
AVANT-SATIE! USING ERIK TO ENCODE TASK-RELEVANT EXPRESSIVITY INTO THE ANIMATION OF AUTONOMOUS SOCIAL ROBOTS

Tiago Ribeiro*
INESC-ID &
Instituto Superior Técnico
University of Lisbon
Portugal
me@tiagoribeiro.pt

Ana Paiva
INESC-ID &
Instituto Superior Técnico
University of Lisbon
Portugal
ana.paiva@inesc-id.pt

March 2, 2022

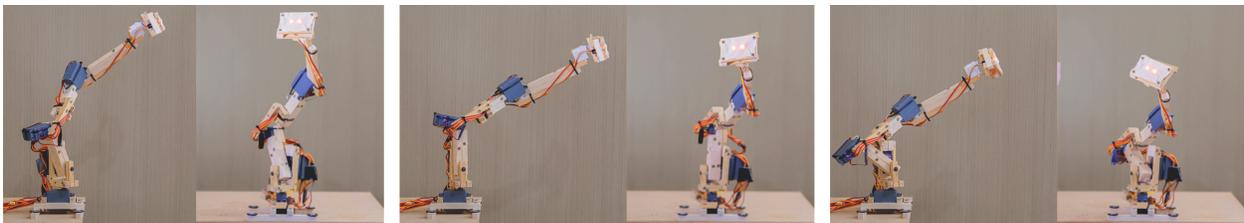


Figure 1: Pictures of the three different postures used by Adelino in the AvantSatie scenario (condition **C-ERIK**). These are the actual output of ERIK as used in the game, while gazing forward, and not a set of static pre-designed postures. Each pair shows a side view and its corresponding frontal view. The postures represented are: **left** - Neutral; **center** - Warm; **right** - Cold.

ABSTRACT

ERIK is an expressive inverse kinematics technique that has been previously presented and evaluated both algorithmically and in a limited user-interaction scenario. It allows autonomous social robots to convey posture-based expressive information while gaze-tracking users. We have developed a new scenario aimed at further validating some of the unsupported claims from the previous scenario. Our experiment features a fully autonomous Adelino robot, and concludes that ERIK can be used to direct a user's choice of actions during execution of a given task, fully through its non-verbal expressive queues.

Keywords Robot Animation, Autonomous Social Robots, Expressive Kinematics, Inverse Kinematics, Intention, Expression of Thought, Illusion of Life

1 Introduction

Character animation is the process of breathing the illusion of life into visual drawings or renderings of the embodiment of a character that takes part in some kind of narrative [1, 2]. The process however becomes much more difficult when applied to the field of autonomous social robots, given that the flow and especially the *timeline* of the story is driven not only by the interaction between users and the artificial intelligence (AI) [3], but also because the spacial dimension of the interaction is also linked to the user's own physical motion and placement.

*www.tiagoribeiro.pt

This challenge is remarkable enough that character animation for robots can be considered a new form of animation, which builds upon and extends the current concepts and practices of both traditional and Computer-Graphics (CGI) animation. Ribeiro & Paiva have defined robot animation as *the workflow and processes that give a robot the ability of expressing identity, emotion and intention during autonomous interaction with human users* [4, 5].

One such common *process* of robot animation that we are interested in is face-tracking, which directs a robot’s gazing towards the face of the human with whom it is interacting. For a simple robot, e.g., a neck with two Degrees of Freedom (DoFs), face-tracking can be easily implemented by extracting a vertical and horizontal angle from the system’s perception components (e.g. camera, Kinect), and using those to directly control individual motors of the robot’s neck. Gazing behaviour can also be compound, featuring not only face-tracking, but also deictic queues towards surrounding objects, and in conjunction with other static or motive expression (e.g. posture of engagement, nodding in agreement), as is the case when using a manipulator-like robot who’s end-effector takes on the expressive role of being the character’s head (e.g. the Adelino robot, depicted in Figure 1). In that case the compound-gazing process can also be used to express e.g. identity, emotion and intention, which are typically desirable for social robots [6].

In this paper we report how an existing inverse kinematics technique (ERIK) can be used in a real-world interactive application in which compound-gazing takes a major role, in order to allow an autonomous social robot to convey a recognizable intention-directed expression, and more importantly, to test its ability to direct the user’s choice of action during execution of a given problem-solving task, fully through non-verbal expressive queues.

2 Related Work

Robot animation builds on existing models and techniques from various fields. This section presents not all of them, but the ones we found most relevant for grounding work and inspiration.

Both Hoffman and Weinberg have created interactive robots that behave in musical and theatrical environments. The AUR is a robotic desk lamp with 5 DoFs and an LED lamp which can illuminate in a range of the RGB color space [7]. It is mounted on a workbench and remotely controlled through a hybrid control system that allows it to be used for live puppeteering, in order to allow the robot to be expressive while also being responsive. Its motion was composed through several layers. The bottom-most one moves each DoF based on a pre-designed animation that was made specifically for the scene of the play. When set to establish eye contact, several specific DoFs would be overridden by an inverse kinematics (IK) solution using CCD [8]. A final *animacy* layer added smoothed sinusoidal noise, akin to breathing, to all DoFs, to provide a more lifelike motion to the robot. Weinberg has also dedicated to the creation of robotic musical companions, such as Shimon, a gesture based musical improvisation robot created along with Hoffman [9]. Shimon plays a real marimba and its behaviour is a mix between his functionality as a musician, for which he plays the instrument in tune and rhythm, and being part of a band, for which he performs expressive behaviour by gazing towards his band mates during the performance.

Various expressive social robots have been created at MIT’s MediaLab [10], in particular the AIDA, which is a friendly driving assistant for the cars of the future. AIDA interestingly delivers an expressive face on top of an articulated neck-like structure to allow to it move and be expressive on a car’s dashboard. [11].

The use of animation principles has been developed and explored by various authors (e.g. [12, 13, 14, 15]). In particular, Takayama, Dooley and Ju [16] explored how the expression of intention leads to a sense of *thought*, using the PR-2 robot [17], and featured a collaboration with a professional animator. Later, Ribeiro & Paiva have also proposed a list of Principles of Robot Animation, in both an early and revised version [18, 5]. From all these authors we collect that *Thought* and *Intention* are two concepts that are central in character animation, and in the portrayal of the illusion of life.

The challenge of providing legible, predictable motion to autonomous collaborative robots was addressed by Dragan et al. [19], who demonstrate the benefits of including such properties into motion planners. Their technique however focuses on these properties in particular, and rely on motion-planning for trajectory generation.

2.1 Expressive Inverse Kinematics

Computing the motion of an articulated structure is commonly done through Forward Kinematics (FK) and Inverse Kinematics (IK). Given a kinematic chain of N segments S_i connected through joints J_i , where each *parent* joint allows its *child* segment to rotate about an axis R_i , the process of Forward Kinematics is to calculate the resulting position and/or orientation of an Effector S which is linked to any of the joints, given a set of angles applied to each joint; Inverse Kinematics is the calculation of the set of angles which, when applied to each joint, would bring the effector S (as close as possible) to a target position and/or orientation T . Due to space constraints we cannot fully describe the inner workings of all existing techniques, however those are already extensively described in both [20]

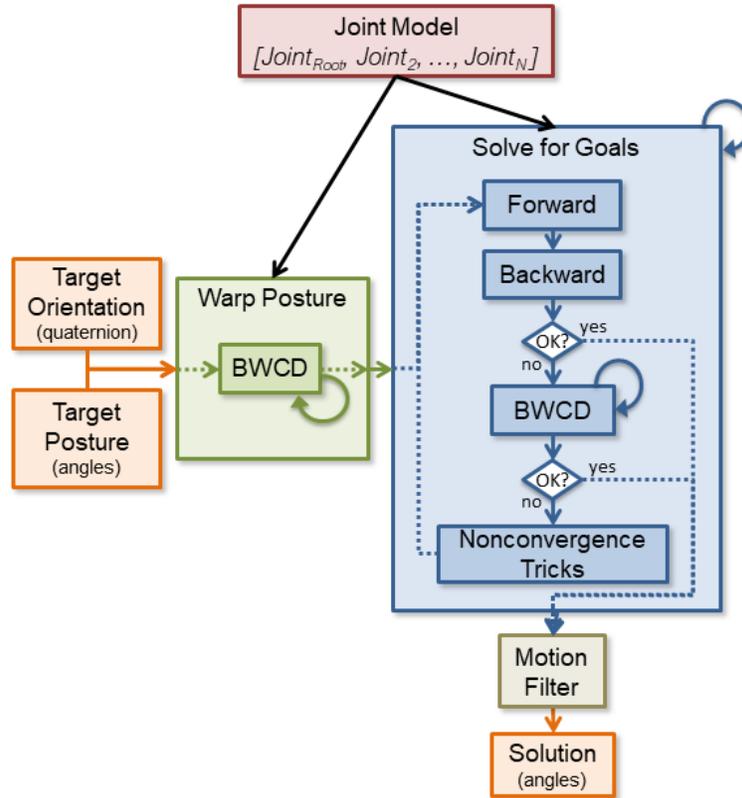


Figure 2: The ERIK Pipeline. The given Target Posture and Orientation are first warped using BWCD, so that the posture’s end-point is aiming towards the Target Orientation, without enforcing joint limits. The result feeds into the Iterative portion, which, through various phases on each iteration, approximates the configuration to the target parameters following a geometric approach based on FABRIK, and returns the final solution (with limits enforced). The Joint Model contains the skeletal information and auxiliary operations. The final solution runs through a motion filter to ensure smooth, continuous output. (figure and text cited from [25])

and [21] which are excellent reads on the topic. The most common techniques are based on iteratively calculating the inversion of the Jacobian matrix of the system, such as in the Jacobian transpose, DLS or SDLS techniques [20, 22]. Other popular techniques are the CCD [8] and FABRIK [23], which take a geometrical approach to the problem instead of solving a matrix system. These two methods in particular became popular in CGI as they produce good results with a low computational cost. FABRIK in particular is highly flexible, extensive to various complex situations, and provides quick and naturally-looking results [24].

Previously, Ribeiro & Paiva detailed and evaluated ERIK [25], an expressive kinematics technique that builds on both FABRIK and a variation of CCD (called *BWCD*), which allows a robot to express a given posture towards a specified direction, given an arbitrary embodiment, in real-time and without requiring motion planning or prior training. The technique is claimed to provide a flexible and extensible solution, untied to specific parameters such as an emotion or an interaction feature. It was designed for real-time applications, with minimal authoring required, given that a single input posture can be expressed towards any direction without drastically compromising its underlying expressive intent, in a way that also allows it to convey the illusion of life. However the technique’s complex formulation also makes it difficult to assure its reliability and consistency, especially given that, besides its algorithmic evaluation performed in that paper, it was initially evaluated with users in *Ahoy!*, a scenario in which participants remained static, and the robot was being controlled, to some extent, through a Wizard-of-Oz mechanism [26]. The ERIK pipeline is shown in Figure 2 which also includes a brief description of its internals. Its full description and algorithm is described in [25], and should be consulted by the interested reader.

3 AvantSatie - The Piano Game Companion

After analysing the previous *Ahoy!* scenario used to evaluate ERIK [26], we became interested in studying its use in a more realistic and fully autonomous setting, by studying not only the robot’s ability to convey a recognizable expression, but more importantly, its ability to direct the user’s choice of action in a problem-solving task. In the previous evaluation, the players were additionally standing still, which failed to validate the algorithm’s ability to express a given posture towards different directions. Furthermore, the queues were used to try to direct the players towards the correct answer in a pantomimic game, which took an open input (through speech), and required the use of *Wizards* to listen and assess the player’s answer, which then triggered the robot’s reactive expression. In that scenario, the robot seemed to be more like part of the riddle, than as part of a solution to it. Our goal is to understand if the technique can actually be used in a collaborative scenario in which users solve a problem that is independent of the robot, but in which the robot’s expressive behaviour may take a role in assisting the user, by providing expressive queues that facilitate the task.

The question we therefore posed to address is:

Can ERIK provide Adelino with the ability to communicate non-verbal hints to a user about what action to take next during a particular task, while also conveying the illusion of life?

For that purpose we established the following requirements for a new evaluation scenario: the robot must be **fully autonomous**, in order to evaluate ERIK based on its actual response-time; users must be forced to **move around** in order to fully evaluate the expressive gaze-tracking behaviour; the task must be a non-subjective problem-solving one, and must be **solvable even without expressive queues** (hints) through a *trial-end-error* method, however such hints should allow to solve it with significantly less *errors*.

The previous scenario already featured Adelino (depicted in Figure 1), a highly expressive, articulated craft robot whose end-effector is used as a face, and was purposely built as a low-fidelity one with an *organo-tech* look, in order to challenge the paradigms of robot design, and to seem more appealing to the arts’n’crafts and DIY² communities. One of our claims for building such a robot is that, although it exhibits some involuntary shakiness, as can be seen in the accompanying video, we consider that it can become part of its own unique *character* and aesthetically even contribute to a sense of lifelikeness, as long as the animation system and techniques used (e.g. ERIK) are sturdy enough and can properly control its conveyed communicative intent beyond that issue. This question of whether the communicative intent of the robot can be consistently conveyed beyond any shakiness or low-fidelity motion is therefore intrinsically part of our research question.

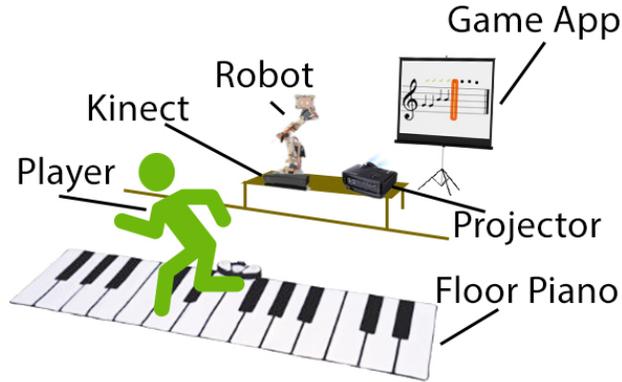
Based on these arguments, and because we thought Adelino’s design was appropriate for our evaluation, we decided to use the same robot as in the previous scenario, given that our major concern was on the actual evaluation’s activity design and conclusions, and not on the use of this particular robot. In fact, if the claims do stand, then our opinion is that Adelino and *AvantSatie* may also represent a breakthrough in robot and application design for HRI, and inspire future generations of robots.

3.1 AvantSatie - Activity Design

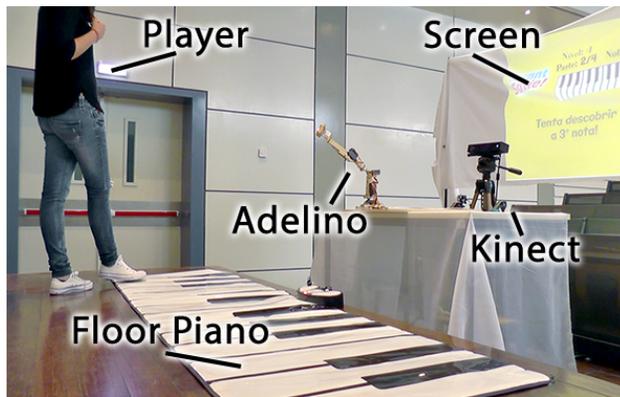
AvantSatie is a pervasive game where players must discover the musical score of a piece using a floor piano where they can step on to play notes. To help them in the game, they interact with an autonomous Adelino robot that will help them discover which notes compose the musical scores of two different pieces (each piece corresponds to a level). By interacting with the piano, observing the robot and following the instructions, participants either adopt a trial-and-error process, or track the robot’s hints, to discover each successive note. The game’s set-up is illustrated in Figure 3a along with a shot of an actual experimental session (Fig. 3b).

The fully autonomous robot enriches the setting of the game by providing the story as well as being socially present through compound-gazing. Its gazing behaviour combines a face-tracking feature, and deictic gazing towards specific piano keys. At the same time, it is expressive and can shape its posture while gazing, in order to convey hints to the player. Through the understanding of such hints, players can play the game while minimizing the amount of mistakes performed through the trial-and-error nature of the gameplay. Yet, if they do not pay attention to the robot, or fail to understand it, their task in the game becomes much more difficult. The game was designed and iteratively tested with users in pilot studies, in order to ensure that the instructions and gameplay were clear, instead of relying on initial instructions given by the experimenters, which could introduce biases.

²Do-It-Yourself http://en.wikipedia.org/wiki/Do_it_yourself



(a) The set-up of AvantSatie.



(b) The shot of a study session.

As the game is about discovery, the scores and composition are initially unknown to the player. They must therefore attempt to play keys on the piano until they find each correct note. The robot's behaviour is fully non-verbal. Only a screen is projected behind it, providing basic instructions and progress (e.g. current level).

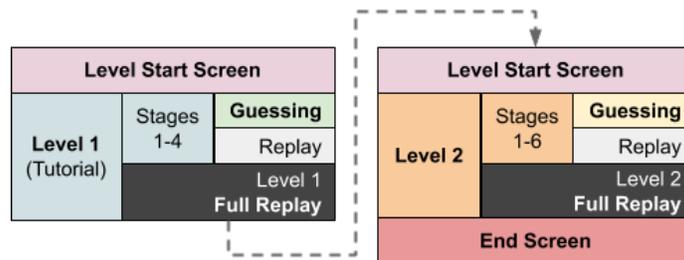


Figure 4: A diagram of the game-flow of Avant Satie. There are two levels, composed of a sequence of stages. Each stage is composed of a challenging activity (Guessing) followed by a rewarding activity (Replay), with each level ending with a greater reward (Full Replay).

The structure of the game-play is outlined in Figure 4. There are two levels in the game, which correspond to two musical compositions, each one with distinct levels of difficulty. Each level (music) is split into stages, which we called Parts, as that term fits better in the context of the game (e.g. the music in Level 1 is composed of 4 Parts/Stages). Each Part contains a sequence of one to six individual notes to be discovered one by one in the correct order.

Figure 5 shows some shots of the AvantSatie screen throughout the game. Here we present shots of the English version, while the study was ran using the Portuguese version (as can be seen by comparing these with Figure 3b).

The start screen requires the player to interpret a piano figure and to interact with the floor piano (fig. 5.a), upon which the robot performs an affirmative animation, i.e., nodding as if trying to say "yes!" (first positive feedback). This ensures that the player understands the basic interaction pattern of the activity i.e., screen displays instructions, playing the piano triggers a reaction on the robot. It then follows with a little storyline and instructions on how to play (fig. 5.b). Because Adelino is designed and animated as a non-verbal character, we rely on the screen to present in-game instructions, which also helps to immerse the player (versus having provided instructions prior to the activity). Whenever an instruction screen is being presented (e.g. fig. 5.b) the robot turns to face the screen, as a mechanism to direct the player's attention to it (otherwise due to enthusiasm, the player might be too focused on the robot and overall set-up). This also adds a feeling of presence - the robot is aware both of the player and of its surroundings (i.e. the screen - a point of shared attention). When the first instruction set is over, the robot turns back towards the user and plays the affirmative animation again. This animation is later used throughout the game, so it was important to initially present and reinforce it as a positive feedback.

On the first stage of all, the player is presented with no information except for the instruction "Discover the 1st note!" (fig. 5.c). While the player is moving around in front of the piano, the robot only performs face-tracking behaviour. Upon playing some note, the robot assesses it as the player's guess. If it is correct, the robot performs the affirmative animation, after which the screen progress updates to e.g. "Discover the 2nd note!", and the robot goes back to face-tracking. By having the robot provide feedback before the screen does, we manage to keep the player's visual attention focused on the message that the robot is communicating, instead of instigating them to search for feedback and new information on the screen (which we found to be the *instinct* of most people, during our pilot tests and iterative scenario design process).

These steps repeat until all the notes of the current Stage are found. After that, the robot replays all the Stage's notes, while pointing at each corresponding piano key, and then instructs the user to repeat it, with the screen exhibiting an illustration of the piano, highlighting each note, so that the player can unequivocally follow (fig. 5.d). This Replay phase serves as a reward to the player for having struggled to discover the composition. Each level was purposely built as a sequence of *Challenging* followed by *Rewarding* phases in order to maintain the user's engagement. At the end of each level the player gets to replay the full Level's composition as a bigger reward. The name of the piece and composer is revealed, and the player replays all parts (fig. 5.e-f). The first Level is slightly shorter and easier than the second one (no black keys are used). It is also used as a tutorial, and allows to separate data collection in order to mitigate the effects of learning how to interact with and play the game.

3.2 Experimental Design

The main purpose of this experiment was to study if ERIK successfully endowed Adelino with the ability to communicate task-directed information fully through its posture, while it is gaze-tracking the player. For that, we compared three versions of AvantSatie.

In version **C-ERIK** and **C-EBPS**, Adelino would respond to each of the player's guesses by modifying its posture based on a "warm-cold" heuristic. Upon each wrong guess the robot would therefore shift its expressive posture to either *Warm* or *Cold*, while keeping the gaze-tracking behaviour towards the player. Figure 1 (teaser in the first page) shows how these three postures look like on Adelino. The figure shows each of the neutral, hot and cold posture as output by the ERIK algorithm while gazing forwards, towards an angle of about 25° above the horizon.

We did not initially mention that the robot would perform this type of behaviour to any participant, in any study condition. Instead, it was expected that during the first minutes of the game (i.e. the Tutorial level), the players in these two conditions would notice that the robot was performing some behaviour that seemed congruent with their guesses, and would learn how to interpret the robot's posture in order to score better in the game.

The difference between **C-ERIK** and **C-EBPS** is purely technological - in **C-ERIK** we used ERIK to perform the gaze-tracking with an expressive posture. EBPS, is a non-IK example-based posture synthesis technique for which we previously authored a large number of postures for each *Warm*, *Neutral* and *Cold* expressions, each posture representing a pair (*expression*, *direction*), which sets the robot facing through a range of directions that are expected for this game, and produces a final posture through interpolation. In run-time, given the face-tracking information, we take the two postures that represent the directions that are closest to the target one and interpolate them. In our pilot studies we verified that this technique provided very smooth and acceptable results, by creating postures for each horizontal direction from -70° to 70° with 10° interval between them, which leads to 15 postures per expression, for each vertical direction. We initially considered to use 3 or 4 vertical directions, but upon testing realized that 2 would be enough, and that the interpolation between the two extreme vertical positions yielded acceptable results. The two vertical directions for which the postures were created were at 0° (looking straight ahead, i.e., to the horizon) and 80° (upwards). The



Figure 5: Various screenshots of the projected screen of the AvantSatie game: **a)** start screen; **b)** (part of) instruction screens; **c)** guessing phase in Level #1; **d)** replay phase level #1 (including replay phase instruction); **e)** Full-level replay and completion; **f)** Excerpts of replay phases for level #2.

postures used in **C-EBPS** were additionally designed with special attention in order to look very similar to the ones generated by ERIK.

We note that, while the end-result looked similar to the one we wanted to achieve with ERIK, it required a considerable amount of work to create all those postures; Furthermore if we wanted to modify one of the expressions, we would be required to re-create all the postures for that expression and rebuild the interpolation model. Using ERIK only requires to author one single posture per intended expression.

Finally, **C-Control** differs from the other two conditions by not performing any posture-expressive behaviour at all. The whole game is exactly the same, and the robot also performs gaze-tracking behaviour, using ERIK to perform gaze-tracking, but always holds the same, *Neutral* posture throughout the game. Therefore in this condition players rely solely on trial-and-error to discover the notes.

The study followed a between subjects design with random assignment within the three different conditions. In all conditions the game design follows the same structure, with the same levels and tutorial information.

In order to address our research question, we established the following hypotheses:

- **H1:** Players in **C-Control** will commit more mistakes than in **C-ERIK** and in **C-EBPS**.
- **H2:** Players in **C-ERIK** will play as well as in **C-EBPS**.
- **H3:** The robot is perceived to convey the illusion of life in all conditions.
- **H4:** The robot is perceived to convey the illusion of life more in **C-ERIK** and **C-EBPS** than in **C-Control**.
- **H5:** The players are able to perceive the robot's intention and motivation as being towards helping them in both **C-ERIK** and **C-EBPS** but not in **C-Control**.
- **H6:** The game is understandable and the robot is perceived to understand and play well in all conditions.

Upon arrival, the participants filled out the consent form in a separate room before being led to the game room. There they were given the same initial instructions, without revealing that the robot would indicate the result of their guesses through a change in posture. Instead they were solely informed that there would be two simple compositions to discover, that they should perform each guess and observe the robot, until they were able to discover all the notes, and otherwise just follow the instructions on screen. The researcher would direct them to enter the room and start interacting without following them, as the robot was already active and would start face-tracking them once they stepped into the Kinect's field of view. This ensured that the participants also noticed it immediately as an autonomous entity and would become immersed into the game. The screen provided the starting instruction, which was to play a D note on the floor piano, along with an icon showing explicitly which key that was. Therefore it was the participant who took the step to initiate the game, while also ensuring that they understood the piano to be a controller for it. When finished, the participants were taken back to the initial room, administered a set of questionnaires, and received a thank-you gift (a movie ticket) at the end.

The following questionnaires were used as subjective measures:

- *Networked Minds* [27] scale (**NM**), from which we took the *Perceived Message Understanding* (**PMU**) and the *Co-Presence* (**CP**) dimensions;
- *Inclusion of Other in Self* [28] scale to measure the closeness that the participants felt with the robot (measure **IOS**);
- **RoSAS** [29] scale to measure the perception of the robot's social attributes Competence (**RC**), Warmth (**RW**) and Discomfort (**RD**);
- *Perceived Adaptability* (**PA**) from the *Almere model*;
- *Robot's Performance & Usability* (**RPU**) scale to measure how well the participants felt the robot to be able to play the game, and how well the overall gameplay was designed;
- *Robot's Intention & Motivation* (**RIM**) scale to measure how much participants felt the robot was there to provide tips and how much it wanted to succeed in helping them;
- *Animation Illusion of Life* (**AIL**) scale to measure the illusion of life of the robot.

The questionnaires for the RPU, RIM and AIL scales were specifically designed to address our research question. The RPU scale in particular is composed of two dimensions:

Task Performance (**TP**) measures how well the participants perceived the robot to know the game and perform the task well;

Task Usability (TU) measures how easy and intuitive the participants felt it was to understand the task and the game-play interaction with the robot and the screen.

The RIM scale is also composed of two dimensions:

Robot's Intention (RI) measures how much the participants felt that the robot was providing hints to them throughout the task;

Robot's Motivation (RM) measures how much the participants felt that the robot's intrinsic motivation (i.e. *purpose*) was to help them (by providing hints).

Table 1 lists the questions used for each of these measures.

All the questionnaires were answered in a 6-point Likert scale except for the RoSAS, which was answered in a 9-point Likert scale, and the IOS measure which was answered in a 7-point pictorial scale. The data presented in the results has already been corrected by inverting the scores in negative scales. All items were shuffled to mask for their dimensions.

RPU (TP + TU) - Robot's Performance & Task Usability	
TP1.	The robot knew where each note of each music was.
TP2.	The robot always understood what note I had played.
TP3.	The robot knew each music very well.
TP4.	The robot knew every music by heart.
TU1.	I had to look at the screen to know what happened at each moment.
TU2.	I wouldn't understand the game without looking at the screen.
TU3.	The game screen had all the info I needed to understand the game.
TU4.	I had to follow the screen to know what to do.
RIM (RI + RM) - Robot's Intention & Motivation	
RI1.	I wouldn't have understood the game without the robot.
RI2.	I managed to find the correct notes thanks to the robot.
RI3.	I wouldn't have discovered the musics without the robot's help.
RI4.	The tips that the robot gave me helped me to find the correct notes.
RI5.	The robot's tips were consistent with my attempts to find each correct note.
RI6.	The robot gave me tips while I was trying to find each correct note.
RI7.	I was able to understand if I was close or far from the correct note thanks to the robot's tips.
RM1.	The robot wanted me to find the correct notes.
RM2.	The robot wanted me to discover all of the musics.
RM3.	The robot was enthusiastic with my attempts to find the correct notes.
RM4.	The robot thought about helping me.
AIL - Animation Illusion of Life	
AIL1.	The robot's movement was smooth and natural.
AIL2.	The robot seemed to be alive.
AIL3.	The robot reminded me of characters I know from movies.
AIL4.	The robot's motion followed my rhythm.

Table 1: RPU RIM and AIL questionnaires.

The following objective data was also collected during each session, and measured only during the *Guessing* phases of the game:

- **Time** spent guessing;
- **WrongHot** number of incorrect guesses which were however assessed as Hot;
- **WrongCold** number of incorrect guesses which were however assessed as Cold;
- **WrongTotal** aggregates WrongHot and WrongCold.

3.3 Experimental Results

A total of 59 university students (30 females and 29 males) with ages ranging from 18 to 35 ($M = 22.78$; $SD = 3.96$) were recruited. From these, two were excluded due to not complying with the instructions, thus yielding a total of 57 valid participants, which resulted in a balanced distribution of 19 participants per condition. 19% of the participants had already interacted with a robot before once, and 37% more than once. 42% reported a low to no level of expertise playing some musical instrument, while 39% reported an intermediate expertise, and 19% an advanced expertise. As for experience reading sheet music, 63% reported a low score, 25% an intermediate score, and 12% declared to be experts.

Throughout this analysis, we will be considering an additional **C-Enhanced** group which averages the measures of both **C-ERIK** and **C-EBPS** in order to treat them both as a single group.

We started by using the Shapiro-Wilk test of normality to verify for which measures the data was normally distributed ($\rho > 0.05$). Although we have four groups, in our analysis we will compare the means of only two groups at a time, thus we used the Student's t-Test when the data is normally distributed in both groups being tested, and the Mann-Whitney U test otherwise.

Table 2 shows the results of comparison of the means of all the subjective and objective measures, including sub-dimensions, between the different groups, with Figure 6 providing a clearer illustration. Within it we present each

Measure Type		ERIK-Control	EBPS-Control	Enhanced-Control	ERIK-EBPS
Subjective	AIL	(Tukey, $\rho = 0.046*$)	(Tukey, $\rho > 0.05$)	(Tukey, $\rho = 0.022*$)	(Tukey, $\rho > 0.05$)
Subjective	RIM	(Dunn, $\rho = 0.003*$)	(Dunn, $\rho = 0.001*$)	(Dunn, $\rho = 0.000*$)	(Dunn, $\rho > 0.05$)
Subjective	RPU	(TukeyHSD, $\rho > 0.05$)	(Dunn, $\rho = 0.030*$)	(Dunn, $\rho > 0.05$)	(Dunn, $\rho > 0.05$)
Subjective	IOS	(Tukey, $\rho > 0.05$)	(Dunn, $\rho = 0.009*$)	(Dunn, $\rho = 0.044*$)	(Dunn, $\rho > 0.05$)
Subjective (sub)	ri	(Dunn, $\rho = 0.006*$)	(Tukey, $\rho = 0.002*$)	(Dunn, $\rho = 0.001*$)	(Dunn, $\rho > 0.05$)
Subjective (sub)	rm	(Welch, $\rho = 0.002*$)	(Welch, $\rho = 0.005*$)	(Welch, $\rho = 0.002*$)	(Welch, $\rho > 0.05$)
Subjective (sub)	tp	(Dunn, $\rho > 0.05$)	(Dunn, $\rho > 0.05$)	(Dunn, $\rho = 0.044*$)	(Dunn, $\rho > 0.05$)
Subjective (sub)	tu	(Dunn, $\rho > 0.05$)	(Dunn, $\rho > 0.05$)	(Dunn, $\rho > 0.05$)	(Tukey, $\rho > 0.05$)
Objective	Time	(Tukey, $\rho > 0.05$)	(Tukey, $\rho > 0.05$)	(Dunn, $\rho > 0.05$)	(Tukey, $\rho > 0.05$)
Objective	WrongTotal	(Dunn, $\rho = 0.037*$)	(Dunn, $\rho > 0.05$)	(Dunn, $\rho = 0.015*$)	(Dunn, $\rho > 0.05$)
Objective	WrongHot	(Dunn, $\rho > 0.05$)	(Dunn, $\rho = 0.037*$)	(Dunn, $\rho = 0.012*$)	(Dunn, $\rho > 0.05$)

Table 2: Results of the subjective scales and sub-dimensions, and objective measures, including the post-hoc used in each case. Green with an asterisk* marks comparisons which are significantly different. Omitted measures (RW, RD and WrongCold) do not show any significant difference.

measure's comparison of means (Figures 6a,6b,6d) along with Figure 6c which shows how each subjective measure in each group compares to the scale's median value, i.e., is the average score significantly *positive* (above median), *negative* (below median), or *neutral* (inconclusive). For that we took each scale's median value (e.g., for the 6-point likert scale 'AIL' the median is $\frac{6+1}{2} = 3.5$), and compared the distribution of each group's results to it. We used the One-Sample t-Test when the data followed a normal distribution, and the One-Sample Wilcoxon Signed Rank test otherwise. In order to make the presentation of results more summarized and easier to follow, we have also gathered the following comparison cases:

Strong Expressivity Difference (*Strong E-D*) : Significant difference between **C-Control** and all **C-ERIK**, **C-EBPS** and **C-Enhanced** groups, with no difference between **C-ERIK** and **C-EBPS**. These findings will be attributed to the inclusion of the postural/intention-expressive behaviour in the activity, regardless of its technical implementation.

Soft Expressivity Difference (*Soft E-D*) : Significant difference between **C-Control** and **C-Enhanced**, and also between **C-Control** and either **C-ERIK** or **C-EBPS**. These findings will also be attributed to the postural/intention-expressive behaviour, but with an indication that one of the technological implementations may have performed better.

Technological Difference (*T-D*) : Significant difference between the **C-Control** and either the **C-ERIK** or the **C-EBPS**, but not between the **C-Control** and the aggregated **C-Enhanced** group. These findings will be attributed to some difference in the technological implementation only.

3.3.1 Regarding the Subjective Measures

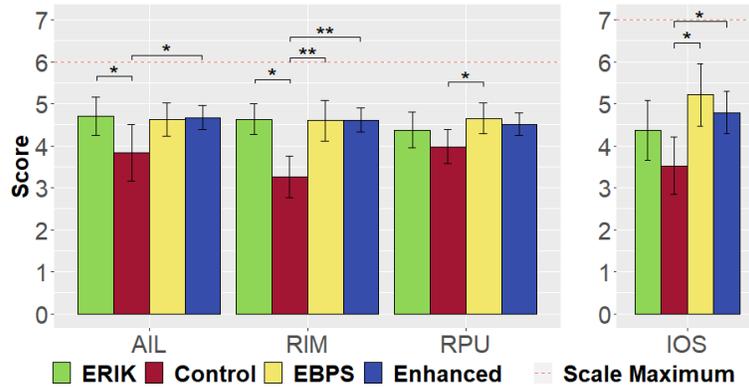
Within the various subjective measures used in our study, we draw the following:

AIL, RIM, NM and PA all reported significantly lower scores in **C-Control** with a *Strong E-D*.

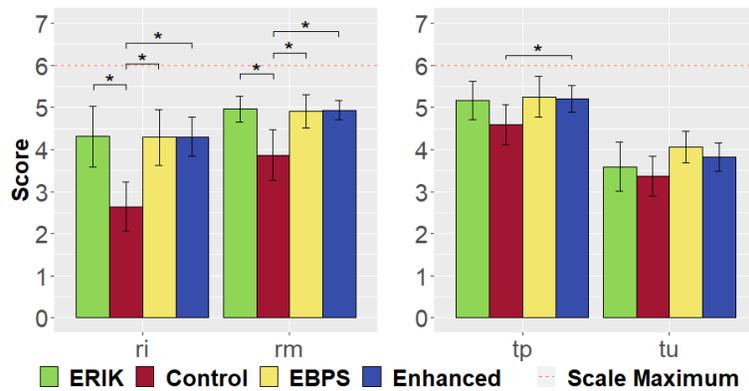
RPU reported a significantly lower score in **C-Control** with a *Soft E-D*.

IOS, RC reported a significantly lower score in **C-Control** compared to **C-EBPS** with a *T-D*.

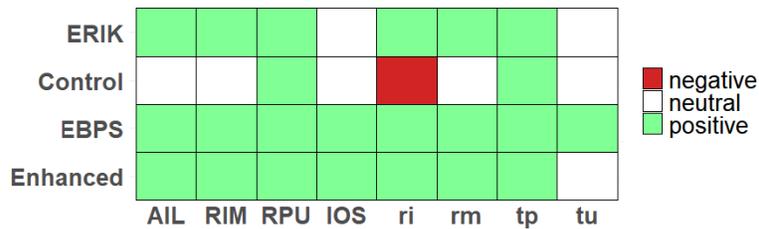
RC reported a significantly lower score in **C-Control** compared to **C-ERIK** with a *T-D*.



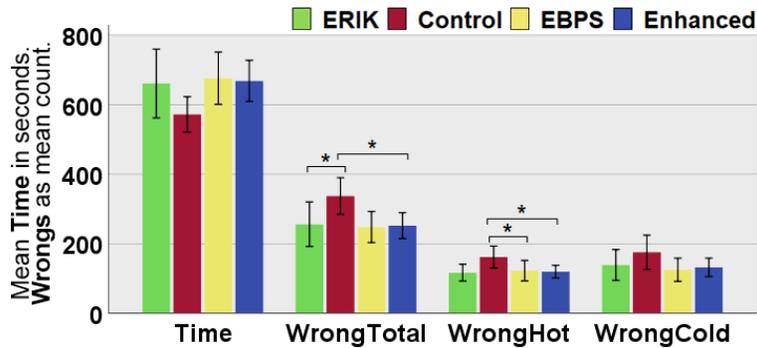
(a)



(b)



(c)



(d)

Figure 6: **a)** Comparison of the scores of the subjective measures' scales. **b)** Comparison of the dimensions used to compose some of the subjective measures. **c)** Comparison of means to the scale's median value on the subjective measures. **d)** Comparison of the objective measures. All error bars represent a 95% C.I.

RPU, NM, RC and RD all reported positive scores in all groups.

AIL and RIM reported positive scores in all enhanced groups, and neutral in the **C-Control** group.

PA, IOS and RW are inconclusive regarding the polarity of the scores, given that they show mixed results.

Within the dimensions that compose these scales, we can verify as expected based on the **RIM** and **NM** results, that the **ri**, **rm**, **cp** and **pmu** dimensions all report the same **Strong E-D**. However the **RPU** scale had reported only a **Soft E-D**. Analysing further, we can see that the **tp** dimension (robot's Task Performance) did exhibit the expected **Strong E-D**, while the **tu** dimension (Task Usability) reported only a significant **T-D** between the **C-Control** and the **C-EBPS** conditions.

We also have a very interesting finding within the **RIM** scale. Regarding the Robot's Intention (**ri**), the **C-Control** scored significantly negative while all the other groups scored positively (and not neutral). This is a very strong difference (which had already been pointed out in the comparison of means). Similarly, albeit with a smaller difference, the Robot's motivation (**rm**) and the Networked Minds' Perceived Message Understanding (**pmu**) were also perceived to be neutral in **C-Control**, while in all the others it was positive. Finally it is important to note that while the **RPU** scale was positive across all groups, we found that this was mostly due to the perceived robot's Task Performance (**tp**), as the Task Usability (**tu**) dimension scored significantly neutral on all groups, except on **C-EBPS** where it scored positively.

3.3.2 Regarding the Objective Measures

Figure 6d shows how the four objective measures performed across the different groups. The measure of **WrongCold** did not reveal any differences between conditions. However both the measures of **WrongHot** and **WrongTotal** show a **Strong E-D** with the **C-Control** condition containing a significantly higher amount of wrong answers than the other groups. The Time measure shows a **Soft E-D** as the **C-Control** group performed significantly faster than the **C-EBPS** group ($t=2.406$, $\rho=0.022$) and the aggregated **C-Enhanced** group ($U=242.0$, $\rho=0.044$).

3.4 Discussion and Conclusion

The results collected and analysed show us in general that our research question is supported. Looking into each of our initial hypotheses:

H1 - Players in **C-Control** will play worse than in **C-ERIK** and in **C-EBPS**: **True**. Participants in both **C-ERIK** and **C-EBPS** committed less mistakes than in the **C-Control**, by decoding and exploiting the hints given by the robot through its full-body posture.

H2 - Players in **C-ERIK** will play at least as well as in **C-EBPS**: **True**. As there was no significant difference between the total amount of wrong answers given by the players in **C-ERIK** and in the **C-EBPS**, thus the expressive postures provided by ERIK were as relevant and legible as the manually designed ones.

H3 - The robot is perceived to convey the illusion of life in all conditions: **Partially True**. All conditions except for **C-Control** reported a positive **AIL** average score. **C-Control** was neutral however, and not negative. Our interpretation is that the robot's additional intentional-directed expressive behaviour in the enhanced groups contributed significantly to convey they illusion of thought (which is core to the illusion of life).

H4 - The robot is perceived to convey the illusion of life more in **C-ERIK** and **C-EBPS** than in **C-Control**: **True**. Both **C-ERIK** and **C-EBPS** scored significantly higher in the measure than the **C-Control** group.

H5 - The players are able to perceive the robot's intention and motivation as being towards helping them in both **C-ERIK** and **C-EBPS** but not in **C-Control**: **True**. The score was significantly higher and positive both in **C-ERIK** and **C-EBPS** than in **C-Control**. When analysing **RIM**'s sub-dimensions separately (**ri** and **rm**), we find that the Robot's Intention scored negative in the **C-Control** condition. This was very relevant as it was the only dimension that scored negatively. Regarding the Robot's Motivation, we find a similar pattern, except that in **C-Control** it scored neutral (slightly better). Our guess is that in general the players had a positive feeling about the robot (based on the results from RoSAS) and therefore, maybe considered that the robot did intrinsically want to help them (although they reported negatively on its hint-providing intention). Furthermore, the Networked Minds (**NM**) and Almere's Perceived Adaptability (**PA**) scales both follow the same pattern, significantly higher in the enhanced conditions compared to **C-Control**, thus reinforcing that regardless of the technique used, the intention-directed behaviour of the robot, as designed and integrated into the gameplay, had a positive effect on various measures regarding the perception of the robot's intention, motivation and closeness towards the player.

H6 - The game is understandable and the robot is perceived to play well in all condition: **Partially True**. The **RPU** measure was positive in all groups. However when breaking down the scale, we find the Task's Usability (**tu**) was scored as neutral in all except the **C-EBPS** group, and that this difference is actually significant compared to **C-Control**. The game was designed in order to ensure that the participants would not become too affected by the lack of the posture-expressive behaviour (in **C-Control**) that they would not understand the task at all. While the **tu** measure reports a **T-D** on **C-EBPS**, it was not reported in the whole **C-Enhanced**, which means that we fail to refute that the inclusion or absence of the robot's intention-directed expressive behaviour does not cause a significant effect on the participants' understanding of the game and the task. Therefore the iterative game-design method (with 3 pilot tests) allowed to tweak the usability of the game to an acceptable level, even in the absence of the robot's full expressive behaviour, while also noting that the overall game-design and/or interaction design could have still been made better.

We further noted additional findings such as that participants in **C-Enhanced** took more time to complete the task than in **C-Control**. Inspection of the audio-video data captured from the study revealed that participants in the former conditions, having understood that the robot was giving tips, would try notes at a lower pace in order to inspect the robot's response, while in **C-Control**, after a while they would quickly play random notes, driving quicker through the task with a significantly higher number of mistakes.

The RoSAS scale shows only a significant difference in the Robot Competence (**RC**) measure between **C-ERIK** and **C-Control**. However the difference did not hold for the whole of the **C-Enhanced** group. We suspect that the ERIK algorithm may have yielded a higher feeling of competence, because the use of that algorithm is prone to result in more dynamic/responsive motion, which players may have attributed to a higher sense of acknowledgement of the other, and capability of attention, on the robot's part. In overall however, no concrete difference may be concluded between **C-ERIK** and **C-EBPS**, given that on comparing the various scales, there were either none, or mixed differences (e.g. in contrast to the previous remarks, for the IOS scale, the **C-EBPS** scores significantly higher than **C-Control**, but here **C-ERIK** does not). Although the RoSAS scale seems not to have added any relevant information, that fact may be used to also hypothesize that across all groups, the robot was perceived as being nearly the same *character* - which was desirable for our study.

We have found evidence that our research question (Section 3) is positively supported, given that:

- When ERIK was used, the participants noted its intentiondirected postural behaviour, and were able to intuitively understand it without having been given any information about its existence in order to perform better on a task that required it.
- The effect of using ERIK was similar to that of a manually tailored (and laborious) alternative technique EBPS, in that no significant differences were found for any of the measures between those two groups, while significant ones were found especially on key measures when compared to the Control condition. This means that ERIK can be used in substitution of such currently existing manually-tailored and arduously worked techniques (such as ones based on *learning-by-examples*).
- The difference between the various conditions did not hinder the player's understanding and playability of the game, impacting only on their performance, which reveals that the selected task was properly designed to answer our question.
- Despite the shakiness of the robot due to the fact of it being a low-fidelity craft robot, results show that the algorithm succeeded in making it convey both the intended expressivity, and the illusion of life, meaning that it is likely to work on both similar or more solid robots.

In addition, we highlighted the importance of the *illusion of thought* to the overall *illusion of life in robots*, as already had been initially proposed by Takayama et al. [16]. In our case we further demonstrated that such illusion can also be expressed through fully interactive expressive postures that are computed in real-time, and are therefore most appropriate for use-cases involving autonomous social robots.

ACKNOWLEDGEMENTS

This work was supported by national funds through FCT - Fundação para a Ciência e a Tecnologia with references UID/CEC/50021/2019 and SFRH/BD/97150/2013.

References

- [1] Frank Thomas and Ollie Johnston. *The Illusion of Life: Disney Animation*. Hyperion, 1995.

- [2] Eric Goldberg. *Character Animation Crash Course!* Silman-James Press, 2008.
- [3] Bill Tomlinson. From linear to interactive animation: how autonomous characters change the process and product of animating. *Computers in Entertainment (CIE)*, 3(1):1–20, 2005.
- [4] Tiago Ribeiro. *Creating the Illusion of Life in Autonomous Social Robots*. PhD thesis, Instituto Superior Técnico, University of Lisbon, 2020.
- [5] Tiago Ribeiro and Ana Paiva. Nutty-based robot animation – principles and practices, 2019.
- [6] Cynthia Breazeal. Towards Sociable Robots. *Robotics and Autonomous Systems*, 42(3–4):167–175, 2008.
- [7] Guy Hoffman, Rony Kubat, and Cynthia Breazeal. A hybrid control system for puppeteering a live robotic stage actor. *IEEE International Symposium on Robot and Human Interactive Communication - RO-MAN '08*, pages 354–359, aug 2008.
- [8] L.-C.T. Wang and C.C. Chen. A combined optimization method for solving the inverse kinematics problems of mechanical manipulators. *IEEE Transactions on Robotics and Automation*, 7(4):489–499, 1991.
- [9] Guy Hoffman and Gil Weinberg. Gesture-based human-robot Jazz improvisation. In *IEEE International Conference on Robotics and Automation - ICRA '10*, pages 582–587, 2010.
- [10] Jesse Gray, Guy Hoffman, Sigurdur Orn Adalgeirsson, Matt Berlin, and Cynthia Breazeal. Expressive, interactive robots: Tools, techniques, and insights based on collaborations. In *ACM/IEEE International Conference on Human-Robot Interaction - HRI '10 - Workshop on What do Collaborations with the Arts Have to Say About Human-Robot Interaction*, 2010.
- [11] Personal Robots Group. MIT Media Lab. AIDA robot.
- [12] Albert Van Breemen. Animation engine for believable interactive user-interface robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems - IROS '04*, volume 3, pages 2873–2878, 2004.
- [13] Ryan Wistort. Only robots on the inside. *interactions*, 17(2):72–74, 2010.
- [14] Ross Mead and Maja J Mataric. Automated Caricature of Robot Expressions in Socially Assistive Human-Robot Interaction. *ACM/IEEE International Conference on Human-Robot Interaction - HRI '10 - Workshop on What do Collaborations with the Arts Have to Say About Human-Robot Interaction*, mar 2010.
- [15] Michael J Gielniak and Andrea L Thomaz. Enhancing interaction through exaggerated motion synthesis. *ACM/IEEE International Conference on Human-Robot Interaction - HRI '12*, page 375, 2012.
- [16] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought. In *ACM/IEEE International Conference on Human-Robot Interaction - HRI '11*, page 69, 2011.
- [17] Willow Garage. PR-2 robot.
- [18] Tiago Ribeiro and Ana Paiva. The Illusion of Robotic Life Principles and Practices of Animation for Robots. In *ACM/IEEE International Conference on Human-Robot Interaction - HRI '12*, number 1937, pages 383–390, 2012.
- [19] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. Effects of Robot Motion on Human-Robot Collaboration. *ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 1:51–58, 2015.
- [20] Samuel R. Buss. Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods. *University of California, San Diego, Technical Reports.*, 2009.
- [21] Andreas Aristidou, J. Lasenby, Y. Chrysanthou, and A. Shamir. Inverse Kinematics Techniques in Computer Graphics: A Survey. *Computer Graphics Forum*, 37(6):35–58, 2018.
- [22] Samuel R. Buss and Jin-Su Kim. Selectively Damped Least Squares for Inverse Kinematics. *Journal of Graphics, GPU, and Game Tools*, 10(3):37–49, 2005.
- [23] Andreas Aristidou and Joan Lasenby. FABRIK: A fast, iterative solver for the Inverse Kinematics problem. *Graphical Models*, 73(5):243–260, 2011.
- [24] Andreas Aristidou, Yiorgos Chrysanthou, and Joan Lasenby. Extending FABRIK with model constraints. *Computer Animation and Virtual Worlds*, 27(1):35–57, 2016.
- [25] Tiago Ribeiro and Ana Paiva. Expressive Inverse Kinematics Solving in Real-time for Virtual and Robotic Interactive Characters, 2019.
- [26] Tiago Ribeiro and Ana Paiva. Animating the Adelino Robot with ERIK. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 388–396, Glasgow, UK, 2017. ACM.
- [27] Chad Harms and Frank Biocca. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004*, pages 246–251, 2004.

- [28] Arthur Aron, Elaine N. Aron, and Danny Smollan. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596–612, 1992.
- [29] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. The Robotic Social Attributes Scale (RoSAS). *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, (March):254–262, 2017.