Communicating Assertiveness in Robotic Storytellers

Raul Paradeda[1,2], Maria José Ferreira[1,3], Carlos Martinho[1], and Ana Paiva[1]

[1] INESC-ID and Instituto Superior Técnico, University of Lisbon - Portugal
[2] State University of Rio Grande do Norte, Natal - Brazil
[3] Madeira Interactive Technologies Institute, Madeira - Portugal
{raul.paradeda, maria.jose.ferreira}@tecnico.ulisboa.pt,
{carlos.martinho, ana.paiva}@gaips.inesc-id.pt

Abstract. Social robots have been used to perform the role of storytellers in areas like education and pediatric rehabilitation. With the use of this technology, it is possible to setup different voices, simulate emotional states and even personalities for the same robot. However, finding the best setting that might define a trait for a storyteller robot, is not an easy task. What elements should be manipulated? Should it have a personality? If yes, which one? In this work, we try to answer these questions by studying several setups that will allow us to create an assertive social robot to act as a storyteller. We evaluate the assertiveness impression by manipulating three robot characteristics: posture, pitch, and speech rate. A within-subject study was conducted with 37 participants watching eight videos in which a social robot tells a short story. In each video, the robot presents a different setup, and the participant reports the level of assertiveness of that robot. We found a significant difference between the setups of pitch and posture as well as an acceptable assertive robot's configuration using a combination of those three characteristics.

Keywords: Storytelling · Social Robots · Assertiveness.

1 Introduction

Emotions and voice changes are well-known characteristics that storytellers apply while telling stories to give a pleasant experience to the listeners. Changes in speech can be used to create the character's personality and emotional state [15], and appropriate voice manipulations can even give the illusion of more than one character, and depending on the narrative, create an exciting and engaging atmosphere. We all know that a monotonous voice in a storyteller can lead to audience disengagement whereas speech modulations can boost engagement, motivation, pleasure and immersion on the plot. A storyteller establishes the mood of the story flow. Further, through his/her expression and posture, a storyteller can convey enjoyment [22], happiness, hope, sadness, and many other feelings. A storyteller's facial expressions and posture, can enrich and colour the ambience of each scene. For example, in a joyful scene, a storyteller can express happiness through the voice, expressions, etc [10].

As the field of social robotics develops, robots are now sufficiently developed to act as storytellers. Storyteller robots have been employed in learning environments [8], pediatric rehabilitation [21], persuasion [13] and games [5]. Yet, to build a robot (or agent) that acts as a storyteller, its voice features, combined with its expressions are crucial as they can help mitigate the uncanny valley problem [12] and generate engaging experiences.

In this paper we will consider that a robotic storyteller should exhibit a posture and voice in a way that is believable and simultaneously engaging for an audience. To achieve one needs to simulate specific human like characteristics, in particular personalities and traits. For example, by exhibiting personality traits, such as extroverted, introvert, assertive, among others, gives a consistency to the robots' behaviour that will make it more believable and engaging. In particular, through the capability of expressing emotions and beliefs in direct, honest and appropriate ways, we can convey an assertive personality type [2,11,14].

In this study we show how to use verbal and non-verbal cues in a social robot that convey assertiveness in order to, in future studies, influence a user interacting with a storyteller robot [19]. Persuasion strategies to implement into the storyteller robot is the assertive personality and thus boost the effects of persuasion and engagement in a story told by a social robot. To do this we parameterised a set of features in a robots storyteller and created videos of the robot telling a story. We evaluated the videos and the results have shown that there is a significant difference in the perception of assertiveness when features like voice and posture are changed.

2 Related Work

Different features can characterise voice, researchers have considered that pitch, timbre, volume and speech rate are the most relevant ones. Pitch is related to the rate of vibration of the vocal cords. As the number of vibrations per seconds increases, the pitch increases as well, and the voice would sound higher[16].

In the work of Page and Balloun [18], the participants perceived the speaker as most aggressive and lacking in self-assurance when speaking in a high voice volume. They heard audio recorded voices in either a low, moderate or high voice volume. In another work [7], the authors found evidence that subjects are recognised as assertive when they spoke louder. The conclusions were taken from the feedback of participants that measured the assertiveness from tapes where subjects answer trivia questions. Furthermore, the perception of emotions in dialogues can be affected by the speaking rate. In Devillers and Vasilescu [4], from the analysis of recorded audios, the participants judged that a faster speech rate is characteristic of irritation and satisfaction. The voice manipulations reported provide us indications that it might be possible to create different voice perceptions regarding the speaker's personality and emotional state.

Voice manipulation has been investigated in different ways, and most researchers resort to specific software to perform this manipulation. For example, in [9] and [24] PRAAT software was the selection of researchers to analyse voice features. In our work, we try to examine if using just the available resources of a standard computer it can achieve reliable results in a more natural/human voice. To do so, we focused our attention on a Text-to-Speech (TTS) from Windows and then examine which Speech Synthesis Markup Language (SSML) features could manipulate prosody.

From the human-robot interaction in a narrative scenario having the robot acting as the storyteller, it is possible to produce some benefits[17]. For example, in the work of Plaisant et al. [21], the authors developed a robot to act in a storytelling scenario with children in rehabilitation. The children could teach the robot to act out emotions like sad, happy and excited. In that study, the emotional expressions performed by the robot regarding the story flow aid to improve children interest as well as support the rehabilitation experience.

Striepe and Lugrin [26], compared the effects of emotions and non-verbal behaviours using an emotional social robot storyteller. The authors reported that the emotional robot could transport the participants equally well as the traditional audiobook. Besides, they advised that the use of social robots performing the role of storytellers needs to be linked with the display of emotions and non-verbal behaviours.

In general, storytelling is all about sharing stories, and the way we choose to tell them makes them more or less remarkable. In spite of the fact that people anthropomorphise technology [23] would a robot storyteller makes us have the same type of experience as we have with a human storyteller? Given this state of the art, in this paper we explore the assumption that through the manipulation of different features in a robot it is possible to create different perceptions regarding the storyteller's personality traits (e.g. assertiveness) and emotional state. As such, we test this assumption by creating different robotic storytellers, varying such features and comparing how they are perceived by users.

3 Research Methods

This paper presents a quantitative study conducted to find evidence that may provide a better robot configuration to convey assertiveness as it acts as a storyteller. To address it we recorded a social robot telling a short story with different setups. Then, participants filled in a questionnaire stating how much assertive s/he considered the robot in each configuration. The surveys and videos were run online via Google Forms. The study tested the following hypotheses:

- H1: Specific vocal manipulations can influence the perception of assertiveness in a social storyteller robot.
- H2: A fast speech rate of a storyteller robot can increase a person's understanding of assertiveness on the robot.
- H3: The posture of a social storyteller robot can induce a person's perception of assertiveness of the robot.
- H4: The use of vocal manipulations combined with posture boosts the perception of assertiveness in a social storyteller robot.

3.1 The Narrative

A short story named *"The Wise Rabbit"*[4] was narrated by the robot for one minute and ten seconds. In some parts of the story, depending on the narrative flow, the robot exhibited facial expressions, such as Joy and Anger.

[4] The Wise Rabbit. Born 1980, M, from Islamabad, Pakistan. `https://www.storystar.com/story/11499/talha/fiction/drama-interest-2`

As mentioned, the robot was programmed to tell the story by the manipulation of pitch and speech rate, since they are some aspects of voice that can indicate personality [1].

3.2 Materials

The EMYS robot was chosen to be the storyteller mainly because of its capability of exhibiting recognisable emotions through facial expressions (Fig 1 (a to g)). In each setting, the manipulated parameters in the robot's voice were pitch (with values x-low, default and x-high) and speech rate (values set as medium and +20%)[5]. The rate values were defined based on preliminary tests, where the rate faster than +20% speech understanding could be affected. Also, the robot's posture settings were changed to: (1) neutral, (2) pride and (3) shame. The neutral posture exhibits the robot with head and eyebrows in a levelled position. Differently, the pride posture (Fig 1 (h)) the robot presents the head in a higher place, and the eyebrows are more open than in the neutral pose. In the shame posture (Fig 1 (i)), the robot's head is tilted down, and the eyebrows are almost closed. Videos were recorded with the robot in each of those setups.

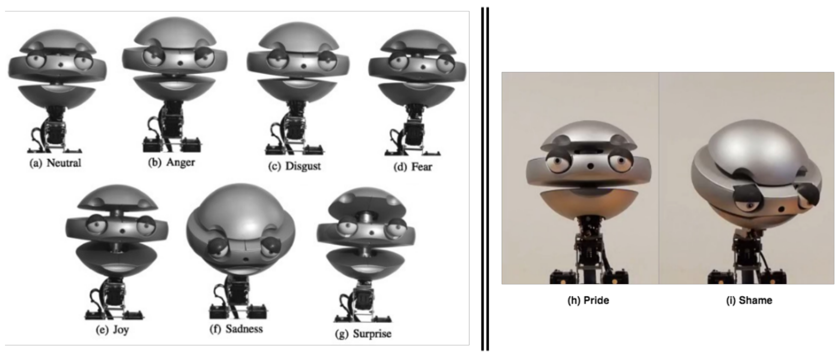

Fig. 1. Facial expression performed by EMYS robot (a to g)[6] and Postures pride (h) and shame (i).

3.3 Questionnaire

A questionnaire, with ten questions[3], was used to measure the level of assertiveness. The questions measured how a person sees her/himself regarding some statements. Each statement is divided into positive and negative keys. The positive ones are: *Take charge; Try to lead others; Can talk others into doing things; Seek to influence others; Take control of things.* The negatives are *Wait for others to lead the way; Keep in the background; Have little to say; Don't like to draw attention to myself; Hold back my opinions.* The participant informed how much s/he agrees with each statement using a 5-point Likert scale (1 - strongly disagree, 5 - strongly agree). Furthermore, we also inquired about the level of assertiveness in the different possible names given to the robot (Emys, Glin and none of those).

[5] Pitch and speech rate values were chosen from the prosody elements. https://www.w3.org/TR/speech-synthesis/#S3.2.4

3.4 Procedure

As participants start the experiment they provide some personal data, like gender and age, and then answer a set of questions measuring his/her assertiveness. Then, participants are asked to choose from two different names (Emys and Glin) pointing which one seems to be more assertive. Participants also inform if they ever interacted with the robot used in the study. Finally, a set of eight videos (presented in pairs) measuring a specific setup is shown to the participant. Videos were pairwise because we seek to identify which value of each setting influence more the perception of assertiveness individually.

In this way, each participant is going to view four comparisons each of which contains two videos. Each video presents the same assertiveness questionnaire that the user filled in with his data, but now regarding the storyteller robot's performance. Moreover, it was asked which emotion the participant perceived in the video from the options: joy, surprise, sadness, anger and none. As already discussed, the eight videos have a different robot's setting; the first six settings can be seen in Table 1.

Table 1. Robot's setting for each video.

Comparison	Video	Pitch	Rate	Posture
C1	V1	x-low	medium	neutral
	V2	x-high		
C2	V3	default	medium	neutral
	V4		+20%	
C3	V5	default	medium	pride
	V6			shame

In Table 2, the last two configurations for the robot's settings are presented. These two setups were intended to represent what we believe to be a more assertive robot and a less assertive robot (V7 and V8 respectively). Those configurations were obtained according to our literature research and our perception when testing each feature in the robot before the participants' tests.

Table 2. Robot's setting to our predefined assertive and non-assertive configuration.

Comparison	Video	Pitch	Rate	Posture
C4	V7	x-low	+20%	pride
	V8	x-high	medium	shame

Besides, to avoid the tiredness of evaluation, the comparisons were randomised at every ten participants employing a Latin square design. For example, the first ten participants watched the videos in the comparisons order C1-C2-C3-C4; then, the next order is C4-C3-C2-C1, and so on.

4 Results

To perform the statistical analysis of the data we used the SPSS software, and p-values of 0.05 or less as being considered significant. The analysis covers demographic information to statistical analysis of parametric and non-parametric data. Further, to measure the perception of the assertiveness trait from the answers of each video, it was calculated the average of all positive statements.

4.1 Demographic

Considering the described scenario, the questionnaires with the videos were available online for 15 days, and a total of 37 people answered it. Table 3 presents information about participants gender, age and if they ever interacted or not with the EMYS robot.

Table 3. Statistics of the Participants.

Gender	Age Avg (Std.)	Quantity	Interacted EMYS	Not Interacted EMYS
Female	$34_{(11.75)}$	26	8	18
Male	$29.7_{(3.25)}$	11	1	10
Total	$32.7_{(10.19)}$	37	9	28

4.2 Assertiveness of Name

We questioned participants about which name seemed to be more assertive between three options Emys, Glin or none of them. Based on the results, 18 people recognised the Emys name as more confident, while 13 thought that Glin was more assertive and six none of them.

4.3 Normality Test

To assess the normality of our data, we first perform a normality test using the Shapiro-Wilk test. The results reveal that for video 6 (V6) and video 8 (V8) the data significantly deviate from a normal distribution with p=0.005 and p=0.008 respectively. For all the other videos the data presented a normal distribution.

4.4 Significant Difference

After the normality test, we ran a paired t-test to see if there were significant differences between the pitch, V1 (Pitch-Low) and V2 (Pitch-High) and the rate, V3 (Rate-Medium) and V4 (Rate+20%). We found out that Pitch-Low and Pitch-High scores were not correlated (r=0.290, p=0.081), but the scores for Rate-Medium and Rate+20% were weakly and positively correlated (r=0.365, p=0.026). On the opposite, the paired sample test reveals that there was a significant average difference between Pitch-Low and Pitch-High (t_{36} = 2.343, p=0.025) but not for Rate-Medium and Rate+20%.

For the non-parametric data, the posture, V5 (Posture-Pride) and V6 (Posture-Shame), we applied the Wilcoxon Signed Rank Test. The test indicated that there was a statistical difference between the postures, Z=-2.534, p=0.011. Regarding our predefined configurations for assertiveness (V7) and non-assertiveness (V8),we applied a Wilcoxon Signed Rank Test. Statistical differences were detected among the two predefined configurations, Z=-3.305, p=0.001.

To understand what was the impact that the dimensions pitch and posture had over the predefined configuration for assertiveness (V7) an ANOVA with repeated measures (Pitch-Low, Posture-Pride and V7, designated as behaviour) were considered. The Mauchly's Test of Sphericity indicated that the assumption of sphericity was not violated, $X^2(2)$=0.089, p=0.956. Having this in mind, the tests of within-subjects effects revealed that the mean scores for behaviour were statistically significant different F(2,72)=4.066, p=0.021. Since we wanted to understand where those differences occur, we analysed the post hoc test using the Bonferroni correction. The differences detected revealed that there were only significant differences in our configuration of assertiveness and the Posture-Pride (p=0.027; η^2=0.101). This value of η means that 10% of the difference in the perception of assertiveness, the predefined configuration explained it, indicating a difference between our configuration and pride posture. The result can indicate that when we combine the pitch with value x-Low and the posture pride, we have a higher perception of assertiveness.

To know whichever is the influence of the dimensions pitch and posture had over the predefined configuration for non-assertiveness (V8) we ran the Friedman Test for Pitch-High, Posture-Shame and V8. The results revealed that there is no significant differences between them with $p>0.05$.

In summary, the results reveal the perception that participants had over the video comparisons made regarding their assertiveness for the pitch, rate, posture and the predefined configurations. For two of the individual comparisons, pitch and posture, there were significant differences. Regarding the two predefined configurations for assertiveness and non-assertiveness, they were also considered statistically different. The results report that the posture pride has an impact on the predefined configuration of assertiveness when comparing both.

5 Discussion and Future Work

Comparing the robot settings with the respective value, we found that there is a significant difference in the perception of assertiveness. This difference occurred when comparing two values in pitch and posture dimensions, making them essential dimensions to not ignore when conveying assertiveness. In those cases, the manipulation between the pitch values x-low and x-high lead people to feel that the x-low value fitted more an assertive robot. This fact can be noticed by the mean of the assertiveness questionnaire for each pitch video, see Fig 2 columns Pitch Low and Pitch High the mean and standard deviations. This perception can be justified by some studies that show that people tend to respond better to speakers with a lower pitch [25]. In this sense, our first hypothesis was validated, as the voice manipulation influenced the perception of assertiveness in a social storyteller robot by the participants.

A similar effect happened in the videos where the posture is assessed. The pride posture is distinguished as more assertive than a shame posture. The mean of pride posture video is higher than the shame posture video, see Fig 2 columns Posture Pride and Posture Shame. The shame posture can be characterised as more non-assertive mainly because it is a posture where the robot is gazing less at the person, as the robot's head is most of the time tilt down. The pride posture, on the contrary, had the robot's gaze directed most of the time at the participant. In this sense, the eye contact is an essential factor to identify and be more assertive [20]. Further, the third hypothesis was also validated, once the robot's posture gives the perception of assertiveness in the participants.

However, the speed at which the robot tells the story did not create the perception of assertiveness. The videos where the robot tells the story at a medium

		Pitch Low	Pitch High	Rate Medium	Rate + 20%	Posture Pride	Posture Shame	Assertive	Non-Assertive
N	Valid	37	37	37	37	37	37	37	37
	Missing	0	0	0	0	0	0	0	0
Mean		3.0000	2.5135	2.8919	2.7243	2.8162	2.3081	3.3622	2.3622
Median		3.0000	2.6000	3.0000	2.8000	3.0000	2.0000	3.4000	2.0000
Std. Deviation		.98658	1.12501	1.05892	1.10589	.93853	1.18565	1.06023	1.14002

Fig. 2. Descriptive statistics about all eight videos rated by participants.

rate and with +20% did not obtain significance difference. A possible justification of this might be due the +20% value not being enough to perceive the differences between each storytelling rate. Another argument is that this factor might not be perceived individually, it might need to be measured in consonance with others features. In this case, our second hypothesis could not be considered valid. It is necessary to perform more studies to investigate the effect of this setting using a different range of values. In fact, our results also show that the speed rate did not seem to influence the perception of assertiveness. However, we measured the robot telling the story +20% faster than a medium rate. In the future, it would be desirable to investigate the influence of a faster or a slower rate in storytelling by a robot.

However, the most significant difference happens in the case of the storytelling while performing the predefined configurations. This effect could be due to the combination of the robot's features to simulate an assertive and non-assertive behaviour. Based on the results, our last hypothesis was validated, the combination of vocal manipulations and posture can boost the perception of assertiveness in a social storyteller robot.

Finally, based on the results, we believe that it is possible to find a suitable setting for a social robot to play the role of a storyteller with assertiveness trait. We evaluate only three features in a universe of characteristics that can be measured. In fact, our study contributes with a valid configuration to use in scenarios where the assertiveness trait is important in a storyteller robot. With the increased use of social robots in storytelling scenarios and the fact that the robots configurations are dependent on several factors, such as the embodiment, more research is required to determine a suitable general configuration for assertiveness. The assertiveness trait is indeed considered essential in many contexts, such as for example in leadership roles, and giving orders. Because of that, we intend to use those configurations in a future scenario where it can be explored together with persuasion.

Acknowledgments. We would like to thank the National Council for Scientific and Technological Development (CNPq) program Science without border:201833/2014-0 - Brazil and Agência Regional para o Desenvolvimento e Tecnologia (ARDITI) - M1420-09-5369-000001, for PhD grants to first and second authors respectively. This work was also supported by Fundação para a Ciência e a Tecnologia: (FCT) - UID/CEC/50021/2013 and the project AMIGOS:PTDC/EEISII/7174/2014.

References

1. Apple, W., Streeter, L.A., Krauss, R.M.: Effects of pitch and speech rate on personal attributions. Journal of Personality and Social Psychology **37**(5), 715 (1979)

2. Bradley, J.H., Hebert, F.J.: The effect of personality type on team performance. Journal of Management Development **16**(5), 337–353 (jul 1997). https://doi.org/10.1108/02621719710174525, `https://www.emeraldinsight.com/doi/10.1108/02621719710174525`
3. Costa, P.T., McCrae, R.R.: Revised NEO personality inventory (NEO PI-R) and NEP five-factor inventory (NEO-FFI): professional manual. Psychological Assessment Resources Lutz, FL (1992)
4. Devillers, L., Vasilescu, I.: Prosodic cues for emotion characterization in real-life spoken dialogs. In: Eighth European Conference on Speech Communication and Technology (2003)
5. Figueiredo, R., Brisson, A., Aylett, R., Paiva, A.: Emergent stories facilitated. In: Joint International Conference on Interactive Digital Storytelling. pp. 218–229. Springer (2008)
6. Kedzierski, J., Muszyński, R., Zoll, C., Oleksy, A., Frontkiewicz, M.: Emysemotive head of a social robot. International Journal of Social Robotics **5**(2), 237–249 (2013)
7. Kimble, C.E., Seidel, S.D.: Vocal signs of confidence. Journal of Nonverbal Behavior **15**(2), 99–105 (1991)
8. Kory, J.J.M.: Storytelling with robots: Effects of robot language level on children's language learning. Ph.D. thesis, Massachusetts Institute of Technology (2014)
9. Lubold, N., Pon-Barry, H., Walker, E.: Naturalness and rapport in a pitch adaptive learning companion. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). pp. 103–110. IEEE (dec 2015). https://doi.org/10.1109/ASRU.2015.7404781
10. Martinez, L., Falvello, V.B., Aviezer, H., Todorov, A.: Contributions of facial expressions and body language to the rapid perception of dynamic emotions. Cognition and Emotion **30**(5), 939–952 (2016)
11. MOON, Y., NASS, C.: How Real Are Computer Personalities? Communication Research **23**(6), 651–674 (dec 1996). https://doi.org/10.1177/009365096023006002, `http://journals.sagepub.com/doi/10.1177/009365096023006002`
12. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. IEEE Robotics Automation Magazine **19**(2), 98–100 (June 2012). https://doi.org/10.1109/MRA.2012.2192811
13. Mutlu, B., Forlizzi, J., Hodgins, J.: A storytelling robot: Modeling and evaluation of human-like gaze behavior. In: Humanoid robots, 2006 6th IEEE-RAS international conference on. pp. 518–523. IEEE (2006)
14. Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, C.: Can computer personalities be human personalities? In: Conference Companion on Human Factors in Computing Systems - CHI '95. pp. 228–229. ACM Press, New York, New York, USA (1995). https://doi.org/10.1145/223355.223538, `http://portal.acm.org/citation.cfm?doid=223355.223538`
15. Niculescu, A., van Dijk, B., Nijholt, A., See, S.L.: The influence of voice pitch on the evaluation of a social robot receptionist. In: 2011 International Conference on User Science and Engineering (i-USEr). pp. 18–23 (Nov 2011). https://doi.org/10.1109/iUSEr.2011.6150529
16. Niculescu, A., Van Dijk, B., Nijholt, A., See, S.L.: The influence of voice pitch on the evaluation of a social robot receptionist. In: User Science and Engineering (i-USEr), 2011 International Conference on. pp. 18–23. IEEE (2011)
17. Ozaeta, L., Graña, M.: On intelligent systems for storytelling. In: The 13th International Conference on Soft Computing Models in Industrial and Environmental Applications. pp. 571–578. Springer (2018)

18. Page, R.A., Balloun, J.L.: The effect of voice volume on the perception of personality. The journal of social psychology **105**(1), 65–72 (1978)
19. Paradeda, R.B., Martinho, C., Paiva, A.: Persuasion based on personality traits: Using a social robot as storyteller. In: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. pp. 367–368. HRI '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3029798.3034824, `http://doi.acm.org/10.1145/3029798.3034824`
20. Paterson, R.: The Assertiveness Workbook: How to Express Your Ideas and Stand Up for Yourself at Work and in Relationships. New Harbinger Publications (2000)
21. Plaisant, C., Druin, A., Lathan, C., Dakhane, K., Edwards, K., Vice, J.M., Montemayor, J.: A storytelling robot for pediatric rehabilitation. In: Proceedings of the fourth international ACM conference on Assistive technologies. pp. 50–55. ACM (2000)
22. Ramirez, M.A.: Interactive storytelling in the library. (Chophayom Journal) **27**, 87–94 (2016)
23. Reeves, B., Nass, C.I.: The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press, New York, NY, US (1996)
24. Rodero, E.: Intonation and Emotion: Influence of Pitch Levels andContour Type on Creating Emotions. Journal of Voice **25**, e25–e34 (2011). https://doi.org/10.1016/j.jvoice.2010.02.002
25. Sprole, S.: 5 principles of speaking with more assertiveness - accent artisan, `http://accentartisan.com/article/assertiveness/`
26. Striepe, H., Lugrin, B.: There once was a robot storyteller: Measuring the effects of emotion and non-verbal behaviour. In: International Conference on Social Robotics. pp. 126–136. Springer (2017)