Mining Interaction Log Data in a Creative Arts Practice using Transition Matrices

1st Author Name
Affiliation
City, Country
e-mail address

3rd Author Name
Affiliation
City, Country
e-mail address

4th Author Name
Affiliation
City, Country
e-mail address

ABSTRACT
In this note we present an analysis of an interaction data archive associated with a long-term practice-led research program investigating ensemble musical performance practice on touch screens. We use transition matrix measures to distinguish between sessions of different contexts, and the internal structure of these sessions. The data analysis proceeds by first classifying these data into higher level gesture types and summarising the performers' transitions between these gestures. We find that two transition-matrix measures, flux, and entropy, can be used to distinguish different performance types and contexts, as well as the internal structure of the performances in our corpus.

Author Keywords
Creativity Support Tools; Collaborative Interaction; Methodology

ACM Classification Keywords
H.5.5. Information Interfaces and Presentation (e.g. HCI): Sound and Music Computing—*Methodologies and techniques*

INTRODUCTION
Modern computing devices and interfaces continue to present opportunities to support new kinds of creative expression and collaborative interaction [10]. However, it has been acknowledged that understanding and evaluating interfaces in creative contexts can be very challenging [12]. This problem is even more pronounced in artistic *performance*, where subjective interface evaluation through questionnaires or interviews may be both practically difficult and culturally inappropriate. However, when an artistic interface is digital, it becomes possible to log artistic performances and these logs can be used for curation and preservation [3]. A question we address is: can this log data be used to understand and support a broader creative-practice-led research program?

In this note we show how an archive of collaborative performance logs can be quantitatively analysed using estimated transition probabilities between gestural states. We will demonstrate that two measures on transition matrices, flux and entropy, show significant differences between different kinds of performances, and yield insight into the internal structure of these performances.

Our archive of interaction data-logs has been accumulated to support a long-term, practice-led, research program into simple iPad interfaces for group music-making. Over a period of more than two years (April 2013–June 2015), we collected data from group music-making sessions where participants used our custom-built, touch-screen apps on Apple iPad devices. Although several different apps have been included in our corpus, they shared a common mode of free-form touch interaction where tapping produced short sounds and swirling or swiping produced long sounds. Apart from these similarities, the apps featured a variety of sound palettes, different kinds of visual feedback, and different kinds of networked interactions (between iPads in an ensemble and with a central server).

In all of our 95 interaction sessions, every touch interaction (touches begin, touches move, touches end) was timestamped and logged to disk for later analysis. These sessions included *rehearsals* without an audience, as well as *concerts*, in front of a live audience, where performative goals were paramount. They also included a number of *recording* sessions for the purposes of formally-structured HCI experiments. In the recording sessions, the performers were also interviewed and surveyed– something which would have been disruptive and inappropriate for the concerts. Survey and interview data were not part of the corpus. Some of the sessions were iPad-only and others used iPads together with other acoustic percussion instruments. Although the majority of the sessions were free-form improvisations, a number of rehearsals and performances involved composed pieces where the musicians read and performed percussion notation using the apps. Some of these session contexts are are shown in Figure 1 and other descriptive statistics about the corpus are shown in Table 1. In our corpus, the median length of sessions is 7m26s and the median number of participants per ensemble is four. Each session record is labelled with metadata about its *context* (concert, rehearsal, recording), its *type* (improvisation or composition), and its *instrumentation* (iPads-only, or iPads with acoustic percussion). The distribution of recorded sessions by context and type is shown in Figure 2.

Figure 1: Collaborative musical interaction with our touch-screen interfaces in two different *contexts* and two different *instrumentations*: (left to right) a rehearsal with iPads only, concert performance with iPads and percussion instruments, and concert performance with iPads only.

	Total	Min	Median	Max
N	95			
Length	13H9M19S	1M19S	7M26S	22M20S
Participants	353	2	4	9
Flux		0.02239	0.1445	0.3246
Entropy		0.7683	3.342	4.442

Table 1: Representative statistics from the corpus of musical touch-screen interaction sessions. Each record in the corpus consists of low level touch-data recorded in CSV format. Many of the participants were present in multiple sessions, but this data was not tracked.

MACRO GESTURES AND TRANSITION MATRICES

A useful way of modelling user interfaces is to consider their theoretical configurations as state vectors and to describe transitions between these states during an interaction sequence [15, 14]. An inverted approach is to record users' interactions and manually code interaction states to produce a transition model. This method has been applied to usability analysis of many applications outside of the creative arts, for example [4] studied resource-planning software for security applications and used the transition probability matrix of these state-sequences to draw conclusions about the nature of interactions with that software. As an example from the creative arts, protocols obtained from a computer interface for "live coding" computer music have been coded as states and analysed using transition matrices [13] in both the textual (editing) and musical dimensions, in order to arrive at conclusions regarding artistic style.

In our corpus of iPad musical performances, logs of low-level touch-data record each performer's micro-gestures. Because it has been established that musical performances on touch-screens define vocabularies of continuous macro gestures that unfold over several seconds of touch interaction [6], and that gestural performances can be divided into sequences of such states [9], these logs can be used to identify a sequence of such states for each performer. If a performers' touch interactions can be classified as belonging to one of m distinct states, then, over the course of a session, each participant X generates a length N sequence of these interaction states X_n ($n = 1, \ldots, N$) which can be modelled as a Markov process. If there is an (automated) classification process for identifying each participant's state sequence from their interaction data, then we can estimate the $m \times m$ transition matrix (TM), P, which characterises this process by setting each entry, p_{ij}, to be the proportion of times state j follows state i in the sequence. The transition activity of the whole ensemble can be calculated by averaging the TM for each performer. The matrix can be normalised so that the sum of all elements is equal to one.

To generate state sequences for our corpus we have used an automated method to sample the data at regular intervals and to classify the logs of touch events into the nine continuous touch-screen gestures identified by Martin et al [7]. A Random Forest [2] classifier[1] was applied to 5-second windows of touch-data in our corpus to provide chains of gesture-states for each participant in each session. These chains can be used to create 9×9 transition matrices for individual performers and the whole ensemble in each session.

QUANTITATIVE TRANSITION MATRIX MEASURES

While the interaction sessions recorded in our corpus were performed by many different participants in different contexts and on different instrumental setups, the common interface and method of recording the data allows all the sessions to be transformed into gesture sequences, and transition matrices. While these matrices can be used to visualise the structure of individual performances, as has been done in [13], we also wish to extract quantitative measures from these TMs to compare the large number of sessions in our corpus. To do this we calculate two high-level quantities from each session: "flux" and "entropy". The flux of a sequence is defined as the ratio of state transitions (e.g. $A \to B$ where $A \neq B$), to self transitions (e.g. $A \to A$). The flux of a TM P is given by: $\text{flux}(P) = 1 - \text{tr}(P)$ where the trace $\text{tr}(P)$ is the sum of the diagonal entries of P. Intuitively, flux is a measure of how frequently the participant/group changes state. Flux returns a value in the interval $[0, 1]$ and will return 0 when participants

[1]The Random Forest classifier used was from Python's Scikit-learn package[8].

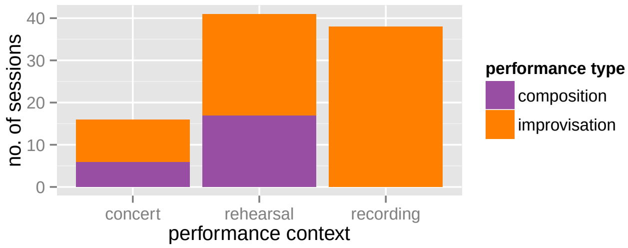


Figure 2: Distribution of the sessions in our interaction corpus by session *context*. Improvisations or compositions were repeated several times in rehearsals and recording sessions but only performed once in concerts.

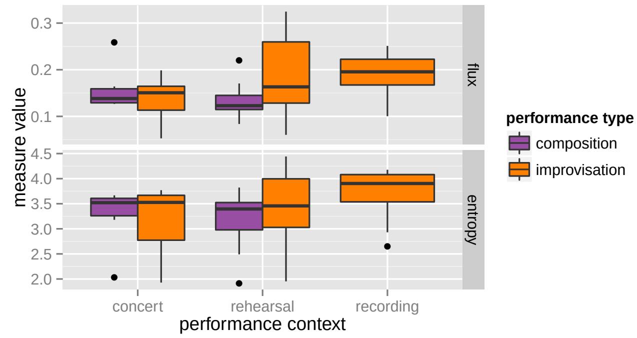


Figure 3: Distributions of flux and entropy values by performance context. Recording and rehearsal sessions had the highest flux for improvisations. Compositions had lower flux in rehearsals than concerts.

never change state, and 1 when participants never stay in the same state for two consecutive measurements.

Another, evocative measure that can be defined on a TM is its entropy, defined in the information-theoretic[11] sense: $H(P) = -\sum_{i,j} p_{ij} \log_2(p_{ij})$. This measure is small when the matrix is sparse, and largest when each matrix element is equal. It offers a different perspective on collaborative interactions than the flux measure by capturing the breadth of the gestural space explored by the participants throughout the course of an interaction. Consider a degenerate case of a participant who alternates between just two states for a whole performance: the flux in this case will be maximal ($\text{flux}(P) = 1$) since there are no self-transitions, (only $A \to B$ or $B \to A$) even though the participant has only used a small subset of the state space. The entropy of this interaction, on the other hand, will be low. Entropy, therefore, is a measure of how broad the participant's exploration of the state space is in a given interaction.

We use these matrix measures to answer two questions. First, do the transition matrix measures allow us to distinguish between different performance contexts, types and instrumental setups? Secondly, do these measures allow us further insight into the internal structure of performances?

Differentiating interaction sessions

The distributions of flux and entropy with respect to session *contexts* (concert, rehearsal, recording), *types* (improvisation or composition) and *instrumentation* (iPads-only, or iPads with acoustic percussion) are shown in Figure 3. A three-way ANOVA procedure was performed to find whether the metadata could successfully predict the entropy and flux values.

All main effects on flux were found to be significant: session context ($F(2, 86) = 6.28, p < 0.01$), session type ($F(1, 86) = 7.21, p < 0.01$), and instrumentation ($F(1, 86) = 20.06, p < 0.001$). Significant interaction effects were also found for session context and type ($F(1, 86) = 6.18, p < 0.05$), and session context and instrumentation ($F(1, 86) = 10.48, p < 0.01$).

For entropy, significant main effects were found due to session context ($F(2, 86) = 12.10, p < 0.001$), and instrumentation ($F(1, 86) = 25.86, p < 0.001$). There was no main effect for session type although the interaction of session type and instrumentation was found to be significant ($F(1, 86) = 9.93, p < 0.01$).

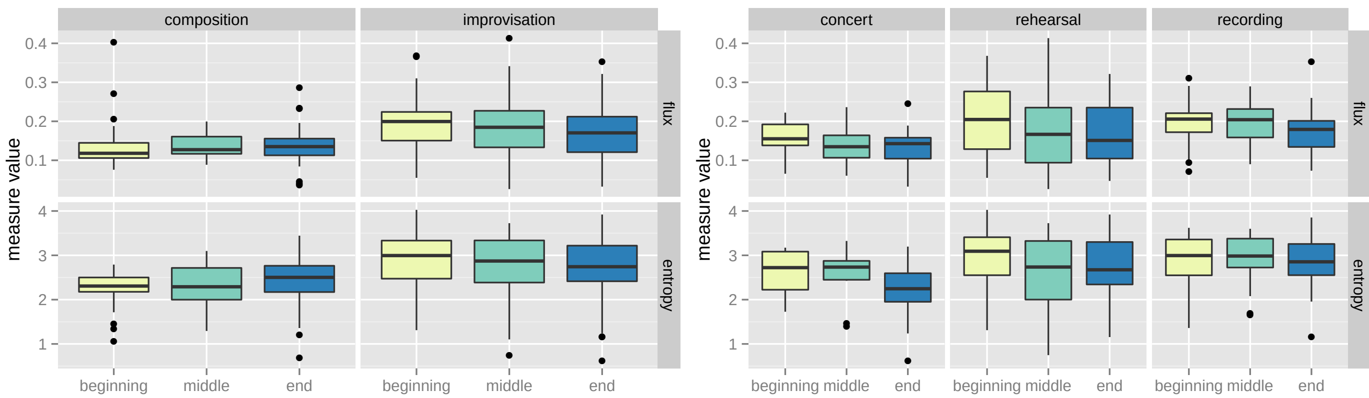
These tests suggest that the different interaction styles that might be present in the various session contexts, types and instrumentations in our corpus appear to be discriminated by values of the flux and entropy of their transition matrices.

Breaking sessions into sections

It appears from our data-mining results that each of flux and entropy can be used to differentiate between session styles and types, but can they be used to understand some of the internal structure of sessions? Since Aristotle, a classical model for understanding temporal art-forms is the three-phase "beginning, middle, end" structure [1]. To investigate how this structure fits our corpus, we divided the gestural-sequences of each session into three equal sections and calculated the transition matrices given by each section.

Figure 4a shows that the flux and entropy of compositions are much more stable across the three sections than are improvisations. This stability of compositions is to be expected since performers use the same gestures in each rehearsal and concert of written scores, however it may be that different kinds of compositions could result in different structural patterns. Both measures have a downward trend through sections in improvisations with Kruskal-Wallis tests showing a significant effect of section on flux ($\chi^2(2) = 5.8, p = 0.05$). This could be explained by performers entering an exploration stage at the start of an improvisation, where they change gesture frequently to experiment with different ideas. At the end of an improvisation, performers may change gesture less frequently while "winding up" the concert performance.

The gestural change in improvisation can be further explored by dividing the sessions by session context. Figure 4b shows this division for improvisations only. In rehearsals and recordings the distributions of flux and entropy over sections has a similar shape, both measures dropping in the ending section of recordings and after the beginning section in rehearsals. In concerts however, flux drops after the beginning while entropy drops substantially in the ending section. This indicates a more constrained performance style in this closing



(a) Sessions divided by performance type.

(b) Improvisation performances divided by session context.

Figure 4: The flux and entropy of different interaction sessions divided into beginning, middle, and end.

section, where performers transition between fewer gestures in the final section of concert performances.

Performance styles over time

While the interaction sessions in our corpus have many different contexts, they are all part of a program of artistic and HCI research that has evolved over time. Figure 5 shows the values of flux and entropy for each session in the corpus. Over the course of this data collection, different ensembles and activities have been emphasised and some of these differences can be seen as clusters of performances having similar fluxes and entropies in the two panels. For instance, the earliest series of rehearsals and concerts, from early 2013 until July 2014, had fluxes that were between approximately 0.05 and 0.17. Just before and after July 2014, fluxes were between 0.1 and 0.26. Finally, between early and mid 2015, fluxes were between 0.08 and 0.32. Within these blocks of performances, distinct clusters can be identified. The higher upper bounds of fluxes close to July 2015 relative to those a year earlier may indicate the development of a more "fluctuating" style of performance over time. In the same period, the bound of entropies remained more consistent indicating that the total space of available gestures continued to be explored in a relatively stable way.

CONCLUSIONS

In this note we estimated transition matrices for a data corpus of 95 collaborative musical performances using touch screens. We have applied two simple matrix measures, flux and entropy, to the data and have shown that both of these measures appear to have discriminatory power. Not only do these measures differentiate between different types of sessions, and the internal structure of these sessions, but they provide qualitative insight into the performances. For improvised performances, our results showed that concerts had the lowest values of both flux and entropy, while recording sessions had the highest. This may be due to the increased pressure on performers to take fewer risks when improvising in front of an audience at a concert and indicates that perhaps rehearsals should emphasise the development of such risk-taking. Our results for the three phases of performance (beginning, middle, end) revealed more variation in improvisations than compositions. As a creative response to this finding, new works might be composed that specifically explore and emphasise different gestural sequences across the three phases.

Figure 5: Distribution of flux and entropy values of sessions through time. Different performance styles are visible as clusters of similar flux regardless of session context.

Our data-mining of interaction-logs from creative interfaces suggests further approaches for future work. Our flux and entropy measures have proven useful in the analysis of our data corpus, but many other transition matrix metrics might be found to be useful for these and different interaction data. Our assumption of equal intervals for the beginning, middle and end of performances could be optimised with a goal to identifying a common, or even optimal, apportionment of time to each phase. A wide dissemination of corpuses such as ours could better support the replication of findings in HCI research [16]. The macro-gestures that these processes reveal could be used as the basis for "transcoding" new works of new media art [5] from these data corpuses.

REFERENCES

1. Aristotle. 350 B.C.E. *Poetics*. The Internet Classics Archive. http://classics.mit.edu/Aristotle/poetics.1.1.html Translated by S. H. Butcher.
2. Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32. DOI: http://dx.doi.org/10.1023/A:1010933404324
3. David England, Jocelyn Spence, Celine Latulipe, Ernest Edmonds, Linda Candy, Thecla Schiphorst, Nick Bryan-Kinns, and Kirk Woolford. 2014. Curating the Digital: Spaces for Art and Interaction. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 21–24. DOI: http://dx.doi.org/10.1145/2559206.2559222
4. Thomas Kannampallil and Steven Haynes. 2007. Using Event Based Markov Model Simulation to Analyze Interactive Human Behavior. In *Americas Conference on Information Systems*. Article 142. http://aisel.aisnet.org/amcis2007/142
5. Lev Manovich. 2002. *The Language of New Media*. MIT Press, Cambridge, MA, USA.
6. Charles Martin, Henry Gardner, and Ben Swift. 2014. Exploring Percussive Gesture on iPads with Ensemble Metatone. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1025–1028. DOI: http://dx.doi.org/10.1145/2556288.2557226
7. Charles Martin, Henry Gardner, and Ben Swift. 2015. Tracking Ensemble Performance on Touch-Screens with Gesture Classification and Transition Matrices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Edgar Berdahl and Jesse Allison (Eds.). Louisiana State University, Baton Rouge, Louisiana, USA, 359–364. http://www.nime.org/proceedings/2015/nime2015_242.pdf
8. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12 (Nov. 2011), 2825–2830. http://dl.acm.org/citation.cfm?id=1953048.2078195
9. Jeff Pressing. 1988. Improvisation: Methods and Models. In *Generative Processes in Music*, J. Sloboda (Ed.). Oxford University Press, Oxford, UK.
10. Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, Ted Selker, and Mike Eisenberg. 2005. *Design principles for tools to support creative thinking*. Technical Report. Institute for Software Research, School of Computer Science, Carnegie Mellon University. http://repository.cmu.edu/isr/816
11. C. E. Shannon. 1948. A Mathematical Theory of Communcation. *The Bells Systems Technical Journal* 27 (October 1948), 379–423.
12. Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. DOI: http://dx.doi.org/10.1145/1323688.1323689
13. Ben Swift, Andrew Sorensen, Michael Martin, and Henry Gardner. 2014. Coding Livecoding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1021–1024. DOI: http://dx.doi.org/10.1145/2556288.2557049
14. Harold Thimbleby. 2004. User Interface Design with Matrix Algebra. *ACM Trans. Comput.-Hum. Interact.* 11, 2 (June 2004), 181–236. DOI: http://dx.doi.org/10.1145/1005361.1005364
15. Harold Thimbleby, Paul Cairns, and Matt Jones. 2001. Usability Analysis with Markov Models. *ACM Trans. Comput.-Hum. Interact.* 8, 2 (June 2001), 99–132. DOI: http://dx.doi.org/10.1145/376929.376941
16. Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI Should Be Replicating and Validating Results More: Discuss. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 463–466. DOI: http://dx.doi.org/10.1145/1979742.1979491