis (drumming generation)	intelligent musical system, Latent Drummer	Complete
Wyse et al. [93]	Audio synthesis	model	Complete
Graf et al. [30]	Sound-Gesture Mapping	mixed reality musical interface (MRMI)	Partial
Nuttall et al. [61]	MIDI synthesis	Model	Complete
Lee et al. [51]	Sound-Gesture Mapping	Entangled, a multi-modal instrument	Complete
Xambó et al. [94]	Sound annotation	Agent, companion	Partial
Lupker et al. [53]	MIDI synthesis	an artificially intelligent assistant	Complete
Gillick et al. [27]	MIDI synthesis	AI	Complete
Murray-Browne et al. [59]	Sound-Gesture Mapping	project Sonified Body, a system	Complete
Gioti et al. [28]	Sound annotation	Bias, a computer music system	None
Tahiroğlu et al. [74]	Audio synthesis	AI-terity instrument	Complete
Gregorio et al. [32]	Parameter-Sound Mapping	Instrument prototype	Complete
Thelle et al. [77]	Sound-parameter mapping	Musical agent	Complete
Benetatos et al. [5]	MIDI synthesis	BachDuet, computer agent	Complete
McCallum et al. [56]	Parameter-Sound Mapping	IML systems, MaxiInstruments, musical system	Not specified
Kobayashi et al. [48]	Sound label	Deep Learning, deep learning inferences	Complete
Tahiroğlu et al. [75]	Audio synthesis	AI-terity instrument	Complete
DeSmith et al. [15]	Gesture-Sound Mapping	SQUISHBOI, controller	Not specified
Nyström et al. [62]	Sound-Gesture Mapping	Continuous MIDI controllers	Not specified
Tsiros et al. [80]	Intrument recognition, settings generation	Channel-AI (product name)	Not specified
Proctor et al. [66]	MIDI synthesis, volume	an assistance tool	None
Næss et al. [60]	MIDI synthesis	intelligent interactive instrument	Complete
Yaremchuk et al. [98]	Gesture classification	The Rulers, DMI	Complete
Weber et al. [88]	MIDI synthesis	machine learning technique	Not specified
Tanaka et al. [76]	Sound-Gesture Mapping	system	None
Parke-Wolfe et al. [64]	Sound-Gesture Mapping	Toolkit	None
Martin et al. [55]	Parameter-Gesture Mapping	DMI	Complete
Faitas et al. [18]	MIDI synthesis	System	Complete
Schedel et al. [68]	Parameter-Gesture Mapping	NIME	Not specified
Gregorio et al. [31]	Parameter-Sound Mapping	System	Complete
Yang et al. [97]	MIDI synthesis	method	Complete
Macionis et al. [54]	Timbre et speech recognition	Sansa, hyper-instrument	Not specified
Ianigro et al. [38]	Audio synthesis	algorithm, search algorithm	Not specified
Hantrakul et al. [33]	Synthesis of position/parameter	System	Complete
Visi et al. [84]	Parameter-Gesture Mapping	modosc	Complete
Bennett et al. [6]	Audio synthesis (Rythms)	Neurythmic, interactive system	Not specified
Beyls et al. [7]	MIDI synthesis	Pock, system	Not specified
Wu et al. [92]	Sound-Gesture Mapping	Embodied Sonic Meditation	Complete
Zbyszyński et al. [100]	Sound-Gesture Mapping	CodeCircle	Not specified
Visi et al. [83]	Sound-Gesture Mapping	Method	Complete
Vogl et al. [85]	MIDI synthesis rythmic patterns	Prototype interface	Complete
Scurto et al. [70]	Sound-Gesture Mapping	Tool, system	Not specified
Jordà et al. [40]	MIDI synthesis rythmic patterns	generative drumming agent	Not specified
Volioti et al. [86]	Sound gesture mapping	x2Gesture, engine	None
Schacher et al. [67]	Sound-Gesture Mapping	workbench	Not specified
Bullock et al. [8]	Open-source toolkit	ml:lib (set of opensource tools)	Not specified
Xia et al. [95]	Timing prediction	an artificial performer	Complete
Wikstr^{f}om et al. [90]	Sound-Gesture Mapping	System	Not specified
Fried et al. [24]	Sound annotation	AudioQuilt , system	Complete
Kl^{f}ugel et al. [46]	Parameter-Sound Mapping	prototype	Complete
Françoise et al. [22]	Sound-Gesture Mapping	prototype application	None
Zandt-Escobar et al. [99]	Sound-Gesture Mapping	PiaF, augmented piano	Complete
Kiefer et al. [43]	Parameter-Gesture Mapping	open source library	Complete
Snyder et al. [73]	Parameter-Gesture Mapping	The Birl, instrument	None
Hashimoto et al. [34]	Parameter synthesis	system	Complete
Fried et al. [23]	Sound-Gesture Mapping	System	Complete
Thorogood et al. [78]	Sound annotation	Impress system	Not specified
Kl^{f}ugel et al. [47]	Sound annotation	Model	Complete
Kim et al. [44]	Gesture Detection	Model	Complete
Ben-Asher et al. [4]	Emotion-Gesture Mapping	System	Complete
Fasciani et al. [19]	Parameter-Gesture Mapping	method	None
Van Nort et al. [82]	Sound-Gesture Mapping	FILTER, system	Complete
Derbinsky et al. [14]	MIDI synthesis	System	Partial
Gillian et al. [26]	Sound-Gesture Mapping	Digito, virtual musical instrument	Not specified
Smith et al. [72]	ML toolkit	toolbox	Not specified
Hochenbaum et al. [35]	Gesture recognition	System	Complete
Levy et al. [52]	Audio synthesis et MIDI	OMax, software	Not specified
Astrinaki et al. [2]	speech synthesis	MAGE platform + instrument	Complete

Table 6: Summary of results according to criteria of target user, user intervention, modifiable parameters, user training

Author	Target user	User intervention	Modifiable parameters	User training
Erdem et al. [16]	Author's practice	Not specified	No	No
Warren et al. [87]	Author's practice	Not specified	Not specified	Not specified
Wyse et al. [93]	Sound designers, engineers	End of process	Yes	Not specified
Graf et al. [30]	Musicians	Not specified	No	Yes (fine tuning)
Nuttall et al. [61]	Musician , No expert en ML	End of process	No	No
Lee et al. [51]	Not specified	End of process	No	No
Xambó et al. [94]	Livecoder	Middle of process	Yes	Yes
Lupker et al. [53]	Musician	Middle of process	Yes	Yes (fine tuning)
Gillick et al. [27]	Musician	Not specified	No	No
Murray-Browne et al. [59]	Musician, Composer, artist	Middle of process	No	No
Gioti et al. [28]	Musician	Middle of process	No	Yes
Tahiroğlu et al. [74]	Author's practice	Not specified	No	No
Gregorio et al. [32]	Sound designer and musicians	End of process	No	No
Thelle et al. [77]	Instrumentalist	End of process	Yes	No
Benetatos et al. [5]	Musician	End of process	No	No
McCallum et al. [56]	Musician	Not specified	Yes	Yes
Kobayashi et al. [48]	Performers	No user in the process	No	No
Tahiroğlu et al. [75]	Author's practice	Not specified	No	No
DeSmith et al. [15]	DMI practitioners	No user in the process	Yes	Yes
Nyström et al. [62]	Author's practice	Not specified	No	No
Tsiros et al. [80]	Sound engineers	End of process	No	No
Proctor et al. [66]	Group of performer	End of process	Yes	No
Næss et al. [60]	DMI practitioners	End of process	Yes	No
Yaremchuk et al. [98]	Not specified	Not specified	No	No
Weber et al. [88]	Pianist	End of process	Yes	No
Tanaka et al. [76]	Performer	End of process	Yes	Yes
Parke-Wolfe et al. [64]	Teacher and therapist	Beginning of process	Yes	Yes
Martin et al. [55]	DMIers	End of process	No	Yes
Faitas et al. [18]	Musician	No user in the process	No	No
Schedel et al. [68]	Author's practice	Beginning of process	Yes	Yes
Gregorio et al. [31]	Researchers	Not specified	Not specified	Not specified
Yang et al. [97]	Not specified	Not specified	Not specified	Not specified
Macionis et al. [54]	Not specified	End of process	No	Not specified
Ianigro et al. [38]	Not specified	Not specified	No	Not specified
Hantrakul et al. [33]	Not specified	Middle of process	No	No
Visi et al. [84]	Musician	End of process	No	Yes
Bennett et al. [6]	Composer, sound designer	End of process	No	No
Beyls et al. [7]	Not specified	Not specified	Yes	Not specified
Wu et al. [92]	Performer	No user in the process	No	No
Zbyszyński et al. [100]	Sound designers, NIMErs	Not specified	Not specified	Not specified
Visi et al. [83]	Musician	Beginning of process	No	No
Vogl et al. [85]	Musician and producer (EDM)	Middle of process	Yes	No
Scurto et al. [70]	Novice et expert	Fin de process	Yes	Yes
Jordà et al. [40]	Musician and producer (EDM)	Beginning of process	Yes	No
Volioti et al. [86]	Pianist	End of process	No	Yes
Schacher et al. [67]	Instrumentalist	End of process	No	Not specified
Bullock et al. [8]	Novices	Fin de process	Yes	Yes (offline)
Xia et al. [95]	Musician	Not specified	No	No
Wikstr´om et al. [90]	Not specified	None	No	No
Fried et al. [24]	Musician	Middle of process	No	Yes
Kl´ugel et al. [46]	Sound designer	Not specified	No	No
Françoise et al. [22]	Musician	End of process	No	Yes
Zandt-Escobar et al. [99]	Pianist	End of process	No	No
Kiefer et al. [43]	DMIers	Not specified	No	No
Snyder et al. [73]	Saxophonist	Middle of process	No	Yes
Hashimoto et al. [34]	Author's practice	Not specified	Yes	Yes
Fried et al. [23]	Not specified	Not specified	No	No
Thorogood et al. [78]	Composers	Not specified	No	No
Kl´ugel et al. [47]	Novices	Not specified	No	No
Kim et al. [44]	Not specified	None	No	No
Ben-Asher et al. [4]	Pianist	End of process	Yes	No
Fasciani et al. [19]	Not specified	End of process	Yes	Yes
Van Nort et al. [82]	Author's practice	Not specified	Not specified	Not specified
Derbinsky et al. [14]	Drummer	Not specified	Yes	Yes
Gillian et al. [26]	Musician	Not specified	Not specified	Not specified
Smith et al. [72]	No expert users	Not specified	Not specified	Not specified
Hochenbaum et al. [35]	Drummer	Not specified	Not specified	Not specified
Levy et al. [52]	Musician	Middle of process	Not specified	Not specified
Astrinaki et al. [2]	Not specified	End of process	Yes	No