

## RESEARCH

# Evaluating Creativity in Automatic Reactive Accompaniment of Jazz Improvisation

Fabian Ostermann, Igor Vatulkin and Günter Rudolph

Music generating computer programs can support jazz musicians and students during performance and practice, for instance by providing accompaniment for solo improvisation. However, such software typically plays sequences of static precomposed snippets and does not react to the user. In that context, it is hardly possible to determine whether such a system has any of its own creative powers. Within the scope of a user study with 20 participants, we evaluate and compare the mobile application iReal Pro to our own system, the evolutionary automatic and reactive system called ‘EAR Drummer’ that generates drum patterns as accompaniment to jazz solo improvisation. It adapts its behaviour in real-time by heuristic rules based on music properties derived from the user’s melodies. The user-based evaluation is performed by following the standardised procedure for evaluating creative systems (SPECS). The analysis of the results is based on a Linear Mixed Effects Model to consider fixed and random effects on the survey data. The model reveals that our system outperforms iReal Pro in all of SPECS’s partial components of creativity and significantly outperforms it for 7 of those 14 components including variety, originality, emotional involvement, and social interaction. Further, it is characterised as “better” and “more interesting” in the user survey. A conflicting observation is that while 70% of the study participants tend to prefer our more “creative” system as support for stage performances, only 40% find it more suitable for practice. Further analysis addresses differences between user groups defined by their played instrument, age, and musical experience.

**Keywords:** Evolutionary music accompaniment generation; drum patterns; evaluation of creative music systems; Linear Mixed Effects Model; SPECS user study; learning jazz music

## 1. Introduction

The art of musical improvisation is a natural way of expressing human inner creativity. “The activity of instantaneous creation is as ordinary to us as breathing” (Nachmanovitch, 1990, p.17). It is found as part of many cultures in different areas around the world and in different eras, with diverse traditions and aesthetics that are most often developed independently (Bailey, 1993). Each piece of music, each piece of art is a reflection of our own mind (Nachmanovitch, 1990, p.25). It requires highly *creative* efforts that are hard to measure, or even to describe in a formal way.

“[Jazz] Students face enormous challenges in mastering both their respective instruments and the complex musical language” (Berliner, 1994, p.51). Many students are in constant search for literature like harmonic theory books that help in theoretical understanding. Also, they look for more and more practice exercises that promise faster improvements of instrumental skills. These are

educational *tools* aiming to guide the students and support their individual learning curve.

In jazz, music students want to learn the art of musical interaction and spontaneous creation of compositions. Effective learning thereof is difficult to achieve by self-reliant studying. To address this problem, we already proposed a new tool, our *reactive* music system EAR DRUMMER (Ostermann et al., 2017). It is capable of simulating a virtual practice partner. The inventive music generation process is driven by Evolutionary Computing (EC).

Bringing computers into the creative domain of jazz music is no novel idea. Attempts have been made to develop algorithms that generate jazz solos or automatic accompaniment. Especially when it comes to real-time composition, computers have advantages by their speed of calculation. Our aim is to determine how those computer music performances are perceived with special regard to the domain of jazz improvisation. To what extent can their output be described as creative? How can a creative program support human musicians? How do musicians like and think about being accompanied by creative machines? Are there scenarios where they

appreciate an automatic composer as creative partner, or even accept it as creative individual? Could the artificial creativity, if it exists, boost the training success of jazz students?

To shed light on these questions, a user study is proposed. Its primary objective is a deeper understanding of the way the creative potential of musical systems is perceived by humans. The *Standardised Procedure for Evaluating Creative Systems* (SPECS) was chosen as a *measure of creativity* to quantitatively evaluate EAR DRUMMER by human users. As a baseline competitor, we chose the proprietary system *iReal Pro* (Technimo, 2021), a non-reactive mobile practice application that is quite popular among jazz students. The user study targets the musical improvisation domain of “non-free” jazz. Improvisers are bound to the harmonic progression and rhythm of a given composition. They are used as “vehicles for improvisation” (Berliner, 1994, p.63).

The benefit a creative system provides for the training success of jazz students is a secondary objective that is discussed, because both systems were initially designed to fulfill this specific task. However, this topic is even more difficult to measure quantitatively and would require studies over a longer period of time. This is why we primarily focus on measuring and discussing the existence, the amount, the quality, and the value of artificial creativity within the systems before we can fully address further objectives. Obviously, tools that provide no interaction can hardly help in developing skills of musical interaction. However, tools that are perceived as truly creative could boost the students’ creativity during training. Generally, the acceptance of intelligent reactive music programs for training purposes is vital for the success of their training.

To define the area of investigation, we start with an introduction to the tradition of jazz improvisation in Section 2. A discussion on how computers can be integrated in that tradition including thoughts on acceptance and benefit of artificial musicians from John Al Biles and George E. Lewis follows in Section 3. Previous work on jazz accompaniment generation and similar approaches are presented in Section 4.1. Section 4.2 shows an overview of *iReal Pro*’s technical features. Music generation with the help of EC is introduced in Section 4.3. Our EAR DRUMMER system is briefly presented in Section 5. A detailed explanation of the SPECS methodology and comparison to related approaches are provided in Section 6. Sections 7 and 8 treat the explanation of the study’s outline and its evaluation by applying a linear mixed effects model to the survey data, respectively. Finally, we summarise the most relevant findings in Section 9.

## 2. The Jazz Solo and its Accompaniment

In its long history starting in the 1890s, jazz music is predominated by the element of improvisation (Gioia, 2011, Chapters 2–5). In the 1940s, a style called *bebop* lifted the amount of improvisation and technical virtuosity to a previously unknown level. The needs for training, musical understanding, and mental capabilities increased (Berliner, 1994, p.51). Later styles are based on the same concept of improvisation in turns, which we outline in the following. The literature summarises them as *modern jazz* (Gioia, 2011, Chapter 6).

In the *modern jazz* tradition, composed pieces consisting of a *head* melody, and an accompanying harmonic progression called *changes*, provide the structure for improvisation (Berliner, 1994, p.63). The music starts by playing the *head*. As accompaniment, the *rhythm section* (typically drums, bass, and piano or guitar) provides the chord *changes*. When the *head* has been presented, the progression of the *changes* is repeated. The *head*, however, is replaced with spontaneously invented melodies by one of the musicians. This musician takes the role as improvising *soloist* presenting melodic ideas over the continuously repeating cycle of *changes* (Berliner, 1994, p.63ff). After one soloist signals to take turns, another musician takes the role of the soloist. Finally, the *head* is presented again.

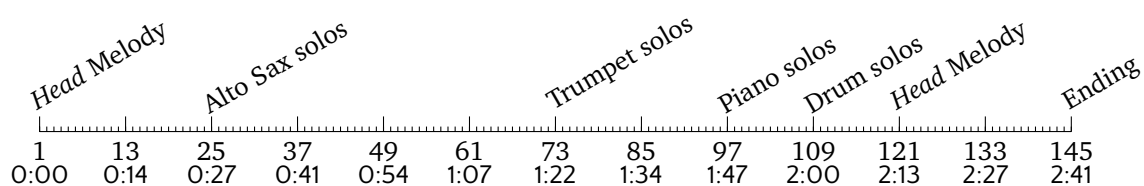
As hands-on example, the structure of “Au Privave”<sup>1</sup> by Charlie Parker (1920–1955) is presented in **Figure 1**. We marked the points at which different soloists improvise. The repetition of the *changes* binds the improvisation together. All parts start at multiples of twelve, because the composition itself builds upon a twelve-bar structure.

This traditional procedure has its strengths in its clarity and comprehensibility. It is therefore a good starting point for students. More complex forms of improvisation like *free jazz* are beyond the scope of our study. For further explanations see Bailey (1993, Part 5) and Gioia (2011, Chapters 7,8).

## 3. Creative Computers and Jazz Improvisation

Many attempts have been made to implement systems that enter the domain of musical improvisation. They led to diverse responses from researchers, musicians, and audiences. For many, the fact that machines are unable to perform emotional involvement categorically reduces the value of such systems.

However, some pioneers implemented systems intended to improvise jazz solos. Regardless of potential failure, they wanted to explore new opportunities. One is *Voyager* (Lewis, 2000). Although based on rather simple statistical rules with properties like pitch, volume, duration, or



**Figure 1:** The parts of “Au Privave” by Charlie Parker. The upper scale measures the bars, lower scale the time (m : s s).

rhythm regularity (Collins, 2010, pp.219–221), *Voyager* independently produces musical phrases while interacting with human improvisers in free improvisations (no *changes*).

One focus of Lewis's work is the nature of virtual musicians. He discusses jazz in the context of African-American culture. He states, that “one's own sound” is important when judging improvisers. “Own” means a dimension of uniqueness. “Sound” is not only referred to as timbre, but as the whole of an improvisational performance manifesting in the “expression of personality, the assertion of agency, the assumption of responsibility and an encounter with history, memory and identity” (Lewis, 2000, p.37). He argues that *Voyager* has shown all those elements when performing in concertante settings at various venues. He mentions, that in “African musical traditions a musical instrument ‘is often regarded as a human being’” (Lewis, 2000, p.37). Following that tradition, *Voyager*, which is at least an instrument, should be treated as a vivid participant.

Generally, Lewis's argumentation loosens the ties of conventional definitions. He defines improvisation as adding “new material [...] to the overall piece” (Lewis, 2000, p.38). The “relatedness of particular materials need not be and quite often cannot be ‘objectively’ demonstrable” (Lewis, 2000, p.38). That means it is irrelevant whether the musician adding material is human or machine. The existence of improvisation is undeniable. Further, Lewis identifies the process of improvisation as “under the general heading of ‘creativity’” (Lewis, 2000, p.38). Consequently, systems that improvise are somehow creative themselves.

A system that generates non-free jazz melodies is *GenJam*, a “model of a novice jazz musician learning to improvise” (Biles, 1994). The model consists of encoded musical ideas that get mapped to jazz-typical chords. These ideas are improved by EC with manual evaluation by a human judge. When *GenJam* performs, it picks ideas from its stock to produce melodies which follow the *changes* of a given composition.

Biles, like Lewis, points out that “he has performed a few hundred gigs with *GenJam* and has at least some anecdotal evidence from listeners that *GenJam* is a convincing improviser” (Biles, 2007b, p.164). He notes that humans and technology can influence each other. In his case, “there is no question that [he himself] is now a much stronger musician in general and improviser in particular than before he began taking *GenJam* seriously as a musical collaborator” (Biles, 2007b, p.168). He claims that he learned a lot while playing with *GenJam* as well as while working on its exact implementation. This is an important finding, because Biles demonstrated a potential gain of human experience by creatively interacting with a musical machine.

That way, Lewis and Biles showed that computers can take the leading role as soloist in a jazz band to an adequate degree of success and acceptance. The results were mostly enjoyed by the audience, or if not, at least interest was aroused. Consequently, the existence of creative powers within machines can be considered reasonable and should be evaluated further.

## 4. Related Work

In order to learn and improve the musical skills needed to play jazz music as described in Section 2, many proposals have been made. Besides practising the instrument and studying harmonic theory (Levine, 1995), practical experience must be gained. Soloing accompanied by a rhythm section has special importance. Students must meet and play together. Since this is not always possible, technical aids simulating band situations were developed. Attempts range from special audio records to computer programs. We provide an overview in the following.

### 4.1 Previous Work on Generating Jazz Accompaniment

The first attempt on simulating ensemble playing in jazz was the *Aebersold Playalong Series* (Aebersold, 1967). Aebersold recorded rhythm sections playing without a soloist. On playback, a jazz student can solo and thereby vaguely experience how improvising with a real band feels. Today, many free *backing track* collections like *LearnJazzStandards.com* (Vaartstra, 2010) are available.

The recording approach has limitations. It is time-consuming and cost-intensive. Once a track is recorded, it can hardly be changed. Key, tempo, and instrumentation are fixed. By using single recordings multiple times, students experience monotony. That reduces the creative moment.

Therefore, automatic music generation with variable parameters like chord progression, time signature, genre, or tempo was proposed. The first practicable implementation was *Band-in-a-Box* (Gannon, 1990), which became popular among musicians.<sup>2</sup> Fein (2017) gives an introduction to its functionalities.

Today, many similar tools like *ChordPulse* (Flextron, 2001), *JamStudio* (ChordStudio, 2008), or *SessionBand* (UK Music Apps, 2012) exist. The open-source program *Impro-Visor* (Keller et al., 2005), that teaches jazz melody construction, is able to provide backing tracks based on style parameters.

But all of these systems are a one-way street: they do not react to the student's solo, like real musicians would do. To fill that gap, a rather unique system named *Music Plus One* was proposed by Raphael (2001). It accompanies human instrumentalists playing sheet music by adding remaining parts of a composition in time. However, Raphael (2001) targets classical non-improvised music only.

Attempts targeting improvised interaction between human musicians and computers in jazz are the aforementioned *Voyager* (Lewis, 2000), *GenJam* (Biles, 1998) and *Impro-Visor* (Kondak et al., 2016). Other reactive non-jazz systems exist, e.g. *MuseBots* (Brown et al., 2018) or a marimba-playing robot (Hoffman and Weinberg, 2011). All these systems do not just provide accompaniment, but are *soloists* by themselves. Collins (2010, Chapter 6, esp. 6.4) provides further information on musical human-computer interaction.

### 4.2 Popular Software: *iReal Pro*

*iReal Pro* (Technimo, 2021) is a mobile application for automatic backing track generation and quite popular among jazz students. It offers flexible configuration of

many parameters. Initially, chords are entered or chosen from a library. 61 chord structures are available. Chords are assignable to quarter-note positions. 14 time signatures are available, e.g.,  $\frac{4}{4}$ ,  $\frac{3}{4}$ , or  $\frac{7}{8}$ . The music is synthesized by up to three changeable instruments; default is piano, bass, and drums. The tempo range is 40–360 BPM. The music-composing algorithm is mainly affected by a style parameter. iReal offers 50 different styles, e.g., *Medium Swing*, *Uptempo*, *Bossa Nova*, or *Afro Cuban*. They are further grouped in *Jazz*, *Latin*, and *Pop*.

The style algorithms in iReal are (presumably) using precomposed phrases that are concatenated to form continuous compositions. Dias and Guedes (2013) provide an easy-to-understand example for automatic composition of a walking bassline. Collins (2010, Chapter 8) provides a discussion of algorithmic composition along further examples. A full overview of functionalities and suggestions on how to integrate iReal in the daily practice routine are provided by Fein (2017).

#### 4.3 Evolutionary Computer Music

For the generation of art by algorithms, *Evolutionary Computing* (EC) is a common and well researched approach. Bäck et al. (1997) provide an in-depth explanation of EC. For musicians, we suggest the introduction by Husbands et al. (2007) and the summary of music as an application domain by Biles (2007a).

Miranda and Biles (2007) provided the first systematic summary of related work, covering the fields of audio synthesis, musical composition, and generative performance. *GenJam* by Biles (1994) is considered the first work applying EC to jazz improvisation. A recent review of EC applied to music composition is Loughran and O'Neill (2020).

Another source of related work is the International Conference on Computational Intelligence in Music, Sound, Art and Design (EvoMUSART).<sup>3</sup> It presents works on EC-based real-time composing and accompaniment tools, e.g., *Musicblox* (Gartland-Jones, 2003); see also Santarosa et al. (2006) and De Prisco et al. (2016). For jazz-related papers see, e.g., Bäckman and Dahlstedt (2008) and Hutchings and McCormack (2017).

A closely related work of drum pattern variation using EC is *evoDrummer* (Kaliakatsos-Papakostas et al., 2013). It demonstrates that novel rhythm patterns can be created from given “base rhythms”. Further, it provides an overview to percussive rhythm generation and drum loop altering methods less closely related. The main difference to EAR DRUMMER is that *evoDrummer* does not handle musical input. Therefore, the proposed measure of rhythmic divergence and the drum features, which seem similar to ours at first sight, could not be adopted. Another related work on dueting with an artificial jazz drummer beyond EC is by McCormack et al. (2019). A neural network learns appropriate musical responses from specially manufactured demo recordings. Beside the musical information, biometric data was collected. Despite shared goals, this appealing attempt is less practical for real users than EAR DRUMMER, because of its more complex experimental environment.

#### 5. EAR Drummer

None of the aforementioned systems combines reactivity with the jazz music domain in the way EAR DRUMMER does. The accompaniment generating systems presented in Section 4.1 are able to support a jazz solo practitioner. But, all of them lack the ability of reacting to the soloists' melodies and *improvising* an accompaniment themselves. But this is essential for real jazz music.

To address this limitation, an improved system must deal with musical input. It should be influenced by soloists' melodies and interact like human musicians would. Therefore, the reactive accompaniment system EAR DRUMMER was implemented. EAR DRUMMER targets to support improvisation in a *modern jazz* solo context as described in Section 2. It analyses melodies in real-time using statistical measures and follows the curve of musical tension.

EAR is an acronym for *Evolutionary*, *Autonomous*, and *Reactive*. The core component is an *evolutionary* algorithm. It *autonomously* generates solutions in musical contexts. And it *reacts* musically to soloists it accompanies. Generated solutions are synthesized by drumset sounds. Consequently, it is called *drummer*.

EAR DRUMMER uses statistical analysis of rhythm and harmony of soloists' melodies in order to generate rhythmic output. It is inspired by human jazz drummers responding to musical structures. The underlying evolutionary algorithm handles drum patterns as individuals. The patterns are constantly altered by random mutation in an evolutionary loop. All new patterns emerge from a prototype pattern that has to be manually entered in advance to the improvisation. This *initial pattern* represents the desired music style (genre). The fitness function considers 14 heuristic rules. Some of them use the initial pattern as reference in the evaluation to (re)establish the desired style. Others react to different musical properties of the melodies and try to alter new patterns by specific operating principles. The impact of each rule to the overall fitness value can be manually changed. As a result, users can influence the drummer's reactive behaviour to focus more on desired aspects. During the study (see Section 7), however, users were not allowed to change the weightings (nor to know about them) to preserve comparability.

Since the scope of the present paper is limited and focuses on the evaluation of EAR DRUMMER, we refer to our previous publication for detailed explanations (Ostermann et al., 2017). To demonstrate EAR DRUMMER's performance abilities, audio recordings of the system in action (including recordings from the user study) and instructional videos presenting the system's GUI are available online.<sup>4</sup> With those demonstrations in mind, it will be considerably easier to follow the reasoning of the study's outline and evaluation (Sections 7 and 8). Source code and compiled Java binaries of the EAR DRUMMER system are also available online.<sup>5</sup>

#### 6. Evaluating Creativity with SPECS

The evaluation of creative systems faces difficulties of definition, differentiation, and comparability. In artificial intelligence research, solutions are badly needed.

Computers have recently entered domains of high-level artistic tasks. Researchers must be able to measure their success reasonably in order to interpret their improvements correctly.

The first attempts to identify creative qualities of musical machines are the *Musical Directive Toy Test*, the *Musical Output Toy Test*, and the *Discrimination Test* (Ariza, 2009). These are variations of the *Imitation Game* (Turing, 1950) and, therefore, do not provide a quantitative comparable measure.

The first attempt to propose formal empirical criteria was made by Ritchie (2001). He defined 14 statements that call for a more qualitative understanding of artificial creativity, but lack a systematic evaluation procedure. The same applies to the *Creative Tripod* framework (Colton, 2008) and the FACE and IDEA models (Colton et al., 2011). However, both made further improvements in the qualitative definitions by modeling the impact of a machine performance on the audience.

The *Standardised Procedure for Evaluating Creative Systems* (SPECS) was introduced by Jordanous (2012b) and fully presented by Jordanous (2012a). The author says that

“SPECS is a standardised and systematic methodology for evaluating computational creativity. It is flexible enough to be applied to a variety of different types of creative systems and adaptable to specific demands in different types of creativity.” (Jordanous, 2012a, p.iv)

Because of its standardisation and systematics, SPECS is chosen over the other approaches. SPECS suggests a three-stage process of evaluation (Jordanous, 2012a, Section 5.3):

1. A definition of creativity must be identified that suits the system to be evaluated.
2. The measuring criteria have to be clarified, so that the process of evaluation gains unambiguousness.
3. The system has to be tested against the defined criteria.

Jordanous identifies 14 components of creativity to use as criteria. They suit the domain of creative interactive music systems. In particular, SPECS proved its applicability to improvisation systems in an example case study (Jordanous, 2012a, Chapter 6). We apply SPECS analogously on our evaluation of EAR DRUMMER and comparison to iReal. Thereby, we directly satisfy the requirements of the first two stages.

The methodology of SPECS builds upon the identification of 14 components as sub-aspects of creativity. Those components were derived from computational linguistics analysis: 30 academic papers treating creativity and 60 other papers were gathered. By applying *Log Likelihood Ratio* (LLR), words were identified that appeared significantly more often in papers about creativity. LLR calculates the difference between observed and expected occurrence of words (Jordanous, 2012a, Equation 4.1). The 694 identified words were grouped using the *semantic*

*similarity measure* (Lin, 1998) and the *Chinese Whispers clustering* algorithm (Biemann, 2006). A manual review of the papers led to 14 title labels for the resulting word clusters. Because the linguistic analysis was performed on papers across various domains, these labels represent basic concepts of domain-independent creativity.

If creativity is to be measured in a specific domain, weighting is suggested. Jordanous (2012a, Section 6.3.2) already performed a relative importance analysis on musical improvisation creativity as an exemplary case study. Questionnaires from 34 participants with mixed skill levels identified the weighting. Written surveys about reactions to the term of musical improvisational creativity were conducted. The statements were manually assigned to the 14 components. Because the participants were unaware of them, a weighting of importance to the domain of musical improvisation was derived.

We decided to follow the proposal of Jordanous and reuse the well-grounded weighting in our evaluation. However, the unweighted sum of the components is also considered in comparison. All components with recommended weights are presented in **Table 1**.

In Jordanous's exemplary case study, a jury of three experts judged the quality of three systems on the 14 components. They had 30 minutes to learn about one system and listen to audio examples. Jordanous's own system was compared to *GenJam* and *Voyager*. Ratings of the latter two will be discussed in Section 9. A crowd-based evaluation was also proposed but not conducted. Because we are interested in the opinions of jazz students with different skill levels, we decided to evaluate EAR DRUMMER

**Table 1:** The 14 components of creativity and assigned weights (as percentages) suiting the domain of musical improvisational creativity. This table is adopted from Jordanous (2012a, p.169).

No.	Component	Weight
1	Social Interaction and Communication	14.9
2	Domain Competence	12.5
3	Intention and Emotional Involvement	13.9
4	Active Involvement and Persistence	7.8
5	Variety, Divergence, and Experimentation	7.1
6	Dealing with Uncertainty	6.4
7	Originality	5.8
8	Spontaneity/Subconscious Processing	5.4
9	Independence and Freedom	5.4
10	Progression and Development	5.4
11	Thinking and Evaluation	5.1
12	Value	5.1
13	Generation of Results	3.7
14	General Intellect	1.4
		100.0

by a larger group of participants. Furthermore, we want to provide hands-on experience to our judges.

## 7. Study Outline

To evaluate EAR DRUMMER, a human user study is realised. The study aims to reveal insights on value, performance, attractiveness, and advantages of EAR DRUMMER, which are also vital elements for training success for jazz students. Therefore, comparison to a similar system is considered. The idea of EAR DRUMMER sprang from the desire to have reactive support during practice. iReal is chosen for comparison, because it has a similar intention, but without providing reactive abilities. However, its use is popular among jazz students.

Simple questions about preferences between the systems are useful and may provide essential results, but do not help to unravel their *inner creativity*. How do the users experience the systems while playing with them? And how do they interpret musical structures they produce? Maybe as intelligent, correct, profound, sensible, logical, reasonable, meaningful, or even *creative*?

The use of SPECS is supposed to reveal benefits of reactive accompaniment over backing track generation with only little variation. The question is about the degree of creativity that is awarded to the systems by human users. Since the term “creativity” is rather abstract, we use SPECS component-wise analysis in a blind study.

To reduce influences, both systems were neither technically explained nor visually shown. The participants just played their instruments accompanied by music sounding from a loudspeaker. Because EAR DRUMMER generates drums only, unreactive basslines with small freedoms were added to provide a minimal harmonic accompaniment. iReal also was restricted to drums and bass. Thereby, both systems differed only in the way the drums were generated: iReal by precomposed patterns, EAR DRUMMER by EC. Furthermore, the participants were unaware of being faced with reactivity at all.

Instruments were restricted to piano and guitar. On the one hand, this was because of the easy and (nearly<sup>6</sup>) lossless possibilities of converting instruments' output to MIDI data. On the other hand, pianists and guitarists usually have knowledge and experience as both jazz soloist and accompanist. Furthermore, the chance of a negative impact on the study results due to incomparable instrumentalists is reduced.

The systems were presented to the participants in random order. They were asked to improvise to one system's output. The *changes*, tempo, and style were determined in advance. Two compositions were presented to each participant. “Autumn Leaves” by Joseph Kosma (1905–1969) was the first. The participants were allowed to improvise for five minutes. EAR DRUMMER as well as iReal were set up to play a simple jazz swing groove at a slow tempo of 100 BPM. During the following five minutes, the same setup was presented but the tempo was raised to 200 BPM. In the final five minutes, the *changes* of “Blue Bossa” by Kenny Dorham (1924–1972) were played. The rhythmic style was changed to *Bossa Nova*.<sup>7</sup> The tempo was

140 BPM. All parameters of EAR DRUMMER were determined prior to the user study (see Appendix B).

After testing the first system, the participants were asked to answer a questionnaire with the 14 SPECS components (see Appendix A). We hereinafter refer to those ratings on first sight as *initial ratings*. The components are presented with title and three short descriptive phrases, which may guide the participants' decision. Those were derived from the recommended questions from Jordanous (2012a, Appendix D) and selected by best fitting the domain of musical improvisation systems. Each component is assigned an integer rating from 0 to 10, which describes its relevance to the system tested.

After initial rating, the participants tested and rated the other system (*final rating*) following the exact same procedure as before. Additionally, it was possible to change the rating on the first system, in case the impression had changed after comparison of both. Consequently, there are twice as many collected final ratings for EAR DRUMMER and iReal as for the initial one. We show that this procedure had no effect on the evaluation in Section 8.1.

The questionnaire's second part consists of four questions we hereinafter refer to as *statements of preference*. They will be used as comparison to the results of the component-wise analysis in Section 8.3:

1. Which system did you find better?
2. Which system was more interesting?
3. Which system would you preferably use for practising?
4. Which system would you preferably use for a stage performance?

The questionnaire's third part gathers participants' background data on age, years of experience, additional instrument skills, self-considered level of experience, and familiarity with other music genres.

## 8. Study Evaluation

The primary objective of the study's evaluation is a component-wise analysis using SPECS's components. Section 8.1 presents our general approach of statistical significance analysis based on a Linear Mixed Effects Model (LMEM). Section 8.2 provides component-wise interpretation of the statistics. In Section 8.3, the results are discussed in comparison to the *statements of preference*. Finally, correlations between the ratings of the systems and participants' background statistics are given in Section 8.4.

### 8.1 LMEMs of SPECS's Components

The data gathered by the questionnaire contains both *fixed effects* (variables which do not vary for the comparison of the both systems, e.g., the experience level of a participant) and fully *random effects* (e.g., the identity of the participants). Therefore, a LMEM is considered most suitable for evaluation. The model calculations were implemented in R (R Core Team, 2019) using `lme4` (Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2017).



**Table 3:** Results of the ratings on both systems sorted by adjusted  $p$ -value.

Component	$\beta(\text{System} = \text{ED})$	$p$ -value	adj. $p$ -value
Variety, Divergence, and Experimentation	4.3000	<0.0001	0.0005
Originality	4.1500	0.0002	0.0028
Intention and Emotional Involvement	2.8500	0.0004	0.0050
Generation of Results	3.4500	0.0005	0.0059
Spontaneity and Subconscious Processing	3.6000	0.0006	0.0068
Progression and Development	3.2000	0.0018	0.0183
Social Interaction and Communication	1.8000	0.0061	0.0485
Dealing with Uncertainty	1.7500	0.0122	0.0834
General Intellect	2.4500	0.0119	0.0834
Independence and Freedom	2.1000	0.0275	0.1373
Domain Competence	0.8500	0.1216	0.4864
Active Involvement and Persistence	0.6000	0.2246	0.4864
Thinking and Evaluation	1.3500	0.1313	0.4864
Value	1.0000	0.2633	0.4864
Mean Rating	2.3893	0.0020	0.0183
Weighted Mean Rating	1.5984	0.0016	0.0171

perception of involvement in the musical interaction, in contrast to iReal.

**Variety, Divergence, and Experimentation** For this component, EAR DRUMMER shows the greatest difference to iReal. It is outperformed by 4.3 points on average. The boxplots in **Figure 2** show a clear separation of the distributions. The inventive manner caused by the EC approach in EAR DRUMMER is identified in contrast to the monotonous outputs of iReal. Further, those qualities are essential to break the monotonous experience of playing with a machine. They were clearly attributed to EAR DRUMMER.

**Originality** EAR DRUMMER outperforms iReal by 4.15 points, which is the second largest difference. The concept of new, surprising, and unexpected ideas was clearly identified within EAR DRUMMER. The constantly mutated population of musical ideas inside the evolutionary loop are indeed perceived as original.

**Spontaneity and Subconscious Processing** EAR DRUMMER outperforms iReal by 3.6 points. Spontaneity is accredited to EAR DRUMMER. Unreactive iReal cannot support this ability. Since spontaneity is an essential element in a successful jazz performance, this is a serious advantage and crucial demand for creative systems in a jazz context.

**Progression and Development** EAR DRUMMER outperforms iReal by 3.2 points. But the average rating on this component for EAR DRUMMER is 6.0 and thus the lowest of all. Based on oral responses, the participants were unsatisfied by the short periods in which EAR DRUMMER refers to its musical ideas. Since EAR DRUMMER does not have a long-term memory, this is perfectly understandable. A more extensive understanding of improvisation is requested and thus provides a promising

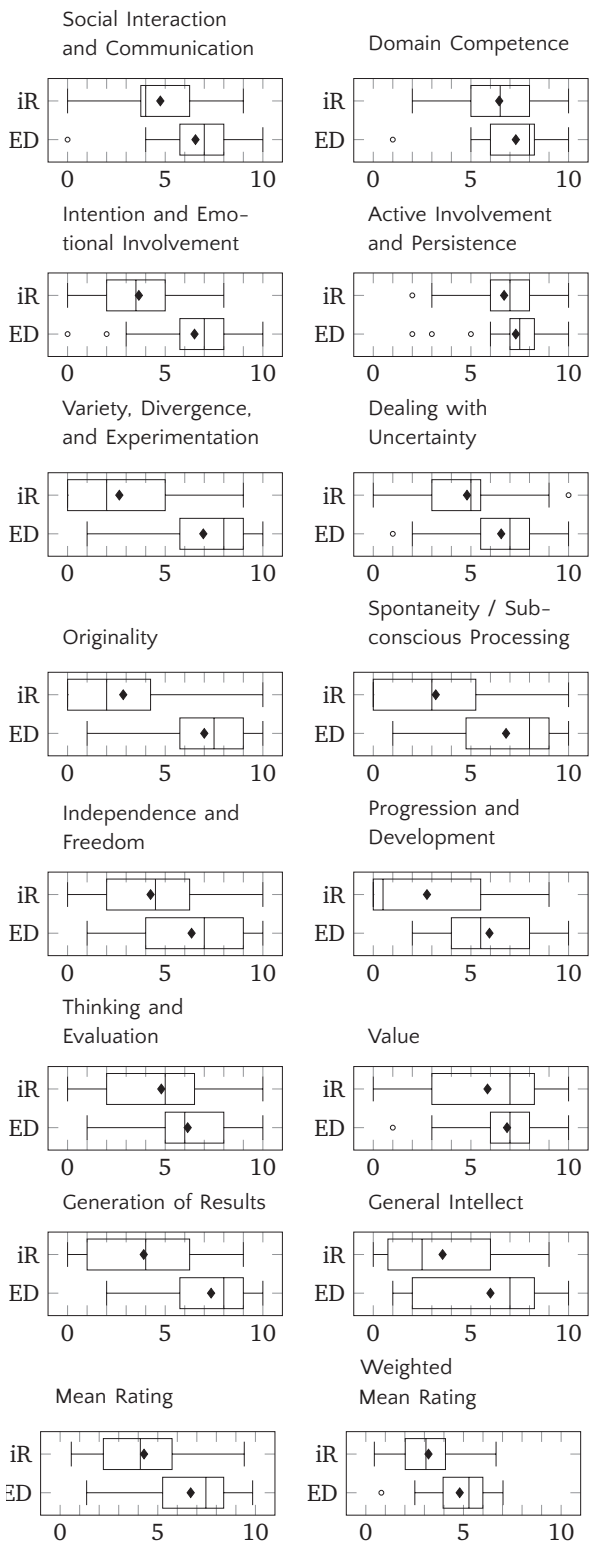
field for future research. The significant difference is explained by the even lower ratings for iReal which is perceived as even more monotonous.

**Generation of Results** Asked about the actual results of the systems, EAR DRUMMER outperforms iReal by 3.45. This component has the highest rating for EAR DRUMMER with 7.3. The fact that EAR DRUMMER produces *new* improvisations is rewarded in contrast to iReal's precomposed phrases. EAR DRUMMER's results were identified by the participants as sensible independent musical improvisation.

**(Weighted) Mean Rating** The average of all 14 components' averages is considered by SPECS as an overall estimator of creativity. EAR DRUMMER with 6.7 is significantly superior to iReal with 4.3 by 2.4 points. When applying the weighting for the domain of musical improvisational creativity from **Table 1**, the ratings of both systems decrease. EAR DRUMMER falls to 4.8, but remains superior to iReal by 1.6 points. Besides, the significance of the adjusted  $p$ -value even increases for the weighted criterion.

Because of the fact that iReal was rated lower on average than EAR DRUMMER for all components in the final rating, the question of *fairness* for the comparison must be raised. Because iReal was not developed to be either interactive nor intelligent while EAR DRUMMER was, there is no big surprise that the results appear to be biased. However, the primary objective of the study was on the question, whether the potential creativity of EAR DRUMMER could be identified by the users at all. In a worst case scenario, the deterministic system iReal could have been rated equal or even better in comparison. That would have led to the question, whether an interactive music system can be termed "creative" at all. Encouragingly, the results are clear: the creativity within a system can evidentially





**Figure 2:** Boxplots for the ratings on each SPECS component for iReal (iR) and EAR DRUMMER (ED).

be identified by human users in the domain of musical improvisation when interacting and communicating musically through their instruments.

**8.3 Evaluation of Statements of Preference**

To justify the component-wise SPECS analysis on the creative properties of the systems, we performed further evaluation of more direct questions on subjective

preferences. The answers may provide information on how the SPECS findings relate to the actual personal opinions. The queries contained votes by the users on the system that was considered the “better” one, the more “interesting” one, that would be preferred for “practising” or even for live performances (“stage”). Below, we include oral feedback from the study in our considerations.

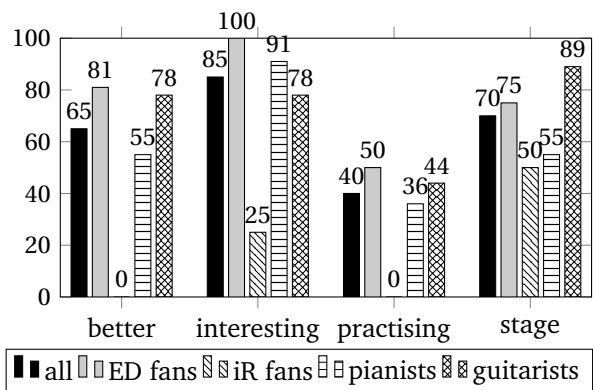
**Figure 3** shows bar plots of the votes of all participants and of identified subgroups “EAR fans” vs. “iReal fans” (explanation follows) and pianists vs. guitarists. Each bar indicates the percentage of participants that voted for EAR DRUMMER. Complementarily, all other votes were for iReal.

First of all, EAR DRUMMER is accepted to be the better system in the final comparison by 65% of participants (left black bar). Further, 85% of the participants found it more interesting and 70% prefer it for use in live performances. However, only 40% of the participants prefer it for practising. In contrast to these answers, the weighted SPECS value and its component analysis in Section 8.1 indicated clearer opinions on the systems in comparison. The question is: to what degree is the creativity of a system relevant for the final preference? And which other qualities are desired by users in this context?

To examine the relationship between SPECS ratings and the statements of preference, the voting of participants who rated a higher weighted SPECS value on iReal than EAR DRUMMER was evaluated separately. They are hereinafter referred to as *iReal fans* and *EAR fans*. Thereby, 4 of the 20 participants fall under the category iReal fans.

All iReal fans (diagonal-striped bars) voted for iReal to be the better system, but also 19% of the EAR fans (solid-grey bars) voted for iReal. However, all EAR fans voted for EAR DRUMMER to be the more interesting system, and also 25% of the iReal fans shared that opinion. Therefore, we assume that the participants appreciated the inventive and spontaneous behavior of EAR DRUMMER. But iReal with its characteristics like stability, comprehension, and predictability is also attractive for certain users.

A look at the third and fourth question strengthens these arguments. When practising, 100% of the iReal fans and 50% of the EAR fans prefer iReal. It was explained in oral responses that EAR DRUMMER would distract by its massive independent and unpredictable behaviour.



**Figure 3:** Statistics of the statements of preference. The bars show the percentage of user votes for EAR DRUMMER.

Others, however, stated that they were driven into a creative mood by the improvisations of EAR DRUMMER. But the majority prefer a simplified setting when practising. This is because typical goals are internalising theoretic and motoric abilities by countless repetitions. Consequently, that attitude changes for stage performances with an audience. The participants feel more comfortable with a system that produces ideas and catches attention. It has been said that instead of playing together with iReal, one could as well perform solo. iReal does not add any creative contribution to a performance. The voting for EAR DRUMMER on “stage” by 50% of the iReal fans, in comparison to 0% on “practising”, is an indicator for the validity of this explanation.

When comparing pianists with guitarists, the latter (cross-hatched bars) are slightly more positive about EAR DRUMMER. However, for “more interesting”, 91% of the pianists (horizontal-striped bars) voted EAR DRUMMER higher, while only 78% of the guitarists did. Two conclusions appear plausible: first, there is no significant difference between the groups. Second, guitarists think more positively about more complex systems. That could be explained by the fact that guitarists are generally more sophisticated in technology (during the experiments all used electric guitars). Future studies with other instrumentalists could verify this assumption.

Other groups based on age, years, and level of experience were tested but no relevant results were identified. Theoretically, it could have been advantageous to ask the users directly which system they experienced as the more creative one. This question was omitted from the questionnaire because it could have revealed too much about the actual research focus. Anyway, the interpretation of “more interesting” against “better” votes and the results of the SPECS component analysis imply an unambiguous conclusion that reactive systems are determined to be more valuable and beneficial for creative tasks. Therefore, such systems are also better suited to help jazz students practicing creative improvisation in the context of musical interaction, but not for simple repetitive exercises.

#### 8.4 Background Correlations

In this section, correlations of the attributes of *age* and *years as improviser* with the values of SPECS components and the Mean Ratings are examined. The attribute *years as musician* revealed no significant correlation. With the number of participants  $n = 20$ , the condition  $p \leq \alpha = 0.05$  is fulfilled if the correlation coefficient  $|r| \geq 0.38$ . In the following, we report significant correlations only.

In most cases, the attribute *years as improviser* correlates negatively to iReal’s SPECS components (lower rating the more years of improvising experience). The significant correlations are presented in descending order:

- $r = -0.54$  for “Variety, Divergence, and Experimentation”
- $r = -0.48$  for “Dealing with Uncertainty”
- $r = -0.47$  for “Thinking and Evaluation”
- $r = -0.45$  for “Progression and Development”
- $r = -0.41$  for Weighted Mean Rating

- $r = -0.40$  for “Spontaneity and Subconscious Processing”
- $r = -0.39$  for “Intention and Emotional Involvement”
- $r = -0.39$  for “General Intellect”

The *age* of the users correlates for iReal ratings with following components:

- $r = -0.45$  for “General Intellect”
- $r = -0.39$  for “Intention and Emotional Involvement”
- $r = -0.38$  for “Originality”

The *age* correlates for EAR DRUMMER ratings with:

- $r = -0.44$  for “Value”
- $r = -0.41$  for “General Intellect”
- $r = -0.38$  for “Domain Competence”

The results could be interpreted as follows: the younger and the less experienced the participants are, the less they are able to identify the intelligence and creativity of the systems. They tend to rate the systems more equally. However, testing the year attribute against the mathematical difference between the SPECS values of EAR DRUMMER and iReal gives correlations between  $r = 0.11$  and  $r = 0.13$ . They are slightly positive, as expected, but not significant.

## 9. Conclusions and Outlook

The primary objective of our study was to estimate how improvising human musicians generally respond to a more or less creative music system. Our reactive system EAR DRUMMER, which generates drum patterns with the help of an evolutionary algorithm, was compared by human users to the popular jazz-practice accompaniment generator iReal Pro. This revealed the general assumption of a greater creativity within the reactive system. When conducting a user study on creativity using the *Standardised Procedure for Evaluating Creative Systems* (SPECS), the components “Variety, Divergence, and Experimentation”, “Originality”, and “Intention and Emotional Involvement” were identified as the most significant creative aspects of EAR DRUMMER in contrast to iReal Pro. On the downside, the users showed greatest uncertainty for the components “Domain Competence”, “Active Involvement and Persistence”, “Thinking and Evaluation”, and “Value”.

Direct questions on user preferences showed that 65% of all participants found EAR DRUMMER “better” and 85% found it “more interesting” than iReal Pro. However, the preferences were very different with regards to application scenarios. When asked about the attractiveness of use for stage performances, the participants mostly agreed in preferring the creative system. But in a plain practice environment, too much creative support was judged distracting, especially when repetitive exercises were to be performed. A completely deterministic and non-reactive accompaniment system like iReal Pro then provides more suitable assistance. The reasons lie in the diverse ways of individual practising. For instance, exercises for specific playing techniques or mastering faster tempos require

completely different forms of assistance than exploring creative improvisation and musical interaction. Here, the advantage of EAR DRUMMER is the possibility to adjust the weights of its reactivity rules, so that it may strongly reduce its non-deterministic and creative behaviour with respect to the user's demands. This possibility was not tested during the study and, therefore, is not part of the evaluation. Potential future improvements to EAR DRUMMER will include more advanced fitness functions. Its framework can easily be upgraded by just replacing the rule-based core inside the evolutionary loop.

We have selected SPECS as a verified and established tool to directly compare creativity. However, we should not interpret the results as an absolute measure of creativity. The improvisational systems *GenJam* and *Voyager* (see Section 3) got weighted ratings of 5.0 and 3.3, respectively, by expert judgment (Jordanous, 2012a, p.179). It is highly questionable whether EAR DRUMMER with a weighted rating of 4.8 can be considered “more creative” than *Voyager*, or whether iReal with a weighted rating of 3.2 as “nearly as creative”. To estimate how strong the impact of the rating procedure and concrete application scenario is on the average level of the rating itself remains a future research topic. Applying our proposed user study procedure on other systems and with a larger number of participants would reveal further insights.

The collected impressions and responses assume an existing open-mindedness and fascination for computer musicianship. A deeper understanding of what it means to model creativity and spontaneously compose music can be supported and encouraged by further research as already proposed by Biles and Lewis (see Section 3). The research focus should be oriented towards the order of the SPECS components: the highly rated components imply a high productivity of computers when massively generating novel ideas and highly unorthodox music is required. Those abilities can be considered rather guaranteed, because of the ease with which computers can provide them. The challenging task is to improve on the low rated components. These include skills like matching various established musical styles, being persistent within an improvisational process, or acting more “thoughtfully” to gain acceptance and appreciation from an audience.

Encouragingly, many oral responses of the study participants already showed appreciation for the attempt of developing creative musical systems for the purpose of improving and expanding the variety of practice possibilities as well as to enhance and innovate the variety of improvisational stage performances, whether done by humans, artificial musicians, or both in collaboration.

## Notes

- <sup>1</sup> Available on album “The Genius Of Charlie Parker, #8 – Swedish Schnapps” (1951), Verve Records Catalog No.: MG 8010, 489-2, listen online at: <https://www.youtube.com/watch?v=3ZyHNyVgaqA> (Accessed: 13/10/2021).
- <sup>2</sup> Biles (1994) uses Band-in-a-Box to generate backing music for GenJam.
- <sup>3</sup> Held since 2003, EvoMUSART is part of EvoStar: <http://www.evostar.org> (Accessed: 13/10/2021).

<sup>4</sup> Demo material: <https://doi.org/10.5281/zenodo.5564676>.

<sup>5</sup> EAR DRUMMER's Github repository: <https://github.com/OysterSandwich/EAR-Drummer>.

<sup>6</sup> For guitars, the pitch-to-MIDI converter *Sonus G2M V3* (Sonus, 2018) was used.

<sup>7</sup> *Bossa Nova* originates from Brazilian samba music, that jazz-influenced musicians like Antônio Carlos Jobim (1927–1994) and João Gilberto (1931–2019) interpreted far more slowly in the 1950's (Castro, 2003).

## Additional File

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

- **Appendix A–D.** Includes: The questionnaire (A), EAR-Drummer system settings (B), all 16 LMEMs (C), and participants' statistical information (D).. DOI: <https://doi.org/10.5334/tismir.90.s1>

## Competing Interests

The authors have no competing interests to declare.

## References

- Aebersold, J.** (1967). *Volume 1 – How To Play Jazz & Improvise*. Jamey Aebersold Play-A-Long Series. Jamey Aebersold Jazz.
- Ariza, C.** (2009). The interrogator as critic: The Turing test and the evaluation of generative music systems. *Computer Music Journal*, 33(2): 48–70. DOI: <https://doi.org/10.1162/comj.2009.33.2.48>
- Bäck, T., Fogel, D. B., and Michalewicz, Z.** (1997). *Handbook of Evolutionary Computation*. CRC Press, Boca Raton, USA. DOI: <https://doi.org/10.1201/9780367802486>
- Bäckman, K., and Dahlstedt, P.** (2008). A generative representation for the evolution of jazz solos. In *Applications of Evolutionary Computing*, pages 371–380. Springer. DOI: [https://doi.org/10.1007/978-3-540-78761-7\\_40](https://doi.org/10.1007/978-3-540-78761-7_40)
- Bailey, D.** (1993). *Improvisation: Its Nature And Practice In Music*. Hachette Books, New York, USA.
- Bates, D., Mächler, M., Bolker, B., and Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1): 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Berliner, P. F.** (1994). *Thinking in Jazz: The Infinite Art of Improvisation*. University of Chicago Press, Chicago, USA. DOI: <https://doi.org/10.7208/chicago/9780226044521.001.0001>
- Biemann, C.** (2006). Chinese Whispers: An efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing*, pages 73–80, Morristown, USA. Association for Computational Linguistics. DOI: <https://doi.org/10.3115/1654758.1654774>
- Biles, J. A.** (1994). GenJam: A genetic algorithm for generating jazz solos. In *Proceedings of the International*

- Computer Music Conference (ICMC 1994)*, pages 131–137. International Computer Association.
- Biles, J. A.** (1998). Interactive GenJam: Integrating real-time performance with a genetic algorithm. In *Proceedings of the International Computer Music Conference (ICMC 1998)*, pages 131–137. International Computer Association.
- Biles, J. A.** (2007a). Evolutionary computation for musical tasks. In (Miranda and Biles, 2007), pages 28–51. DOI: [https://doi.org/10.1007/978-1-84628-600-1\\_2](https://doi.org/10.1007/978-1-84628-600-1_2)
- Biles, J. A.** (2007b). Improvising with genetic algorithms: GenJam. In (Miranda and Biles, 2007), pages 137–169. DOI: [https://doi.org/10.1007/978-1-84628-600-1\\_7](https://doi.org/10.1007/978-1-84628-600-1_7)
- Brown, A., Horrigan, M., Eigenfeldt, A., Gifford, T., Field, D., and McCormack, J.** (2018). Interacting with Musebots. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 19–24.
- Castro, R.** (2003). *Bossa Nova: The Story of the Brazilian Music That Seduced the World*. Chicago Review Press.
- ChordStudio.** (2008). *Jam Studio.com – The online music factory*. Chord Studio, Inc., <http://www.jamstudio.com/Studio/index.htm>. Accessed: 13/10/2021.
- Collins, N.** (2010). *Introduction to Computer Music*. Wiley, Chichester, UK.
- Colton, S.** (2008). Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Symposium on Creative Systems*, pages 14–20.
- Colton, S., Charnley, J., and Pease, A.** (2011). Computational creativity theory: The FACE and IDEA descriptive models. In *Proceedings of the 2nd International Conference on Computational Creativity*, pages 90–95.
- De Prisco, R., Malandrino, D., Zaccagnino, G., and Zaccagnino, R.** (2016). An evolutionary composer for real-time background music. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design*, pages 135–151. Springer International Publishing. DOI: [https://doi.org/10.1007/978-3-319-31008-4\\_10](https://doi.org/10.1007/978-3-319-31008-4_10)
- Dias, R., and Guedes, C.** (2013). A contour-based jazz walking bass generator. In *Proceedings of the Sound and Music Computing Conference*, pages 305–308.
- Fein, M.** (2017). *Teaching Music Improvisation with Technology*. Oxford University Press, New York, USA.
- Flextron.** (2001). *Chord Pulse – The handy virtual backing band*. Flextron Bt., <http://www.chordpulse.com/index.html>. Accessed: 13/10/2021.
- Gannon, P.** (1990). Band-in-a-Box. PG Music, Inc.
- Gartland-Jones, A.** (2003). MusicBlox: A real-time algorithmic composition system incorporating a distributed interactive genetic algorithm. In *Applications of Evolutionary Computing*, pages 490–501, Berlin, Heidelberg. Springer. DOI: [https://doi.org/10.1007/3-540-36605-9\\_45](https://doi.org/10.1007/3-540-36605-9_45)
- Gioia, T.** (2011). *The History of Jazz*. Oxford University Press, Oxford, UK.
- Hoffman, G., and Weinberg, G.** (2011). Interactive improvisation with a robotic marimba player. *Autonomous Robots*, 31: 133–153. DOI: <https://doi.org/10.1007/s10514-011-9237-0>
- Holm, S.** (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, pages 65–70.
- Husbands, P., Copley, P., Eldridge, A., and Mandelis, J.** (2007). An introduction to evolutionary computing for musicians. In (Miranda and Biles, 2007), pages 1–27. DOI: [https://doi.org/10.1007/978-1-84628-600-1\\_1](https://doi.org/10.1007/978-1-84628-600-1_1)
- Hutchings, P., and McCormack, J.** (2017). Using autonomous agents to improvise music compositions in real-time. In *Computational Intelligence in Music, Sound, Art and Design*, pages 114–127. Springer International Publishing. DOI: [https://doi.org/10.1007/978-3-319-55750-2\\_8](https://doi.org/10.1007/978-3-319-55750-2_8)
- Jordanous, A. K.** (2012a). *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*. PhD thesis, Department of Informatics, University of Sussex.
- Jordanous, A. K.** (2012b). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3): 246–279. DOI: <https://doi.org/10.1007/s12559-012-9156-1>
- Kaliakatsos-Papakostas, M. A., Floros, A., and Vrahatis, M. N.** (2013). evoDrummer: Deriving rhythmic patterns through interactive genetic algorithms. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design*, volume 7834, pages 25–36, Berlin/Heidelberg, Germany. Springer. DOI: [https://doi.org/10.1007/978-3-642-36955-1\\_3](https://doi.org/10.1007/978-3-642-36955-1_3)
- Keller, B., Jones, S., Thom, B., and Wolin, A.** (2005). An interactive tool for learning improvisation through composition. Technical Report Tech Report HMC-CS-2005-02, Harvey Mudd College, Claremont, USA.
- Kondak, Z., Konst, M., Lessard, C., Siah, D., and Keller, R. M.** (2016). Active trading with Impro-Visor. In *Proceedings 4th International Workshop on Musical Metacreation (MUME 2016)*.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13): 1–26. DOI: <https://doi.org/10.18637/jss.v082.i13>
- Levine, M.** (1995). *The Jazz Theory Book*. Sher Music Co., Petaluma, USA.
- Lewis, G. E.** (2000). Too many notes: Computer, complexity and culture in Voyager. *Leonardo Music Journal*, 10: 33–39. DOI: <https://doi.org/10.1162/096112100570585>
- Lin, D.** (1998). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304, Madison, USA.
- Loughran, R., and O’Neill, M.** (2020). Evolutionary music: Applying evolutionary computation to the art of creating music. *Genetic Programming and Evolvable Machines*, 21: 55–85. DOI: <https://doi.org/10.1007/s10710-020-09380-7>

- McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M. T., Yee-King, M., and d'Inverno, M.** (2019). In a silent way: Communication between ai and improvising musicians beyond sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11. Association for Computing Machinery. DOI: <https://doi.org/10.1145/3290605.3300268>
- Miranda, E. R., and Biles, J. A.,** editors (2007). *Evolutionary Computer Music*. Springer, London, UK. DOI: <https://doi.org/10.1007/978-1-84628-600-1>
- Nachmanovitch, S.** (1990). *Free Play: Improvisation in Life and Art*. Penguin Putnam, New York, USA.
- Ostermann, F., Vatolkin, I., and Rudolph, G.** (2017). Evaluation rules for evolutionary generation of drum patterns in jazz solos. In *Computational Intelligence in Music, Sound, Art and Design (EvoMUSART 2017)*, pages 246–261. Springer. DOI: [https://doi.org/10.1007/978-3-319-55750-2\\_17](https://doi.org/10.1007/978-3-319-55750-2_17)
- R Core Team.** (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.Rproject.org/>.
- Raphael, C.** (2001). Synthesizing musical accompaniments with Bayesian belief networks. *Journal of New Music Research*, 30(1): 59–67. DOI: <https://doi.org/10.1076/jnmr.30.1.59.7121>
- Ritchie, G.** (2001). Assessing creativity. In *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*, pages 3–11, York, UK.
- Santarosa, R., Moroni, A., and Manzolli, J.** (2006). Layered genetical algorithms evolving into musical accompaniment generation. In *Applications of Evolutionary Computing*, pages 722–726, Berlin, Heidelberg. Springer. DOI: [https://doi.org/10.1007/11732242\\_70](https://doi.org/10.1007/11732242_70)
- Sonus.** (2018). *g2MV3–Universal Guitar-to-MIDI Converter Version 3*. Sonuus Limited, <https://www.iconnectivity.com/sonuusshop/g2m-v3-universal-guitar-to-midiconverter-version-3>. Accessed: 13/10/2021.
- Technimo.** (2012–2021). *iReal Pro: Tutorials – Learn the Ropes*. New York, USA. <https://www.irealpro.com/video-tutorials/>. Accessed: 13/10/2021.
- Turing, A. M.** (1950). Computing machinery and intelligence. *Mind*, 59(236): 433–460. DOI: <https://doi.org/10.1093/mind/LIX.236.433>
- UK Music Apps.** (2012). *Session Band*. UK Music Apps Ltd. Company, England. <http://sessionbandapp.com/index.html>. Accessed: 13/10/2021.
- Vaartstra, B.** (2010). Learn jazz standards. <https://www.learnjazzstandards.com/about/>. Accessed: 13/10/2021.

**How to cite this article:** Ostermann, F., Vatolkin, I., and Rudolph, G. (2021). Evaluating Creativity in Automatic Reactive Accompaniment of Jazz Improvisation *Transactions of the International Society for Music Information Retrieval*, 4(1), pp. 210–222. DOI: <https://doi.org/10.5334/tismir.90>

**Submitted:** 28 February 2021

**Accepted:** 06 October 2021

**Published:** 30 November 2021

**Copyright:** © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**][ *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.**

**OPEN ACCESS** 