The Role of Gestures and Movement in Computational, Embodied Storytelling

Philipp Wicke[1,*]

[1]*Centre for Language and Information Processing (CIS), Ludwig-Maximilians-Universität München (LMU), Oettingenstraße 67, Munich, 80538, Germany*

Abstract

With advances of deep learning systems, robust and convincing models for text, image and movement generation are revolutionising many multi-modal forms of generative art. The domain of artificial story-generation has seen this revolution through the development of generative pretrained transformer models. Moreover, storytelling robots can benefit from these deep learning systems not only relying on high-quality textual artefacts, but also on movement and voice synthesis provided by these systems. Yet, the development from symbolic systems to neural systems has burdened the perception of artificial creativity with an opaqueness. Where symbolic systems expressed the intentional use of a conceptual system, neural models bury these conceptual systems in their intractable statistical depth. This research proposes the use of symbolic gestures to augment artificial, robotic storytelling in order to explicate the hidden conceptual structures of artificial storytelling. Building on gesture research and image schemas, we show how the symbolic nature of schematic movement and spatial metaphor can aid the perception of artificial creativity.

Keywords

Computational Creativity, Gestures, Embodiment, Storytelling

1. Introduction

The current advances in Artificial Intelligence (AI) research are largely caused by developments in Machine Learning which enables ever more complex, ever deeper neural network architectures [1]. These architectures produce systems of unprecedented performance in a variety of domains [2, 3, 4]. Most notably, the effectiveness of these systems relies on great amounts of data, large quantities of computing power and neural networks which are optimised to learn the complex correlations hidden in the data. One major breakthrough in language modelling has been the generative pretrained transformer (GPT) architecture [2]. Its attention mechanism and its ability to scale its performance with an increasing number of statistical parameters makes these systems multitask, few-shot learners [3]. This means that these language models are able to generate text, answer questions or translate without being trained specifically on certain tasks. Consequently, large language models have been used to implement a variety

of "creative" language systems [4, 5, 6]. The biggest improvement of these systems is their capability to create long-range dependencies and entity coherence within the created texts whilst forming non-generic plotlines. Yet, advancements of large statistical models come with a cost. Not only are these systems computationally expensive in training and data (making them irreproducible for conventional research labs), they are also - as of now - incapable of explaining their reasoning or underlying conceptual structures. Our research argues that this deficiency questions whether or not these systems can be evaluated in terms of artificial creativity and proposes an embodied approach for robotic storytelling which relies on the use of symbolic gestures in order to recover some power over the analysis of these opaque statistical systems.

2. Understanding Creative Processes

To define the role of embodiment in the perception of human and artificial creativity, this research adopts the theory of embodied cognition [7]. Under this premise, many features of cognition are embodied in that they are critically dependent upon characteristics of the body of an agent [8]. Investigating creativity, for both machines and humans, therefore requires an investigation of those bodily characteristics which can influence (or enable) creative cognition. The field of Computational Creativity (CC) explores human and machine creativity by constructing autonomous systems capable of producing novel and useful artefacts [9]. Hence, the acknowledgment of embodiment has given rise to various research efforts in the field of CC (for an overview, see [10]). With respect to embodiment and artificial creativity, our current research investigates the role of gestures in the domain of robotic storytelling. The next section presents a short overview of related works bridging gesture studies and creative storytelling.

3. Related Works

Computational Storytelling An embodied storyteller requires a story to be told and an instruction of how the story is supposed to be told. The former is the *fabula* ("what is told") and the latter is the *discourse* ("how it is told") [11]. Tasking machines to generate a fabula has been studied since the 1960's [12] and in the following decades, symbolic systems have modelled the intricate deep structures of plot generation to achieve this task. In contrast, generative neural networks entered the field only in the past decade and, whilst abandoning a focus on deep structures, have brought forth high quality textual outputs [2, 3]. Three selected systems will briefly be reviewed to explain our choice for the *Scéalextric* story generator [13] in our gesture studies. The *MEXICA* story generator by Pérez y Pérez [14] has been in development for more than 20 years and models various mechanisms of plot development, e.g. rule systems to describe a story grammar, implementing the Engagement-Reflection model of creative writing [15] and establishing long-range dependencies through emotional links between the characters. These very rich deep structures that incorporate mechanisms informed by human writers provide insights into creative computation, yet the resulting plots are too reflective of the static deep structures and cannot give rise to rich and diverse stories on the surface level of text. In contrast, large language models such as GPT-3 [3] are argued to excel in various language tasks and possess writing skills that can compete with basic human writing skills [16]. Nonetheless,

neural models fail to make underlying creative processes explicit or accessible. Moreover, a lack of symbolic hooks does not allow performers to retrieve or insinuate further instructions in the generated stories. In short, GPT models provide excellent surface forms, but no deep structures.

The *Scéalextric* story generator [13] can be placed conceptually between these systems. It is a knowledge-base system with declarative knowledge structures. The system has an easily extensible, symbolic knowledge base that can provide access points for performers to include mixed modalities such as emotion, movement and gestures. Its deep structure is based on causally connected plot points, which have been derived from Cook's *PLOTTO* book [17] which contains a large set of plausible plot elements linked in a coherent fashion. These basic plot elements could be further augmented with idiomatic renderings retrieved from large language models. Providing easily accessible deep structures and an extensible surface structure, a system such as *Scéalextric* is a viable candidate to catalyse a creative fusion between computational (neural) story generation and coherent embodied performances with robots as actors [18].

Gestures and Movement in Computational Storytelling   So far, we have identified a suitable system to generate the *fabula*, but storytelling requires the plot to be enacted. Guckelsberger et al. [10] review how various studies in CC have used robotic or virtual agents to enact creative performances such as storytelling, improv or dance. Many modalities can be chosen to enact a textual plot, e.g. movement of the body, sound, images, facial expressions or gestures. Among the non-verbal expressions, gestures and body movement have been studied by Rond et al. [19] through *ImprovBot*. Their robot uses simple schematic movements in a narrative context to enhance an improv performance. Likewise, the rapping robot *Shimon* [20] uses gestures as building blocks for musical creation and synchronisation with human collaborators. In robotic storytelling, Ham et al. [21] craft 21 gestures along with 8 gazing behaviours in collaboration with a professional stage actor. These non-verbal movements are assessed in a storytelling context and the authors provide empirical evidence for a positive, synergistic effect of gesture and gaze. Similarly, studies by Sugimoto et al. [22] and Catala et al. [23] highlight the different positive effects of moving actors/robots in different storytelling contexts.

4. Arguments for Robotic Movement

Our research argues that gestures and spatial movement can not only enhance artificial storytelling performances, but that they can and should be used in any robotic performance that can afford to do so. Here, we provide and outline three core arguments derived from our research:

1. **Conceptual structures** can be explicated and support the study of creative performance.
2. **Interpretation** of performance can be controlled in more detail.
3. **Coherence** in gestures and movement can only enhance performances.

Conceptual Structures   Whether improv, dance or storytelling, a performing agent uses its means of verbal and bodily expression to enact concepts. Those concepts must undergo changes and abstractions in order to be translated onto the stage. In line with the performative motto "Show, don't tell", spatial movement and gestures can show to explicate, underline or contradict

what is told. Spatial movement and gestures can be highly schematic and their meaning is grounded in physical experience. Various studies have shown how humans are able to interpret and appreciate the meaning and use of simple spatial movements in robot performances (for drones [24]) [25, 26]. In order to explicate the conceptual structures of an underlying idea (e.g. a plot element), the image-schematic nature of spatial movement and gestures needs to be applied. This relies on the assumption that schemas such as UP, DOWN, NEAR or FAR are fundamental semantic building blocks of human reasoning and concept formation [27]. Image schemas structure spatial movement and evidence suggests they also structure gestures [28, 29]. McNeill [30] describes them as fundamental asset to study human's conceptualising capacities.

The related works in textual story generation show that most symbolic story generators produce rich conceptual structures, but cannot compete with neural language models' high quality surface level texts, which in return miss access to low-level representations. Spatial movement and gestures can insert themselves between generated conceptual structures and surface level text. We have shown this ability in various studies in the context of storytelling [26, 31, 32] and argue here that this can also translate to other performative domains.

Interpretation Spatial movement and gestures allow the performing agent to connect with the human observer's knowledge of physical experience. As an example, the robotic agent relies on the human's experience of gravity, proximity, depth and locomotion. Intuition of these experiences links the performer with the observer. The performer can show actions which would otherwise only be told, for example "She will give him a warm welcome" can be depicted with a simulated hug or open arms for one actor or an actual hug for two performing agents. The same message can be construed using spatial movements or gestures. Alternatively, the actor might repeatedly slam their fist into its open hand. This metaphorical gesture makes use of the FORCE/CONTACT schema, which construes the message to mean a violent welcome. Lastly, two actors can move further apart or closer together in order to indicate whether the message is meant to be threatening or affectionate. In that case, the spatial movement becomes the metaphor PHYSICAL DISTANCE is EMOTIONAL DISTANCE. Further examples of how the interpretation can be construed using movement/gestures are described in the *Scéalextric* storytelling performance in the final section.

Coherence Adding multiple modalities (e.g. gestures, movement) to a performance arguably distributes the audience's attention. Hence, it is important that additional augmentation of the underlying script only enhances the expressivity on stage and does not diminish or dilute it. Moreover, multiple modalities should work in synchrony and coherence in order to support or modulate each other. In fact, gestures depend on and provide context for accompanying speech acts [33]. With these properties, gestures can improve the performance to create meaningful connections between verbal and non-verbal acts. Likewise, if spatial movement is used in accordance with verbal acts, even in metaphorical ways (e.g. PHYSICAL DISTANCE is EMOTIONAL DISTANCE), our studies have shown that the audience appreciates the combination of coherent spatial movement and gestures over one of these movement types alone [26]. Ultimately, movements and gestures allow the actor to tap into the audience's embodied intuitions in expressing conceptual structures that hide behind a mere textual presentation of a plot.
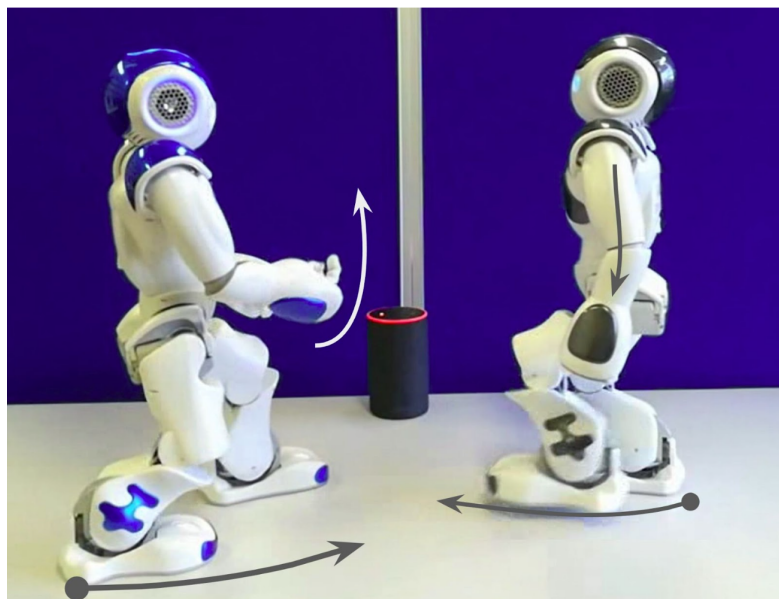
Figure 1: The two robotic actors augment the generated plot with gestures and spatial movement in order to explicate conceptual structures, coherence and facilitate metaphorical interpretation.

5. Evidence from Computational Storytelling Performances

The implementation of a system that produces robotic storytelling performances, which can be scaled to include multiple agents with varying degrees of embodiment and movability, has been presented in our work on Scéalability [34]. This research exemplifies the proposed strategy of utilising spatial movement and gestures in order to express coherence, explicate conceptual structures and provide pathways for metaphorical interpretation for audience and actors.

The underlying story generator combines plot elements in a causally coherent manner and provides a chain of these elements as plot sequences. Each plot element is an action triplet A-action-B with A and B as placeholders for the two characters in each plot. A resulting 3-element plot might be: A-gave-kiss-to-B, B-fell-in-love-with-A, A-married-B. This underlying conceptual structure of the plot is then augmented with various modalities. A and B become characters from the same fictional universe by consulting Veale's NOC knowledge-base [35]. The actions (e.g. gave-kiss) are rendered with idiomatic sentences (e.g. "A gave B a romantic kiss on the lips"), augmented with appropriate movements (e.g. One actor gives the other a kiss) and the emotional valence of the action is registered (e.g. A forms a romantic bond with B). Over the course of the performance semantic insertions are tracked and allow the actors to choose alternative modes of enactment. For example, a character in the story can be repeatedly rebuffed by the other character and might grow increasingly frustrated. This increasing emotional pressure should be reflected in the actors' motion and mode of enactment. In [35], we show how access to the emotional valence allows the performing actors to insert irony or metaphorical expressions in their movement. For example, a deteriorating relationship can be reflected by increasing physical distance between the actors (e.g. "repeatedly slams fist

into hand" accompanying a "greeting" to construe its meaning). A video of the performance is available at bit.ly/2Ud0GYx and a snapshot can be seen in Figure 1, which shows the schematic, spatial movement. In the Figure, schemas within the gestures have been marked with arrows.

6. Conclusion

In this paper, we argued that the use of gestures and spatial movement in embodied performances can connect discrete conceptual structures of plot generation with more abstract means of performance. The use of the moving body enables performing agents to explore alternative interpretations, manifestations of conceptual structures and overall coherence, which is known to enhance their expressivity. Novel approaches in AI are distorting the perception of creative text generation (i.e. high-quality creative text generation becomes increasingly creative), but the true understanding of deeper semantic structures remains hidden within their computational intractability. Ultimately, this research suggests that embodiment (here: spatial movement and gestures) can play a vital role in explaining hidden structures of meaning. While the presented example is built upon a symbolic approach for storytelling, the principles of coherence, interpretation and conceptual structures are applicable to other forms of performances (e.g. dance, improv) and other types of generative systems (e.g. neural language models). The implementation of these suggestions is the study of future work.

Acknowledgments

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.
[2] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (2019) 9.
[3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.
[4] T. Winters, P. Delobelle, Survival of the wittiest: Evolving satire with language models., in: ICCC, 2021, pp. 82–86.
[5] Y. Agafonova, A. Tikhonov, I. P. Yamshchikov, Paranoid transformer: Reading narrative of madness as computational approach to creativity, Future Internet 12 (2020) 182.
[6] A. Calderwood, V. Qiu, K. I. Gero, L. B. Chilton, How novelists use generative language models: An exploratory user study., in: HAI-GEN+ user2agent@ IUI, 2020.
[7] F. J. Varela, E. Thompson, E. Rosch, The embodied mind, revised edition: Cognitive science and human experience, MIT press, 2017.
[8] L. Foglia, R. A. Wilson, Embodied cognition, Wiley Interdisciplinary Reviews: Cognitive Science 4 (2013) 319–325.

[9] T. Veale, F. A. Cardoso, Computational creativity: The philosophy and engineering of autonomously creative systems, Springer, 2019.

[10] C. Guckelsberger, A. Kantosalo, S. Negrete-Yankelevich, T. Takala, et al., Embodiment and computational creativity, in: International Conference on Computational Creativity, 2021.

[11] P. Gervás, Computational approaches to storytelling and creativity, AI Magazine 30 (2009) 49–49.

[12] J. E. Grimes, Linguistic and anthropological projects using the computer, The Use of Computers in Anthropology (1965) 515–516.

[13] T. Veale, Déja vu all over again, in: Proceedings of the International Conference on Computational Creativity, 2017.

[14] R. Pérez y Pérez, MEXICA: a computer model of creativity in writing., Ph.D. thesis, University of Sussex, 1999.

[15] M. Sharples, How we write: Writing as creative design, Routledge, 2002.

[16] K. Elkins, J. Chun, Can gpt-3 pass a writer's turing test?, Journal of Cultural Analytics 5 (2020) 17212.

[17] W. Cook, PLOTTO: the master book of all plots, Tin House Books, 2011.

[18] P. Wicke, Computational storytelling as an embodied robot performance with gesture and spatial metaphor, in: PhD Thesis, University College Dublin. School of Computer Science, 2021.

[19] J. Rond, A. Sanchez, J. Berger, H. Knight, Improv with robots: creativity, inspiration, co-performance, in: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, 2019, pp. 1–8.

[20] R. Savery, L. Zahray, G. Weinberg, Shimon the rapper: A real-time system for human-robot interactive rap battles, arXiv preprint arXiv:2009.09234 (2020).

[21] J. Ham, R. Bokhorst, R. Cuijpers, D. v. d. Pol, J.-J. Cabibihan, Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power, in: International conference on social robotics, Springer, 2011, pp. 71–83.

[22] M. Sugimoto, T. Ito, T. N. Nguyen, S. Inagaki, Gentoro: a system for supporting children's storytelling using handheld projectors and a robot, in: Proceedings of the 8th International Conference on Interaction Design and Children, 2009, pp. 214–217.

[23] A. Catala, M. Theune, H. Gijlers, D. Heylen, Storytelling as a creative activity in the classroom, in: Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition, 2017, pp. 237–242.

[24] A. Bevins, B. A. Duncan, Aerial flight paths for communication: How participants perceive and intend to respond to drone movements, in: Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 2021, pp. 16–23.

[25] H. Nakanishi, Y. Murakami, D. Nogami, H. Ishiguro, Minimum movement matters: impact of robot-mounted cameras on social telepresence, in: Proceedings of the 2008 ACM conference on Computer supported cooperative work, 2008, pp. 303–312.

[26] P. Wicke, T. Veale, The show must go on: On the use of embodiment, space and gesture in computational storytelling, New Generation Computing 38 (2020) 565–592.

[27] G. Lakoff, Women, fire, and dangerous things: What categories reveal about the mind, University of Chicago press, 2008.

[28] A. Cienki, Image schemas and mimetic schemas in cognitive linguistics and gesture studies, Review of Cognitive Linguistics. Published Under the Auspices of the Spanish Cognitive Linguistics Association 11 (2013) 417–432.

[29] I. Mittelberg, Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses, Cognitive semiotics 11 (2018).

[30] D. McNeill, Hand and mind1, Advances in Visual Semiotics (1992) 351.

[31] P. Wicke, T. Veale, Show, don't (just) tell: Embodiment and spatial metaphor in computational story-telling., in: ICCC, 2020, pp. 268–275.

[32] P. Wicke, T. Veale, Walk the line: Digital storytelling as embodied spatial performance, in: 7th Computational Creativity Symposium at AISB, 2020.

[33] S. D. Kelly, D. J. Barr, R. B. Church, K. Lynch, Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory, Journal of memory and Language 40 (1999) 577–592.

[34] P. Wicke, T. Veale, Interview with the robot: Question-guided collaboration in a storytelling system., in: ICCC, 2018, pp. 56–63.

[35] T. Veale, P. Wicke, Metaphor, blending and irony in action: Creative performance as interpretation and emotionally-grounded choice., in: ICCC, 2021, pp. 319–326.