Designing a Robot Which Paints With a Human: Visual Metaphors to Convey Contingency and Artistry*

Martin Cooney and Peter Berck[1]

***Abstract*— Socially assistive robots could contribute to fulfilling an important need for interaction in contexts where human caregivers are scarce–such as art therapy, where peers, or patients and therapists, can make art together. However, current art-making robots typically generate art either by themselves, or as tools under the control of a human artist; how to make art together with a human in a good way has not yet received much attention, possibly because some concepts related to art, such as emotion and creativity, are not yet well understood. The current work reports on our use of a collaborative prototyping approach to explore this concept of a robot which can paint together with people. The result is a proposed design, based on an idea of using visual metaphors to convey contingency and artistry. Our aim is that the identified considerations will help support next steps, toward supporting positive experiences for people through art-making with a robot.**

I. INTRODUCTION

The current paper explores the concept of a robot that can paint together with people: In connection with a shortage of human caregivers, there is a need to provide mental and cognitive support, e.g., to persons with depression, traumas, dementia, or autism. One form of supportive therapy involves making art such as paintings, which has been reported to improve vital signs, self-image, rest, and stress tolerance. In such situations, there is a growing feeling that robots can be helpful for humans, not as judges, but as partners–but, current art robots typically make art either alone by themselves, or as a tool under the control of a human artist. Thus, it would be useful to explore how to design a semi-autonomous robot that can paint together with a person–we refer the reader to our previous work for a detailed discussion of the above points with various supporting references [1].

II. DESIGN

To gain insight into how to design a robot to paint together with a human, we employed a collaborative prototyping approach [2]–starting from a basic scenario from our previous work, extended by collaboration with two professional artists and two engineering students, and followed by some brainstorming and implementation: Our basic scenario was dyadic (one human and one robot), visual (non-verbal), held over a single session (e.g., ten minutes), and allowed free choice of what to paint. For a safe platform with cameras and sufficient reach to paint, we used the Baxter robot.

*The authors received funding for this work from the Swedish Knowledge Foundation (Sidus AIR no. 20140220 and CAISR 2010/0271).

[1]The authors are with the School of Information Technology, Halmstad University, 301 18 Halmstad, Sweden martin.daniel.cooney@gmail.com

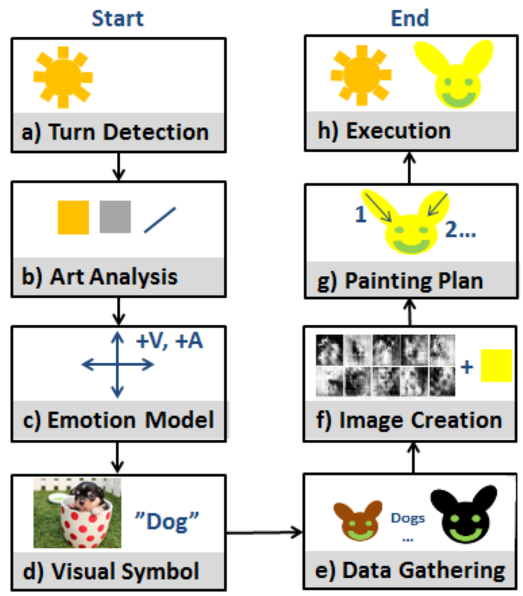

Fig. 1. Basic concept: use of visual metaphors to convey contingency and artistry in an interactive art robot

Although we started with a separate canvas for the human and robot, an enjoyable experience occurred when our artists tried painting on a shared canvas, face-to-face with the robot, which was remotely controlled by a student to paint in a contingent way.[1] Next we explored how the robot could automatically accompany a person, starting with the simplest strategy of mimicking, as part of an ongoing student project, but we observed that although nice as a demo, direct mimicking is predictable and thus not so artistically engaging [3]. Thus, we believe it is important that a robot's painting seems appropriate, by appearing to be both *contingent* with respect to a person's art and *artistic* in line with people's expectations for art.

To balance contingency and artistry, we propose a design based on our concept of a *visual metaphor*. Merriam-Webster defines a metaphor as "a figure of speech in which a word or phrase literally denoting one kind of object or idea is used in place of another to suggest a likeness or analogy between them."[2] We believe that a similar phenomenon could

[1]In doing so, the robot often painted over the human's work, which for us resulted in complexity and interest, but might not be desirable for art therapy–an unpleasant impression could result from ignoring a human, who should be the center of attention–so we suggest that in a basic scenario a robot and human leave some space for one another to paint.

[2]https://www.merriam-webster.com/dictionary/metaphor

underlie some perceptions in visual art: For example, calm water viewed from the side forms a horizontal line, which can evoke a calm impression. A dynamic, toppling object can appear to be diagonally oriented, where diagonal lines can evoke a dynamic impression. Thus, water and toppling objects can act as visual metaphors. We believe that such a concept can be used by a robot to create art which is contingent to, but different from, a person's art. For example, if a person paints something calm, like a green meadow, the robot could paint something else which is calm, like a lake, beside the meadow. Our design is summarized in Fig. 1 and in a video[3], and detailed below:

(a) Turn detection. A robot could paint at the same time as a human, or wait for a turn. For a simplified prototype, to avoid interfering while the human paints, we suggest turn-based painting with haptic control as a first step.[4]

(b) Art analysis. To analyze a person's painting, computer vision can be used to detect colors, lines/shapes, composition, and depicted objects. For our prototype we explored converting an image of a person's art to HSV space to color-pick six basic hues and calculate average intensity, and used Hough Transform to detect lines, although we noted some sensitivity to parameters (e.g., in detecting rounded shapes).

(c) Emotion Model. To infer emotional meaning from visual analysis, some heuristics can be used in conjunction with a categorical or dimensional emotion model. For our prototype, a linear combination was used to calculate valence and arousal, where Ståhl's model was used to calculate a contribution of pixels of each hue to both valence and arousal, intensity affected valence (light being positive), and the prevalence of diagonal lines influenced arousal.

(d) Visual Symbol. To move from abstract emotions to a more concrete target, large sentiment lexicons such as Affective Norms for English Words (ANEW), SentiWordNet, or WordNet-Affect, could be used in conjunction with concreteness ratings. For our prototype, a simplified capability was implemented to search affective image databases; for example, searching for happy, relaxed, sad, and angry emotions turned up images involving skydiving, nature, a cemetery, and mutilation (IAPS), and dogs, flowers, gray yarn, and injuries (OASIS), respectively.[5]

(e) Data Gathering. To generate a new image, current algorithms can require many training examples; search engines or image datasets, such as Google Image Search and ImageNet, offer access to billions and millions of images respectively. For our prototype, we encountered some limitations with numbers of images and time requirements in web-scraping, and space requirements for image datasets; thus we selected the Stanford Dogs Dataset with 20500 images, to investigate one example of a happy image of a dog from (d) above.

(f) Image Creation. To try to make "unique" art, Deep Convolutional Generative Adversarial Networks (DC-GANs) such as StyleGAN can be used. For our prototype, we experimented with a DC-GAN on Google Colaboratory to develop small MNIST-sized grayscale compositions over 200 epochs for rapid training, in order to acquire results in hours rather than days or months. The results could then be abstracted, tweaked using the heuristics from (c) such as hue, and distance maximized from previous art to appear creative. A challenge is how to select images which are good and not uncanny, without human intervention; as well training time currently prohibits online interaction.

(g-h) Painting Plan and Execution. To break down generated images into strokes to paint, existing approaches for painting robots can be followed which are not specific to the current interactive scenario (e.g., [4]); one modification could be to predict time requirements and only conduct some strokes to avoid making a person wait long. For our prototype, we investigated having our robot also convey emotions through a facial expression in its display, its motions (curvature and velocity), and voice, as it painted.[6]

III. DISCUSSION

Thus, we explored the design space for a robot capable of painting with a person, proposing a design to convey contingency and artistry based on visual metaphors. Much work remains to be done in this area–our design, although limited to the designated scenario, could be modified for use with other contexts such as music[7] or writing; moreover, our next steps will involve refining[8] and integrating modules, as well as conducting user studies to verify that the system appears contingent and artistic, and can support positive user experiences.

ACKNOWLEDGMENT

We thank Dan Koon, Peter Wahlberg, Erik Westberg, Nils Lindhqvist and others who kindly contributed thoughts!

REFERENCES

[1] M. Cooney, and M. Menezes, Design for an Art Therapy Robot: An Explorative Review of the Theoretical Foundations for Engaging in Emotional and Creative Painting with a Robot, Multimodal Technologies Interact, vol. 2, no. 3, pp. 52, 2018.

[2] D. St-Onge, C. Gosselin, and N. Reeves, A comparison of collaborative approaches in robotic artworks, In Proceedings of ICRA 2011.

[3] C. Kroos, and D. C. Herath, Evoking agency: attention model and behavior control in a robotic art installation, Leonardo, vol. 45, no. 5, pp. 401-407, 2012.

[4] T. Lindemeier, J. Metzner, L. Pollak, and O. Deussen, Hardware-Based Non-Photorealistic Rendering Using a Painting Robot, In Computer graphics forum, vol. 34, no. 2, pp. 311-323, May 2015.

[3]https://youtu.be/yeNXg2UKA-8

[4]Each modality that could be used to detect turns has potential demerits: Users with cognitive impairments could have trouble haptically controlling a robot with buttons. Audio control, e.g., by detecting keywords with CMU PocketSphinx or Google Speech, requires sufficient volume, clear enunciation, and quiet motors. Visual control using, e.g., background subtraction through OpenCV and timers, is vulnerable to illumination changes and a robot's ego-motion. Thus, multimodal control could afford more robustness.

[5]We note however that there can be bias and variance in the categories and images represented–e.g., not all dog images will seem happy.

[6]For example, https://www.youtube.com/watch?v=01AC2JuefC8

[7]For example, if a human plays a sad melody, a robot could look up sad metaphors such as "cemetery" and related musical data (e.g., motifs in pieces incorporating this word in titles, lyrics, or associated explanations) to generate a contingent but artistic accompaniment.

[8]Refinement will include incorporating "monosemic" metaphor selection considering variance in emotional meaning, faster image generation enabling online interactions, and a model for the degree to which a metaphor can be abstracted before its meaning is lost.