Towards Lifelong Interactive Learning For Open-ended Embodied Narrative Improvisation

Mikhail Jacob
Georgia Institute of Technology
Atlanta, GA 30305, USA
mikhail.jacob@gatech.edu

ACM ISBN 978-1-4503-4403-6/17/06.
http://dx.doi.org/10.1145/3059454.3078699

Abstract
This paper describes a doctoral research plan modeling collaborative embodied narrative improvisation, using lifelong interactive learning to mitigate the knowledge-authoring bottleneck. Research methodology involves building interactive system models of the improvisation process, public installation/exhibition, user experience studies in public/lab settings, and ablation experiments. The article concludes with a research timeline.

Author Keywords
Improvisation; embodied narrative; VR; imitation learning; generalized hypergraph; modeling creativity.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
The creative cognition community has formally studied creativity [8] and attempted to computationally model it [25] in various domains. These studies have made significant contributions to both the study of creative cognition and the computational modeling of creativity. However, the recent popularity of virtual reality (VR) as a storytelling medium has emphasized the lack of computational models of co-creativity in embodied

Embodied Narrative: Narratives constructed using the affordances and constraints of bodies situated within their environment Broadly, embodied narrative includes the interaction modalities of speech and touch, but is used in this article exclusively to mean full-body free-form gestural interaction with an environment.

Knowledge-authoring Bottleneck: The difficulty, cost, and delay in acquisition of expert knowledge followed by its subsequent structuring and storage so as to enable efficient future utilization are referred to as the knowledge-authoring bottleneck [4].

Uncanny Valley Effect: When robots are gradually made to look more humanlike, they progressively appear more realistic until a point at which the subtle remaining blemishes start to make them look creepier progressively [19].

creative domains such as *embodied narrative* improvisation (embodied narrative improv).

VR games and experiences are a fast-growing segment of the digital entertainment [26] as well as training and simulation industries [2, 21]. Naturalistic interaction in VR, such as walking and manipulating the virtual world using the body, supports the open-endedness of improv experiences. Users can physically mime flipping burgers [14] or scaling Mt. Everest [23] to do so in-game. In contrast, non-player characters (NPCs) are limited to small sets of animations due to the *knowledge-authoring bottleneck* [4] to expanding them.

Various techniques exist in the game AI literature to mitigate the knowledge-authoring bottleneck. However, the flexibility and expressivity of interaction required for embodied narrative improv makes them less applicable. NPC movement can be learned/generated from motion capture datasets. However, few datasets exist for a broad set of human action and few (if any) have formal semantics annotating how they affect the environment. Generated movements can also appear unnatural, exhibiting the *uncanny valley effect* [19].

One way to mitigate the knowledge-authoring bottleneck for an open-ended improv system is to learn actions and action selection knowledge from all human users over its lifetime. This would be facilitated by a knowledge representation that could be updated incrementally, integrating new and prior learned knowledge in real-time. Additionally, the user could interactively correct learning errors.

This article describes a doctoral research plan for modeling the collaborative improvisation of open-ended embodied narratives. The research focuses on solving the knowledge-authoring bottleneck involved using lifelong interactive learning from observation and experience. Research methodology involves modeling the improv process computationally, interactive system building, public installation/exhibition, user experience studies in public/lab settings, and ablation experiments. The article concludes with a research timeline.

LuminAI

LuminAI (formerly *Viewpoints AI*) is a dome-based interactive art installation [11] where human participants can interact expressively with virtual characters in order to collaboratively improvise movement-based performances. LuminAI learns what actions it can perform and how to sequence them in response to a collaborator's actions [12]. Learned actions can be flexibly reused in many different contexts. It also learns sequences of actions by learning Markov chains [3] from observation. This models the improv process as a mutual coupling where the characters and users are responding to each other.

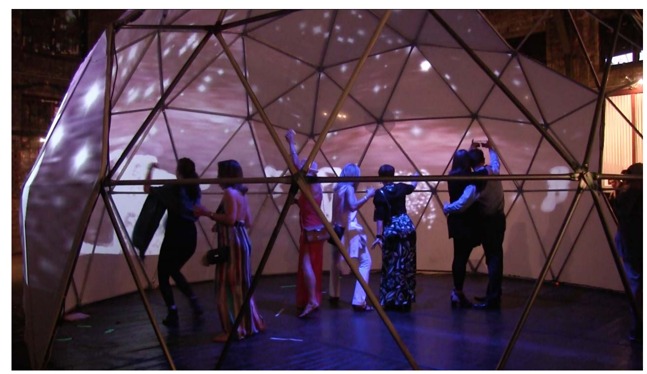

Figure 1: LuminAI dome installation with people improvising open-ended embodied proto-narratives with each other and virtual characters (humanoid figures made of motes of light)

Markov Chain: A stochastic model for representing a sequence of possible events where the probability of each event happening depends only on the state caused by the previous event [3].

Viewpoints: Viewpoints theory analyzes movement and embodied performance along various dimensions: tempo, duration, repetition, kinesthetic response (naturally timed responses to stimuli), shape, gesture, movement topography (patterns of movements over time), spatial relationships, and architecture (the set) [1].

Cognitive Architecture: A software architecture that models the fixed structures of the mind in order to produce intelligent behavior given the right knowledge [15].

Episodic Memory: A type of human memory that remembers autobiographical events [24].

In addition, the installation utilizes a computational formalization of *Viewpoints* movement theory [1] to analyze the improvised scene, organize its experiences, and aesthetically transform learned actions [12]. The system chooses responses using multiple response strategies within a cognitive architecture called Soar [15] that performs the reasoning. These strategies include repetition (to create rapport), transformation along Viewpoints dimensions (to create novelty), retrieval of observed movements from episodic memory [24] that are similar in Viewpoints dimensions (to generate aesthetically appropriate novelty), and using the learned action Markov chains discussed before (to follow expected patterns).

LuminAI has been publically exhibited in informal, invited, and peer-reviewed showings. Iterative design methodology was used, with participant feedback incorporated after each showing. The installation has also been formally evaluated using theoretical analysis of authorial leverage and ablative user study [13] as well as qualitative study during a public exhibition [16].

Given a definition of narrative as a sequence of causally and temporally related actions, the improvised performance in LuminAI constitutes a proto-narrative. This is an initial step along the stated research direction because the actions performed in it are temporally and aesthetically related but not fully causally related. The MImE & GENII projects address this issue.

MImE & GENII

MImE (Movement-based Improvisation Environment) is a VR experience where a person and a virtual character can collaboratively improvise open-ended embodied narratives. The initial research prototype being built enables embodied narrative improv within a larger narrative of the protagonist and their service android crash landing on an alien planet, damaging the android's memory. In order to survive, explore, and escape, the player and their android have to perform collaborative tasks using various objects in the environment with the player teaching the android the necessary skills to help complete the task. A future version with less structure is planned, but out of scope for the current research plan.

GENII (Graphical Embodied Narrative Intelligence & Improvisation) is the intelligent agent for lifelong interactive learning and virtual character control in open-ended embodied narrative experiences such as MImE. In order for the GENII-controlled android in MImE to learn how to interact with the environment, it observes the player's interactions with it, reasons about how the player's physical body movements change the state of objects in the environment and uses that to build an inductive model of the effects of the player's action. It uses this model to perform an action in the environment during collaborative tasks with feedback from the player. GENII agents thus perform interactive imitation learning [20] of action models [27].

The learned action models are stored in a generalized hypergraph [10] knowledge representation. This allows GENII to hierarchically connect the grounded percepts it experiences to abstract concepts that are derived from them. Sequences of human body position data observed by the agent are stored at the lowest layer, segmented at the next layer by the changes they cause to the environment. The movements between each change in the environment are sub-actions. They are clustered together based on similarity of motion and

Interactive Imitation Learning: Learning skills by imitating the movements of a teacher, reconstructing the goals of the example, and using feedback to improve [13].

Action Model: The creation of models that predict causes and effects of actions [20].

Generalized Hypergraph: An extension of hypergraphs and graphs where single edges can connect an arbitrary number of vertices and edges themselves [10].

Metacognition: Regulatory awareness and reasoning that seeks to modulate the agent's own cognition [7].

effect. Labels for prototypes of each sub-action are crowdsourced. They are then cleaned and processed. This is repeated for the action, composed of sub-actions. Given a goal state, this hierarchical action model learned by the agent can be used to act in the world through action planning [5]. It can also be connected to knowledge bases for common sense [22] and linguistic reasoning [18], or used for analogical reasoning [9] and conceptual blending [6] in the future, but that is out of scope for this research.

GENII also has a set of metacognitive [7] models of creativity that evaluate the novelty, value, surprise, and aesthetic description of knowledge (like in [17]) at each layer of the hypergraph. This is used to filter out low quality knowledge from the agent. This includes ignoring bad percepts during interactive learning, forgetting action models during system idle time, and generating only appropriately creative responses from sequences or combinations of learned knowledge.

Research Plan

After initial investigations into embodied narrative improv with the LuminAI project, the author's focus is now on the MImE/GENII project that is ongoing with initial prototypes being completed for the start of summer 2017. At the time of writing, there is one partial user task complete in MImE with associated assets. In GENII, the motion layer learning has been added and evaluated along with a model of motion dissimilarity. By summer, MImE will include 2-3 open-ended user tasks within the crashed spaceship interior. In GENII, this will include initial models of novelty and value; tracking of partial object state; and motion layer segmentation using object state changes. This will allow baseline user evaluations and system evaluations

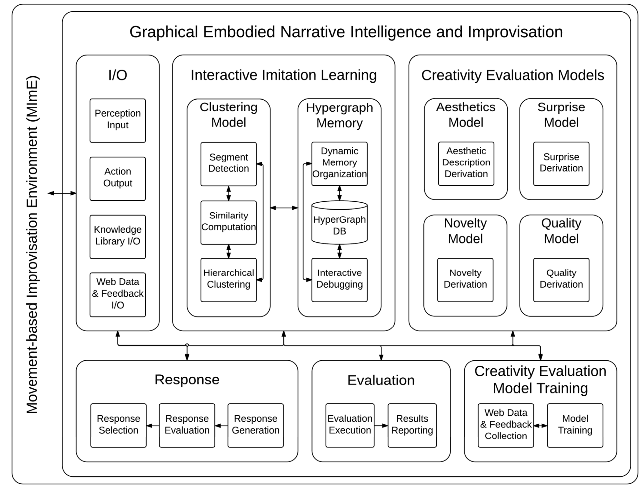

Figure 2: The GENII agent architecture.

in summer 2017. With results and feedback from them, in fall 2017, the remaining MImE scenario will be added (base camp). In GENII, novelty, surprise, value, and aesthetic description models; full object state tracking; and response generation with learned action models will be completed. Spring 2018 will be spent evaluating the system, refining it, and writing the dissertation. The defense will be in summer 2018.

Acknowledgements

I would like to thank my advisor, Dr. Brian Magerko, for his support and the rest of the ADAM Lab at Georgia Tech for their help. This material is based upon work supported by the National Science Foundation under Grant No. 1320520.

References

1. A. Bogart and T. Landau. 2006. The Viewpoints Book: A Practical Guide to Viewpoints and Composition. Theatre Communications Group.

2. Neil Cooke and Robert Stone. 2013. RORSIM: a warship collision avoidance 3D simulation designed to complement existing Junior Warfare Officer training. *Virtual Reality* 17, 3: 169–179.
3. David Cope and Melanie J Mayer. 1996. *Experiments in musical intelligence*. AR editions Madison, WI.
4. Andrew Csinger, Kellogg S Booth, and David Poole. 1994. AI meets authoring: User models for intelligent multimedia. *Artificial Intelligence Review* 8, 5-6: 447–468.
5. Kutluhan Erol, James Hendler, and Dana S Nau. 1994. HTN planning: Complexity and expressivity. *AAAI*, 1123–1128.
6. Gilles Fauconnier and Mark Turner. 2008. The way we think: Conceptual blending and the mind's hidden complexities. Basic Books.
7. John H Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist* 34, 10: 906.
8. Howard Gardner. 2008. Art, mind, and brain: A cognitive approach to creativity. Basic Books.
9. Dedre Gentner and Linsey Smith. 2012. Analogical reasoning. *Encyclopedia of human behavior* 130: 130.
10. Borislav Iordanov. 2010. HyperGraphDB: a generalized graph database. *International Conference on Web-Age Information Management*, Springer, 25–36.
11. Mikhail Jacob, Alexander Zook, and Brian Magerko. 2013. Viewpoints AI: Procedurally Representing and Reasoning about Gestures. *Proceedings of the Digital Games Research Association*.
12. Mikhail Jacob, Gaëtan Coisne, Akshay Gupta, Ivan Sysoev, Gaurav Gav Verma, and Brian Magerko. 2013. Viewpoints AI. *Proceedings of the Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.
13. Mikhail Jacob and Brian Magerko. 2015. Interaction-based Authoring for Scalable Co-creative Agents. *Proceedings of International Conference on Computational Creativity*.
14. Owlchemy Labs. 2016. *Job Simulator: The 2050 Archives*. Owlchemy Labs.
15. Pat Langley, John E Laird, and Seth Rogers. 2009. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research* 10, 2: 141–160.
16. Duri Long, Mikhail Jacob, Nicholas M. Davis, and Brian Magerko. 2017. Designing for Socially Interactive Systems. *Proceedings of the Eleventh ACM Conference on Creativity and Cognition*.
17. Mary Lou Maher. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, Desire Network, 22–28.
18. Palmer Martha, Gildea Dan, and Kingsbury Paul. 2005. The proposition bank: a corpus annotated with semantic roles. *Computational Linguistics Journal* 31, 1.
19. Masahiro Mori, Kari F MacDorman, and Norri Kageki. 2012. *The Uncanny Valley: The Original Essay by Masahiro Mori*. .
20. Manuel Mühlig, Michael Gienger, and Jochen J Steil. 2012. Interactive imitation learning of object movement skills. *Autonomous Robots* 32, 2: 97–114.
21. Neal E Seymour, Anthony G Gallagher, Sanziana A Roman, et al. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery* 236, 4: 458–464.
22. Robert Speer and Catherine Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. *LREC*, 3679–3686.

23. Sólfar Studios. 2016. *Everest VR*. Sólfar Studios.
24. Endel Tulving. 1985. Elements of episodic memory.
25. Tony Veale, Kurt Feyaerts, and Charles Forceville. 2013, Creativity and the Agile Mind: A Multi-disciplinary Study of a Multi-faceted Phenomenon. Walter de Gruyter.
26. BUSINESS WIRE. 2017. Worldwide Spending on Augmented and Virtual Reality Forecast to Reach $13.9 Billion in 2017, According to IDC. .
27. Jie Yang, Yangsheng Xu, and Chiou S Chen. 1997. Human action learning via hidden Markov model. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 27, 1: 34–44.