Influence of Environmental Context on Recognition Rates of Stylized Walking Sequences

Madison Heimerdinger(✉) and Amy LaViers(✉)

Mechanical Science and Engineering Department,
University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
{heimerd2,alaviers}@illinois.edu

Abstract. Affective movement will likely be an important component of robotic interaction as more and more robots move into human-facing scenarios where humans are (consciously or unconsciously) constantly monitoring the motion profile of counterparts in order to make judgments about the state of their counterpart. Many current studies in affective movement recognition and generation seek to either increase a machine's ability to correctly identify human affect or to identify and create components of robotic movement that enhance human perception. However, very few of these studies investigate the influence of environmental context on a machine's ability to correctly identity human affect or a human's ability to correctly identify the affective intent of a robot. This paper presents the results of a user study that investigated how human perception of stylized walking sequences (created in [1]) varied based on the environment where they were portrayed. The results show that environment context can impact a person's ability to correctly perceive the intended style of a movement.

Keywords: Style · Affect · Gait · Context · Human-robot interaction

1 Introduction

In the coming years, automated systems will interact with human users in increasingly unstructured tasks. These systems, which may manifest as animated avatars or moving machines, need to tap into human movement patterns and conventions in order to operate successfully. For example, an in-home, care-giving robot must move with the soft, gentle, and confident patterns that are attributed to the movement and vocal inflection of doctors and nurses in care-giving, "bedside" scenarios. In this paper, we note the role of *context* in the relationship between internal state and external movement patterns in order to understand how to prescribe it in simulated systems. Specifically, we consider *style* (movement parameters or features) and *affect* (human response to movement patterns with respect to context) to be important, separate components of this process.

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 272–282, 2017.
https://doi.org/10.1007/978-3-319-70022-9_27

Internal human affect is externally manifested through multiple mediums including gestures, postures, and facial expressions [2,3]. Stylized body movements are nonverbal ways in which affective information about the state of the mover can be recognized and understood by other agents in their environment – as researchers in computer vision note in [4]. The human perception of affect generated from stylized movements is heavily dependent on the situation being experienced or imagined. If no situational or environmental context is provided, there is a significant chance that the observer will misinterpret the intended affect of the movement [5]. Understanding the relationship between movement, environment, and affect could enhance the communication abilities that are currently limited in human-robot interactions [5].

Stylistic movement recognition and generation studies can be broadly categorized as either machine-driven perception studies or human-driven perception studies. The machine-driven perception studies often utilize machine learning algorithms to analyze and categorize the style of a movement sequence using information previously obtained from a training set [6,7]. Generally, these studies focus on classifying a small set of well defined action inputs into a finite set of qualitative descriptions using kinematic motion parameters. Human-driven perception studies generally consist of stylized movement generation that is validated by means of a user study [1,8]. Some of these studies use animation tools such as Disney's Twelve Principles of Animation as guidelines for creating believable stylized robotic movement [9,10]. Other studies have evaluated whether human viewers noticed differences in motion patterns, termed styles, as in [11], where surveys on whether viewers noticed distinct patterns, without corresponding affective labeling, in a robotic performance. Again, these studies often focus on a small set of well defined actions that embody a finite list of styles (e.g. gait as happy, sad, feminine, masculine, energetic, or tired in [1].

This paper presents the results of a user study that examined the relationships between style selection/recognition rates and environmental context. Section 2 presents a more detailed view of related work that this paper extends. Section 3 includes a detailed outline of the study on affect recognition in various contexts performed for this paper. Section 5 includes a discussion of the results. Section 6 summarizes the findings of the user study and outlines future work stemming from the results of this study.

2 Previous Work in Style and Affect

The input and output walking sequences produced by Etemad and Arya in [1] were used directly to generate the stimuli for the user study presented in this paper. This section will summarize the process for creating and evaluating the stylized walking sequences that was presented in that work. The next section will describe how we modified their stimuli to examine contextual effects.

Etemad and Arya use a linear motion model to describe the various movement parameters involved in a 54 degree-of-freedom (DOF) skeleton walking. The linear model Etemad and Arya derive describes a complete motion sequence as

a linear combination of primary features and secondary features. The primary features are associated with the base action of the movement sequence, such as throwing or walking. The secondary features are the components associated with the style in which the action is performed.

The primary features used in [1] were based on neutral walking sequences that were captured using a motion capture system. The secondary features were created by 11 experienced animators. The animators were tasked with creating secondary features using radial basis functions to transform a neutral walking sequence into various stylized walking sequences. The animators were allowed to used up to three radial basis functions (RBFs) per DOF in the model to create the secondary feature sets for the secondary themes. The feature sets created by all of the animators for a particular secondary theme were averaged together to create one secondary feature set. The secondary themes that the animators created were labeled as *energetic*, *feminine*, *happy*, *masculine*, *sad* and *tired*.

The styles of the secondary themes were evaluated by 16 new participants. These participants were asked to rate the appropriateness of the given style (used interchangeably with "affect" in this work) label for various walking sequences. In this experimental design, participants were thus primed with the styles of the walking sequences that were intended by the animators. Please refer to [1] for further details on the generation and labeling of the dataset leveraged here.

The affective domain of modern psychology encompasses the experience of feeling or emotion. Affective states can be decomposed into three dimensions: valence, motivational intensity, and arousal. Valence is a subjective positive-to-negative evaluation of a stimuli; motivational intensity is a subjective evaluation of the urge to approach a stimuli; and arousal can be subjective or measured as activation of the sympathetic nervous system [12]. Russell developed the circumplex model of affect that is used to map affective states to a two dimensional chart using ratings of valence and arousal [13]. This model has been used in an array of applications, including affective classification of blog posts [14] and images with affective ratings of valence and arousal [15,16].

Databases containing images with affective ratings of valence and arousal have been created to assist in emotional research studies. The International Affective Picture System (IAPS) is the most widely reference affective visual stimulus database. IAPS images are subject to copyright restrictions to help ensure the integrity of the ratings. The Geneva Affective PicturE Database (GAPED) and the Open Affective Standardized Image Set (OASIS) are open-access databases that contain affective ratings for 730 and 900 color images, respectively.

3 Methodology

A user study was developed to include environmental context as a variable in style recognition of animated walking sequences. The two goals of the user study were to gain insight into the effects of environmental contexts on the style recognition rates of the walking sequences and to determine if environmental contexts effect style selection rates of the walking sequences.

The walking sequences that were used in this study were produced by Etemad and Arya in [1]. Video clips of the neutral input sequence and the stylized output sequences were provided as additional media attachments to the online version of [1]. The stylized walking figures were extracted from the original walking sequences and superimposed onto seven different backgrounds (49 new animation videos) in an attempt to visually simulate an environmental context. The environmental context for the original videos was labeled as *white background* (7 animation videos). The specifics of the video creation process and the structure of the user study questionnaire will be outlined later in this section.

The style labels (excluding *neutral*) were categorized into three different categories: *discrete emotions*, *moods*, and *physical states*. *Discrete emotions* and *moods* both fall under the veil of affective terms that are used in research and are formally defined in [17]. It should be noted that some style labels may fit under more than one category. However, for the purpose of this study, each style was only assigned to one category. The categorization of the style labels is shown in Table 1. The affective and stylistic dimensions of the styles within a category were assumed to be bipolar. For this reason, Table 1 includes a column that describes the styles within a category as either *positive* or *negative*.

Table 1. Categorized list of of style labels

	Discrete emotions	*Moods*	*Physical states*
Positive	Happy	Energetic	Feminine
Negative	Sad	Tired	Masculine

The images in OASIS were used as inspiration when choosing the environmental contexts for this study. For this study, the valence and arousal values of applicable images in OASIS were mapped to a circumplex model. This mapping was then used to select four images that were located in regions of the circumplex model that were associated with the style labels in the *discrete emotions* and *moods* categories. The mapping was not considered when selecting the images that were associated with the style labels of the *physical states* or *neutral* because these labels do not describe affective states. A mapping of the OASIS videos that were used for reference is shown in Fig. 1.

Videos of environmental contexts that contained similar characteristics to the chosen OASIS images were found on YouTube and used to create the stylized walking animations that were used in this user study. The actual valence and arousal ratings of the videos used in the study are unknown. However, the similarities between the videos and OASIS images were initially used to estimate the general locations of each video in the circumplex model. The OASIS images that were used as reference and snapshots of the environmental contexts used in our study are shown in Fig. 2.

All video editing was performed using iMovie. As mentioned in Sect. 3, the baseline walking sequences utilized in this study were a product of [1]. The

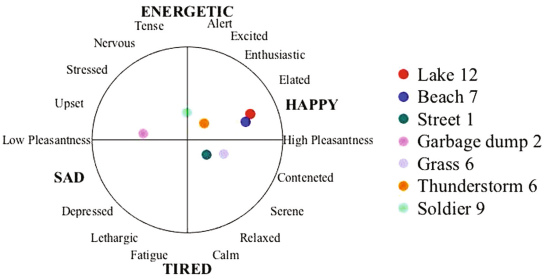

Fig. 1. A mapping of the affective ratings of the referenced OASIS images [15] to the circumplex model of affect.

Fig. 2. The OASIS images used, from right to left, are *Lake 12*, *Beach 7*, *Street 1*, *Garbage Dump 2*, *Grass 6*, *Thunderstorm 6*, and *Soldier 9* [15]. The second row contains snapshots of the videos used to create environmental contexts for walking animations in the user study. The labels of the snapshots, from right to left, are *aspens*, *beach*, *city street*, *garbage*, *grass field*, *lightning storm*, and *war zone*. The OASIS images and the environmental contexts that are in the same column have similar features and thus may have a similar valence and arousal rating. (Color figure online)

original video clips for each stylized walking sequence were approximately 2 s long and consisted of a blue stick figure model completing roughly one full gait cycle on a white background. Snapshots from the final animation videos are provided in Fig. 2.

The first step was to increase the duration of each walking sequence. To accomplish this, each walking sequence was looped roughly 5 times to create stylized walking animations that lasted approximately 10 s. Additionally, color correction was performed on the walking sequences to change the color of the stick figure model to black. The stick figure model color was altered so that it would appear to be the same color in all of the environmental contexts. The walking animations with the *white background* context required no further editing after this step. The stylized walking animations were then created for each environmental context by overlaying the stick figure motion onto the various environmental contexts and then using iMovie's background removal feature.

In some cases, the timing or orientation of the environmental context was edited to match the gait cycle timings of the walking sequences. The timings of the walking sequences were never altered to ensure that the duration of each gait cycle remained constant. It should be noted that the size and placement of the stick figure models was not constant across animations containing different

environmental contexts. However, the sizes and locations of the stick figure models in each animation containing the same environmental context, regardless of affective label, were roughly the same.

The user study questionnaire consisted of 64 questions. The first eight questions were standard demographics questions. The remaining 56 questions were linked to the 56 stylized walking animation videos described in Sect. 3. There was only one question per page and the order that these pages were presented to each participant was randomized. Each question asked the participants to choose the style that they felt best described the stylized walking animation that appeared at the top of the page. The participants were allowed to rewatch each stylized walking animation as many times as they deemed necessary. However, participants were not allowed to return to questions that they had previously answered and submitted. The style choices the participants could choose from were *neutral*, *energetic*, *feminine*, *happy*, *masculine*, *sad* and *tired*. The order that the style choices were displayed was randomized for each question.

A total of twenty participants were recruited at the University of Illinois Urbana-Champaign. The study included twelve female and eight male participants ranging between the ages of 19 and 40. The average age of all participants was 21.85 with a standard deviation of 4.65 years. Although English was not the native language of every participants, all participants did considered themselves fluent in English. Participants were compensated for their time.

4 Results

This section reviews the results of this initial user study. The green values in Table 2 correspond to the selection rates that have 95% confidence interval lower-bound estimates larger than the upper-bound of the 95% confidence interval statistically expected for random selection rates and suggest that the styles intended by the animators were highly perceptible in the simulated environments. The 95% confidence interval estimations were calculated using Eq. 1. In Eq. 1, the variable $\hat{p}$ is the sample statistic that measures the percentage of the sample that meet the criteria being investigated, N is the sample size, and 1.96 is the multiplication factor associated with 95% confidence.

$$\hat{p} \pm 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \tag{1}$$

The correct style recognition rates for each walking animation are presented in Table 2. The "correct" style for each animation was determined by the label that was assigned to the associated walking sequence in [1]. If participants were to randomly select the style choices, the expected correct selection rate would be 1/7, or 14.3%. In Table 2, $\hat{p}$ was defined as the portion of a style selection that was correct for a single walking animation video and the sample-size N was 20.

The values in Fig. 3 represent the percent of the time each style was selected in an environmental context. If participants were to randomly select the style choices, the expected selection rate would be 1/7, or 14.3%. The sample statistic

Table 2. Correct style recognition rates of walking animations

Context	Original walking sequence styles						
	Energetic	*Feminine*	*Happy*	*Masculine*	*Neutral*	*Sad*	*Tired*
Aspen	25.0%	30.0%	10.0%	50.0%	50.0%	30.0%	35.0%
Beach	20.0%	30.0%	35.0%	35.0%	50.0%	30.0%	10.0%
City street	15.0%	35.0%	5.0%	45.0%	30.0%	35.0%	20.0%
Garbage	25.0%	25.0%	15.0%	45.0%	40.0%	40.0%	40.0%
Grass field	15.0%	30.0%	20.0%	35.0%	25.0%	35.0%	30.0%
Lightning	25.0%	35.0%	10.0%	50.0%	35.0%	60.0%	15.0%
War zone	35.0%	35.0%	10.0%	60.0%	40.0%	35.0%	15.0%
White	30.0%	45.0%	15.0%	70.0%	25.0%	10.0%	20.0%

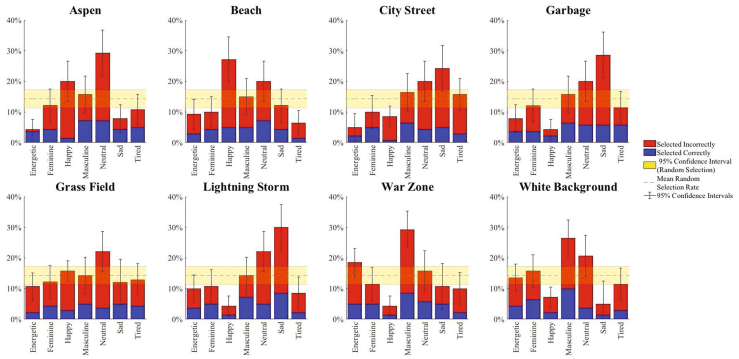

Fig. 3. These bar graphs show the percent of the time each style was selected in a given environmental context, the number of times each style was selected correctly and incorrectly for the environmental contexts, and the corresponding 95% confidence interval estimations.

for Fig. 3 was equal to the percent of the time a style was selected in the 7 walking animations that included the same environmental context and the sample-size N was 140.

The values in Fig. 4 represent the percent of the time each style was selected for a walking sequence. If participants were to randomly select the style choices, the expected selection rate would be 1/7 or 14.3%. The sample statistic was defined as the overall style selection rate (correct or incorrect) for 8 animations that contained the same walking sequence and the sample-size N was 160.

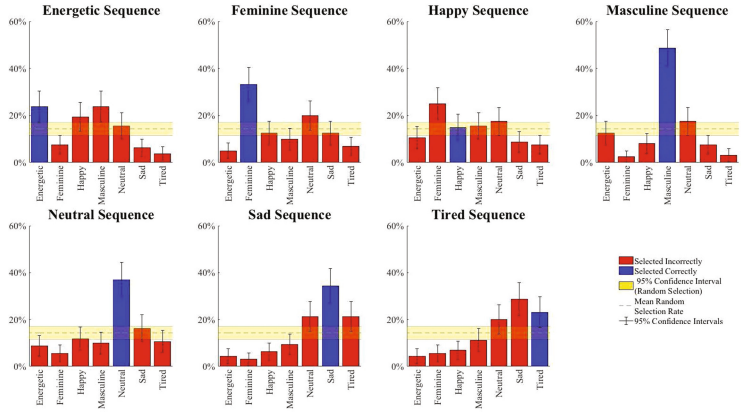

Fig. 4. Bar graphs indicating the percentage of the time each style was selected for the stylized walking animations, regardless of environmental contexts.

5 Discussion

In [1], participants were asked to use a forced-choice scale to rate the magnitude of the style that was displayed in a walking sequence. The type of question asked in the previous user study primed the participants by telling them the intended style of the walking sequence. The questions that were used in this study were multiple choice and did not prime the participants with labels describing either the environmental contexts or the walking sequences. For these reasons, the results of this study expanded on and did not overlap with the results of the previous user study. The goals of this study were to investigate the influence of environmental context on style selection rates of the walking animations in [1] in the absence of this priming. The environments utilized in this study were chosen to have features similar to images from OASIS [15] so that affective ratings of the environments could be estimated. In future studies, it may be beneficial to include environments that are typically associated with human-robot interactions.

The results shown in Table 2 show that for the sample population, 50 out of the 56 walking animation videos had selection rates that were larger than what would be expected for random selection. These results suggest that for our participant pool, the style intent of the animators in [1] remained perceptible in the majority of the simulated environmental contexts. However, only 3 of those videos had 95% confidence interval estimates strictly larger than random selection expectations. Additionally, the 95% confidence interval estimates for this data varied as much as $\pm 21.9\%$ from the sample results.

The results in Fig. 3 indicate that the perceived style of a walking animation can be influenced based on the environment where it is viewed. The environmental contexts *aspen*, *garbage*, *lightning storm*, and *war zone* all had results that suggest a positive association with one style label and a negative association

with one style label. The *garbage* and the *lightning storm* contexts both had a positive association with the *sad* selection label and a negative association with the *happy* selection label. Since *happy* and *sad* were defined as bipolar quantities in Sect. 3, these results suggest that the *garbage* and *lightning storm* contexts augmented the perception of negative *discrete emotions* in the stylized walking sequences.

The results in Fig. 4 suggest that on average, the style intent of the animators in [1] remained evident for the *feminine*, *masculine*, *neutral*, and *sad* walking sequences. It is worth noting that the three walking sequences with the highest style ratings in [1] were also *feminine*, *masculine*, and *sad*. This relationship suggests that on average, style intent remains recognizable across various environmental contexts when the style of a walking sequence is rated highly in the absence of environmental context.

An interesting observation that should be noted is that the *energetic*, *masculine*, and *tired* walking sequences all had style selection rates for the associated bipolar styles that suggest negative associations. These results suggest that on average, the environmental contexts did not cause the perceived styles to be inverted for the *energetic*, *masculine*, and *tired* walking sequences. The *tired* walking sequence was the only sequence to have a positive association with a style label (*sad*) that did not match the original style intended by the animators. This positive association is likely correlated to the large number of similarities between the *sad* and *tired* walking sequences that were presented in [1].

6 Conclusion

This paper expands on the conclusions made in [1] by evaluating style recognition/selection rates of the walking sequences in the absence of priming and in the presence of environmental contexts. The user study presented in this paper consisted of a set of 56 stylized walking animations that featured seven stylized walking sequences and eight environmental contexts. The results in this paper also provide data indicating that environmental context can manipulate the perceived style of a walking animation. These results highlight the importance of the role of environmental context in providing critical information that is used by humans when trying to assess the affect of stylized movement.

We are conducting a parallel study where we probe to understand whether viewers envision their own environmental contexts to make the primed style context fit the movement they are seeing. For example, a jaunty march through a park could be labeled “happy” while that same march through a war zone might be labeled “aggressive” or even “afraid”.

Understanding the relationship between environment, stylized movement, and perceived affect is critical for the creation of efficient nonverbal communication channels between humans and robots. An increase in affective nonverbal communication that is initiated by robots and understood by humans may increase the rate at which robots are integrated and accepted into home environments. In order to create effective nonverbal communication, we must

first determine how stylized robotic movements generate affective responses in humans across environmental contexts.

Acknowledgments. This work was supported by the Mechanical Science and Engineering Department and NSF grants #1701295 and #1528036.

References

1. Etemad, S.A., Arya, A.: Expert-driven perceptual features for modeling style and affect in human motion. IEEE Trans. Hum. Mach. Syst. **46**(4), 534–545 (2016)
2. Kleinsmith, A., Bianchi-Berthouze, N.: Affective body expression perception and recognition: a survey. IEEE Trans. Affect. Comput. **4**(1), 15–33 (2013)
3. Russell, J.A., Fehr, B.: Relativity in the perception of emotion in facial expressions. J. Exp. Psychol. Gen. **116**(3), 223 (1987)
4. Zacharatos, H., Gatzoulis, C., Chrysanthou, Y.L.: Automatic emotion recognition based on body movement analysis: a survey. IEEE Comput. Graph. Appl. **34**(6), 35–45 (2014)
5. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. **31**(1), 39–58 (2009)
6. Bernhardt, D., Robinson, P.: Detecting affect from non-stylised body motions. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 59–70. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74889-2_6
7. Bernhardt, D., Robinson, P.: Detecting emotions from connected action sequences. In: Badioze Zaman, H., Robinson, P., Petrou, M., Olivier, P., Schröder, H., Shih, T.K. (eds.) IVIC 2009. LNCS, vol. 5857, pp. 1–11. Springer, Heidelberg (2009). doi:10.1007/978-3-642-05036-7_1
8. Etemad, S.A., Arya, A.: Modeling and transformation of 3D human motion. In: GRAPP, pp. 307–315 (2010)
9. Ribeiro, T., Paiva, A.: The illusion of robotic life: principles and practices of animation for robots. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 383–390. IEEE (2012)
10. Van Breemen, A.: Bringing robots to life: applying principles of animation to robots. In: Proceedings of Shapping Human-Robot Interaction Workshop Held at CHI 2004, pp. 143–144 (2004)
11. LaViers, A., Egerstedt, M.: Controls and Art: Inquiries at the Intersection of the Subjective and the Objective. Springer, Cham (2014). doi:10.1007/978-3-319-03904-6_1
12. Harmon-Jones, E., Gable, P.A., Price, T.F.: Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. Curr. Dir. Psychol. Sci. **22**(4), 301–307 (2013)
13. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
14. Paltoglou, G., Thelwall, M.: Seeing stars of valence and arousal in blog posts. IEEE Trans. Affect. Comput. **4**(1), 116–123 (2013)
15. Kurdi, B., Lozano, S., Banaji, M.R.: Introducing the open affective standardized image set (OASIS). Behav. Res. Methods **49**(2), 457–470 (2017)

16. Dan-Glauser, E.S., Scherer, K.R.: The geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behav. Res. Methods **43**(2), 468 (2011)
17. Barsade, S.G., Gibson, D.E.: Why does affect matter in organizations? Acad. Manag. Perspect. **21**(1), 36–59 (2007)