Reflecting on the Musicality of Machine Learning based Music Generators in Real-Time Jazz Improvisation: A case study of OMax-ImproteK-Djazz

Dimitrios Smailis[1], Areti Andreopoulou[1], and Anastasia Georgaki[1] *

Laboratory of Music Acoustics and Technology (LabMAT), National and Kapodistrian University of Athens

Abstract. This work looks at improvisations produced by the OMax, ImproteK, and Djazz ML generators, through the lens of the elements of music and suggests a musically-oriented evaluation methodology. This idiomatic music analysis is presented from a jazz performer's point of view, reflecting upon cognitive foundations of emotion and meaning. The analysis, based mainly on the evaluation of already published material and on the authors' own experiments, shows musical drawbacks in terms of tension and release of the resulting melodic lines, voice leading of chords, rhythm, groove, dynamic control, and structure.

Keywords: musical analysis, machine learning, jazz improvisation technologies, machine improvisation,

1 Introduction

This work reflects on the improvisation capabilities of the OMax-ImproteK-Djazz system, from now on referred to as OID. OID is a technology which allows humans and machines to improvise real-time music side by side. OID is not an autonomous agent. It needs a human to control its parameters; hence it is regarded as co-creating system. It uses Factor Oracles (Allauzen, Crochemore, & Raffinot, 1999), a finite state automaton for machine improvisation (Assayag & Dubnov, 2004). The focus of this paper is on the responsiveness of the OID audio capture and generation engines in regards to the elements of music in real-time performances. Apart from the evaluation of already published material, the authors have also experimented with OID using OMax4x, OMax5 beta (Assayag, Bloch, Chemillier, Cont, & Dubnov, 2006) and DYCI2 (Nika, Déguernel, et al., 2017).

* The authors would like to thank the developers of OID, Gerard Assayag, Benjamin Levy, Jerome Nika and Marc Chemillier, for the induction of OMax and DYCI2 into LabMAT of National and Kapodistrian University of Athens.

1.1 Overview of the OMax-Improtek-Djazz Music Generators

The OID is a style modelling system (Bloch, Dubnov, & Assayag, 2008) that generates real-time improvisations based on a corpus of real-time acquired audio or MIDI samples or predefined data. It uses two input modes, live and play-file. In OMax, musical material from the corpus is segmented into different phrases using silence detection (Bonnasse-Gahot, 2014). Information extraction from the input audio stream is performed on the basis of predefined descriptors (user-defined audio features). Following that, the learning algorithm builds a statistical model from the acquired samples and the generation algorithm walks through the model generating an interleaved and recombined musical stream (Assayag & Dubnov, 2004).

ImproteK was created as a variant of OMax to handle improvisations on chord progressions (Ayad, Chemillier, & Pissis, 2018), by introducing a temporal specification (scenario) in the music generation process (Nika, Chemillier, & Assayag, 2017). Its memory is built on a labeled sequence of musical content in a way that can follow both the beat (through tap-tempo) and the chord progression. When a chord progression acts as the given scenario it becomes a navigation leader. A score follower (Antescofo) acts as a sequencer by emitting in real-time the current position in the harmonic grid (Cont, 2008). To deal with time stretching, pitch-shifting, and filtering the system uses a phase-vocoder algorithm (SuperVP) (Depalle & Poirot, 1991).

Djazz is the state of the art in the OID system, implementing techniques for indexing and creating improvisations using a given chord progression. It relies on a dictionary, built with musical sequences (audio or MIDI) associated with known chord changes. This dictionary can be built either from on-the-fly live recordings, or a corpus of pre-selected jazz solos. New improvisations can be generated by combing the sequences found within these solos (Ayad et al., 2018).

2 Music Expectations in the OID System

The developers of the OID system seem to have considered the psychology behind music expectation as the core inspiration in modelling its architecture (Cont, Dubnov, & Assayag, 2006). Their concept was also based on the anticipation mechanism (Huron, 2008), claiming that such a modelling approach constitutes a complex musical behaviour such as long term planning and generation of learned formal shapes.

According to Leonard Meyer (1956), musical emotions derive from the arousal, suspension, and fulfilment of expectations. Tension and release are suggested to be key factors in providing momentum and defining structure in music and are therefore important in the music sense of emotion. Expectations arise through implicit schematic and dynamic knowledge of musical regularities, acquired through repeated exposure to a particular style, such as jazz music (Tillmann, Bharucha, & Bigand, 2000). Bianco et al. (2020), suggests that the listeners' arousal and attention are intentionally manipulated by composers and improvisers through modulation of predictability and schematic violations. Additionally, Huron (2008) and Bianco et al. (2020) suggest that composers and improvisers are able to emotionally design experiences through preparation and control. The belief of having control over a situation is a vital aspect of musical expectations theory (Meyer, 1956).

3 The Musicality of the OID Computer Music System

"[...]As should be clear to any musical reader, assessing a music generator in an objective manner, if not impossible, would set along disputable measures of goodness. On the other hand, in most music practices and styles, what is considered as wrong can be constituted as a feature depending on the context[...]" (Cont et al., 2006). Defining assumptions and intentions towards improvising machines, is essential if one is to improvise with them effectively (Lewis, 2019). This section evaluates the musical outcomes of the OID system drawing upon jazz language conventions.

3.1 Beat, Meter, Rhythm, Groove and Silence

Due to the expressive micro-timing nature of jazz performances, automatic beat-tracking mechanisms (Bonnasse-Gahot, 2014) were withdrawn from OID. Manual tapping is favored instead, as it leads to more reliable results. However, this method is prone to distractions (Dannenberg, 2012). Precision between tap times and true beat times can be problematic depending on the tasks the human operator of OID executes during a performance.

To the authors' knowledge, throughout the development of the OID system, no sufficient research was conducted regarding it's response to beat accentuation and time signature. Nevertheless, these features are style defining in jazz performance and composition. Beat accentuation and syncopation is in the heart of jazz music. For example, in a 4/4 time signature, norm for most of the jazz repertoire, beats 2 and 4 are important not only in regards to velocity, but for chord and scale-tone placements. This is true for both traditional and modern jazz styles. An improviser, for example, may place chord tones on beats 1 and 3 and non-chord tones on beats 2 and 4 (Ligon, 1996)(Figure 1). The opposite is also possible. Having control over this placement can reaffirm the form of a tune, or provide a rhythmic counterpoint to the form. In addition, it extends the musicians' pallet to include more colors and suspense (Berliner, 2009).

When compound or odd meters come into play, several problems arise. First, the division of the meter is laid out in a specific way i.e. 3+2+2 for a 7/8 odd meter (Figure 2). In such cases, the accents of several elements of music i.e. melody, rhythm, harmony, timbre, dynamics, fall onto specific places within the meter. Second, and this is true for simple meters too, multiple time signatures often coexist within a meter (cross rhythms and rhythmic hemiolas), i.e. 3 over 2 or 5 over 4 etc. The work of Srinivasamurthy et al (Srinivasamurthy, Holzapfel,

Fig. 1. Chord-tone placement on beats 1 and 3

& Serra, 2014) regarding rhythm may prove to be helpful to the development of the rhythmic abilities of the OID system.

Fig. 2. Example of an odd meter with accents on beats 1, 4, 6

Rhythm, a defining characteristic of jazz, is improvised during performance. OID plays the segments with no real-time listening capabilities, hence not able to follow its human co-improviser. Consequently, this lack of real-time rhythm awareness affects the groove. Even when the tempo stays constant, slight rhythm fluctuations can cause misalignment of the previously recorded material with the human improviser. Furthermore, fluidity with the tap-tempo function is necessary. Watching videos uploaded on the official Djazz website, one can draw useful conclusions in regards to the ways Djazz handles rhythm and groove[1].

Silence is a standalone musical element for many musicians and educators. Experimentation showed that OID does not introduce pauses in its musical output by default. However, this can be balanced out by the fact that the OID is a co-creative agent, reserving this aspect of musicality for the humans operators interacting with it. Nevertheless, considering empty feature vectors as valid "states" for building the indexed graph, conditioned on certain constraints, may be an interesting addition.

3.2 Harmony

Machine estimation of complex chords, as seen in jazz, is a daunting task and a known problem in Music Information Retrieval (MIR) (Odekerken, Koops, & Volk, 2020). To date, Automatic Chord Estimation (ACE) algorithms are not able to feed real-time music making systems.

As presented earlier, a scenario in OID, can be a chord progression, annotated by the user, which will become the navigation leader of the Factor Oracle. From a jazz performer's point of view, constraining the performance around

[1] Some indicative examples can be found in:
a) `https://www.youtube.com/watch?v=tsTI2MOOBWg\&t=217s` ,
and b) `https://www.youtube.com/watch?v=NTWjYJj0LX8\&t=147s`

a static chord progression raises several issues. Credibility is important when sourcing chord progressions as Real Books, ACE and human transcriptions can vary significantly. Moreover, chord progressions in jazz constitute only a part of the lead-sheets. During a jazz performance, these chord progressions are improvised by altering or completely replacing one or several of the "original" chords with others. Additionally, music accompaniment in jazz, also known as comping, follows voice leading conventions, such as the ones seen in figure 3. While ImproteK is able to improvise the accompaniment (Nika & Chemillier, 2012), it is not able to apply such voice leading rules to the accompaniment tracks it generates.

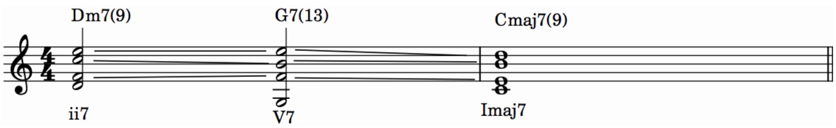

Fig. 3. Voice leading conventions of chords as seen in Jazz: every tone is moving smoothly to the next by either maintaining the same note or by a step-ward movement

3.3 Melody

Traditionally, jazz musicians improvise by playing a series of notes within a harmonic framework. The improvisation techniques listed below are some popular processes that strongly address the tension-release concept, which is associated with universal music-making methods:

- Targeting 3^{rds} through 7^{ths} and vice versa: the melody moves over the bar-line in a stepwise manner from the 7^{th} of the current chord to the 3^{rd} of the following (Figure 4).
- Resolution colors: resolutions by step or a fifth, over the bar-line, are common in jazz. Other melodic moves provide weaker resolutions signaling the arrival of the next chord in a less definitive manner (Figure 5).
- Contour: emphasizing horizontal melodic constructions that can carry listeners over the bar-lines towards longer-range goals (Figure 6).

Fig. 4. 7^{th} to the 3^{rd} over a barline, Charlie Parker on 'Donna Lee'

Fig. 5. Resolving by stepwise movement over the bar-line, John Coltrane on 'Lazy Bird'

Fig. 6. Ascending melodic contour, Michael Brecker on 'Softly as In a Morning Sunrise'

These are just a few examples of idiomatic melodic jazz improvisation techniques governed by conventions. These conventions constitute fundamental melodic constructions developed over hundreds of years (Crook, 1991). They are part of what is called the language of jazz. That is melodic devices that efficiently and stylistically address the tension and release concept. Melodic patterns over chord changes are defined by over-the-barline movement, rhythm (placement within the meter), and direction (ascent, descent, arch, inverted arch, stationery). To the best understanding of the authors, the aforementioned techniques are not addressed by the OID system. OID lacks harmonic anticipation in terms of melodic leading to the next chord. The re-injection of captured material in random places within the meter, a process defined by either the scenario or/and the descriptors chosen by the user, does not lead to jazz lines governed by conventions. This is evident in videos uploaded on the dedicated ImproteK channel[2].

Additionally, in a recent experimental work Deguernel et al, (2018) augmented the OMax Factor Oracle paradigm by adding a multidimensional (harmonic and melodic) probabilistic module, trained over the Real Book content. This experimentation, which can be found on the OID website[3], includes improvisations generated by just a Factor Oracle and improvisations generated by a Factor Oracle combined with a probabilistic model. When those two options are compared, one can hear obvious random melodic leaps in the first case, which become smoother, albeit present, in the second. This analysis is further supported by musicians collaborating on this prospective work. Double bassist Louis Bourhis evaluating generated improvisations on two tunes "Anthropology" and "Donna Lee" characterized the improvisations as melodic patchworks of Parker without a feeling of consistency (Déguernel et al., 2018). Pascal Mabit added: "[...]Harmony makes sense in a continuity.[...]At the moment, it doesn't take that into account, or it is juxtaposing them in a random manner. We don't really hear harmony. We hear note after note, or phrases after phrases. And even

[2] Some indicative examples can be found in:
a) `https://vimeo.com/153605886` b) `https://vimeo.com/158559825` c) `https://vimeo.com/153605441`

[3] `http://repmus.ircam.fr/dyci2/demos/`

inside phrases, there is not necessarily any harmonic sense[...]" (Déguernel et al., 2018).

3.4 Timbre

Compared to other algorithmic improvisers, OID is capable of handling audio. However, in both input modes, live and play-file, fluctuations of tempo raise the need for time-stretching. In the live mode this can happen due to possible tempo fluctuations between the moment of the audio recording and the moment of the performance, while in the play-file mode tempo will have to be adjusted to accommodate the current beat. Furthermore, pitch shifting, which is used to accommodate new tonal centers, will be needed for play-file mode. Time stretching and pitch shifting applied to an audio signal introduce an unavoidable small latency as well as artifacts in the transformed sound, compromising the quality of the produced audio and therefore affecting timbre (Liuni & Röbel, 2013). Nevertheless, it must be acknowledged that there exist certain algorithms which appear to perform better than others on these tasks, such as the one discussed in (Pruuvsa & Holighaus, 2017).

3.5 Dynamics

As already mentioned, the OID output re-injects a recombination of previously captured audio excerpts into the performance. To the authors' best knowledge, while in OMax the generation algorithm walks through the model and generates musical streams according to the given descriptors and in ImproteK its navigation is restricted by the scenario, in both cases, it does not consider the amplitude of the recombined audio slices. Listening sessions, carried out as part of this work, revealed that the OID generation algorithm produced dynamically unbalanced melodic phrases characterized by random, extreme and sudden dynamic changes.

3.6 Structure – Form

The notion "when jazz tells no story, it is simply not good" is widespread among jazz musicians (Bjerstedt, 2015). Stan Kenton suggests that "[...]the problem today is that good improvisers are so rare. There are many people who can make sense out of their improvisation, but very few are really saying anything[...]" (Bjerstedt, 2015). Putting aside the emotional aspect of the term Storytelling, the evaluation of the structured improvisation of the OID system becomes difficult. The term calls for metaphors that are affected by, social, national, racial, and elitist characteristics.

The capabilities of the OID system in storytelling, or in other words, it's long-term structures of improvisation are limited. This is evident in some early experiments with ImproteK[4] as well as in recent experimental work by Deguernel et al. (2018). Bouhris states that OID lacks understanding of the global form

[4] available at https://vimeo.com/153605886

of the chord chart, adding that: "[...]I feel as if it just takes the chords one after the other[...]what it does works with the chords but it doesn't always make sense[...]" (Déguernel et al., 2018). It could be interesting to examine how the notion of attention (Huang et al., 2018) and concepts such as that of jazz mapping (Vassilakis, Georgaki, & Anagnostopoulou, 2019) could be adapted towards effective story telling of real-time ML improvisational systems such as the OID. The lack of structured outputs in ML systems is currently one of the major problems beyond the OID in other major projects, such as the ML driven work Magenta by Google (Roberts, Engel, Raffel, Hawthorne, & Eck, 2018).

4 Conclusions

This paper attempted to provide a basic functionality description coupled with a musicological evaluation of the OID system, in the context of jazz music. It was argued that the performance of the OID system, within the scope of the jazz musical idiom, sets along disputable measures of musicality.

Modulation of predictability through schematic violations constitute an integral part of the overall charm of jazz music. Emotional designation in jazz composition and improvisation is achieved both, through preparation and control and instinctively through schematic violations of music regularities, acquired through the repeated exposure to a particular style. Continuous, uncontrollable schematic violations as seen in OID do not provide a balance between expectation and surprise. Let alone groove, silence, timbre and dynamics, through the analysis provided in section three, one can conclude that although stylistically the OID sounds "jazz-like", scenarios and descriptors used to constrain the generation algorithm cannot guarantee jazz lines governed by idiomatic conventions such as rhythm and beat placement, over the bar-line movement, contour, structure, and conventional voice leading in chord progressions.

The narrative of the OID system is both linear and fragmented. Its methodological attitudes are characterized by both essential and anti-essential as well as foundational and anti-foundational attitudes. This binarity creates, an inconsistent blend of modernist and post-modernist aesthetics within the jazz idiom.

5 Future Work

There is great artistic potential and creative controversy in articulating meaning and intent in performances using ML. Future work will include research in understanding how ML is being, or can be used to support innovation in live music performance, composition and interactive installations. The authors will be reaching beyond witticism to scrutinize musicological, cultural and philosophical aspects of such activities and their consistency in social life and aesthetics.

References

Allauzen, C., Crochemore, M., & Raffinot, M. (1999). Factor oracle: A new structure for pattern matching. In *International conference on current trends in theory and practice of computer science* (pp. 295–310).

Assayag, G., Bloch, G., Chemillier, M., Cont, A., & Dubnov, S. (2006). Omax brothers: a dynamic yopology of agents for improvization learning. In *Proceedings of the 1st ACM workshop on audio and music computing multimedia* (pp. 125–132).

Assayag, G., & Dubnov, S. (2004). Using factor oracles for machine improvisation. In *Soft Computing* (Vol. 8, pp. 604–610). Springer.

Ayad, L. A., Chemillier, M., & Pissis, S. P. (2018). Creating improvisations on chord progressions using suffix trees. In *Journal of Mathematics and Music* (Vol. 12, pp. 233–247). Taylor & Francis.

Berliner, P. F. (2009). *Thinking in jazz: The infinite art of improvisation.* University of Chicago Press.

Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition*, *138*, 103621.

Bjerstedt, S. (2015). Storytelling in jazz improvisation: implications of a rich intermedial metaphor. In *19th Conference of Nordic Network for Research in Music Education (NNMPF), 2015.*

Bloch, G., Dubnov, S., & Assayag, G. (2008). Introducing video features and spectral descriptors in the omax improvisation system. In *ICMC* (pp. 1–1).

Bonnasse-Gahot, L. (2014). An update on the somax project. *IRCAM-STMS, Tech. Rep.*

Cont, A. (2008). Antescofo: Anticipatory synchronization and control of interactive parameters in computer music. In *International Computer Music Conference (ICMC)* (pp. 33–40).

Cont, A., Dubnov, S., & Assayag, G. (2006). Anticipatory model of musical style imitation using collaborative and competitive reinforcement learning. In *Workshop on Anticipatory Behavior in Adaptive Learning Systems* (pp. 285–306).

Crook, H. (1991). *How to improvise: an approach to practicing improvisation.* Advance music.

Dannenberg, R. B. (2012). Human computer music performance. In *Dagstuhl follow-ups* (Vol. 3).

Déguernel, K., Vincent, E., & Assayag, G. (2018). Probabilistic factor oracles for multidimensional machine improvisation. In *Computer Music Journal* (Vol. 42, pp. 52–66). MITP.

Depalle, P., & Poirot, G. (1991). A modular system for analysis, processing and synthesis of sound signals. In *Proceedings of the International Computer Music Conference* (pp. 161–161).

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., ... Eck, D. (2018). Music transformer. *arXiv preprint arXiv:1809.04281*.

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation.* MIT press.

Lewis, E. (2019). *Intents and purposes: Philosophy and the aesthetics of improvisation.* University of Michigan Press.

Ligon, B. (1996). *Connecting chords with linear harmony.* Hal Leonard Corporation.

Liuni, M., & Röbel, A. (2013). Phase vocoder and beyond. *Musica/Tecnologia*, 73–89.

Meyer, L. (1956). *Emotion and meaning in music.* Chicago: University of Chicago Press.

Nika, J., & Chemillier, M. (2012). ImproteK: Integrating harmonic controls into improvisation in the filiation of omax. In *International Computer Music Conference (ICMC)* (pp. 180–187).

Nika, J., Chemillier, M., & Assayag, G. (2017). ImproteK: Introducing scenarios into human-computer music improvisation. *Computers in Entertainment (CIE)*, *14*(2), 1–27.

Nika, J., Déguernel, K., Chemla, A., Vincent, E., Assayag, G., et al. (2017). Dyci2 agents: merging the" free"," reactive", and" scenario-based" music generation paradigms. In *International Computer Music Conference.*

Odekerken, D., Koops, H. V., & Volk, A. (2020). Decibel: Improving audio chord estimation for popular music by alignment and integration of crowd-sourced symbolic representations. *arXiv preprint arXiv:2002.09748*.

Pruuvsa, Z., & Holighaus, N. (2017). Phase vocoder done right. In *2017 25th European Signal Processing Conference (EUSIPCO)* (pp. 976–980).

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. In *International Conference on Machine Learning* (pp. 4364–4373).

Srinivasamurthy, A., Holzapfel, A., & Serra, X. (2014). In search of automatic rhythm analysis methods for turkish and indian art music. *Journal of New Music Research*, *43*(1), 94–114.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological review*, *107*(4), 885.

Vassilakis, D., Georgaki, A., & Anagnostopoulou, C. (2019). "jazz mapping" an analytical and computational approach to jazz improvisation. *SMC*.