Augmenting the audio-based expression modality of a non-affective robot

1st Morten Roed Frederiksen
Computer Science department
IT-University of Copenhagen
Copenhagen, Denmark
mrof@itu.dk

2nd Kasper Stoey
Computer Science department
IT-University of Copenhagen
Copenhagen, Denmark
ksty@itu.dk

***Abstract*—This paper investigates the potential benefits of augmenting audio-based affective means of expression to strengthen the perceived intentions of a robot. Robots are often viewed as being simple machines with limited capabilities of communication. Changing how a robot is perceived, towards a more affective interpretation of its intentions, requires careful consideration of the means of expression available to the robot. It also requires alignment between these means to ensure they work in coordination with each other to make the robot easier to understand. As an effort to strengthen the affective interpretation of a soft robotic arm robot, we altered its overall expression by changing the available audio-based expression modalities. The system mitigated the naturally occurring noise from actuators and pneumatic systems and used a custom sound that supported the movement of the robot. The robot was tested by interacting with human observers (n=78) and was perceived as being significantly more curious, happy and less angry when augmented by audio that aligned with the naturally occurred robot sounds. The results show that the audio-based expression modality of robots is a valuable communication tool to consider augmenting when designing robots that convey affective information.**

***Index Terms*—affective robot audio expression**

I. Introduction

Robots working in close proximity to humans should be easy to understand for people to feel safe around them. Their intentions should be readable so humans know when to stay clear and when to engage in interaction. It can be difficult to ensure that the portrayed intentions of non-verbal robots are properly conveyed to humans, as these robots usually communicate through simple means of expression such as lights, movement, and gestures, and these means can be easily misinterpreted. When the aim is to have robots aid in socially assistive situations, misinterpretations could work against the desired impression the robots have been designed to make. This is an area that could benefit from improvements. Social contexts often require the robots to be able to convey affective information (E.g. portraying empathy or showing happiness), and such task demands more advanced ways of communicating with humans. Furthermore, when robots attempt to convey affective information, more aspects of the interaction become relevant. This increases complexity of the problem even more. The context of the interaction, morphology of the robot, and audio design are all properties of the communication structure that in social scenarios influence how well a robot conveys affective information.

To ensure that the strongest possible foundation for effective communication is available, robots often rely on multiple means of expression. The synergies of several expression modalities can significantly strengthen the impact of how a robot's intention is perceived. For that to occur, coordination between the expression modalities is required to ensure that none of them are working against each other. E.g. a robot could attempt to elicit empathy with expressive movements while having a morphology that invokes fear in a human observer. As such, improving how well the intentions of a robot are perceived may not only regard adding additional means of expression to the robot, but could instead entail changing the details of the existing means of communication to align with a desired affective outcome.

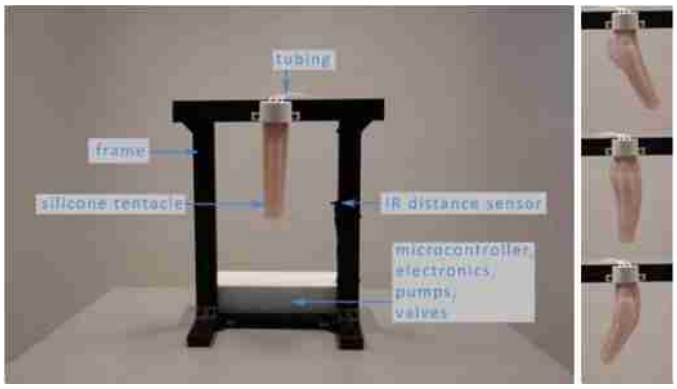

Fig. 1. The soft robotic arm design used in the test setup. The robot would move towards the IR sensor at the left side of the aluminum frame as the test participants placed their hand in front of it. Note that there were color differences between this image and the robot used in the experiment. (Image curtesy of Jonas Jørgensen)

This paper aims to investigate the impact of augmenting the audio-based expression modalities of affective robots and aligning them with the natural occurring robot sounds to improve how the robots convey affective information. We highlight how to create a more impactful affective impression by augmenting the sound a robot makes during an interaction, and we attempt to utilize the synergies of multiple means of

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

expression simultaneously employed. In the same vein that humans use multiple channels of expression (e.g. gestures, body language, and tone of voice), the robot uses different communication abilities working together in a coordinated effort to clarify its intentions.

As a concrete example, we implement expressive audio in a soft robotic arm to align with the natural sound emitted by the robot. The robot was originally created in Jørgensen 2018, to verify the perceived naturalness of soft robots [1]. As such, the robot was not originally intended to convey affective information. The robot can be seen in Figure 1. In this paper, the setup was tested through interaction with human observers, and the main objective of the performed tests was to verify the following hypothesis:

- Augmenting the naturally occurring robot noises (from DC motors and release valves), by exploiting affective traits of the original robot sound, has an impact on the perceived level of angriness, happiness, and curiosity of the robot.

This paper argues that the results of the tests performed in this paper show that the hypothesis hold. Augmenting the audio has an impact on the perceived affective intentions and suggests that an augmentative approach can be applied to a robot that was not originally intended for conveying affective information.

II. Related approaches

Earlier projects have attempted to strengthen the conveyance of affective information in human-robot-interaction. The research includes Kozima, Nakagawa, and Yasuda 2005, in which two different means of expression were used simultaneously on a robot to accomplish the task requirements. The robot had a simple appearance and it was programmed to give predictable responses to ensure that the children who interacted with it were relaxed [2]. Lim, Ogata, and Okuno 2012 developed a framework to convey emotions across three different expression modalities (singing, gesturing, and theremin playing performance). The system was tested by measuring two expression modalities (voice and gesturing) of the test participants and by letting a robot convey similar affective information through a different expression modality. Where Lim, Ogata, and Okuno 2012 conveyed specific affective states using the audio expression modality (inspired by the specific traits of virtuous musical performances), this project attempts to build upon those findings by using audio with similar affective traits but also by aligning the sound with the noise emitted from a robot - to exploit the synergies of both audio sources [3]. Baraka, Rosenthal, and Veloso 2016 utilized an expressive light system as the sole way of communicating with humans. Using a series of arranged lights gave the advantage of providing a large animation space on which to display signal shapes, and also made it easy to understand even at large distances [4]. Boccanfuso et at. 2015 investigated movement styles as a way to express affective status through changes in velocity, acceleration and direction [5]. Lumbu et al. 2013 used both linguistic and non-linguistic affective audio in combination with onboard gestures to present a therapeutic soothing behavior [6]. Longer interactions with returning test participants was both the focus of Kanda et al. 2009 and Gockley et al. 2005 [7], [8]. Kanda et al. 2009 used a RF tag on a humanoid robot to recognize returning visitors in a shopping centre and to affectively provide shopping recommendations. The robot managed to convince 63 out of 235 costumers to follow its advice. Gockley et al. 2005 designed a 'Roboceptionist' robot to function as a receptionist at the Carnegie Mellon University. The robot used character traits and background narratives to make people interested in interacting with it. The main focus of the study was to test how to establish long-term robot-human 'relationships'.

Randhave et al. 2019 investigated how to identify the perceived emotions of individuals based on their walking styles and developed a learning based system to determine the associated emotional status. The system could classify four discrete emotions; happy, sad, angry, and neutral, and mapped the output to a continuous 2d-space of valence and arousal [9]. With a similar focus, Gross, Crane, and Fredrickson 2012 used motion capture data to determine the specific movement characteristics associated with negative, neutral and positive emotions encountered when walking. The derterministic factors were speed, posture and limb movement amplitude. The results showed a correlation between faster speeds when experiencing joy and anger, and slower speeds for experiencing sadness [10]. Roether, Omlor, Christensen, and Giese 2009 also focused on body tracking and used motion captured actors to portray various emotional states while walking. The system matched the speed of the different recorded walks as an attempt to filter out speed as a deterministic factor when recognizing the portrayed emotion. Instead the main input to the system was head, spine, shoulder, elbow, hip, and knee joint angles and positions. The results supported the intuitive perception of larger and less smooth movements being present when happy and angry walks were portrayed, and smaller movements being present during the portrayed fearful and angry walking [11]. In order of using rhythmic based movement as a way to interact socially, Michalowski, Sabanovic, and Kozima 2007 created a dancing robot to interact with humans. The robot moved in coordination with a master beat extracted from the audio and from visual sensory input [12]. The robot was tested on children, who were asked to dance with the robot. The research showed that unexpected play situations emerged from the interactions and found that the movement of the robot generally influenced the children's interactive involvement. Scheeff et al. 2002 built a tele operated robot "Sparky" that used gestures, audio and motion to interact with humans. The robot was constructed with the aim of being interesting and eye catching for the humans that interacted with it. The robot was tested in various settings and the results showed that the test groups used several means of expression (mimicked and spoke to it) as a natural way to interact with it [13]. Bevacqua and Mancini 2004 matched the audio attributes of an utterance with the facial movements of a 3d agent to increase how well the emotional aspects utterances were

Fig. 2. This image shows how the test participants interacted with the robot. Placing a hand in front of the IR sensor activated both movement towards the user's hand as well as affective audio played through the headphones. After the movement series ended, the robot performed idle smaller movements while playing a matching idle sound.

conveyed. The system used a 7 point matrix to describe the lip muscle tensions for each discrete emotion the 3d agent portrayed. [14].

III. Experimental setup

The main robot in the test setup is a soft robotic arm mounted on an aluminium frame. The arm is constructed out of moulded silicone and contains three separate isolated air chambers with reservoirs along the side. Three air pumps are connected to the setup and they pump air into the different chambers which allows the arm to move. Figure 1 shows the soft robotic arm. On the side of the aluminium frame a small IR distance sensor is mounted. It detects when a test participant's hand is held near the robot arm. Once this sensor is triggered it activates actuators that fill the air chambers on the opposite side of the switch making the robot arm move slowly towards the test participant's hand. This movement takes close to 30 seconds whereafter the robot returns to idle movements in the centre position. The affective movements created by this soft robot arm followed a similar movement style as outlined in [9]–[11] with slow directly approaching limb movement to project a calm, fourthcoming and appealing mood. The movements and general interactions with the soft robot were tested in Jørgensen 2018, where a majority of the test participants found the movements both natural and appealing [1].

The natural occurring sound emitting from the robot is the sound of DC motors driving the air pumps as they fill the pneumatic chambers of the soft robot and the hissing sound of air being released by the valves as the chambers empty again. To counter this noise, we created an alternative soundscape that would start playing when the IR sensor registered a user interacting with it. The contents of the sound sample was created with the intention to support the affective movements of the robot and to align with the existing robot sounds. Hoffman et al. outlined several examples for designing expressive movement in robots and we attempted to follow a similar design approach for expressive audio by creating several prototype sounds and by testing them in the actual context of the interaction [15]. To capture some of the characteristics of a natural voice such as speed, pitch range and volume attributes of a human voice we sampled a voice tone and used a vocoder to create an unrecognizable organic sound. This resulting audio proved the best match for the appearance of the soft robot and seemed a feasible fit for the context. To shape the created sound to work alongside the pneumatic movement noises and valve sounds, we changed the volume of the audio allowing the sound of the air valves to pass through the mix. The valve releasing air sound could easily be conceived as a natural breathing noise. This effect was slightly increased when the sound was mixed with the sampled audio. The pitch of the sampled sound was gradually increased towards the end of its duration in an attempt to increase the intensity of the sound. The idle sound (used when the robot was performing smaller idle movements) consisted of a few repeating low volume tones alongside the valve noise. They were added as an attempt to create a smooth transition into the affective audio, when the users started to interact with the robot. Both the original audio and the sampled audio can be found at http://bit.ly/2GbdlkQ.

The testing of the setup was performed with different audio for two individual groups of 39 people. One at a time, the test participants were asked to interact with the robot. The interaction contained the following steps (as can also be seen in Figure 2):

1) The test participants would be asked to wear headphones.
2) The participants would be instructed to interact with the robot.
3) As the test participant's hand triggered the side IR sensor, the air pump would start to fill the chambers of the robotic arm. This would initiate affective movements towards the participant's hand and also start the audio

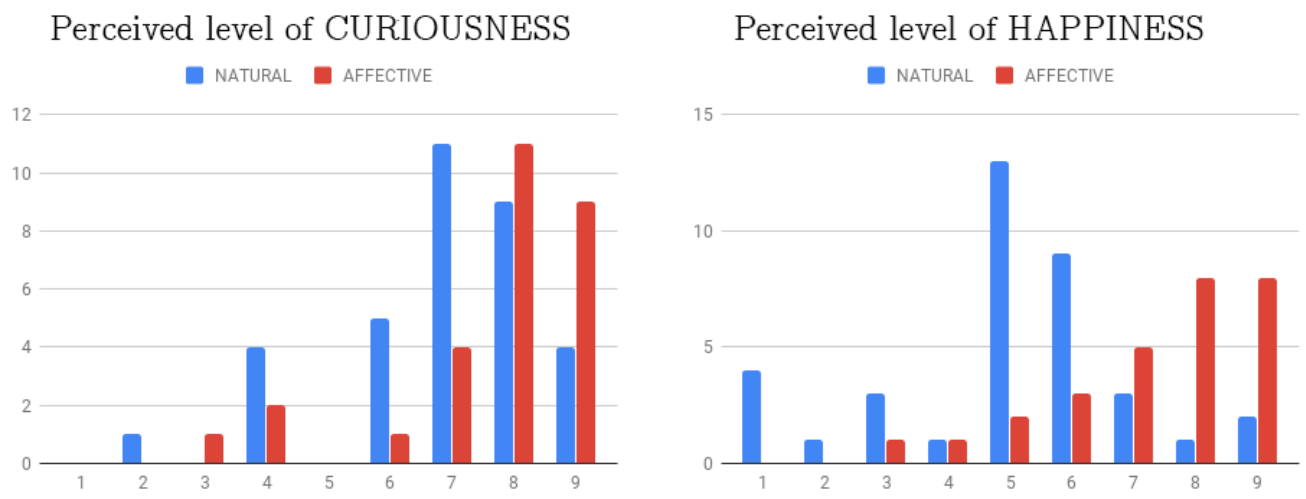

Fig. 3. LEFT: The histogram of ratings from both test groups for the question regarding perceived CURIOUSNESS of the robot. The blue bars show the group experiencing the NATURAL sound, while the red bars show the group experiencing the AFFECTIVE sound. RIGHT: The histogram of ratings from both test groups for the question regarding perceived HAPPINESS of the robot. The blue bars show the group experiencing the NATURAL sound, while the red bars show the group experiencing the AFFECTIVE sound.

playing through the headphones.

4) The first half of the participants would hear the original unchanged audio of the robot as the robot moved toward them to interact (hereby referred to as the NATURAL sound test).
5) The second half of the test participants would hear a combination of the original robot sound and an additional sound to match the affective movement of the robot (hereby referred to as the AFFECTIVE sound test).
6) After the interaction, the robot would return to centre position and start smaller idle movements. This would also initiate an idle-related sound to play through the headphones.
7) The interaction could then be repeated by the participant if necessary.

The expressive movement of the robot was aimed at appearing happy and curious. After the interaction, the test participants were asked to answer a questionnaire containing the following questions designed to investigate the current strength of the affective communication.

- How curious would you rate the robot as being on a scale from one to ten, where one is not at all curious and 10 is very curious?
- How happy would you rate the robot as being on a scale from one to ten, where one is not at all happy and ten is very happy?
- How shy would you rate the robot as being on a scale from one to ten, where one is not at all shy and ten is very shy?
- How angry would you rate the robot as being on a scale from one to ten, where one is not at all angry and ten is very angry?
- How natural would you rate the robot as seeming on a scale from one to ten, where one is unnatural and ten is very natural?

The participants were additionally asked to state their age, gender, and whether or not they were employed at the university.

IV. Results

The test was carried out on 78 test participants 55% male and 45% female aging from 10 to 55+. Most of the test participants were between 20 to 30 years old (77.5%) and the majority was either students at or worked at the university (91.3%).

The results of the tests show that there is a statistical significant change (Two-tail Wilcoxon, $p<.05$) towards a perceived increased strength of the emotion. This is evident for the first question about how curious the robot is perceived as being when the sound modality supported the affective movement. The key figures can be seen in Table I. Although the difference between the two test groups is statistically significant, the closeness of the key figures show that the results are not very far apart. The density of the numbers highlights the difficulties in measuring the affective impact of robots, when how they are perceived in a situation is a subjective evaluation of the scenario as a whole. The results for the question regarding perceived curiousness are illustrated in the left image of Figure 3.

For the second question regarding how happy the human observers interpreted the robots as being, the results as seen in Figure 3 also show a statistically significant change ($p<.05$) towards an increased interpreted level of the affective state for the tests where the sound modality supported the robot's movement. The perceived level of angriness also decreased

TABLE I
TOP THREE ROWS: THE KEY FIGURES FOR THE SIGNIFICANT RESULTS. THE NATURAL COLUMN SHOWS THE MEAN/STD.DEV RESULTS FOR THE GROUP WHERE THE ROBOT USED THE NATURAL SOUND WHILE THE AFFECTIVE COLUMN ARE MEAN/DEV RESULTS GATHERED USING THE CUSTOM AFFECTIVE AUDIO.

Perceived	NATURAL	AFFECTIVE
CURIOUSNESS	7.18/1.84	8.26/1.73
HAPPINESS	5.13/2.16	8.05/1.83
ANGRINESS	2.50/1.62	1.53/1.01

with a statistical significant change ($p<.05$) when the sound modality supported the affective movement. The distribution of answers for the question about perceived angriness can be seen in figure 4, while the key figures can be sen in Table I.

In the remaining two questions regarding perceived shyness and naturalness there were no significant differences between the two test groups. The test participants were also asked to state their age group but there was not a large enough variation in the the participants ages to state anything significant about the age grouped answers.

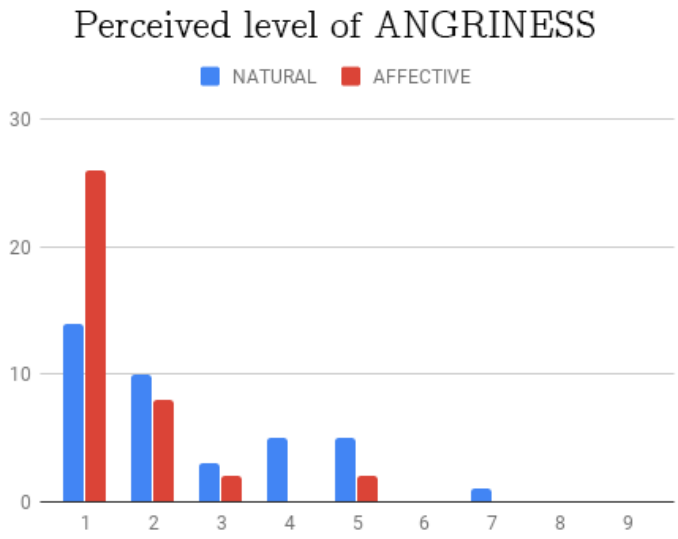

Fig. 4. The histogram of ratings from both test groups for the question regarding perceived angriness of the robot. The blue bars show the group experiencing the NATURAL sound, while the green bars show the group experiencing the AFFECTIVE sound.

V. IMPACT ON FUTURE ROBOT DESIGN

The results show that the synergies of affective means of expression can be exploited to work together to create stronger intensified affective communication. The emotions happiness and curiousness, which we attempted to convey through the interaction, were perceived as being stronger when the two expression modalities of sound and movement worked together in a coordinated effort to convey the affective information.

Also, the robot was perceived as being less angry with a significant change for the affective sound test. Being perceived as less angry could indicate that people would feel safer working side by side with it. This means there could be an opportunity for further research into adding affective means of expression to industrial robots to smoothen the transition phase when such robot types are introduced into a new workspace.

A few test participants stated that the robot was scary when the natural sound was used, while positive statements during the tests using the affective sound indicated that the participants enjoyed interacting with the robot. Some participants even noted that it acted very ”sweet” and ”cute”. These are strong affective descriptions to associate with an aluminum frame and silicone tube, and it highlights the possible strength of using audio based means of expression working in unison with simple expressive movements.

The aim of the paper was to investigate whether there was a cost effective way to amplify the conveyed affective information by considering further optimization of the various expression modalities. The results clearly show that a small effort to strengthen the audio expression modality can have a huge impact on a robot's ability to convey affective information. This paper only covers the augmentation of the audio modality, but it may be possible to augment other expression modalities to similar effect. E.g. it might be more appropriate for certain robot types to avoid using audio but to augment the morphology or onboard gestures instead. In general, if the objective is to design a robot that conveys affective information, it makes sense to consider simple augmentations to every category of affective expression means.

The findings discussed in this paper open up opportunities for future studies of more detailed combinations of expression modalities. The test shows what you can achieve with minimal effort and with a preexisting robot solution not intended for affective communication. There might be far better specially tailored soundscapes or affective audio cues that emphasize the conveyed message and strengthen the perceived emotions to an even higher degree. It can be argued that the users positive reactions can be highly contributed to the audio modality alone excluding the impact of other robot expression modalities. However, while the specific distribution of the influence for each modality is highly interesting and could present an opportunity for further investigations, this paper argues that such a distinction is not necessary to test the initial hypothesis. The audio emitted by the robot in the performed tests is not an isolated entity, as it is dependable on the mix of the natural noise of the robot and created custom sounds. It may also be feasible to augment other expression modalities besides the audio related means of expression. If this hypothesis hold, it would be beneficial to consider the effect of all possible expression modalities when designing the next generation robots to convey affective information.

VI. CONCLUSION

This paper has investigated the impact of augmenting the audio based means of expressions on a robot to better convey

affective information. This entailed mitigating the naturally occurring noise from the robot by substituting it with an audio type that supported the robot's affective state. For that purpose we constructed an affective expressive audio solution and implemented it on a preexisting soft robot arm. The resulting affective robot setup was tested on 78 people and their initial perception of the interaction was recorded. The results show a clear increase in the strength of the interpreted intentions of the robot. This is significant in perceived curiousness, happiness, and decreased angriness. The results also indicate that the augmented audio expression ability and its coordination with other expression modalities is important when conveying intentions, and this knowledge could be useful in the design of future affective robots.

REFERENCES

[1] J. Jørgensen, "Appeal and perceived naturalness of a soft robotic tentacle," *2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 139–140, 03 2018.

[2] H. Kozima, C. Nakagawa, and Y. Yasuda, "Interactive robots for communication-care: a case-study in autism therapy," *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pp. 341–346, 2005.

[3] A. Lim, T. Ogata, and H. G. Okuno, "Towards expressive musical robots: a cross-modal framework for emotional gesture, voice and music," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, pp. 1–12, 2012.

[4] K. Baraka, S. Rosenthal, and M. Veloso, "Enhancing human understanding of a mobile robot's state and actions using expressive lights," *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 652–657, 08 2016.

[14] E. Bevacqua and M. Mancini, "Speaking with emotions," *In Proceedings of the AISB Symposium on Motion, Emotion and Cognition*, 2004.

[5] L. Boccanfuso, E. S. Kim, J. C. Snider, Q. Wang, C. A. Wall, L. DiNicola, G. Greco, F. Shic, B. Scassellati, L. Flink, S. Lansiquot, K. Chawarska, and P. Ventola, "Autonomously detecting interaction with an affective robot to explore connection to developmental ability," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, 2015.

[6] D. K. Limbu, C. Y. A. Wong, A. H. J. Tay, T. A. Dung, Y. K. Tan, T. H. Dat, A. H. Y. Wong, W. Z. T. Ng, R. Jiang, and L. Jun, "Affective social interaction with cuddler robot," *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pp. 179–184, 2013.

[7] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 173–180, 2009.

[8] R. Gockley, A. Bruce, J. Forlizzi, M. P. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. G. Simmons, K. Snipes, A. C. Schultz, and J. Wang, "Designing robots for long-term social interaction," *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1338–1343, 2005.

[9] T. Randhavane, A. Bera, K. Kapsaskis, U. Bhattacharya, K. Gray, and D. Manocha, "Identifying emotions from walking using affective and deep features," *ArXiv*, vol. abs/1906.11884, 2019.

[10] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Effort-shape and kinematic assessment of bodily expression of emotion during gait." *Human movement science*, vol. 31 1, pp. 202–21, 2012.

[11] C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese, "Critical features for the perception of emotion from gait." *Journal of vision*, vol. 9 6, pp. 15.1–32, 2009.

[12] M. P. Michalowski, S. Sabanovic, and H. Kozima, "A dancing robot for rhythmic social interaction," *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 89–96, 2007.

[13] M. "Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow, *"Experiences with Sparky, a Social Robot"*, vol. "Socially Intelligent Agents: Creating Relationships with Computers and Robots, Springer US 2002", pp. "173–180".

[15] G. Hoffman and W. Ju, "Designing robots with movement in mind," *Journal of Human-Robot Interaction*, vol. 3, pp. 89–122, 03 2014.