Tahiroğlu, Koray; Kastemaa, Miranda; Koli, Oskar

AI-terity: Non-Rigid Musical Instrument with Artificial Intelligence Applied to Real-Time Audio Synthesis

Published in:
Proceedings of the International Conference on New Interfaces for Musical Expression

Published: 25/07/2020

Document Version
Publisher's PDF, also known as Version of record

AI-terity: Non-Rigid Musical Instrument with Artificial Intelligence Applied to Real-Time Audio Synthesis

Koray Tahiroğlu
Department of Media
Aalto University
School of ARTS
FI 00076 AALTO Finland
koray.tahiroglu@aalto.fi

Miranda Kastemaa
Department of Media
Aalto University
School of ARTS
FI 00076 AALTO Finland
miranda.kastemaa@aalto.fi

Oskar Koli
Department of Media
Aalto University
School of ARTS
FI 00076 AALTO Finland
oskar.koli@aalto.fi

ABSTRACT

A deformable musical instrument can take numerous distinct shapes with its non-rigid features. Building audio synthesis module for such an interface behaviour can be challenging. In this paper, we present the *AI-terity*, a non-rigid musical instrument that comprises a deep learning model with generative adversarial network architecture and use it for generating audio samples for real-time audio synthesis. The particular deep learning model we use for this instrument was trained with existing data set as input for purposes of further experimentation. The main benefits of the model used are the ability to produce the realistic range of timbre of the trained data set and the ability to generate *new audio samples* in real-time, in the moment of playing, with the characteristics of sounds that the performer ever heard before. We argue that these advanced intelligence features on the audio synthesis level could allow us to explore performing music with particular response features that define the instrument's *digital idiomaticity* and allow us reinvent the instrument in the act of music performance.

Author Keywords

NIME, artificial intelligence, GANSynth, non-rigid instruments, deep learning with audio

CCS Concepts

•Applied computing → Sound and music computing; Performing arts; **•Computing methodologies → Artificial intelligence;**

1. INTRODUCTION

Music is an outstanding example of cultural and mediated practices that has often been at the forefront of research in emerging technologies. Today, digital technologies and advanced computational features, e.g. deep learning and artificial intelligence (AI) tools are shaping our relationships with music as well as enabling new possibilities of utilising new musical instruments and interfaces. The complex nature of these technologies, now commonplace in our daily lives, intimately connected with machine learning models that are commonly used in practices with new musical instruments. In a domain of research and practice of NIME, the set of machine learning models applied might be referred to a general set of techniques that are used to deal with the challenges of musical interactions with a new musical instrument, intended to help us play it better, to play faster, or even to perform better in other respects. The current developments of tools are advancing such existing models, further lead to modeling new sounds, providing possibilities to generate audio efficiently and faster but at the same time interesting results from the trained data set of samples in audio domain. Deep learning with audio shifts the focus to next level of real-time synthesis of sound by creating completely *new-sounding sounds*.

Figure 1: AI-terity instrument

In this paper, we introduce *AI-terity*, a deformable, non-rigid musical instrument that comprise computational features of a particular AI model for generating relevant audio samples for real-time audio synthesis. Stiffness and physical deformability becomes an opening of the instrument's folded shape (Figure 1), generating audio samples when handheld physical action is applied. This physical manipulation causes control parameter changes in sample-based granular synthesis and new audio samples are distributed around the surface when the performer starts a direct interaction with the instrument. Being able to move through timber-changes in sonic space allows performer to access one way to *idiomatic digital relationship* [26] with sound making and control actions with AI-terity instrument. Our main focus in this paper is the real-time sound synthesis features built in with the generative AI model and the particular idiomaticity of such generative-behaviour in music making experience.

2. RELATED WORK

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

Artificial Intelligence (AI) methods applied to music can be traced back even centuries in relation to its automated algorithmic features [5, 18]. Today, the growing use of AI in music practices is well known by which the AI is employed through more advanced computational features that determine the implications of its usage and it is not limited to only automation. Since the beginning researchers have long sought how musical activity, musical cognition and knowledge can be modeled using AI applications and techniques. Some aimed at understanding the nature of compositional knowledge acquisition and listening [13], possibilities for creating melodies based on chosen emotive impact [17], replications of set of instructions to create new instances of music [4] as well as computer accompaniment music performance systems [5]. In a similar context, Roads [18] introduced *artificial musical intelligence* concept which implies looking at music though increasing body of AI strategies and methodologies. This concept, to some degree, has been reflected in NIME community through various aspects of machine learning methods used in triggering musical interactions with NIMEs.

Fiebrink et.al [9] introduced Wekinator, an application of supervised machine learning method, which could learn a model in relation to a set of input/output parameters. This method uses a trained dataset of the input/output pairs and could modify control parameters or mapping functions for sound synthesis. Wekinator has been widely used for discrete classification events in live music performances with NIMEs and in mixed-media installations [14, 21, 19]. There has also been a growing interest to provide machine learning toolbox systems that could evaluate and recognise static postures over time in different contexts - e.g. in real-time gesture transitions in music performances, performing regression and classifying temporal gestures of the performers [11]. Smith & Garnett [20] and Cont et.al [3] also presented toolbox of unsupervised machine learning algorithms, having the goal to make toolbox accessible to anyone who has the computational skills in commonly known real-time dataflow programming environments like Max and Pure Data. Similarly, one particular approach receiving notable attention is the more advanced artificial intelligence models that are able to provide automatic methods of learning for appropriate representations of inputs and outputs using deep learning methods [10], and for simulated human-robot imitation of particular music performance [25] as well as for prediction of musical events to provide non-intrusive counteractions as an accompaniment agent [23].

It would not be too much to argue that the majority of interests from developers across in applying machine learning methods to music are centered around musical gesture analysis and gestural sound control in NIME community [2]. At the same time, the recent research in artificial intelligence and developments of tools provide a new range of opportunities and challenges for audio synthesis. These are not only limited to gestures to trigger change on temporal events, but allowing us to generate *waveforms* for building audio synthesis modules. For instance, WaveNet is one of the well known generative model for raw audio waveforms [16]. This model has been mostly used for speech synthesis, but because the model itself allows training on a dataset of raw audio, it is possible to generate audio samples for musical outcome, rather than MIDI notes that indicate what to play.

Similarly, WaveGAN is another model for generating raw audio waveforms that has a particular generative architecture based on previously developed model for image synthesis [6]. Google Magenta team presents prominent research approach to raw audio generation with deep learning models specifically through their open source projects NSynth [8] and GANSynth [7]. It has been our research interest to explore and integrate generative AI models applied in audio domain into development of new musical instruments. This in turn has the potential to bring an alternative synthesis of knowledge about musical instruments, as well as enhancing the ability to focus on the sounding features of the new instruments. In our current project, our work is taking on the technical aspects of GANSynth model, which places considerable emphasis on generating various audio samples with different timbre characteristics in the moment of performing music.

Figure 2: Pressure pads attached on the inner surface of the instrument

3. AI-TERITY

In our earlier work, we developed a shape-retaining, freely-deformable interface with frequency modulation (FM) synthesis units using complex circular waveforms and continuous phase control [27, 24]. In the present work, we expanded the stiffness and the shape of the interface to include additional sample-based granular synthesis features such as velocity, step size, play-head treads and duration for each concurrent grains in real time. Al-terity is a non-rigid, flexible musical instrument that enables parameter changes for the sound producing events and the creation of multi-layer synthesiser modules. While force / pressure input remained the most effective control method for this instrument, we were able to incorporate a touch sense arrays with capacitive sensing for further control options to the audio synthesis units. Figure 2 shows the inner sensor connections of the instrument.

3.1 Control Interface

The control interface consists of two interlocking halves of 3D printed soft plastic, which fit together to form the final shape seen in Figure 1. This allows sensors to be installed inside the shape while still keeping them easily accessible for maintenance. Inside the shape there are nine sensors, laid out roughly in a grid. These custom built sensors are able to sense both pressure (using Velostat, a pressure-sensitive conductive material) and proximity (using capacitive sensing). The sensors are read by a Teensy microprocessor, which does some smoothing of the noisy signal before sending it over USB to Pure Data program running on a Mac Mini computer.

3.2 Digital Audio Synthesis Module

In this project, creating a particular type of *digital idiomaticy* to the Al-terity instrument has emerged as a critical

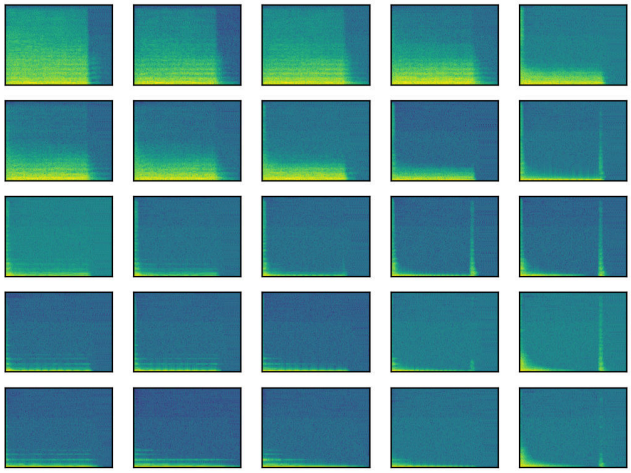

Figure 3: Spectrogram images of the samples interpolating in the latent space

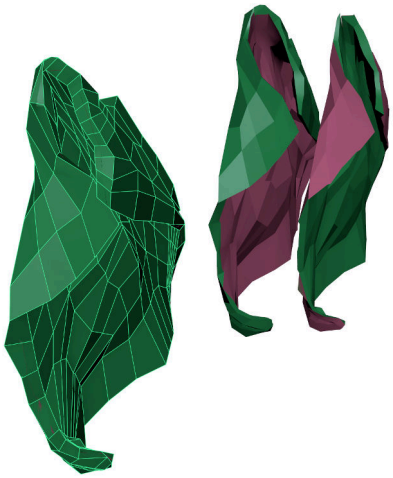

Figure 4: 3D drawings of the AI-terity instrument

factor as we define it here, however this isn't a new process [15, 22], but rather a re-defining of sound making and control actions relation and thus re-envisioning the one of the possible implementation. The ideas surrounding the physicality of the instrument appeared to be analogous to how the latent space in GANSynth model is organised as a spatial distribution of timber features of audio samples in trained data set. Figure 3 shows the interpolation of the audio samples that are organised in random order in GANSynth latent space. The shape of this instrument is based on a 2D visualisation of such high-dimensional space, which is extracted from our approach to represent the latent space as unevenly and geometrically formed surfaces of a folded paper. Unfolding that flexible shaped paper led us to create 3D drawings shown in Figure 4 and then 3D print the object with a mix of white and black photopolymer material which enabled us to set the level of stiffness. Partially folded shape of the instrument allows the performer to access the instrument in a number of different ways. We divided the control surface into three different regions and each region with three different nodes, which allowed us to distribute nine audio samples in relation to the position of the pressure pads on the instrument. These divided regions made it possible to built multi-layer synthesiser modules. It might be important to note here that the audio samples used in these modules are not created or prepared before hand. Instead, GANSynth generates an audio sample for each nine nodes positioned on the instrument each time the instrument receives a direct handheld contact on one of the three regions. Digital audio synthesis module in Al-terity instrument is developed in Pure Data programming environment.

3.2.1 Layered granular synthesis modules

The bottom-layer synthesis module consists of three granular synthesisers, each processing a separate audio sample, manipulating the granular control parameters to generate a mix of sampled signals with different timbers. The duration and the play-head position bring in a more sustained sounds controlled by the amount of the pressure input received on each pressure node in this region. The deformation of the control surface determines the characteristics of the grains to be synthesised and the pressure variation leads to very smooth and lively polyphonic sounds. The temporal features in this synthesis module can be described as a wavy drone effect. This is the result of the duration of the grains between the nodes in this region as they run at a different rate than they would do in some other types of polyphonic sampling.

The middle-layer synthesis module is another granular synthesiser with a significant difference on the output duration and with percussive sound characteristics. Control parameters define the duration of each grain, play-head position and the speed range, transposing the original pitch. The manipulation happens on single grain that is taken on various index points on the audio sample. Three pressure nodes in this mid-region sends the parameters to set the number of the grains. A new grain is defined by extending or receding the current grain index position on the audio sample. While outputting the resulted sounds, the total duration is controlled by a single pressure input variable. This brings up an opportunity to pull and manipulate the generated samples in a relatively short step and have a part of the grains sound slightly longer than the rest, but at the same time ensure that the manipulated grains stay in sync throughout.

The output of the top-layer synthesis module generates a different pulse waveform through its granular synthesiser. The folded shape of the control interface is coupled with smaller sized grains with shorter durations. When the performer opens up the folded parts, audio module responses with subtle but sharp sounds more percussive than the middle region responses. The pressure input also changes the reading index points of the audio sample. Once the performer releases the top layer, after opening up the folded part, the duration of the grains get longer and discrete sounds transforms into more continues sounds until the manipulated folded part gets back to its initial shape.

Figure 5 shows the Pure Data patch in which we integrated GANSynth module into Pure Data environment. The [pyext] external allowed us to run GANSynth python scripts through this patch and read the generated audio samples in Pure Data arrays easily. Once we load the GANSynth model from a trained checkpoint, then for each audio sample generation, we generate random latent vector for timber. Following that an audio sample is generated from each vector with a given pitch variable on a CPU computer within a five-second buffer. In this current model we could be able to condition the pitch for each generated audio sample. Audio samples are in wav format with 16kHz sample rate, mono, 4 seconds long and include 32 bits per sample.

3.2.2 GANSynth: adversarial neural audio synthesis

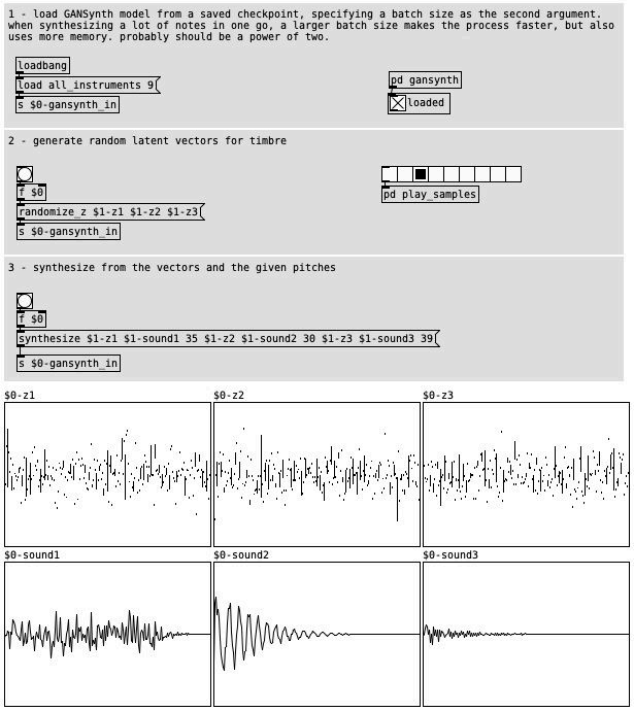

Figure 5: GANSynth integrated in Pure Data

GANSynth is an audio synthesis algorithm based on generative adversarial networks (GANs), introduced by Google Magenta team in January 2019 [7]. Like the earlier NSynth algorithm, GANSynth is designed for generating musical notes at specified pitches, but GANSynth achieves better audio quality and also synthesises thousands of times faster. The vastly improved speed makes the algorithm suitable for interactive purposes, potentially even near-real-time applications for us. Previously, autoregressive models like WaveNet (used in NSynth) represented the state of the art in neural audio synthesis. These models are good at learning the characteristics of sounds over very short time periods (local latent structure) but struggle with longer-term features (global latent structure). They are also very slow, since they have to generate waveforms one sample at a time. In contrast, GANs are capable of modeling global latent structure, as well as synthesising more efficiently [12]. However, adapting GANs to audio generation has proven challenging due to their weakness at capturing local latent structure, producing sounds that lack phase coherence. GANSynth tackles this problem by making several improvements to the network architecture and audio representation.

However, Adapting GANs to audio generation has proven challenging due to their weakness at capturing local latent structure, producing sounds that lack phase coherence. GANSynth tackles this problem by making several improvements to the network architecture and audio representation.

3.2.3 Generative adversarial networks

GANSynth deep learning model is based on generative adversarial networks, a type of generative model where two neural networks – called generator and discriminator – compete against each other [12]. The discriminator network tries to distinguish between real and generated data (e.g. images or audio samples), and is initially trained with a known data set. The generator network aims to produce data that the discriminator cannot tell apart from real data. During training, both networks become better at their respective tasks through backpropagation, resulting in a network that generates very realistic data. A popular analogy given for the generator-discriminator relationship is that of an art forger and investigator. The generator acts as a forger, trying to create convincing counterfeit pieces, while the discriminator acts as an investigator, trying to spot these counterfeits. Often, the generator and discriminator network structures mirror each other. It is common to use a deconvolutional neural network (DNN) as generator and a convolutional neural network (CNN or ConvNet) as the discriminator.

3.2.4 Architecture of GANSynth

The GANSynth network learns to represent timbre as vectors in a 512-dimensional latent space. It is also conditioned on a one-hot representation of pitch, to allow independent control of pitch and timbre in the synthesis process. The generator synthesises audio by sampling a random vector from the latent space according to a spherical Gaussian distribution, and running it through several layers of upsampling convolutions. The discriminator applies corresponding downsampling convolutions and produces an estimate of the divergence between the real and generated data. To encourage the network to make use of the pitch label, the discriminator also attempts to classify the pitch of the generated audio. The divergence estimate and pitch prediction error are combined to form the network's loss function, which is used for backpropagation during training. The network operates on a spectral representation of audio rather than directly synthesising waveforms. The full details can be found in the GANSynth paper [7], but the key elements of the representation are as follows: Magnitude and phase are computed using the short-time Fourier transform (STFT). The logarithm of the magnitude is taken to constrain its range, and the phase is unwrapped and differentiated to give a measure of "instantaneous frequency". Instantaneous frequency expresses the constant relationship between audio frequency and STFT frame frequency, addressing the phase coherence problem. The magnitude and instantaneous frequency are transformed to a mel frequency scale to achieve better separation of low frequencies.

4. REMAIN IN UNCERTAINTY

The ability to generate new audio samples using GANSynth model in real-time, with the characteristics of sounds that the performer ever heard before, and as a result changing digital audio synthesis module's behaviour continuously in the moment of playing is unusual. However, a possible close similarity could be found in the composition *S'offrir* by Emilie Girard-Charest [1]. Tuning discrepancies is considered an important act in the performance of this composition, as the tuning of the cello is continuously modified, manipluated by another musician. The sudden changes of the tuning structure put the performer in an intermittent and unconfident state of performing. At the same time, it creates an intense engagement with the instrument because it allows for the expression of the sound-producing physical structures which are layered in the fine tuners in the tailpiece. This gives performer the possibility to develop and transform a particular relationship that is idiomatic to the instrument's unusual behaviour, which allows massive flexibility and instantaneous exploration of instrument's playability.

Idiomaticity appears in many context in our relationship with musical instruments. We have argued earlier that idiomaticity is not only bounded with composition practice as a way of generating a repertoire for new digital instruments, but also emerges as a chain of influence and rather

[1] `https://vimeo.com/145068700`

a complex interplay between many actors involved in building and performing a digital musical instrument [15]. In the development phase of the Al-terity instrument, our particular assumptions on digital idiomaticity was the possibility to engage in an unusual state of playing in which the performer remains in a state of a continuous *uncertainty*.

Ursula Bertram [1] points out that unpredictable and contradictory thought processes could open up new perspectives of thoughts in art thinking. She discusses whether the uncertainty, even though it is rarely taken beyond the theoretical realm, incorporate an openness to any creative process or not. Her ideas on how norm-based systems or practices more reluctantly react to creative innovation, brings up further discussions on the necessity to develop actual uncertainty in artistic process, abandoning commonly expected notions. For instance, musicians would rather choose the instrument and perform it on the basis of its sounds, its persona, its role in some musical compositions, and their own personal aesthetic taste. In doing so they would get in touch with a physical and psychological aspect that transcends the musician's *musicianship*. If this could be taken any further, if musicians would take it as a self personalised artistic thinking process, remaining in uncertainty will create a particular form of idiomatic relationship between unfamiliar/unusual musical instrument and the musician. The process would open up musician to receive a challenge. There is no doubt that this could result in a collapse of possible options, in the absence of a pre-established practices. Further questions rise here on the contrary whether the musician would prefer to "fail" in the process of seeking to recreate a new set of relationships with an instrument. Perhaps this could be a subject of a follow up future work in our project.

We composed the piece *Uncertainty*[2], idiomatically reflecting the unusual behaviour of the Al-terity instrument, keeping the performer remain in unconfident state of performing (Figure 6). The composition provides a non-rigid but identifiable musical events followed by ever shifting new sounds with new timber. In this state of transformation, our relationship with the instrument, which previously had been intimately connected to ourselves, our musical experience and music practices we only had imagined was possible, could be severely shaken, disrupted and changed. Appearance of new sounds on the synthesis level and being able to move through timber-changes in sonic space could allow performer to explore a whole new range of musical possibilities. Moreover, this could turn into a continuous state of playing, reformulating our relationship and opening up new variety of musical demands. We argue that these advanced intelligence features on the audio synthesis level could allow us to explore performing music with particular response features that define the instrument's digital idiomaticity and allow us reinvent the instrument in the act of music performance. Reinventing the instrument might sound like an overwhelming demand, but surely we can be challenged to consider this. Here what we mean by reinventing is exploring the instrument's potential to re-shape its idiomatic features, discovering new ways of playing that could emerge from the instrument itself. Most likely this re-shaping could happen in audio synthesis modules. For instance, granular synthesis is about modifying the underlying patterns of audio samples. Particular timber of sound in this synthesis could be very fundamental to the characteristic of such digital musical instrument. Changing the characteristics of the audio samples could result in distinctive flavour for output sounds.

[2]`https://vimeo.com/388365942`

Figure 6: Al-terity in performance of the composition *Uncertainty*

5. CONCLUSIONS

In this paper we presented our AI-terity, a deformable musical instrument with an audio synthesis module that uses GANSynth deep learning model for generating audio samples. We described in detail the design, control interface and audio synthesis module of the instrument. We also discussed current architecture and the implementation of the deep learning model that we used in developing Al-terity instrument. Motivated by the results of the implementation of our current project we acknowledge the potential of generative adversarial networks for further developments of NIMEs as they are capable of generating realistic and at the same time with new characteristics of sounds for a given dataset. It is also important to mention here that more advanced research and implementation of AI models are appearing in audio domain and these tools could open up new opportunities for the practices in NIME community.

One possible direction that this project could take in the near future is dealing with the preparation of our own dataset with samples from various alternative musical instruments and train the checkpoint with that dataset to further explore the benefits of these advanced AI models. In this current version of the AI-terity instrument, we used the existing trained checkpoint provided by the Magenta team. This checkpoint sets the timber characteristics for the generated audio samples in GANSynth.

It is also in our research interest to develop further features for conditional generation of audio samples in GAN model. The current GANSynth model is conditioned to a pitch function, and this allows us to only set the pitch variable for audio samples. Even though the relationship between points in the latent space to the generated audio samples is complex, but having an opportunity to organise these points in the latent space, rather than in random order, could allow us to set further timber features to the generated audio samples. At the moment it seems more realistic to develop a conditional model in which both the generator and discriminator are conditioned on further information than pitch, such as noise information, etc..

Applying these advanced technologies in building and performing a new musical instrument, brings in further questions and arguments for music practices. We argue that explicit consideration of using advanced intelligence features on the audio synthesis level allow us to exploit diverse idiomatic features that are more embodied in performative practice of the instrument. Al-terity provides evidence on this argument by re-shaping its idiomatic features in the moment of playing.

The open source code of the project is available at `https:`

//github.com/SopiMlab/DeepLearningWithAudio. It is also important to mention that generation of audio samples with GAN models highly depends on the hardware configuration of the computers. Five-seconds buffer in our work could be reduced significantly even with only onboard or external GPUs to better fulfil real-time response expectations. This is in our near future investment plans.

6. ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (project 319946) and Aalto University A!OLE funding. We would like to acknowledge Janne Ojala's work in this project.

7. REFERENCES

[1] U. Bertram. *Artistic Transfer: Efficiency Through Unruly Thinking*. Columbia University Press, 2019.

[2] B. Caramiaux and A. Tanaka. Machine learning of musical gestures. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 513–518, Daejeon, Republic of Korea, May 2013. Graduate School of Culture Technology, KAIST.

[3] A. Cont, T. Coduys, and C. Henry. Real-time gesture mapping in pd environment using neural networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 39–42, Hamamatsu, Japan, 2004.

[4] D. Cope. Experiments in musical intelligence (emi): Non-linear linguistic-based composition. *Journal of New Music Research*, 18(1-2):117–139, 1989.

[5] R. B. Dannenberg. Artificial intelligence , machine learning , and music understanding. In *Proceedings of the Brazilian Symposium on Computer Music (SBCM2000), Curitiba, Brazil*, 2000.

[6] C. Donahue, J. McAuley, and M. Puckette. Adversarial audio synthesis. *arXiv preprint arXiv:1802.04208*, 2018.

[7] J. Engel, K. K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts. GANSynth: Adversarial neural audio synthesis. In *International Conference on Learning Representations*, 2019.

[8] J. Engel, C. Resnick, A. Roberts, S. Dieleman, M. Norouzi, D. Eck, and K. Simonyan. Neural audio synthesis of musical notes with wavenet autoencoders. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1068–1077. JMLR. org, 2017.

[9] R. Fiebrink, D. Trueman, and P. R. Cook. A meta-instrument for interactive, on-the-fly machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 280–285, Pittsburgh, PA, United States, 2009.

[10] 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. Graduate School of Culture Technology, KAIST, 2013.

[11] N. Gillian, B. Knapp, and S. O'Modhrain. A machine learning toolbox for musician computer interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 343–348, Oslo, Norway, 2011.

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.

[13] O. E. Laske and S. Drummond. Toward an explicit cognitive theory of musical listening. *Computer Music Journal*, 4(2):73–83, 1980.

[14] M. J. Macionis and A. Kapur. Where is the quiet: Immersive experience design using the brain, mechatronics, and machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 335–338, Porto Alegre, Brazil, June 2019. UFRGS.

[15] A. McPherson and K. Tahiroğlu. Idiomatic patterns and aesthetic influence in computer music languages. *Organised Sound*, 25(1):53–63, 2020.

[16] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[17] D. Riecken. Wolfgang-a system using emoting potentials to manage musical design. *Understanding Music with AI: Perspective on Musical Cognition*, 1992.

[18] C. Roads. Artificial intelligence and music. *Computer Music Journal*, 4(2):13–25, 1980.

[19] M. Schedel, P. Perry, and R. Fiebrink. Wekinating 000000Swan : Using machine learning to create and control complex artistic systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 453–456, Oslo, 2011.

[20] B. D. Smith and G. E. Garnett. Unsupervised play: Machine learning toolkit for max. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Ann Arbor, Michigan, 2012. University of Michigan.

[21] 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. Goldsmiths, University of London, 2014.

[22] K. Tahiroglu, M. Gurevich, and R. B. Knapp. Contextualising idiomatic gestures in musical interactions with nimes. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 126–131, Blacksburg, Virginia, USA, 2018.

[23] K. Tahiroglu, J. C. Vasquez, and J. Kildal. Non-intrusive counter-actions: Maintaining progressively engaging interactions for music performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 444–449, Brisbane, Australia, 2016.

[24] K. Tahiroğlu, T. Svedström, V. Wikström, S. Overstall, J. Kildal, and T. Ahmaniemi. Soundflex: designing audio to guide interactions with shape-retaining deformable interfaces. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 267–274, 2014.

[25] A. Tidemann. An artificial intelligence architecture for musical expressiveness that learns by imitation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 268–271, Oslo, Norway, 2011.

[26] S. Waters. panel discussion at `http://dmi.aalto.fi/symposium19`, helsinki, 2019.

[27] V. Wikström, S. Overstall, K. Tahiroğlu, J. Kildal, and T. Ahmaniemi. Reference suppressed for anonymity during peer review. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 3151–3154. 2013.