Do Jazz Improvizers Really Interact?
The Score Effect in Collective Jazz Improvisation

François Pachet, Pierre Roy, Raphaël Foulon; Email: pachetcsl@gmail.com

Sony Computer Science Laboratory, Paris, France

Abstract

Musicians, critics, musicologists and even lay audience seem to agree that improvising musicians *interact* with each other during collective improvisation. However little is known about the nature of such interactions. In particular, a key question is to which extent this interaction involves the *content* of the music (rhythm, harmony, melody, expressiveness)? Such a question is crucial for designing smarter music interaction systems. In this chapter, we propose an analytical framework to identify correlates of content-based interaction. We illustrate the approach with the analysis of interaction in a typical jazz quintet. We extract audio features from the signals of the soloist and the rhythm section. We measure the dependency between those time series with correlation, cross-correlation, mutual information, and Granger causality, both when musicians play concomitantly and when they do not. We identify a significant amount of dependency, but we show it is mostly due to the use of a common musical context, which we call the *score effect*. Therefore, we argue that either content-based interaction in jazz is a myth or that interactions do take place but at unknown musical dimensions.

The Need for Models of Interactive Music

Complexity in nature often results from interactions between individual agents. Music being a by-product of cultural evolution, it is natural to model music form a complex system viewpoint. For instance, the complex productions of a music band can be seen a resulting from a set of interacting agents. Not surprisingly, multi-agent musical systems have been used in many fields of music computing: composition, improvisation, and *interactive systems* in which human musicians play with virtual agents. For instance, Dahlstedt and McBurney (2006) describe an autonomous agent framework for computer-aided composition. Each agent represents a specific compositional process, and agents cooperate to produce music with the goal of triggering various emergent musical properties. *Mama* (Murray-Rust et al., 2006) is an interactive multi-agent architecture, in which the intention of agents is expressed in the context of speech act theory, extended to so-called “musical acts". They argue that the use of a formal set of musical acts allows precise communication between agents, avoiding the complex tasks of analyzing the intentions of other agents from their productions. This system has been extended (Murray-Rust et al, 2011) to a computational model for communication between musical agents and/or humans who engage in dialogues.

Multi-agent architectures have also been applied to real-time music *creation*. Inspired by insect swarms dynamics, Blackwell (2003) designed an improvisational system which self-organizes thanks to the interactions between agents. Similarly, Beyls (2007) describe how a “society of musical agents” may use biological principles to interact via mutual affinities. Other multi-agent systems have attempted to include humans in the loop. For instance, *Genjam* (Biles, 2001) is a system based on a genetic algorithm that learns to improvise over a jazz accompaniment, with the help of a human mentor. Similarly, *Improvagent* (Collins, 2008) discusses how human feedback can be used for reinforcement learning to improve a system performance.

Multi-agent systems are also used to *reproduce* human musical interaction. In those systems, agents representing virtual performers (soloists or accompanying musicians) typically listen to a human user's input by analysing an audio signal or symbolic data such as MIDI, and react with specific behaviors. A desired property of these reactive systems is to let interaction be defined by explicit rules associated to each agent, either programmed or learnt automatically from music material. For instance, Wulfhorst et al. (2003) describe an interactive accompaniment system in which a community of musical agents which uses “fuzzy rules" to interact with a human user. *Chimera* (Gifford and Brown, 2009) triggers various “metric scenarios”, depending on the content played by live performers, and generates a rhythmic accompaniment, and *Jambot* (Gifford and Brown, 2011) is an agent which performs “transformational mimesis”, by producing a percussive accompaniment that imitates the user's input rhythm. In *VirtualBand* (Moreira et al., 2013) and *Reflexive Looper* (Pachet et al., 2013), agents are trained and synchronized to imitate the musical style of real musicians, and produce musical output that exhibits specific audio features, depending of the audio analysis of the other agents' or human user's actions. Finally, Hamanaka et al. (2001) propose a model in which the musical reaction of a musician (a MIDI guitarist) is learnt from a series of examples. However, the system does not address the situation of simultaneous musicians playing together.

A striking aspect of all these multi-agent music systems, is that none of the interactions proposed, to our knowledge, actually relies on a precise analysis of human musical interaction. All of them are based on *a priori* models of music interaction. These models hypothesize (usually implicitly) that music interaction is a complex phenomenon, involving structures that emerge from reactive behaviors. This paper questions this hypothesis by a practical experiment, inspired by studies on interaction as described in the next section.

Analyzing Human Music Interaction

Synchronization

Musical interaction is recognized as an important feature of improvised music, in particular in jazz (Monson, 1996). Interaction includes temporal synchronization between the performers, which aims to "the achievement of a groove or feeling" (ibid., p. 28). The studies by Keller (2008) and Novembre et al. (2012) investigate joint action and musical coordination in musical performance: auditory anticipation, integrative attention, and the distinction between one's own and other's behavior. Temporal coordination, or tempo coordination, have been investigated in duets in jazz improvisations (Schögler, 2000), and piano performances (Shaffer, 1984; Moore and Chen, 2010). Synchronization and coupling phenomena between musicians have been investigated in the context of chamber music ensembles (King, 2004) and string quartets (Goebl, 2009). More recently, Müller et al. (2013) used EEG recordings of couples of guitarists engaged in improvisations in order to identify synchronization patterns at various sampling frequencies. These studies address the analysis of synchronization between musicians, either when they follow a classical music score or when they improvise, but do not investigate the impact of these interactions on musical production. Neural correlates of jazz improvisation have been found during solo performance (Limb and Braun, 2008), as well as during group performance (Donnay et al., 2014). However, specific activities in the brain do not imply that there are correlates of interaction in musical productions.

Content-Based Interaction

Musicians themselves claim to interact beyond mere temporal synchronization by adapting various dimensions of their playing, e.g., expressiveness, melody, harmony, and rhythm, to the music produced by the other performers (Moran et al., 2015). In this chapter, we call these interactions *content-based interaction*. Such interaction is performed by adapting the content of a musical output to that of the other performers (ibid., pp. 21-25).

Several studies of music group performance have been conducted to understand the mechanisms of joint action and implicit synchronization. Schober and Spiro (2013) asked jazz musicians to evaluate the interaction they experienced when they jammed together by answering a series of questions. After an initial playing phase where the musicians jammed together, they were asked to evaluate the interaction they experienced by answering a series of questions. The results revealed a fair amount of discrepancies between the musicians, and the study concludes that fully shared understanding of interactive phenomenon may not be related to the performance's quality.

Marchini et al. (2013) and Papiotis et al. (2012) used computational analysis tools to identify the difference in the productions of musicians playing musical exercises alone versus with another performer. Both audio signals and gesture data have been collected and time-series analysis tools such as Pearson's correlation coefficient and mutual information were used to evaluate the dependency between the musicians. These studies show that musicians tend to synchronize both in intonation and dynamics. However, these studies focus on the situation where musicians play a given score, so the interaction is limited to expressiveness parameters. In D'Ausilio et al. (2012), Granger causality is used to investigate non-verbal communication (gestures) between a conductor and an ensemble. This study also addresses performance of written music and therefore does not address the implicit communication occurring within an improvising ensemble.

The Score Effect

In this study, we focus on the *causal relations* that may appear between the musical productions of two musicians who interact while improvising. We study musical interaction in the context of lead sheet based jazz improvisation, such as bebop, a standard setup in which musicians improvise on a known tune (Limb and Braun, 2008). A lead sheet, such as the one displayed on Figure 1, specifies the melody and the chords. The tune is played several times in a row. A typical performance consists of one statement of the melody specified by the lead sheet, followed by a series of solos, and one final statement of the melody.

<FIGURE 1 ABOUT HERE>

During the solos, the musicians create new melodies freely, but have to stick to the skeleton of the tune, such as the imposed chord sequence, and to synchronize to salient events of the tune. This skeleton creates correlations between the content produced by individual musicians, that we call the *score effect*.

We are interested in correlations between the musician's productions that are *not* due to the score effect. To answer this question, our basic idea is to compare interdependency measures between two instruments when they played together, and when they played at different moments in time, but on the same score. Such an analysis should indeed reveal the added effect caused by actual interaction, as opposed to the score effect.

We study a corpus of jazz multitrack recordings performed during both a live concert and a studio session. In these recordings, each track corresponds to a single instrument. We extract audio feature time-series from the audio tracks, and use various interdependency measures to assess the presence of correlates of content-based interaction in these recordings. We discuss the results in the last section.

Materials and Method

Audio Recordings

Our corpus is composed of 12 multitrack audio files of performances by the "Mark d'Inverno Quintet", a jazz bebop band composed of Mark d'Inverno (piano), François Pachet (guitar), and three critically acclaimed professional jazz musicians: Ed Jones (saxophone), Larry Bartley (upright bass), and Winston Clifford (drums). Recordings were performed in two conditions: in a studio and a concert. In both conditions, each track corresponds to one instrument (saxophone, bass, drums, piano and guitar). The studio tracks were recorded live at Livingstone studios in London, in professional conditions, and musicians played in isolation booths to avoid spills (tracks mixing the sound of several instruments). The concert was recorded in Barcelona in 2011[1]. Instruments were recorded using a live setup (amplifier line out to Direct Input (DI) box for the guitar, dynamic microphones for the other instruments). Due to those live conditions, the concert recordings contain minor sound spills for the bass, so we processed the bass tracks with a low-pass filter[2] to reduce the spills. Both recordings consist of six original songs composed by the band: *I Got It Good and that Ain't Bad*, *I Just Can't Remember*, *May's Dance*, *Song Bouncy*, *Very Late*, and *Why Not*. The songs are denoted by $S_1, \ldots, S_{12}$ as follows: S_1 (resp. S_7) is the live (resp. studio) recording of *I Got It Good and that Ain't Bad*, S_2 (resp. S_8) is the live (resp. studio) recording of *I Just Can't Remember*, etc. Each song contains five tracks, one per instrument, and follows a classical jazz structure: the

[1] A video recording of the concert can be found at http://www.youtube.com/watch?v=5S1WxqR7BJE
[2] The low-pass filter has been set at 48dB per octave with a 500Hz cutoff.

band plays a succession of harmonic grids, from 8 to 16 bars long. First, the band plays the song's melody once, and then musicians take solos.

We focus on the interaction between the saxophone soloing and the bass and drums accompaniments. In the discussion section of this paper, we extend the study to the analysis of other couples of instruments.

Segmentation of the Audio Tracks

We consider the audio signals that correspond to the solos sections. There are from three to 13 solo sections in a single song, for a total of 71 sections[3]. We use the following notation to denote a specific solo in a specific track: $S_{i,t}^{m}$, where m is the index of the song, i is the index of the solo in the song, and t is the instrument, e.g., $S_{2,bass}^{1}$ is the audio signal played by the bass during the second solo on *I Got It Good and that Ain't Bad*. Figure 2 shows a couple of audio signals extracted from the saxophone and bass tracks of the eighth solo section of the song *I Got It Good and that Ain't Bad*. In the following, we investigate the correlations between such couples of signals.

<FIGURE 2 ABOUT HERE>

Feature Extraction

We consider a feature set consisting of 72 features, named f_1 to f_{72}. The set contains the following MPEG-7 low-level timbral descriptors (Chang et al., 2001): spectral centroid, spectral decrease, spectral flatness, spectral kurtosis, spectral rolloff, spectral skewness, spectral spread, harmonic spectral centroid, harmonic spectral deviation, harmonic spectral spread, harmonic spectral variation and two harmonic to noise ratios (computed with 700Hz and 1000Hz as fundamental frequencies). Following Kim and Sikora (2004), we strengthen this feature set by adding 59 low-level features. To take into account melodic and harmonic dimensions, we add the YIN pitch and inharmonicity factor[4] (De Cheveigné and Kawahara, 2002), as well as chroma. We add perceptual descriptors that are computed with psychoacoustic scales: 14 Mel-frequency cepstrum coefficients with their average value and variance, and 24 bark-band amplitudes. We add two features commonly used in music information retrieval (MIR): RMS, crest factor and zero-crossing rate. Finally, we add the

[3] available at http://www.flow-machines.com/InteractionAnalysis
[4] Both YIN features are computed with an absolute threshold of 0.2 for the aperiodic/total ratio.

high frequency content and high frequency ratio, two descriptors that provide additional timbral information.

We segment each audio file $S_{i,t}^{m}$ into beats and we extract the feature set on the raw signals for each segment, to obtain time-series of audio features. We use the shorthand notation $f_k\left(S_{i,t}^{m}\right)$, for the time series that consists of the values of feature f_k on each beat of $S_{i,t}^{m}$. We end up with 72 time-series per solo section and per instrument. Figure 2 shows the waveforms of the saxophone and bass tracks of the first solo of *I Just Can't Remember*. The figure also shows the time-series corresponding to the RMS values of the saxophone track and to the Spectral Centroid of the bass track.

Time-Series Analysis

We evaluate the correlations between couples of time-series corresponding to the same song and different instruments. Following Marchini et al. (2013) and D'Ausilio et al. (2012), we use several methods: time-series analysis (Pearson's correlation coefficient, cross-correlation functions, Granger causality); information theory (windowed mutual information) and information dynamics (Abdallah and Plumbley, 2009).

- Pearson's correlation coefficient measures the instantaneous linear dependency between two time-series, and provides a $[-1,1]$ bounded value: 1 for total positive synchronization, 0 for no correlation and -1 for negative correlation;
- Cross-correlation measures the linear dependency between series, like Pearson's coefficient, but computes it with various time lags applied between them. We use cross-correlation to analyze the non-instantaneous dependency of the time-series. The goal is to capture situations in which a musician (say, the bass player) reacts to another (the saxophone) with some delay. We compute the cross-correlation with different time lags, from 1 to 8 beats long. For each time lag, we consider the absolute value of the cross correlation;
- Mutual information (Giaşu, 1977) measures the amount of common information between two time-series. It provides an unbounded value, expressed in bits. Papiotis et al. (2012) computes mutual information with sliding windows, to assess the dependency over one and five seconds. The average value is eventually considered. We use a similar technique, but with 1-beat overlap, 16-beat long windows, to ensure tempo-independence. We compare shifted time-series: the $[i, i+16]$ beats of the saxophone time-series, with the $[i+8, i+24]$ beats of the bass time-series;

- Granger causality (Granger, 1969) is a statistical hypothesis test to determine if a time-series can predict another one. We use the causal density, which yields an unbounded value that grows with the amount of non-linear dependency. First, we pre-process the couples of time-series by demeaning and detrending them. Then, we build a Granger regression model, with 8 beats as maximal time lag, and compute the causal density.

We investigate the dependencies between the saxophone and the bass tracks for each possible couple of audio features, and for each solo section of the song. These indicators quantify both linear (correlations) and non-linear (mutual information, Granger causality) relations within the solo sections. The next section investigates the nature of these dependencies.

Score Effect or Content-Based Interaction?

For a given song S_m, a given couple of features (f_k, f_l) and an interdependency indicator D (D is either Pearson's correlation coefficient, cross-correlation functions, mutual information, or Granger causality), we compute (1) the total interaction and (2) the amount of score effect. The difference between (1) and (2) is the amount of content-based interaction. For each of the 12 songs in the corpus, we consider a total of 5184 couples (72×72) of features. For each couple of feature, we consider 4 indicators. We create two sets of dependency values:

- The first set, noted $T_{k,l,D}(S^m)$, consists of all the dependency values computed on time-series extracted from the same solo sections, i.e. when musicians played concomitantly. The values in $T_{k,l,D}(S^m)$ estimate the dependency that comes from both content-based interaction and the score effect (the musicians improvise on the same score). Let n be the number of solo sections in the song, $T_{k,l,D}(S^m)$ is defined by: $T_{k,l,D}(S^m) = \left\{D\left(f_k(S^m_{i,sax}), f_l(S^m_{i,bass})\right) \middle| i = 1, \dots, n\right\}$. Note that $T_{k,l,D}(S^m)$ contains n values.
- The second set, $C_{k,l,D}(S^m)$, consists of all the dependency values computed on time-series extracted from crossed solo sections. These values measure only the score effect, as no direct interaction between musicians is at play (they played on the same score but *not* concomitantly). $C_{k,l,D}(S^m)$ is defined by: $C_{k,l,D}(S^m) = \left\{D\left(f_k(S^m_{i,sax}), f_l(S^m_{j,bass})\right) \middle| i \neq j\right\}$. Note that $C_{k,l,D}(S^m)$ contains $\frac{n\times(n-1)}{2}$ values.

For each song S_m, $C_{k,l,D}(S^m)$ consists of all the measures of the score effect and $T_{k,l,D}(S^m)$ consists of all the measures of the total amount of dependency. Before we interpret the values contained in these sets, we assess their statistical significance.

Interpretation

For each song S_m, $m = 1, \dots, 12$ each couple of features (f_k, f_i) and each indicator D, we use the one-way ANOVA (Bartko, 1966) to determine if the data contained in the two sets $T_{k,l,D}(S^m)$ and $C_{k,l,D}(S^m)$ is statistically significant, by analyzing the ratio between the between-sets and within-sets variances. We run the ANOVA with the threshold $\alpha = 0.05$. If the test fails for the couple of features (f_k, f_i) and the indicator D, the dependency measures contained in the sets are not statistically significant over the various combinations of solo sections, and we cannot interpret the data. On the contrary, if $T_{k,l,D}(S^m)$ and $C_{k,l,D}(S^m)$ pass the ANOVA, we interpret the data as follows.

We first compute the *Mean Relative Change* between the sets, defined by:

$$MRC_{k,l,D}(S^m) = \frac{mean\left(T_{k,l,D}(S^m)\right) - mean\left(C_{k,l,D}(S^m)\right)}{mean\left(\left|C_{k,l,D}(S^m)\right|\right)}$$

$MRC_{k,l,D}(S^m) > 0$ indicates that the amount of total interaction is higher than the score effect. We select the couples of features (f_k, f_i) for which this is the case, and discard the other ones. Strong content-based interaction should present a high Mean Relative Change. For instance, .1 indicates that content-based interaction is $1/_{10}$ th of the score effect.

Results and Discussion

This section presents the results obtained for the interaction between saxophone and bass. For each indicator D we select (f_k, f_i) the couple of features for which the highest number of songs have passed the ANOVA test with a positive MRC. Formally, k, l are defined as follows:

$$k, l = \underset{i,j}{\operatorname{argmax}} \left|\left\{m \in \{1, \dots, 12\} : MRC_{i,j,D}(S^m) > 0\right\}\right|$$

If several couples of features (f_k, f_i) maximize the number of songs with a positive $MRC_{i,j,D}$, we consider all of them (see Figure 3). The *main result* of this analysis is that there is no couple of features for which more than 5 songs (out of 12) pass the ANOVA test with a positive *Mean Relative Change*. This absence of a clear correlate is further confirmed by the lack of consistency between indicators, as discussed below.

<FIGURE 3 ABOUT HERE>

Correlation Indicators

For Pearson's correlation coefficient, the best couple of features is $\{RMS, SpectralRolloff\}$. Five songs pass the ANOVA with a positive value of MRC (see Table 1). Furthermore, the maximum value for the correlation coefficient in $T_{k,l,D}$ is 0.26 which is low. The average of the mean relative change values is 0.9, which indicates that the amount of score effect is comparable to that of content-based interaction. However, this happens on a fraction of the corpus.

For the cross-correlation indicator we obtain 20 couples of features (Table 2). For these couples we get significant correlations in at most 4 songs out of 12. The best average correlation value is 0.28, which is higher than the best correlation measured with Pearson's coefficient. However, the best couple for Pearson's ($\{RMS, SpectralRolloff\}$) is absent from those 20 couples of features. Additionally, the maximum number of common songs between these 2 indicators is only 3 (obtained for the couple $\{CrestFactor, SpectralDecrease\}$ with 3 beat offset, and songs S^3, S^{10}, S^{11}).

Mutual Information and Granger Causal Density

For the sliding window mutual interaction indicator, we obtain 3 best couples of features. Only 3 songs pass the ANOVA for these couples (see Table 3). The Granger causal density indicator detected a single couple of features (spectral kurtosis, harmonic spectral spread) for which four songs passed the ANOVA with a positive MRC (0.34). As a summary, we observe no songs and no couples of features shared across the indicators. Marchini and Papiotis (2012) argue that interaction takes place when at least two indicators yield the same couples of features. Therefore, our findings support the claim that the amount of content-based interaction observed is mostly due to spurious data effects. We analyzed 5184 couples of feature time-series, with four interdependency indicators (actually 11 if we count all the time lags for cross-correlation). In such a high-dimensional space, statistical artifacts may be related to the curse of dimensionality.

Interaction between Other Couples of Instruments

We presented the results obtained for saxophone and bass. We investigated relations in a similar way for other couples of instruments: saxophone against drums, bass against drums, and drums against bass. Indeed, these are couples of instrument which are said to interact (see e.g. “musical interaction within the rhythm section and between the rhythm section and the soloist is [a] distinctive process in jazz improvisation” (Monson, 1996, pp. 17-18). The results we obtain for saxophone against drums are similar. Measures reveal less candidates than with the saxophone/bass couple, the maximum average correlation result is 0.26, and, like in the saxophone/bass study, no candidate is detected for more than one interdependency indicator (correlation functions, mutual information or Granger causal density). For the sake of brevity, we do not display the full results here[5].

The dependency we measure between the instruments of the rhythm section (bass against drums and drums against bass) show slightly better results: seven songs passed the statistical significance test, and we measure a maximum average correlation of 0.4. Yet, interaction consistency is also low (about 58% of the dataset), and the dependency values are weak (< 0.5). Finally, no candidate is shared by more than one indicator.

These results seem to further confirm our claim that the amount of content-based interaction, as we define it here, is mostly due to statistical artifacts, and that correlations observed between audio signals are spurious relationships caused by the score effect.

Conclusion

The goal of this study is to assess to which extent human musicians improvising jazz with a shared lead sheet actually interact with each other during solos. To this aim, we proposed a framework for analysis, based on the comparison of correlation estimators computed when musicians play together and when they do not. We illustrated our approach with the analysis of multitrack audio recordings of jazz performances. We studied the correlation between the solo (saxophone) player and the rhythm section (bass and drums), as well as between the members of the rhythm section. We analyzed the audio signals corresponding to performances where the musicians played concomitantly, and when they did not. We used 72 features, and 4 types of correlation indicators. We did not find statistically significant correlates of interaction between audio signals, beyond the score effect. Therefore, we argue that either content-based

[5] They are available at http://flow-machines.com/InteractionAnalysis

interaction in jazz is a myth or that (more probably) interactions do take place but at yet unknown musical dimensions.

The interactive behaviors investigated in this study are somewhat limited: we were looking for low-level correlates of interaction, and not for high level ones such as rhythmic patterns or melodic citations. However, our study questions the use of feature-based mappings for the design of musical agents to simulate interactive behaviors. More generally, it questions the basic, often implicit, hypothesis that complex phenomena (like ensemble jazz improvisation) actually involves complex interactions. Such hypothesis should not be taken for granted, and should be investigated further.

Acknowledgments

This research is conducted within the Flow Machines project which received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 291156.

References

Abdallah S., Plumbley, M. (2009) Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, 21: 89-117.

Bartko, J.J. (1966) The intraclass correlation coefficient as a measure of reliability. *Psychological reports* 19:3-11.

Beyls, P. (2007) Interaction and self-organisation in a society of musical agents. *Proceedings of ECAL 2007 Workshop on Music and Artificial Life (MusicAL 2007)*, Lisbon, Portugal.

Biles, John A. (2001) GenJam: evolution of a jazz improviser. In *Creative evolutionary systems*, Peter J. Bentley and David W. Corne (Eds.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 165-187.

Blackwell, T. M. (2003) Swarm music: improvised music with multi-swarms. *Proc. of the 2003 AISB Symposium on Artificial Intelligence and Creativity in Arts and Science*, pp. 41-49.

Chang, S.F., Sikora, T., Purl, A. (2001) Overview of the mpeg-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688-695.

Collins, N. (2008) Reinforcement learning for live musical agents. *Proceedings of the International Computer Music Conference (ICMC)*, Belfast.

Dahlstedt, P., McBurney, P. (2006) Musical agents: Toward computer-aided music composition using autonomous software agents. *Leonardo* 39:469-470.

D'Ausilio, A., Badino L., Li, Y., Tokay, S., Craighero, L., et al. (2012) Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE* 7(5): e35757.

De Cheveigné, A., Kawahara, H. (2002) Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America* 111(4):1917-30.

Donnay, G.F., Rankin S.K., Lopez-Gonzalez, M., Jiradejvong, P., Limb, C.J. (2014) Neural substrates of interactive musical improvisation: An fMRI study of trading fours in jazz. *PLoS ONE* 9(2): e88665.

Gifford, T., Brown, A.R. (2009) Do androids dream of electric chimera? In Sorensen, Andrew C. (Ed.) *Improvise: The Australasian Computer Music Conference 2009,* Australasian Computer Music Association (ACMA), Brisbane, pp. 56-63.

Gifford, T., Brown, A.R. (2011) Beyond reflexivity: Mediating between imitative and intelligent action in an interactive music system. *Proc. of 25th BCS Conference on Human-Computer Interaction*.

Goebl, W., Palmer, C. (2009) Synchronization of timing and motion among performing musicians. *Music Perception: An Interdisciplinary Journal* 26:427-438.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424-438.

Guiaşu, S. (1977) Information theory with applications. New York: McGraw-Hill.

Hamanaka, M., Goto, M., Otsu, N. (2001) Learning-based jam session system for a guitar trio, *Proceedings of the 2001 International Computer Music Conference*, pp.467-470.

Keller, P.E. (2008) Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 205-221). Amsterdam: IOS Press.

Kim, H.G., Sikora, T. (2004) How efficient is MPEG-7 for general sound recognition? *In: Audio Engineering Society Conference: 25th International Conference: Metadata for Audio*. Audio Engineering Society, London, UK.

King, E.C. (2004) Collaboration and the study of ensemble rehearsal. *Proceedings of Eighth International Conference on Music Perception and Cognition (ICMPC),* Northwestern University, Evanston, Chicago, IL, pp. 11-16.

Limb, C.J., Braun, A.R. (2008) Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. PLoS ONE 3(2): e1679.

Marchini M., Ramirez R., Papiotis P., Maestre E. (2013) Inducing rules of ensemble music performance: A machine learning approach. *Proceedings of the 3rd International Conference on Music & Emotion (ICME),* Jyväskylä, Finland, June 2013. Geoff Luck & Olivier Brabant (Eds.). University of Jyväskylä.

Monson, I. (1996) *Saying Something: Jazz Improvisation and Interaction*. Chicago: University of Chicago Press.

Moore, G.P., Chen, J. (2010) Timings and interactions of skilled musicians. *Biological cybernetics*, 103(5): 401-414.

Murray-Rust, D., Smaill, A., & Edwards, M. (2006). MAMA: An architecture for interactive musical agents. In G. Brewka, S. Coraeschi, A. Perini, & P. Traverso (Eds.), Proceedings of *ECAI 2006,* (pp. 36-40). Amsterdam: IOS Press.

Murray-Rust, D., Smaill, A. (2011) Towards a model of musical interaction and communication. *Artificial Intelligence* 175(9-10): 1697-1721.

Moran, N., Hadley, L.V., Bader, M., Keller, P.E. (2015) Perception of ‘Back-Channeling’ Nonverbal Feedback in Musical Duo Improvisation. *PLoS ONE* 10(6): e0130070.

Moreira, J., Roy, P. and Pachet, F. VirtualBand: Interacting with Stylistically Consistent Agents. *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, pp. 341-346, Curitiba (Brazil).

Müller, V., Sänger J., Lindenberger, U. (2013) Intra- and inter-brain synchronization during musical improvisation on the guitar. *PLoS ONE* 8(9): e73852.

Novembre, G., Ticini, L., Schütz-Bosbach S., Keller P. (2012) Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex* 22(12): 2894-2903.

Pachet, F., Roy, P., Moreira, J. and d'Inverno, M. Reflexive Loopers for Solo Musical Improvization. *Proceedings of SIGCHI ACM Conference on Human Factors in Computing Systems*, pp. 2205-2208, Paris (France), best paper honorable mention award.

Papiotis, P., Marchini, M., Maestre, E. (2012) Computational analysis of solo versus ensemble performance in string quartets: Dynamics and intonation. *Proceedings of the 12th International Conference on Music Perception and Cognition*.

Schober, M.F., Spiro, N. (2013) How much do jazz players share understanding of their performance? A case study. *Proceedings of the International Symposium on Performance Science 2013*, Aaron Williamon and Werner Goebl (Eds), Brussels, Belgium, p p. 257-262.

Schögler, B. (2000) Studying temporal co-ordination in jazz duets. *Musicae Scientiae*, 3:75-91.

Shaffer, L.H. (1984) Timing in solo and duet piano performances. *The Quarterly Journal of Experimental Psychology* 36: 577-595.

Wulfhorst, R.D., Nakayama L., Vicari R.M. (2003) A multiagent approach for musical interactive systems. *Proceedings of AAMAS, the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, pp. 584-591.