Using Autonomous Agents to Improvise Music Compositions in Real-Time

Patrick Hutchings(✉) and Jon McCormack

sensiLab, Faculty of Information Technology,
Monash University, Caulfield East, Australia
{Patrick.Hutchings,Jon.McCormack}@monash.edu

Abstract. This paper outlines an approach to real-time music generation using melody and harmony focused agents in a process inspired by jazz improvisation. A harmony agent employs a Long Short-Term Memory (LSTM) artificial neural network trained on the chord progressions of 2986 jazz 'standard' compositions using a network structure novel to chord sequence analysis. The melody agent uses a rule-based system of manipulating provided, pre-composed melodies to improvise new themes and variations. The agents take turns in leading the direction of the composition based on a rating system that rewards harmonic consistency and melodic flow. In developing the multi-agent system it was found that implementing embedded spaces in the LSTM encoding process resulted in significant improvements to chord sequence learning.

Keywords: Multi-agent systems · Music composition · Artificial neural networks

1 Introduction

1.1 Virtual Improvisers

Generating original music in real-time for live performance or scoring of dynamic media such as games presents many unique challenges. Music should adapt to the changing mood while ensuring musically consistent results. This paper presents an approach to real-time, 'vertical' [10] composition that uses two collaborative virtual agents: a harmony improviser and melody improviser. The approach was developed to create an original performance system that demonstrates common harmonic structures and melodic consistency with real-time adaptability. The specific technique of using improvising role-focused agents comes from observations of co-operative composition in small jazz ensembles where players face the challenges of real-time composition through co-agency [3].

The system uses short melodic themes provided by a human composer to seed the improvisational process. The agents explore variations of the themes in parallel with chord progressions based on patterns found in well-known jazz compositions. Agents self-rate their proposed melodies or chord sequences based on harmonic suitability to select a melody or harmony lead state (see Fig. 1).

J. Correia et al. (Eds.): EvoMUSART 2017, LNCS 10198, pp. 114–127, 2017.
DOI: 10.1007/978-3-319-55750-2_8

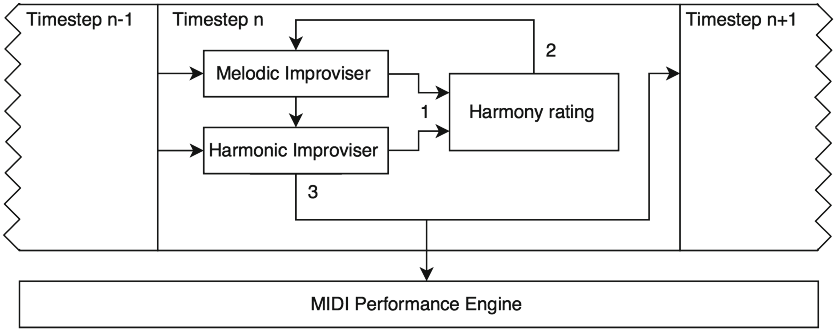

Fig. 1. Overview of composition system by time-step with data flow following the numbered pathways. Improvisation agents propose sequences that are then rated. The agent with the lower rated proposal then moves to an accompanying role and generates a suitable accompaniment for the higher rated proposal. The music is performed and the harmonised sequence is used as input for the next time-step.

The two agents have distinct roles and therefore utilise different techniques. The harmony agent is built on an LSTM neural network trained on the chord progressions of 2986 jazz standards. The melody agent uses pre-seeded melodies and permutations to generate phrases using a knowledge-based model of melodic development.

The basis for the use of these specific techniques comes from an evaluation of the relation between harmony and melody in melodic composition and co-operative composition techniques used by jazz ensembles. These background concepts are presented along with related works and followed by a rationale for specific techniques explored.

2 Background

2.1 Melody and Harmony

In melodic music, melody and harmony typically function in a codependent relationship. Harmony, whether explicit or implied, provides context for melodic phrases and can have a dramatic effect on their affective qualities. Conversely the melody's shape can suggest a harmonic direction or imply a harmony that isn't performed. A composers inspiration may come in the form of a melody, for which they then searches for an effective harmonisation, or a pleasing harmonic progression may inspire new melodic phrases.

The relationship between melody and harmony has proved to be a significant challenge to manage algorithmically and one of the many hurdles on the path to emulating human-like composition skills with computers. The challenge is compounded when generating music in real-time as it involves structures of various length that require consideration beyond the next few moments of music. It is useful to look at music traditions that involve real-time composition to see working systems of dynamically shaping harmony and melody into an unknown future.

2.2 Improvisation Techniques in Jazz

In jazz it is typical for performers to take different roles in a process of collaborative improvisation [3]. Among professional musicians these roles aren't static; much of the excitement of a great performance comes from the shifting dynamic of the relationship *between* the performers. However, at any given time each musician will usually be focused on a melody or accompaniment role. This idea is present in the terminology used by jazz musicians. Percussionists, bassists, guitarists and pianists are often referred to as 'rhythm-section' players and wind instrument players and vocalists are called 'lead' players.

Jazz ensembles typically prioritise harmonic structure. A sequence of chords are repeated and each musician improvises melodic lines or accompaniments to match the sequence [12]. Rhythm-section players often change the tempo, key and rhythmic feel using their ears and intuition. Lead players reshape, or replace melodies, but the fundamental shape of the harmonic structure is usually maintained. By focusing on one role, musicians have the artistic freedom to contribute to the direction of the overall composition structure while avoiding unwanted chaos that can come from overlapping parts.

With professional performers the concept of accompaniment ('comping') goes beyond a simple dynamic of leader and supporter to a relationship of co-agency. The accompanist and soloist work towards a common goal of improvising a creative and original composition while maintaining autonomy to contribute their own artistic signature [2].

3 Related Works

3.1 Computer Improvisation

Various approaches have been used to generate music in real-time both as a standalone composition and for interactive use with human performers. Impro-visor [13] is a software package that can be used to generate jazz melodies using probabilistic grammars. Different grammars trained on solos of well known jazz musicians generate melodies in real-time and demonstrate idiomatic qualities of jazz. Biles' GenJammer [5] uses a genetic algorithm to produce jazz styled melodic phrases.

Pachet's Continuator [17] uses a method of detecting phrase endings, sequence patterns and global properties to trade lines with human improvisers. The sequence pattern analysis uses a Markov model which allows for training to occur during a performance.

Pachet and Roy presented a model of generating lead sheets, including chord sequences, trained on a specific composer to produce new chord sequences in a similar style [18]. Constraints were used to filter procedures to control the bar-relative timing of chords and avoid producing long sequences appearing in the training material. Papadopoulos, Roy and Pachet [19] used Markov constraints for an online lead sheet generation tool that produces music of various styles. The authors pointed out that statistical models, such as Markov models, are more effective at developing short passages than larger, global structures of a composition.

Eigenfeldt and Pasquier's Kinetic Engine [8,9], uses a multi-agent model to generate music using melody, harmony and rhythm agents. Different techniques have been used in different iterations of the software including genetic algorithms and Markov chains.

3.2 LSTM Neural Networks

LSTM neural networks are a class of recurrent neural networks that use gate layers to simulate memory. The memory mechanism in LSTM networks are well suited to learning temporal sequences such as in music. Eck and Schmidhuber [7] used LSTM neural networks to generate chord sequences and melody/chord combinations trained on a jazz-version of the 12 bar blues and simple provided melodies using a single pentatonic scale.

Recently larger datasets have been explored for training LSTM networks. The use of graphics processing units (GPUs) has significantly decreased the training and running time of deep neural networks by running calculations for thousands of artificial neurons in parallel and current hardware is capable of training networks on datasets with millions of tokens in hours or days. Sturm et al. [22] trained an LSTM neural network on 23,958 folk compositions using a large 'deep' network structure. The network has produced thousands of folk compositions published on a regular basis online. The demonstrated effectiveness of LSTM networks for composition in a specific musical style inspired the exploration of LSTM networks in the research presented here.

A study by Choi, Fazekas and Sandler [6] used chord sequences from the jazz Real Books to train a deep LSTM networks to produce new chord sequences. They demonstrated the potential of the approach by generating chord sequences that exhibited structures commonly seen in jazz. Empirical measurements of validation perplexity and details of the model design including method of encoding tokens and justification of network size were not published. These findings invited an exploration of the role of hyper-parameters such as the number of LSTM units per layer, number of hidden layers, learning rate and encoding techniques on perplexity on the same dataset in this paper.

4 Rationale

There are many different possible approaches to real-time music generation. The rationale for the system design presented in this paper is based on the background concepts and analysis of related works.

4.1 Adapting Human-Made Content

The ability to combine the skills of professional composers with the dynamic variability of real-time algorithmically controlled arranging has direct applications in scoring interactive media. A professional composer can develop musical themes based on knowledge of the broader context of the listener, listening environment and associated visual content or narrative, which is difficult to simulate algorithmically.

Augmenting a professionally composed piece of music with an intelligent means of real-time adaptation is desirable if the adaptation does not come at significant cost to the overall quality of the piece. Using common melody manipulations on pre-composed melodies provides a means of adaptation with a high confidence of musical consistency in style and quality. It is also an improvisation device observed in jazz [4].

4.2 Two Agent Approach

Three commonly appearing role types were identified in small jazz ensembles: harmony, melody and percussive rhythm. Harmony and melody were of particular interest in this research as their co-dependent relationship is a challenge to manage in real-time. Rhythmic relationships can also still be explored without percussive rhythm and many ensembles do not contain percussive instruments. For these reasons a system with a harmony and melody agent was developed first and proposals for the development of larger systems with expanded roles are discussed in future works.

For jazz musicians, knowledge of the 'standard' repertoire allows for a shared starting point for a performance and establishes the clichés of the idiom. Deep learning provided a method of training the harmony agent on thousands of jazz standards to gain this familiarization with the harmonic language of jazz. For the melody agent priority was given to adaptation of provided melodies to dynamic harmonic context, so a system based on modifying existing themes through commonly used melody manipulation techniques was tested.

4.3 Long Short-Term Memory Neural Networks

An advantage of using an LSTM model with jazz compositions is that jazz standards typically use a repeated chord sequences of several hundred beats, a length suited to LSTM models that can be trained and run on current hardware. Jazz chord sequences provide new challenges as a training corpus compared to other genres of music such as folk and pop because of the range of chord types and flexible approach to tonality. Jazz also proves an interesting genre for training a system designed for real-time music generation as it is a genre built on improvisation, as already described.

While whole compositions can be created by sampling from a single network, such systems have limited real-time controls and continue to lack the melodic and rhythmic consistency that human composers can achieve with relative ease. Generating chord sequences is a problem where an expansive knowledge of the idiom is of use but intricate temporal detail is not as important as it is in melody generation. An LSTM network was chosen for the harmonic agent to use these strengths and alternate methods were selected for melodic generation where the network would not perform as effectively or flexibly.

Deep artificial neural networks should capture greater stylistic range within a single network due to their larger size and memory mechanisms. The memory mechanism in LSTM networks also increases the potential for learning more distant, timing dependent relationships in composition structures.

Recent advancements in artificial neural network techniques have shown impressive results in language processing [21] that have not yet been applied to music composition or symbolic music analysis. The use of embedded spaces in place of one-hot encoding and peephole mechanisms have been shown to result in significant performance improvements in some cases [11].

5 Harmonic Improviser

5.1 Method

The chord progressions of jazz standards from the Real Book series of jazz books were used to train an LSTM neural network. A script was developed to parse the chord symbols from 3020 Band-in-a-box files of Real Book standards used by musicians as practice accompaniments. As a range of time-signatures are used in jazz and chord changes often happen mid-bar, chords were entered into a database on a per-beat basis. Arrangements with more than 400 beats were removed to focus on simple arrangements leaving 2986 compositions in the database used for training the neural network. Chords were simplified to four note spellings and represented by word tokens. For example C9 and C7b13 would both be simplified to C7.

Analysing tonality in jazz chord progressions is complex as it is common for compositions to move between multiple keys even within a 32 bar melody and atonality was widely embraced in modern jazz. For this reason it was not possible to normalise the scores to a common key. Each composition was transposed in to all 12 keys in the modern western equal temperament tuning system resulting in a corpus of 35832 sequences with a vocabulary of 1668 chord types. The data set covers a wide selection of music styles within the loose 'jazz' label as stylistic flexibility was desirable and 'deep learning' neural networks perform well with large, diverse datasets.

A recurrent neural network (RNN) with LSTM was built using Tensorflow [1]. Initial experiments began with three hidden layers and 512 LSTM units per layer,

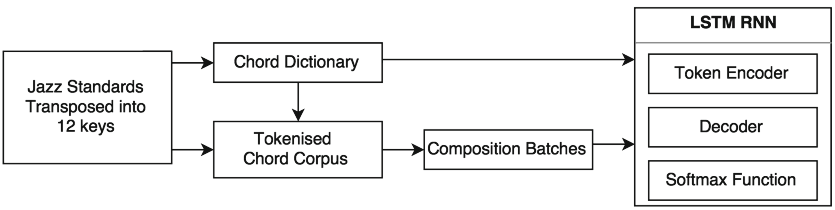

Fig. 2. Training the LSTM neural network on jazz standards

based on the model effectively used by Sturm, Santos and Korshunova [22] to generate folk compositions. A validation set was made from randomly selecting 10% of the tokenised corpus. All of the sequences were padded to have a consistent length of 400 tokens represented with one-hot encoding.

A softmax function was used to output the predicted probability of each token in the vocabulary being the next chord in an input sequence. This was achieved by utilising a sequence-to-sequence LSTM neural network structure. By using a common dictionary for the encoder and decoder and feeding the decoder with the token succeeding the token in the training batch fed to the encoder, the network can be trained to predict the next token in the batch (see Fig. 2). An ADAM [14] optimiser was used to optimise the adaptation of learning rates during training. Experimentation with learning rate and layer size resulted in some reductions to perplexity.

A multiple dimension encoding method was implemented as an alternative to one-hot encoding following the success of Rendel [21] using high dimension encoding in LSTM networks for natural language processing applications. Encoders using 10, 100, 500 and 1000 dimensions were tested.

In chord sequences the number of chords used can be small but establishing a time signature and cadence is important. In highly repetitive sections it is still important to maintain a consistent number of beats per bar, which can be a challenge in beat-by-beat based systems. Improvements were made in this area by implementing a peephole mechanism in the LSTM network where the gate layers depend on the internal state as well as the hidden state of the previous values of the cell. Peephole mechanisms have been shown to be beneficial in learning sequences where precision in timing is important [11], making them particularly relevant to modelling music structures.

Networks with 192, 256 and 512 units per layer and 2 to 3 layers were tested. Networks larger than this were deemed to be unsuited for the task of generating chord sequences in real-time due to the computational cost of running the network for each beat of music and a trend towards better training validation was not observed with increased network size. In larger networks it was observed that validation loss measures stayed above training loss which is a sign of over-fitting.

5.2 Results

Variations to the size and number of hidden layers and the number of dimensions used for token encoding affected the prediction perplexity and speed of training and running the neural network (see Table 1)

The lowest recorded perplexity of 1.61 (see Table 1) was recorded using a 100 dimension encoding with 256 units per layer across two hidden layers and was selected for use in the harmony agent. Larger dimensional representations of tokens also allowed for reductions in the number of LSTM units per layer without significant degradation to perplexity.

When being used as a composition tool the probabilities of each chord from the dictionary coming next in the sequence (C_p) can be sampled using different algorithms. A single chord token can be used to seed a new composition.

Table 1. Best prediction perplexity for different network sizes and token encodings after 20 epochs.

Encoding	One-hot	10 dim	100 dim	500 dim	1000 dim
512 units, 3 layers	2.86	2.30	2.58	1.71	1.77
512 units, 2 layers	2.75	2.05	2.59	1.93	1.74
256 units, 3 layers	3.12	2.07	1.70	1.70	1.71
256 units, 2 layers	3.10	2.59	**1.61**	1.63	1.72
192 units, 3 layers	3.42	2.10	1.83	1.73	1.74
192 units, 2 layers	4.20	1.86	1.80	1.78	1.74

After the seeding token has been used to feed the encoder the most recently selected output chord is fed into the encoder for each beat.

To operate in co-operation with other improvising agents the sampling algorithm should be dependent on the actions of other agents. This is analogous to a performer who has knowledge of the standard practices of the jazz idiom but has the freedom to take detours based on the broader musical context of the performance.

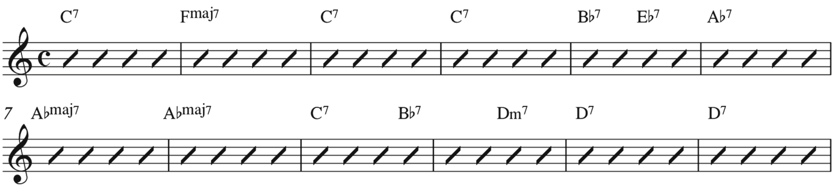

Fig. 3. A chord sequence generated by seeding the network with a single ‘C7’ token and a roulette wheel method of sampling the softmax output.

In the example presented in Fig. 3 the chord changes typically occur once every four beats implying a 4/4 time signature. The entire 12 bar sequence can be analysed in one key and typical jazz harmony techniques are evident. Bar 7 contains a flat-direction major chord, bars 11 and 12 contain secondary dominant chords. Root movements in fourths appear regularly throughout the composition. A selection of other chord progressions are presented in the evaluation.

6 Integrating Melody

6.1 Co-Agency

A melody role agent was introduced to function in co-operation with the harmony agent. The two agents work in a process of co-agency and alternate between leading and accompanying modes. The melody improviser uses a database of phrases based on common manipulations and permutations of human composed melodies and a knowledge-based model of harmony. The two agents use a rating system to find viable candidates for harmonic and melodic direction. To develop co-agency the ratings are compared to select which agent takes on a temporary leading role.

6.2 Phrase Database

The melody agent builds a database of melodic phrases based on seeding melodies of one to four bars length provided as monophonic MusicXML files. Each melody is reversed, inverted, augmented and diminished, separately and in combination. The produced phrases are then sliced, by barline as present in the MusicXML file and by equal divisions of the whole phrase down to four note batches. Each new phrase is then transposed in 12 keys and stored in the database with duplicate entries removed.

For a given phrase (P) each note (P_i) is given a rating for a provided chord sequence. Three categories were established for rating notes. Chord tones, usable tones and avoid notes. Avoid notes were based on Levine [15], extended to include dissonant notes outside of the diatonic scale (see Table 2).

Table 2. Avoid notes and chord tones for example chord types. m = minor, M = major.

Chord type	Avoid notes	Chord tones
Major	m9, M4, M7	M3, M5
Dominant	m9, M4, M8	M3, M5, n7
Minor	m9	m3, M5

A consonance function f(x) rates avoid notes, usable tones and chord tones with consonance ratings of 0, 0.5 and 1 respectively. Variations of these values could be made to adjust tension levels. A rating modifier (b) was introduced to accommodate passing tones by increasing the rating of avoid notes and usable tones by 0.25 each. The total melody rating (M_r) for each melodic phrase of length L is the average of the ratings of each note in the phrase at the time they would be played.

6.3 Lead States

The system has two distinct states. A melody lead state and harmony lead state. In the harmony lead state chord progressions are generated by the harmony agent by selecting the chord symbol with the highest C_p for each beat in the next two bars. The average C_p is recorded and used as the chord rating (C_r) The melody agent then performs a beam search on the melody database to find a melodic line with M_r (see Eq. 1) above a desired threshold to play over the chord progression. If no phrase can be found with a rating above a desired threshold then the melody improviser rests until a phrase is accepted. In the melody lead state the melody agent selects a melody to play and the harmony agent searches the output of the neural network for each upcoming beat from highest to lowest C_p to find a harmonic backing for the melody with Mr above the desired threshold.

$$M_r = \frac{1}{L}\sum_{i=1}^{L} f(P_i) + b \tag{1}$$

When in the harmony lead state, a change of state is triggered when a melodic line is found with Mr greater than C_r for the upcoming sequence or a melodic line is found with M_r above the set threshold that goes longer than two bars. In the melody lead state, a change of state is triggered when the harmonic improviser proposes a chord sequence with C_r greater than M_r for the selected melodic line (see Fig. 4).

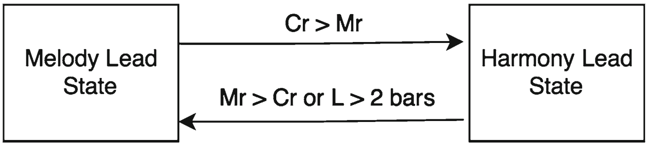

Fig. 4. State changes are triggered by differences in the rating for each agent's proposed improvisation. Melody lead states are also triggered by proposing melodies that go beyond the 2 bar sequences proposed by the harmony agent.

Joining melodic lines end-to-end allows longer melodies to be developed. Restricting the database search to lines with a first note within a specific pitch range from the last performed note reduces computational cost and encourages melodic flow. Results presented in this paper used a maximum pitch leap of a perfect fourth as it produced pleasing results.

Random rests were introduced in between melodic fragments to introduce rhythmic variation and to prevent the agent playing non-stop. With experimentation a probability of 0.1 for each of a crotchet or quaver rest being inserted was deemed pleasant and set as a default value.

By basing the change of state on the rating of the proposed harmonic and melodic direction of the music, both agents are contributing to the structural development of the performance.

6.4 Results

Static chord sequences from the Real Book dataset and the harmony agent were used to test the generation of melodies.

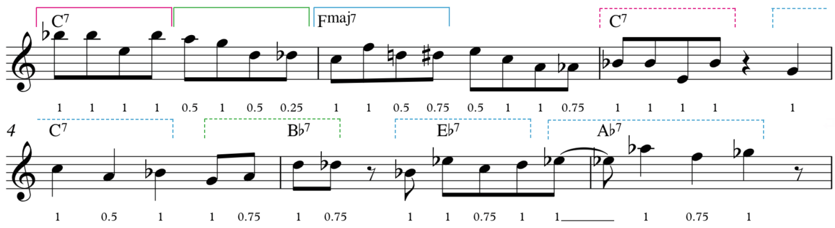

Fig. 5. Melody improvised for a static chord progression. Solid brackets show sources of fragments used, identified by colour. Note consonance ratings appear beneath each note. (Color figure online)

In Fig. 5 three phrases developed over six bars are shown. The melody played in the first two bars was provided by the first author as a seeding melody. In the melody-lead state the chord progression from C7 to Fmaj7 was selected by the harmony agent. The harmony agent proposed a continuation with a C7 chord ($C_p = 0.92$) and triggered a harmony-lead state. Throughout the three phrases fragments of the seeded melody can be seen in various manipulated forms, maintaining high ratings throughout.

The multi-agent system demonstrates commonly seen root movements, consistent time signature and melodic fit with harmony. Examples can be heard at https://dx.doi.org/10.4225/03/58882e3f5521d. The next step was to evaluate the overall quality of the generated music alongside those from existing systems and professional jazz musicians.

6.5 Evaluation

To evaluate the generated music an online survey was developed. The survey was open to participants that self-identified as professional jazz musicians with a bachelor degree in music or at least three years of professional performance experience. Advertising for the survey was done on social media sites groups for jazz musicians and jazz organisations.

In the survey 10 short scores and accompanying music files were presented to participants with chords and melodies pre-generated using different techniques and sources, one at a time (see Table 3). Participants were asked to rate each score using a 5-point Likert scale in categories of: Melodic creativity, thematic

Table 3. Sources of chords and melodic lines for scores presented in the survey. S = jazz standards, I = Impro-visor software package, A = multi-agent system, M = manually composed

Score no:	1	2	3	4	5	6	7	8	9	10
Melody	M	M	I	I	A	A	A	A	I	I
Chords	S	S	S	S	S	S	A	A	A	A

development, harmonic creativity, harmonic consistency, overall musicality and demonstrated knowledge of the jazz idiom.

Music files were produced using a high sample-rate virtual piano to maintain consistent performance characteristics of phrasing and tone for each. Manually composed melodies were produced by the lead author, who has over 10 years experience as a professional jazz musician. The 10 scores used in the survey can be heard at https://dx.doi.org/10.4225/03/58882e3f5521d.

At the time of writing 33 participants have completed the online survey. Calculating the median response for each score in each assessment category shows that improvisations produced with the multi-agent system were assessed at similar levels to those from Impro-visor and jazz standards (see Table 4).

Table 4. Median Likert scale values for each score presented in the online survey. 1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good. Highlighted columns represent scores with both harmony and melody generated by the multi-agent system.

Score no:	1	2	3	4	5	6	7	8	9	10
Harmonic Consistency	4	4	3	4	4	4	3	3	3	2
Chord Sequence Creativity	2	2	2	1	2	2	3	3	3	3
Thematic Development	3	3	3	4	3	3	4	3	3	2
Melodic Creativity	2	3	2	2	3	4	3	3	2	3
Overall Musicality	4	3	4	3	3	3	3	4	3	3
Demonstrated Knowledge of Jazz Idiom	4	4	3	3	3	3	3	3	3	3

7 Discussion

Harmonic consistency for tracks with chords generated by the system was on average rated lower than for those using standard jazz progressions but creativity was rated higher for the same tracks. A higher average rating in chord sequence creativity was likely due to professional musicians being familiar with the standard progressions and therefore deeming them uncreative. The novelty of the sequences produced by the system could be seen as both a sign of creativity and a lack of consistency.

The combination of agents with fundamentally different mechanics results in interesting dynamics and allows the strengths of those mechanics to be focused in an area of most effect. In this system the melodic agent has no training data that would allow for deep explorations of the idiom, but has a high level of musical consistency. The harmonic agent was trained on a large corpus and has a greater knowledge representation of the broader harmonic language of jazz but can be less consistent due to the wider range of possible outputs. Their combination was intended to produce a system that generates music that is identified as being creative with melodies that display a high level of musicality. The results of the survey support the general design of the system as musicality and creativity scores for generated compositions were equal or higher to music from other compared systems.

7.1 Future Work

The multi-agent system presented in this paper utilises only two agents and the database method of melodic development contains only very simple rules of manipulation. Further explorations of more sophisticated melody rules could result in improved results within the multi-agent framework. Other roles such as percussive rhythm and counter-point melody are intended in future iterations. Systems using probabilistic grammars and Markov chains have been effective for learning small time-scale structures in music which could be suited to rhythmic agents.

Jazz standards were used as training data for reasons outline in Sect. 4.3 but data sets from other music genres could also be used. Data collection for other genres such as pop and rock music is currently the subject of investigation.

In recent years, dynamic layering of instrument parts has become common in commercial game releases, where different environment properties, such as avatar movement speed, number of competing agents and player actions, add or remove layers of the score [20]. A pre-recorded instrument layering approach provides a guarantee that adding or removing layers will always work well harmonically, but more flexible, note-level arranging techniques require careful consideration of the harmonic relationships between different instrumental parts. With more stylistic range and perceptual agency [16] of the improvisers spread to non-music content for the purpose of adaptation, the model on which this system is built could have direct applications in this area.

7.2 Conclusion

The results and evaluation support the use of a multi-agent system for real-time music generation. In developing the system a significant finding was made in the application of neural network designs used in language processing for learning chord sequences. Specifically the use of embedded spaces and peephole mechanisms were found to have demonstrable benefits in this area.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: large-scale machine learning on heterogeneous systems (2015). http://tensorflow.org/, software available from tensorflow.org
2. Barrett, F.J.: Coda–creativity and improvisation in jazz and organizations: implications for organizational learning. Organ. Sci. **9**(5), 605–622 (1998)
3. Bastien, D.T., Hostager, T.J.: Jazz as a process of organizational innovation. Commun. Res. **15**(5), 582–602 (1988)

4. Berliner, P.: Thinking in jazz: composing in the moment. Jazz Educ. J. **26**, 241 (1994)
5. Biles, J.A.: Genjam in transition: from genetic jammer to generative jammer. In: Generative Art, vol. 2002 (2002)
6. Choi, K., Fazekas, G., Sandler, M.: Text-based LSTM networks for automatic music composition. arXiv preprint arXiv:1604.05358 (2016)
7. Eck, D., Schmidhuber, J.: A first look at music composition using LSTM recurrent neural networks. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale 103 (2002)
8. Eigenfeldt, A., Pasquier, P.: A realtime generative music system using autonomous melody, harmony, and rhythm agents. In: XIII Internationale Conference on Generative Arts, Milan, Italy (2009)
9. Eigenfeldt, A., Pasquier, P.: Realtime generation of harmonic progressions using controlled Markov selection. In: Proceedings of ICCC-X-Computational Creativity Conference, pp. 16–25 (2010)
10. Folkestad, G., Hargreaves, D.J., Lindström, B.: Compositional strategies in computer-based music-making. Br. J. Music Educ. **15**(01), 83–97 (1998)
11. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. **3**, 115–143 (2002)
12. Johnson-Laird, P.N.: How jazz musicians improvise. Music Percept. Interdisc. J. **19**(3), 415–442 (2002)
13. Keller, R.M., Morrison, D.R.: A grammatical approach to automatic improvisation. In: Proceedings, Fourth Sound and Music Conference, Lefkada, Greece, July. Most of the soloists at Birdland had to wait for Parker's next record in order to find out what to play next. What will they do now (2007)
14. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Levine, M.: The Jazz Theory Book. O'Reilly Media Inc., Sebastopol (2011)
16. Monson, I.: Jazz as political and musical practice. In: Musical Improvisation: Art, Education, and Society, pp. 21–37 (2009)
17. Pachet, F.: Enhancing individual creativity with interactive musical reflexive systems. In: Musical Creativity, pp. 359–375 (2006)
18. Pachet, F., Roy, P.: Imitative leadsheet generation with user constraints. In: ECAI, pp. 1077–1078 (2014)
19. Papadopoulos, A., Roy, P., Pachet, F.: Assisted lead sheet composition using FlowComposer. In: Rueher, M. (ed.) CP 2016. LNCS, vol. 9892, pp. 769–785. Springer, Cham (2016). doi:10.1007/978-3-319-44953-1_48
20. Plans, D., Morelli, D.: Experience-driven procedural music generation for games. IEEE Trans. Comput. Intell. AI Games **4**(3), 192–198 (2012)
21. Rendel, A., Fernandez, R., Hoory, R., Ramabhadran, B.: Using continuous lexical embeddings to improve symbolic-prosody prediction in a text-to-speech front-end. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5655–5659. IEEE (2016)
22. Sturm, B.L., Santos, J.F., Ben-Tal, O., Korshunova, I.: Music transcription modelling and composition using deep learning. arXiv preprint arXiv:1604.08723 (2016)