# From Vocal Prosody to Movement Prosody, from HRI to Understanding Humans

Philip Scales, Véronique Aubergé, Olivier Aycard

# From Vocal Prosody to Movement Prosody, from HRI to Understanding Humans

Philip Scales[1], Véronique Aubergé[2], and Olivier Aycard[3]
Univ. Grenoble Alpes
CNRS
Grenoble INP
LIG
38000 Grenoble
France



## Author Note

[1]Philip Scales, [2]Véronique Aubergé (philip.scales, veronique.auberge@univ-grenoble-alpes.fr), and [3]Olivier Aycard (olivier.aycard@imag.fr)

**Abstract**

Human-Human and Human-Robot Interaction are known to be influenced by a variety of modalities and parameters. Nevertheless, it remains a challenge to anticipate how a given mobile robot's navigation and appearance will impact how it is perceived by humans. Drawing a parallel with vocal prosody, we introduce the notion of movement prosody, which encompasses spatio-temporal and appearance dimensions which are involved in a person's perceptual experience of interacting with a mobile robot. We design a novel robot motion corpus, encompassing variables related to the kinematics, gaze, and appearance of the robot, which we hypothesize are involved in movement prosody. Initial results of three perception experiments suggest that these variables have significant influences on participants' perceptions of robot socio-affects and physical attributes.

*Keywords:* Corpus, HRI, Prosody of Movement

## From Vocal Prosody to Movement Prosody, from HRI to Understanding Humans

### Authors

**Philip Scales** is a third year Computer Science PhD student at the Université Grenoble Alpes (UGA), in the LIG laboratory. He obtained his MSc in Computer Science from UGA-ENSIMAG in Grenoble in 2019. His doctoral work focuses on the study of the mechanisms underlying human-robot interaction, and human perception of mobile agents, as well as the integration of this knowledge into the design of social navigation algorithms for mobile robots.

**Véronique Aubergé** is a CNRS researcher in Human Sciences, she was a research engineer at OROS (1986-91), then PhD in 1992 in Language Sciences and Computer Sciences, teacher at ENSIMAG (1991-94) and CNRS researcher (1992-2012) at ICP Lab, then GIPSA Lab and LIG from 2012. She directed master Idl, Chaire Robo'Ethics, Living Lab Domus, Humanibots etc. She developed cognitive models and algorithms, experimental methods on human interactions in phonetics, spelling, prosody, language learning, expressive text-to speech. She focuses on social robotics as instruments to observe and to design models on the human interactional behaviors focused on the altruism dimension.

**Olivier Aycard** is an Associate Professor in the Computer Science Department at University of Grenoble-Alpes (UGA) and member of the Machine Learning group of the LIG since 1999. He received his PhD in Computer Science from University of Nancy in 1998 and his HDR (Accreditation to Supervise Research) in Computer Science from University of Grenoble in 2010. His research focuses on probabilistic and machine learning techniques for perception for mobile robots. From 2003-13, he was involved in national and european projects related to Advanced Driver Assistant System (ADAS) and Intelligent Vehicles, as well as mobile robotics since 2014.

## 1 Introduction

Mobile robots deployed during field tests tend to be met with mixed reactions from the humans with which they share the environment, who may accept or reject the robot through mechanisms whose details are not yet fully understood (Hebesberger et al., 2017; Mutlu & Forlizzi, 2008). On the one hand, the technical complexity and capabilities of a connected smart device and a robot are quite similar, so one may expect robots to be treated and perceived similarly to machines. On the other hand, some studies point to humans feeling some level of empathy towards robots (Menne & Schwab, 2018; Rosenthal-von der Pütten et al., 2014), and some soldiers have been burying their bomb disposal robots (Carpenter, 2013). At the same time, many researchers and companies seek to deploy mobile robots into human environments to accomplish useful tasks, but they often struggle to explain people's reactions to their robots' behavior. This could be explained by the lack of a unified theory on spatial and navigation behavior in interactions, as pointed out in a recent Social Navigation (SN) survey (C. I. Mavrogiannis et al., 2021) suggesting this lack may be due to the vastness of the domain and fragmented literature focusing on specific sub-problems and variables. This results in an inability to determine how a robot will be perceived given the *combination* of the variables modulating its navigation.

Social Navigation works tend to perform experiments where participants interact with a robot exhibiting a state of the art, complex robot behavior resulting from navigation algorithms combined with the rest of the robot's physical and software design (Carton et al., 2017; Gil et al., 2021; Kamezaki et al., 2019). Questionnaires and interviews allow researchers to perform an *evaluation* of the complex behavior in terms of its acceptability and ability to accomplish a specific task optimally (C. Mavrogiannis et al., 2019). This **top-down** approach enables the evaluation of a given algorithm in a given context, however it makes it difficult to determine which specific aspects of the robot's navigation were responsible for each aspect of the evaluation.

Other works in the field of HRI have explored other interaction modalities such as voice (McGinn & Torre, 2019), gestures (Augustine et al., 2020; Saldien et al., 2014; Zhou & Dragan, 2018), positioning (Brandl et al., 2016), as well as a few works on navigation parameters such as (Saerbeck & Bartneck, 2010). These studies have allowed researchers to determine that each of these modalities play a role in HRI, however most of them consider each modality separately, which makes it difficult to determine whether there are interactions between them when combined into a human's perception of a robot. Studies that consider *combinations* of modalities typically do not include navigation as one of them (Dautenhahn et al., 2009).

One of the goals of our work is to gain an understanding of the holistic perception mechanisms by which humans perceive the navigation of mobile agents (humans, animals, or robots) by building a model linking a robot's navigation and appearance variables to the resulting human perceptions. Such a model could be beneficial in several ways. Firstly, in order to properly integrate robots into their roles and tasks we need to be able to control how they interact with people and how they are perceived. Secondly, studying the underlying mechanisms of HRI could help us to better understand human perception of mobile agents in general. In some sense, robots can be used as a tool to enable systematic and repeatable experiments to explore whether there are common principles in how humans perceive other humans, animals, or robots.

In order to build this model, we take inspiration from prior works by Aubergé et al. on vocal prosody. When considering the vocal interaction modality, most of the communication channel's bandwidth is used in order to convey semantic meaning. But even when using exactly the same words, we can still modify *how* we say them by using the remaining bandwidth and degrees of freedom of the vocal signal such as changing pitch, rhythm, tone, or vocal effort. Prior studies (Sasa & Aubergé, 2017; Tsvetanova et al., 2017) have aimed to study and characterize vocal prosody, and to understand its role in interactions and relations not only among humans, but also between humans and robots.

We aim to expand on this by considering dimensions related to a robot's navigation, or in other words, spatio-temporal interaction. Similarly to the vocal modality, a large part of the degrees of freedom available for navigation are used in order to perform a practical navigation task (e.g. going from point A to point B, or following a person), but there are still ways to modify *how* the task is performed by changing the velocities and accelerations, the shape of velocity profiles, or by adding pauses and saccades. Prosody has been explored in HRI through studies on emotional music prosody (Savery et al., 2021) to generate emotionally expressive non-verbal audio, as well as gestures in (Savery et al., 2019). The influence on perceptions of a robot using voice and hand-over gestures with varying prosody was also studied in (Di Cesare et al., 2017), indicating that this concept may carry over to other modalities. Therefore, in addition to navigation factors, we also consider other dimensions that intervene in the perceptual experience of being near a mobile robot by studying visual appearance factors (presence and shape of eyes, head position, and stability of the robot base) as well as auditive factors through the presence or absence of motor noise. The combination of all of these is encompassed in our use of the term "prosody of movement", describing *how* a navigation motion is performed.

We aim to characterize the prosody of movement by trying to discover and formalize the navigation and appearance primitives which are involved in human perception of this prosody. Here, "primitive" is not used in the sense of "simple", but rather in the sense of a building-block or a fundamental notion upon which more elaborate concepts are built. This is a **bottom-up** approach, where we start by *analyzing* each of the basic fundamental parameters of the robot's behavior, instead of the top-down *evaluation* of a robot behavior generated by a complex state of the art navigation algorithm.

In order to undertake the task of understanding movement prosody, we design a corpus of robot motions which covers a large scope of feasible motions. We then film a mobile robot performing a large number of movement trajectories with varying movement prosody and appearance parameters selected according to the motion corpus design. The parameters of the motion corpus are hypothesized to be responsible for eliciting reactions in people, who tend to interpret robot motions in terms of intentions or attitudes. We also believe that certain motion parameters are analogous to vocal prosody parameters. The first goal of the corpus construction is to systematically categorize and distinguish types of movement. The second goal is to build and provide open access to a novel video corpus, which can be re-used by other researchers to conduct their own studies. The third goal is to design experiments using the robot motions as the stimuli in order to assess their impact on people's perceptions of the robot.

Our hypothesis is that people's attributions of socio-affects and physical attributes to a robot depend on its movement prosody. In this work, we take the following steps to explore this idea:

- we build a novel robot motion corpus encompassing variables derived from the mechanical abilities of the robot, its appearance, and analogies to prior works in vocal prosody (see Table 3);
- we publish and detail the construction of a video corpus representing the combinations of variables of our motion corpus [1];
- we present ten perceptual scales (see Table 1) used to evaluate people's perceptions of robot socio-affects and physical attributes, derived from prior studies on the role of vocal prosody in the construction of relations (Guillaume et al., 2015; Sasa & Aubergé, 2016);
- we use our motion corpus and perceptual scales in three perception experiments, and present preliminary results suggesting that several of the motion corpus variables have significant influences on the perceptual scales.

## 2 Related works

In this section, we present similar works in the field of HRI and Social Navigation which aim to study the impact of navigation and appearance variables on people. Firstly, we discuss which navigation variables have been studied and how they differ from ours. Secondly, we discuss the various approaches used for evaluating the impact on people's perceptions of robots, and describe the basis of our perceptual scales. Thirdly, we discuss the holistic nature of human interactions and perception, motivating our choice to include both navigation and appearance variables in our corpus. Lastly, we motivate the decision to design a corpus of reference motions and videos.

### 2.1 Exploring the variables of social navigation

In this section, we discuss how prior works select navigation parameters to explore, and how the motions are generated. The methods of evaluating the navigation's impact on people and their perceptions of the robot are addressed in the next subsection.

The first method is to develop a full Social Navigation algorithm, either based on machine-learning methods which aim to imitate human navigation (Chen et al., 2017; Ramirez et al., 2016), or by implementing models taken from social psychology such as the Social Force Model (Shiomi et al., 2014). Spatial and proximity factors are the most commonly addressed in earlier works (Rios-Martinez et al., 2015), often being derived from the concept of proxemics (Hall et al., 1968). The algorithm can then be used to control a real or simulated robot, in order to conduct experiments. In (Honour et al., 2021) participants viewed top-down animations of robot trajectories

[1] A video showing examples of the corpus videos can be viewed here https://youtu.be/EiH8o1PjlOw. The full corpus and supplementary materials can be found on our project page https://osf.io/5csrg/

generated via a Socially Aware Navigation (SAN) planner and a traditional planner, in order to compare them. This methodology enables the evaluation of one algorithm, or comparisons between algorithms, but it becomes difficult to determine how each aspect of the trajectories impacts the overall HRI and perception of the robot by humans, regardless of what kind of questionnaire is used.

In (C. Mavrogiannis et al., 2019), participants performed a navigation task while sharing the workspace with a mobile robot using one of three navigation methods (two algorithms, one teleoperated). The authors computed metrics on the robot's trajectories: average acceleration, average energy (defined as the integral of the squared velocity), minimum robot-human distance, path irregularity, efficiency, and topological complexity. Describing the navigation resulting from applying each algorithm may help to understand which variables are important, and what their impact is, but these metrics are relatively global values describing an average measure of the trajectory as a whole. When considering the inverse problem of how to generate a motion that induces a specific perception of the robot, this becomes an issue since the metrics might not uniquely control all of the robot's degrees of freedom.

Other works such as (Sorrentino et al., 2021) take an existing algorithm and alter some of its controllable variables. In this work, participants walked across a room, passing by a robot moving in the opposite direction. The robot avoided the collision using three different minimum obstacle distances and maximal velocities. This facilitates the understanding of the impact of a given parameter which varies systematically within this experiment, but comparisons with other algorithms will still be difficult if there are interaction effects between the controllable variables and non-controllable variables.

Lastly, some works opt to employ hand-crafted trajectories based on a small number of variables, such as curvature and acceleration (Saerbeck & Bartneck, 2010) or acceleration styles (Schulz et al., 2020), tested in basic navigation scenarios. This approach tends to provide a clearer idea of the impact of a given navigation variable, since only the variables of interest are directly manipulated. However each study only deals with one or two variables, once again lacking the power to thoroughly explore interactions between variables.

We take a similar approach by designing our hand-crafted trajectories based on several variables which are systematically combined in order to cover the space of possible trajectories, given our robot's mechanical constraints and abilities. Another principle that guides our corpus design is to select motions in order to be able to study the movement analogs to the body and voice prosody dynamics which have been shown by (Campbell & Mokhtari, 2003; Gobl & Ní Chasaide, 2003; Sasa & Aubergé, 2017; Tsvetanova et al., 2017) to be the support for social interaction dimensions. We use velocity; acceleration; combinations of acceleration, constant, and deceleration phases; as well as smooth, incremental, or saccadic accelerations in order to generate a set of velocity profiles (see Sect.3 for details). These hand-crafted approaches have the benefit of making the trajectory features explicit, and allowing analysis of interactions between variables if they are all combined together when generating trajectories. The drawback is that designing a navigation algorithm capable of generating such trajectories in dynamic environments becomes harder as one adds more variables.

We believe that formulating the social navigation problem in a modular optimization framework similarly to (Khambhaita & Alami, 2020) would enable exploration of the space of possible robot navigation styles though the cost functions and constraints, capturing different metrics and variables involved in people's perception of the robot. Combining this type of modular framework with perception and HRI experiments like ours could be a promising way of studying navigation variables' impact on people's perceptions, as well as designing algorithms that are able to implement navigation styles which generate a given perception of the robot.

### 2.2 Evaluating perceptions of robot socio-affects

Social Navigation studies that deal with evaluating user's perceptions of robots employ a variety of methods, one of which is to use established questionnaires such as the Godspeed Questionnaire Series (GQS) (Bartneck et al., 2009) in (C. Mavrogiannis et al., 2019; Sorrentino et al., 2021), Negative Attitudes towards Robots Scale (NARS) (Nomura et al., 2006), Perceived Social Intelligence scale (PSI) (Barchard et al., 2020) in (Honour et al., 2021), or Robot Social Attitudes Scale (RoSAS) (Carpinella et al., 2017). The items on such scales are often derived from existing theories and paradigms in social sciences.

In contrast, we base our selection of items on adjectives originating from previous studies investigating human vocal and gesture prosody generated during interactions with a robot (Guillaume et al., 2015; Sasa & Aubergé, 2016). The adjectives used in our scales are derived from the participant's self-annotations of their own interaction data. Each scale opposes two adjectives, some related to the physical impression of the motion, others related to perceptions of intentions or attitudes (see Table 1). The adjectives represent vernacular terms that a person may use in their everyday life, as opposed to terms derived from a scientific theory with a specific interpretation within its field. In a sense, these adjectives on the perceptual scales are the tools we are giving the participants in order for them to be able to describe the impression they have of the robot. This is inspired by the impressionistic paradigm, which has been used in prior works to study associations across modalities such as between sounds and shapes in the "kiki, bouba" experiment (Drumm, 2012), or between vocal prosody and colors in one of the series of works by Sagisaka et al. (Watanabe et al., 2014). The experiments were conducted in French, hence the adjectives have been translated to their closest vernacular equivalent in English.

Some SN works focus on the efficiency of the method (often applied to navigating through dense crowds) rather than its impact on people's perceptions of robots (see (C. I. Mavrogiannis et al., 2021) for a recent survey). Other works evaluate the impact of their navigation algorithms using concepts such as acceptability, naturalness, comfort, likability or human-likeness (Kruse et al., 2013). These concepts are important in HRI, however they are general concepts that we

| | | | |
|---|---|---|---|
| Aggressive<br>*Agressif* | Gentle<br>*Doux* | Sturdy<br>*Solide* | Frail<br>*Fragile* |
| Authoritative<br>*Autoritaire* | Polite<br>*Poli* | Strong<br>*Fort* | Weak<br>*Faible* |
| Seems Confident<br>*A l'air confiant en lui-même* | Doubtful, Hesitant<br>*Doute, Hesitant* | Smooth<br>*Lisse* | Abrupt<br>*Rude* |
| Inspires confidence<br>*Inspire confiance* | Doesn't inspire confidence<br>*N'inspire pas confiance* | Rigid<br>*Rigide* | Supple<br>*Souple* |
| Nice<br>*Sympathique* | Disagreeable<br>*Antipathique* | Tender<br>*Tendre* | Insensitive<br>*Insensible* |

**Table 1**
*Perceptual scales (original french wording in italic)*

believe may depend on more specific perceptions such as those we propose to study. Instead of a given style of navigation being inherently acceptable or unacceptable, we explore how a style induces perceptions of robot socio-affects (attitudes, personality), which may explain subsequent judgments on whether the navigation is acceptable within a given task and human-robot relation.

### 2.3 Interactions between HRI modalities

Multimodal Human-Robot Interaction studies which consider the impact of the robot on people such as (Dautenhahn et al., 2009; Saldien et al., 2014; Zecca et al., 2008) have yet to include locomotion and navigation variables. Our aim is to contribute to show that navigation is *intrinsically* social and is an ever-present and unavoidable part of Human-Robot Interaction. In (Chan et al., 2021), the authors designed multi-modal expressive behaviors for a small Cozmo robot, aimed at expressing emotions in the arousal and valence space. The eyes, digger arm, and locomotion modalities were used to design the behaviors. Participants viewed video recordings of the behaviors online, and rated them according to their perception of whether the behavior was energetic and pleasant. This study does integrate several modalities, including locomotion, and evaluates their impact on people's perceptions of the robot. However the dimensions along which these perceptions are expressed differ from ours, and the *locomotion's primary functionality is to express the desired emotion.* In contrast, we study situations where the robot's primary functionality is to perform a navigation task such as moving from point A to point B *while* generating a certain impression as a secondary functionality. Combining stylized motion with task execution was studied in (Zhou & Dragan, 2018), however their focus was on robot arms instead of navigation, and they explored a limited set of styles (happy, sad, and hesitant).

Humans tend to perceive things in a holistic manner, with several studies having found interactions between perception modalities such as audio and shape cues in (Magnani et al., 2017), or speech and lip movement (McGurk effect) in (McGurk & MacDonald, 1976). We believe the social navigation and HRI communities would benefit from holistically considering the appearance of the robot alongside its navigation style. By appearance, we refer to the robot's general aspect (mechanical, bio-evoking, human, animal, cultural references), as well as elements such as its size, color, texture, and structural appearance (sturdy vs. frail). In our motion corpus, we focus on the navigation variables for a given robot (RobAIR "RobAIR mobile robot, designed and built by FabMSTIC, Grenoble," 2021), on which we also have the ability to vary the frail-sturdy appearance dimension, the shape of its eyes, and orientation of its head. Therefore, these appearance variables are also included in the corpus and combined with the navigation variables (see Section 4 for details).

### 2.4 Robot motion and video corpus

In this subsection, we discuss the motivations for building a corpus of motions, and the accompanying video corpus. Designing a corpus of reference motions can help to avoid a common pitfall in Social Navigation which is the dependence on a specific robot platform or navigation algorithm, which makes comparisons between works difficult. To the best of our knowledge, this corpus is the first of its kind. Other researchers may choose to implement the same motions on their own robotic platform, which could help to further study the influence of different robot platforms on HRI. Furthermore, the video corpus allows researchers to conduct studies using exactly the same stimuli, which could help reproducibility of HRI studies (Irfan et al., 2018), and provide insights into the degree of cultural differences in HRI by running studies with participants from different countries. Our video corpus represents a wide range of robot motions, as well as various robot characteristics and motions which are not directly related to navigation. This allows us to study the relations between various perceptual stimuli and also to avoid ceiling effects, where one aspect of the robot's motion or appearance may dominate or nullify the effects of other aspects. The robot is shown moving from left to right in a straight line, using different combinations of acceleration, constant speed and deceleration phases. We take inspiration from previous studies about vocal interactions by choosing to differentiate three different energy levels with which the robot performs each motion. In the next section, we detail how we designed the motions contained in the corpus.

## 3 Robot motion corpus design

In this section, we present the seven variables of our robot motion corpus. In the first subsection, we describe the three variables used to define velocity profiles i.e. curves giving the velocity of the mobile robot over time. In the second subsection, we describe the four variables related to the visual appearance and audio aspects of the mobile robot.

## 3.1 Velocity profile design

The primary goal of the corpus is to enable the study of a robot's motion and kinematics parameters' impact on people. *Velocity profiles* specify the robot's movement, and are built by combining the values of three corpus variables: the *motion sequence*, *kinematics type* and *variant*. The motion sequence determines what we could call the general "shape" of the motion, in terms of speeding up, slowing down, maintaining speed. The kinematics type determines how abruptly the changes in velocity occur, and how fast the robot moves. These dimensions are related to the amount of kinetic energy required to perform the motion, hence the kinematics types represent different energy levels. The variant determines the fine details of the robot's motion, in order to add certain characteristics to it.

In the following subsections, we present the variables which define the velocity profiles, and discuss the factors which had to be accounted for in the design process. These factors arose through five successive cycles of implementing and filming various motions on the robot and testing them in order to determine how the motions looked to a human bystander and on video.

### *3.1.1 Motion sequences*

A *motion sequence* is a succession of motion phases, which can be acceleration phases, constant velocity phases, or deceleration phases. This essentially describes the type of motion performed by the robot. There are several choices for how to design the motion in each phase, mostly in terms of the shape of the velocity curve over time, which could be linear, exponential, logarithmic, sigmoidal, or other types of curves. The impact of using a given type of curve on people's perceptions of the robot has very rarely been studied (see (Schulz et al., 2020) for a comparison of linear and *slow-in, slow-out* velocity profiles). Ideally, we would compare the effect of different curve types, however in this study we limit ourselves to linear curves. We chose to use linear curves since they are the simplest type of curve, both to implement and to analyze.

The six motion sequences are illustrated in Fig. 1. Within the space of all feasible kinds of motions, we aimed to identify the basic building blocks that are representative of most mobile robot motions. The building blocks are:

1. accelerating from a standstill;
2. decelerating to a halt;
3. accelerating, constant velocity, then decelerating;
4. accelerating from a slower velocity;
5. decelerating to a slower velocity;

6. decelerating, constant velocity, then accelerating;
7. accelerating then immediately decelerating;
8. decelerating then immediately accelerating;
9. accelerating from a standstill, then constant velocity;
10. constant velocity, then decelerating to a halt.

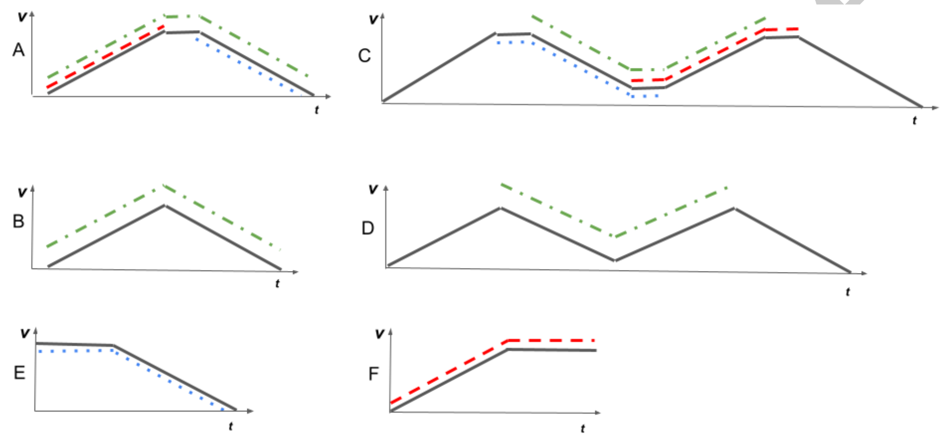


**Figure 1**
*Illustrations of the six motion sequences (solid lines), and ten building blocks (dashed lines).*

Once we had established the building blocks, we determined six motion sequences (A, B, C, D, E, F) which contain the building blocks. Motion sequences A and C are designed to introduce a notion of "pausing" between acceleration and deceleration phases, by inserting small plateaus of constant velocity. The length of these plateaus (300$ms$) was chosen as an analogy to other communication aspects such as duration of a syllable, or of a sign in sign language. We chose a value within this order of magnitude while also making sure that it was long enough to be perceptible when viewing the robot's motion. Motion sequences C and D incorporate a low-velocity phase half-way through the sequence, which may evoke a form of hesitation, as studied in the area of gestures in HRI (Moon et al., 2013).

Of course, choices have to be made regarding the duration of the acceleration and deceleration phases, which we will detail in the following part.

### *3.1.2 Kinematic types*

We use the term *kinematic type* to refer to the bounds which are set on the robot's velocity and acceleration. Velocity and variations of velocity over time (acceleration) are the most basic descriptors of movement, that still allow us to capture a wide enough range of robot motion. Jerk (the time-derivative of acceleration) could be interesting to study, but it is a more advanced notion, and is typically non-trivial to account for in most current social navigation algorithms, so we do not explicitly control or study it in our work. The relevance of acceleration as a factor which influences people's perceptions of robots was stated in a related work (Saerbeck & Bartneck, 2010), which studied the impact of acceleration and curvature on people's attributions of emotions and affects to two different robots.

Our goal was to choose three sets of values of the velocities ($v_{min}$ and $v_{max}$) and accelerations $a$ for the three values of the kinematic type variable. Given these kinematics types are intended to correspond to low, medium or high-energy motion styles, we chose to pair velocities and accelerations in a coherent way. For example, if we choose a *low* acceleration value, we combine it with *slow* $v_{min}$ and $v_{max}$. Pairing a *low* acceleration value with a *fast* $v_{max}$ could be feasible, but it introduces more subtleties when trying to compare two kinematics types in terms of the energy required to perform them.

The decision to use three kinematics types was made firstly for practical reasons, given that using more would imply a much larger corpus. Secondly, in this first study we aim to capture the extreme cases, which should provide us with the most contrast. The low and high energy types are the extreme values, and the medium type serves as a reference point between the extreme values.

In order to select the exact values to specify each kinematics type, there are several aspects one needs to balance and compromise on. Firstly, we have to account for the robot's physical capabilities. Naively using the robot's maximal velocity ($0.8m.s^{-1}$) and acceleration ($2.6m.s^{-2}$) as the high-energy type means the robot reaches its maximal speed within just $0.3s$, far too short to be properly perceived. Other aspects we had to balance were:

1. robot's physical motor limits;
2. minimum perceivable duration of acceleration;
3. duration of the whole motion (impact of exposure time of a stimuli on people's perception);
4. maintain similar duration of acceleration phase;
5. distance traveled (camera field of view and room limitations);
6. distinctness of minimal and maximal velocities;
7. distinctness of kinematic types.

Given all the constraints above, we selected an acceleration time from $v_{min}$ to $v_{max}$ of $1.0s$, so the acceleration phases to or from a zero velocity were between $1.25s$ and $1.5s$, depending on the kinematics type. This gives the motion sufficient length to be properly perceived by the viewer, while also allowing for a clear difference between the three kinematics types. The final values describing the kinematics types are shown in Table 2.

| **Parameter** | Low | Medium | High |
|---|---|---|---|
| $a$ | $0.2m.s^{-2}$ | $0.35m.s^{-2}$ | $0.5m.s^{-2}$ |
| $v_{min}$ | $0.05m.s^{-1}$ | $0.15m.s^{-1}$ | $0.25m.s^{-1}$ |
| $v_{max}$ | $0.25m.s^{-1}$ | $0.50m.s^{-1}$ | $0.75m.s^{-1}$ |
| 0 to $v_{max}$ | $1.25s$ | $1.42s$ | $1.5s$ |
| $v_{min}$ to $v_{max}$ | $1.0s$ | $1.0s$ | $1.0s$ |

**Table 2**
*Kinematics types parameters*

#### *3.1.3 Profile Variants*

We designed two *variants* (incremental and saccade, illustrated in Fig.2), which modify the shape of the velocity profile locally. Once again, the design process for these variants has its roots in analogies to speech production variations which can be used intentionally by people, or originate from health issues or as a consequence of physical characteristics. When people manipulate their speech, it can become a means for them to alter the way other people perceive them.

In this study we aim to evoke two speech variations, hesitant speech and jittery speech. Hesitant speech could be observed when one is unsure of oneself, or not very confident, whereas jittery speech could be associated with frailness due to old age or health issues. Similar characteristics can be observed in people's gestures, and locomotion, which we aim to reproduce.

The *saccade* variant is an attempt to imitate jitter by introducing continuous shaking or stuttering of the robot by rapidly increasing and decreasing the velocity commands to the motors. There were two main conflicting aspects that had to be balanced: on the one hand, the resemblance to the type of dynamics observed in speech or human motion; and on the other hand, the reproducibility of the motion. Applying random perturbations to the velocity profile results in slightly more "natural" jitter, but it also means the motion sequence and kinematics type may become unrecognizable. In order to combat this effect, we perturbed the velocity commands in a deterministic, periodic fashion. Essentially, we determine a period and amplitude of the oscillations of the perturbed velocity profile around the value of the original profile. In our case, the period is $0.2s$, and the amplitude is fixed to different values according to the kinematics type (low: $0.044m.s^{-1}$, medium: $0.090m.s^{-1}$, high: $0.120m.s^{-1}$).

The *incremental* variant is an attempt to imitate hesitation by introducing increments into the acceleration and deceleration phases, meaning that an acceleration phase which is typically a single, constant acceleration applied for one second becomes a succession of three acceleration phases of a third of a second, interleaved by two constant velocity plateaus of $300ms$. These plateaus are a means of conveying pauses in the motion, in a similar fashion to motion sequences A and C, hence the same duration being used. Regarding the number of plateaus, we chose to use two plateaus simply because it allows each acceleration phase to be long enough to be perceptible, while provoking a big enough difference when compared to the original velocity profile.

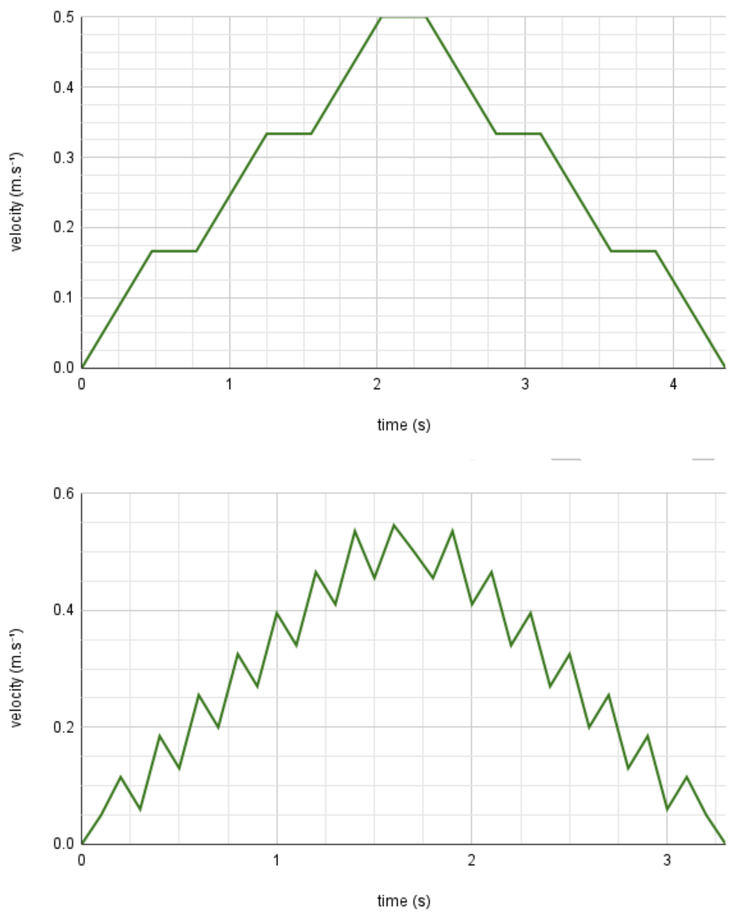


**Figure 2**

*Velocity profiles resulting from applying the incremental (top) and saccade (bottom) variants to motion sequence A, with medium kinematics type.*

In this subsection, we presented the corpus variables which define the velocity of the robot during its movement. In the next subsection, we present the corpus variables which define the audio and visual aspects of the robot.

### 3.2 Beyond Velocity: Robot Appearance and body Dynamics

Although our aim is to study the impact of motion parameters on human's perceptions, we need to keep in mind that other visual factors, as well as the sounds produced by the robot could also have an impact.

People communicate and perceive things holistically, by using and considering several modalities of expression. Therefore, if some variables other than the robot's kinematics also affect people's perceptions (which seems to be the case already for some of them studied independently like gaze or appearance), then their effect will be present in the results. This could become an issue if a variable has a *ceiling effect*, which would make it difficult or impossible to observe the effects of the motion variables we are interested in. We don't know if there are ceiling effects, because velocity, acceleration, hesitations, saccades, eyes, head motion, profile type, stability, and sound have not been studied together yet. Hence, we also manipulate other variables related to the robot's appearance, body motion, and sound to be able to detect and minimize any kind of bias or ceiling effect impacting people's perceptions.

### *3.2.1 Frail or Robust robot*

Typically, robots are designed with robustness in mind, which leads to robots which are stable, with well-mounted parts that do not shake even when the robot is under strain. The apparent physical stability and sturdiness of the mobile robot is an aspect which is rarely investigated, yet it may have an impact on how people perceive the robot's motion. We used two different robots to film our corpus, one typical stable robot and one modified, unstable robot. We modified an existing robot by loosening the front and back balancing wheel assemblies, and by loosely mounting its head on its body. The result is that the whole robot sways back and forth when changing speeds, especially when using the high-energy kinematics type, and its head shakes when using the saccade variant. This gives the robot body a different style of movement dynamics.

### *3.2.2 Eye shape and head movements*

When robot designers include characteristics associated to living beings in their robot, it can impact how much people tend to anthropomorphize it, or how people interpret or perceive the robot's actions. Additionally, the exact shape of the eyes can also convey meaning. We use three eye variants for this corpus: switched off, round, and squinting. The round eye shape is part of the robot's design after a study where it was rated as the most "neutral" eye shape. The squinting eye shape was designed for this corpus in order to convey a colder, more unsettling feeling.

Gaze is a relevant means of interaction for living beings, and studies such as (Breazeal et al., 2005) have also shown it has an important role in HRI, due to its relation to attention, and its implicit signaling of the robot's perception capabilities. Gaze direction can also be tied to navigation and has been used to make motion more legible, as was studied in (Fischer et al., 2016). For these reasons, we complemented the eye shape variants with four head settings: two settings where the head is stationary (facing straight, or facing the side, towards the camera) and two where the head moves during the robot's motion (from the straight position to the side, and vice-versa).

#### *3.2.3 Audio recording*

Most motors used in mobile robot locomotion are noisy, so any variation in the robot's motion also carries an audio signal. In some cases, navigation and control parameters do not cause a significant visual difference, but the change is still clearly audible through the motor noise. Other than motor noise, the chassis and other parts of the robot can also produce sounds which can give information about the structure of the robot. If the robot makes creaking and knocking sounds whenever it moves a bit too suddenly, we might deduce that the robot is not very well built. These types of sounds are called *consequential sounds* and have recently been the object of studies in HRI such as (Tennent et al., 2017) or (Robinson et al., 2021). In both studies, different sounds lead participants to perceive the robot's motion differently, highlighting the necessity of taking sound into account even when studying other dimensions of HRI.

The exposure to the sound produced by the mobile robot can therefore convey information, or be interpreted in various ways by the viewers, even if these sounds are simply direct consequences of the physical properties of the robot. Recording the sound of the mobile robot was therefore necessary, and could be used to contrast people's perception of the same motion when played back with or without the sound.

### 3.3 Summary of corpus variables

In Table 3, we summarize the variables we manipulated in order to obtain each video of the corpus, as well as the different values they can take. The audio variable is adjusted as a post-processing step: we record a given stimuli with sound, duplicate the video and mute one of them to obtain the silent version. In practice, we were unable to include all combinations of the values of these variables due to several limiting factors. Firstly, the stable robot's head was unable to rotate, meaning it was only filmed using the *straight* head setting. Secondly, due to time constraints on the corpus acquisition, we decided to remove certain combinations of variables: the saccades and incremental variants were not combined with round or squinting eye shapes. The resulting corpus contains 450 videos, for a total of 900 combinations of values of the seven corpus variables once we include the audio variable.

In the next subsection, we present the steps taken to ensure that we captured the robot's motion and appearance as faithfully as possible.

## 4 Video corpus acquisition

In the previous subsection, we detailed the design of the robot's motions and appearance variables which constitute the corpus. In this subsection, we detail the considerations and precautions we took in order to produce a high-quality, exploitable corpus of videos. Various prior works have used video-based stimuli for HRI experiments involving moving robots (Carton et al., 2017; Chan et al., 2021; Knight et al., 2016; Torre et al., 2021), some of which validated their results on subsequent

| **Motion sequence** | **Kin. type** | **Profile variant** | **Robot type** | **Eye shape** | **Head setting** | **Audio** |
|---|---|---|---|---|---|---|
| A | low | none | stable | none | straight | with |
| B | medium | saccades | unstable | round | side | without |
| C | high | increment | | squint | straight to side | |
| D | | | | | side to straight | |
| E | | | | | | |
| F | | | | | | |

The first row lists the corpus variable names. Each column lists the values that each variable can take.

**Table 3**
*Robot motion corpus variables.*



in-person experiments (Moon et al., 2013; Reinhardt et al., 2021). A recent study compared video and in-person experimental settings in the context of gestures with similar results, suggesting video studies may be appropriate (Honig & Oron-Gilad, 2020). Some of the variables explored in our corpus result in very slight visual differences in the robot's motion, such as the high-frequency stuttering and shaking induced by the saccade variant, or the swaying of the whole robot body when using the unstable base. For these reasons, we took extra precautions in order to capture the robot's movements as precisely as possible, and to make sure that they are well represented in the videos. We also aim to avoid any differences in the recording conditions that could introduce unwanted biases. All along the motion corpus design and recording, we consulted two experts (a professional videographer and a photographer) to discuss which parameters should be controlled in order to capture the robot's motion and appearance as precisely and truthfully as possible.

## 4.1 Robot Movement Consistency and Framing

In order to minimize differences between two videos showing the same velocity profile, we implemented a method which allows us to execute a selected velocity profile automatically on the mobile robot. This also reduces the chance for errors during the corpus filming which is already a tedious and time-consuming process. In addition, the whole control stack from the velocity profile control code down to the low-level motor controller was analyzed and modified when necessary in order to ensure the robot's motion was as faithful as possible to the velocity profile.

Regarding the camera framing, filming the robot moving towards the camera would give the impression of moving towards the viewer. In the end, we decided against it due to the unclear effect of the lack of depth perception resulting from the use of non-stereo video. Filming a robot motion parallel with the image plane should conserve as much information about the robot's motion as possible, unlike some prior video-based studies which required motion with components which are perpendicular to the image plane (Carton et al., 2017; Knight et al., 2016; Torre et al., 2021). The robot's initial position was also considered, as it may have a priming effect on people's anticipation and interpretation of the robot's motion. For example, if the robot starts very far on the right hand side of the frame, facing right, people could assume the robot will not travel very fast or far. In order to mitigate this, the robot starting point was selected such that the total motion to be performed was centered in the camera's framing. The exact position depends on the motion sequence, kinematics type, and variants; so in the interest of time, the initial positions were approximate and the motions were not always exactly centered.

## 4.2 Environment Characteristics

Regarding the background, it is necessary to make sure that it is *mostly* uniform in order to avoid visual distractions, although when trying to capture movement it can help to have reference points such as vertical lines in order to better perceive the velocity of the robot. The background color should provide a high contrast with the robot's color. The type of ground on which the robot moves should also be considered in combination with the robot's drive assembly, given that any discontinuities in the ground could have repercussions on the robot's motion, and hence, visual appearance. Naturally, there should also be enough space in the environment to perform the longest velocity profile (six meters in our case). We also made sure there were no visible obstacles in the robot's direction of travel, since one could anticipate that the robot will start to slow down before reaching the obstacle.

The experts also highlighted the importance of lighting conditions to get the clearest possible picture of the robot, which is dependent both on the natural and artificial lighting of the room. In our case, strong natural light provided better lighting conditions than indoor artificial lighting, although this meant camera parameters had to be adjusted to compensate for the changing light throughout the day. Good lighting allows the details to be visible, and helps to clearly distinguish background from foreground.

## 4.3 Camera configuration and parameters

Initial filming tests revealed that smartphones are limited in terms of the field of view, and action cameras cause too much distortion with their fish-eye lenses. The experts informed us that the high-framerate recording action cameras provide is also not necessary for our application. We also raised the question of shutter speed, which is tied to the amount of motion blur in an image, but given the speeds of the robot we were dealing with, standard shutter speed settings would suffice. The most important parameters were:

1. lighting conditions to avoid shadows;

2. high resolution to capture details;

3. stability of the camera.

One of the experts performed the final video recordings using a high-quality camera (Canon EOS 5D Mark IV) and advised us on the final framing, positioning, and lighting conditions. The camera and appropriate lenses and settings allowed us to frame wide enough to capture the full movements, while maintaining a good size of the robot in the frame, and high quality capture.

During the filming sessions, routine checks were made to ensure a consistent image over all 450 videos, despite the varying lighting conditions. One adjustment had to be made to the exposure settings of the camera in order for the robot's LED eyes to be clearly distinguishable. The resulting setting (under-exposure) was a compromise between image quality and visibility of the eyes. The exposure also had to be adjusted to compensate for the changing lighting conditions. We provide additional details regarding the camera, lenses, settings, and subsequent video post-processing, encoding, and formats as part of the motion corpus.

## 5 Perception Experiments

In order to analyze the effect of each variable of the corpus on the way in which a mobile robot is perceived we ran a series of three experiments; two online experiments where participants viewed videos of the robot and one embodied experiment, where the robot moved towards the participant, stopping at a pre-determined distance.

The goal of the online experiments was to collect participant's perceptions of the robot for the whole corpus, in order to establish a first baseline regarding which variables had significant influences on how participants perceived the robot. The goal of the embodied experiment was to attempt to replicate the findings of the online experiments for a subset of the corpus containing the variables which were found to be the most influential. All three experiments use similar methodologies.

### 5.1 First online experiment: likert scale

The corpus of 900 stimuli was split into groups of 45 videos within which each value of each variable was represented. Thus, each participant viewed and rated all 45 videos of a given group, meaning they would see all values of all variables several times, but not all combinations. The order of the videos was randomized for each participant, and the number of participants for each video group was roughly balanced. Participants viewed each video once before rating it and moving on to the next. After viewing all of the videos, participants filled in a post-questionnaire, which we do not report on in this paper due to space constraints. At the end of the experiment, participants could choose to leave a free-form comment. In the first online experiment, the rating scales were presented to participants as 5-point likert scales, where the middle option signifies that neither of the adjectives correspond to their perception of the robot. The 900 stimuli were split into 20 groups, half with sound, and half without, meaning the sound variable was a between-subjects variable.

#### *5.1.1 Likert scale results*

A total of $n = 42$ participants completed the first online perception experiment. Participants of all ages were recruited via university mailing lists, experiment recruiting lists, and social media. The first step in our analysis was to determine whether certain corpus variables or certain scales showed wide ranges of responses, or clear-cut bias to one of the opposing adjectives. For each value of each corpus variable, we computed the distribution of participants responses on all videos using that value. With 42 participants, each of the 900 unique combinations of values of variables is only rated four times at most. However in the initial analysis presented in this paper, we do not study unique combinations of values, but rather all combinations that include a given value for a given variable such as all videos using the high kinematics type. In this arrangement, there are at least 300 ratings of videos using each value of each variable of the corpus. The responses followed normal distributions, generally with high mass around the center value as seen in Fig.3. For some scales, different values for certain corpus variables led to distributions that were shifted towards one of the adjectives. For all scales, a high percentage of the responses were on the neutral level (neither one nor the other end of a scale), with half of the scales having 30 to 40% of neutral responses, and the other half having 40 to 45% of neutral responses. Distributions for participants who had sound all showed lower proportions of responses on the neutral level than those without sound.

We performed chi-square association tests to determine if there were significant dependencies between each corpus variable and scale pairing. The resulting chi-square statistic and significance levels are reported in Table 4, with a large proportion of the pairings showing significant dependence.

#### *5.1.2 Likert scale analysis*

The results of the association tests suggest that people's perceptions of a mobile robot along these ten perceptual scales are dependent on several of the motion corpus variables, most notably the kinematics and variant variables. The absence of sound leading to more neutral responses could indicate that the sound was informative and affected people's responses.

The high proportion of neutral responses across most scales and variables could indicate one of two situations:

1. neutral perception: the participant finds neither of the adjectives fitting to describe their perception of the robot;

2. uncertainty: the participant feels unsure of their answer, and prefers to give a neutral response rather than answer in a way that *they perceive* as random.

This was reflected in several participants' comments, stating that they felt like it was difficult to answer, or that they answered randomly. While the association tests did report statistical significance of the dependencies, further investigation into the

| | | Aggressive (-) Gentle (+) | | | | | Authoritative (-) Polite (+) | | | | | Condifent (-) Doubtful (+) | | | | | Inspires conf. (-) Doesn't insp. (+) | | | | | Kind (-) Disagreeable (+) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 |
| Kinematics | high | 12 | 25 | 42 | 17 | 4 | 10 | 26 | 43 | 18 | 3 | 18 | 29 | 23 | 19 | 10 | 6 | 22 | 34 | 23 | 15 | 6 | 19 | 46 | 20 | 9 |
| | medium | 3 | 15 | 45 | 26 | 11 | 3 | 15 | 47 | 27 | 8 | 7 | 25 | 30 | 24 | 14 | 5 | 25 | 36 | 20 | 14 | 7 | 25 | 43 | 16 | 8 |
| | low | 2 | 8 | 39 | 30 | 21 | 2 | 8 | 44 | 30 | 16 | 6 | 19 | 28 | 27 | 20 | 7 | 21 | 34 | 24 | 14 | 7 | 28 | 42 | 16 | 7 |
| Motion Sequence | A | 3 | 16 | 47 | 24 | 10 | 4 | 15 | 49 | 24 | 8 | 9 | 28 | 30 | 24 | 10 | 5 | 26 | 35 | 21 | 13 | 5 | 25 | 46 | 17 | 7 |
| | B | 6 | 18 | 40 | 24 | 13 | 4 | 16 | 47 | 24 | 9 | 6 | 25 | 30 | 22 | 16 | 4 | 22 | 36 | 23 | 15 | 8 | 22 | 43 | 18 | 9 |
| | C | 5 | 15 | 45 | 23 | 12 | 3 | 16 | 45 | 25 | 10 | 6 | 20 | 23 | 27 | 23 | 6 | 20 | 35 | 24 | 15 | 6 | 28 | 40 | 19 | 7 |
| | D | 9 | 16 | 41 | 22 | 12 | 5 | 14 | 47 | 27 | 8 | 5 | 12 | 24 | 33 | 27 | 3 | 14 | 30 | 31 | 22 | 6 | 19 | 41 | 22 | 12 |
| | E | 6 | 15 | 41 | 26 | 13 | 7 | 18 | 41 | 25 | 9 | 22 | 31 | 28 | 14 | 5 | 8 | 31 | 36 | 16 | 9 | 8 | 25 | 48 | 11 | 8 |
| | F | 5 | 16 | 39 | 27 | 13 | 8 | 19 | 38 | 25 | 10 | 15 | 30 | 28 | 20 | 8 | 9 | 23 | 35 | 21 | 12 | 7 | 24 | 44 | 18 | 6 |
| Variant | none | 5 | 17 | 40 | 24 | 14 | 6 | 17 | 41 | 26 | 10 | 14 | 30 | 30 | 19 | 7 | 9 | 27 | 35 | 19 | 10 | 8 | 25 | 42 | 16 | 8 |
| | saccade | 10 | 19 | 43 | 22 | 6 | 5 | 17 | 54 | 20 | 5 | 4 | 13 | 20 | 30 | 33 | 1 | 10 | 31 | 27 | 31 | 4 | 17 | 46 | 21 | 12 |
| | increment | 4 | 11 | 46 | 27 | 12 | 3 | 13 | 46 | 28 | 9 | 7 | 18 | 25 | 30 | 20 | 3 | 22 | 36 | 28 | 12 | 4 | 26 | 47 | 17 | 5 |
| Eyes | none | 6 | 15 | 45 | 24 | 11 | 4 | 15 | 49 | 25 | 8 | 7 | 19 | 26 | 27 | 20 | 4 | 20 | 35 | 24 | 17 | 4 | 22 | 49 | 17 | 8 |
| | round | 4 | 11 | 36 | 31 | 18 | 5 | 13 | 37 | 32 | 13 | 13 | 33 | 29 | 20 | 5 | 11 | 33 | 32 | 16 | 7 | 16 | 33 | 34 | 12 | 5 |
| | squint | 7 | 24 | 40 | 19 | 11 | 9 | 24 | 39 | 20 | 8 | 17 | 32 | 27 | 16 | 7 | 8 | 21 | 35 | 23 | 13 | 5 | 21 | 37 | 25 | 12 |
| Head | straight | 6 | 14 | 44 | 24 | 13 | 6 | 16 | 45 | 25 | 9 | 14 | 23 | 27 | 23 | 13 | 7 | 24 | 37 | 18 | 13 | 6 | 24 | 47 | 15 | 9 |
| | side | 5 | 15 | 40 | 25 | 14 | 4 | 16 | 42 | 27 | 10 | 8 | 24 | 25 | 23 | 19 | 4 | 24 | 31 | 26 | 15 | 7 | 27 | 42 | 17 | 7 |
| | side to straight | 4 | 18 | 44 | 24 | 10 | 4 | 18 | 46 | 25 | 7 | 6 | 26 | 25 | 27 | 16 | 4 | 20 | 36 | 26 | 14 | 6 | 23 | 42 | 22 | 6 |
| | straight to side | 7 | 19 | 39 | 23 | 12 | 7 | 17 | 43 | 24 | 10 | 9 | 27 | 30 | 20 | 14 | 7 | 21 | 32 | 24 | 17 | 10 | 21 | 39 | 19 | 10 |
| Base | unstable | 5 | 16 | 42 | 25 | 12 | 5 | 16 | 45 | 26 | 8 | 8 | 24 | 26 | 24 | 17 | 5 | 21 | 34 | 25 | 16 | 7 | 24 | 43 | 18 | 8 |
| | stable | 6 | 15 | 43 | 22 | 15 | 6 | 18 | 43 | 22 | 11 | 19 | 26 | 30 | 19 | 7 | 11 | 28 | 38 | 14 | 9 | 7 | 23 | 47 | 14 | 9 |
| Sound | sound | 7 | 19 | 34 | 24 | 16 | 6 | 19 | 37 | 27 | 12 | 11 | 27 | 18 | 24 | 20 | 8 | 23 | 27 | 23 | 19 | 7 | 24 | 39 | 20 | 10 |
| | no-sound | 4 | 13 | 51 | 24 | 8 | 4 | 13 | 53 | 24 | 6 | 9 | 22 | 37 | 22 | 9 | 4 | 22 | 43 | 21 | 9 | 6 | 24 | 49 | 15 | 6 |

| | | Sturdy (-) Frail (+) | | | | | Strong (-) Weak (+) | | | | | Smooth (-) Abrupt (+) | | | | | Rigid (-) Supple(+) | | | | | Tender (-) Insensitive (+) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 | -2 | -1 | 0 | 1 | 2 |
| Kinematics | high | 12 | 31 | 27 | 20 | 9 | 13 | 34 | 34 | 15 | 5 | 14 | 20 | 37 | 22 | 8 | 19 | 36 | 33 | 10 | 2 | 2 | 13 | 47 | 19 | 19 |
| | medium | 6 | 26 | 33 | 23 | 12 | 6 | 26 | 40 | 21 | 7 | 13 | 25 | 42 | 15 | 5 | 14 | 33 | 39 | 12 | 3 | 1 | 14 | 57 | 15 | 13 |
| | low | 5 | 24 | 34 | 28 | 10 | 3 | 21 | 38 | 27 | 10 | 15 | 27 | 44 | 13 | 1 | 11 | 26 | 43 | 16 | 4 | 3 | 22 | 48 | 13 | 13 |
| Motion Sequence | A | 6 | 30 | 31 | 24 | 9 | 7 | 28 | 37 | 21 | 7 | 13 | 25 | 45 | 16 | 2 | 16 | 28 | 42 | 12 | 2 | 1 | 19 | 49 | 17 | 15 |
| | B | 5 | 26 | 35 | 23 | 11 | 6 | 28 | 39 | 22 | 6 | 13 | 26 | 40 | 17 | 4 | 13 | 34 | 38 | 13 | 1 | 3 | 15 | 51 | 17 | 14 |
| | C | 4 | 27 | 29 | 28 | 12 | 4 | 26 | 38 | 21 | 10 | 14 | 22 | 41 | 18 | 5 | 12 | 33 | 39 | 12 | 4 | 2 | 20 | 48 | 15 | 16 |
| | D | 5 | 17 | 31 | 30 | 18 | 4 | 19 | 35 | 29 | 13 | 12 | 17 | 41 | 20 | 10 | 16 | 33 | 36 | 12 | 3 | 2 | 15 | 51 | 18 | 14 |
| | E | 15 | 33 | 29 | 16 | 7 | 13 | 33 | 36 | 13 | 5 | 17 | 27 | 40 | 12 | 4 | 16 | 32 | 35 | 15 | 2 | 3 | 15 | 49 | 14 | 18 |
| | F | 10 | 30 | 34 | 20 | 6 | 11 | 30 | 37 | 18 | 4 | 14 | 27 | 39 | 16 | 4 | 16 | 30 | 38 | 14 | 3 | 3 | 16 | 54 | 16 | 12 |
| Variant | none | 10 | 34 | 36 | 17 | 4 | 9 | 33 | 39 | 15 | 3 | 15 | 28 | 41 | 13 | 3 | 13 | 32 | 39 | 14 | 3 | 3 | 18 | 50 | 15 | 14 |
| | saccade | 3 | 13 | 21 | 34 | 29 | 3 | 15 | 29 | 33 | 19 | 11 | 14 | 39 | 24 | 12 | 20 | 33 | 34 | 10 | 3 | 1 | 12 | 49 | 21 | 17 |
| | increment | 6 | 21 | 29 | 33 | 10 | 7 | 21 | 39 | 25 | 8 | 14 | 23 | 41 | 20 | 2 | 14 | 30 | 40 | 14 | 2 | 1 | 16 | 52 | 15 | 15 |
| Eyes | none | 6 | 22 | 29 | 28 | 14 | 6 | 22 | 36 | 25 | 11 | 13 | 21 | 41 | 19 | 6 | 16 | 31 | 38 | 12 | 3 | 1 | 15 | 52 | 16 | 16 |
| | round | 9 | 34 | 36 | 17 | 5 | 9 | 30 | 42 | 16 | 3 | 16 | 30 | 40 | 12 | 2 | 12 | 29 | 39 | 16 | 3 | 5 | 26 | 49 | 9 | 10 |
| | squint | 11 | 35 | 35 | 16 | 4 | 10 | 39 | 36 | 12 | 3 | 13 | 27 | 40 | 14 | 6 | 14 | 36 | 36 | 12 | 2 | 2 | 13 | 47 | 21 | 17 |
| Head | straight | 11 | 28 | 31 | 21 | 8 | 10 | 27 | 37 | 20 | 6 | 15 | 26 | 40 | 15 | 4 | 15 | 32 | 38 | 13 | 3 | 2 | 14 | 52 | 16 | 16 |
| | side | 5 | 24 | 31 | 26 | 14 | 6 | 26 | 37 | 21 | 9 | 15 | 25 | 41 | 15 | 5 | 13 | 31 | 40 | 13 | 2 | 1 | 20 | 55 | 12 | 12 |
| | side to straight | 4 | 28 | 34 | 24 | 10 | 5 | 27 | 38 | 22 | 8 | 12 | 23 | 41 | 19 | 5 | 14 | 31 | 40 | 13 | 2 | 2 | 15 | 49 | 19 | 15 |
| | straight to side | 7 | 27 | 31 | 24 | 11 | 7 | 29 | 37 | 20 | 7 | 14 | 20 | 41 | 19 | 5 | 18 | 32 | 34 | 13 | 3 | 2 | 20 | 45 | 16 | 16 |
| Base | unstable | 6 | 25 | 30 | 26 | 12 | 6 | 26 | 37 | 23 | 9 | 13 | 23 | 41 | 17 | 5 | 14 | 32 | 38 | 13 | 2 | 2 | 17 | 51 | 15 | 15 |
| | stable | 15 | 34 | 37 | 12 | 2 | 12 | 34 | 39 | 12 | 3 | 17 | 28 | 38 | 14 | 3 | 16 | 31 | 37 | 13 | 12 | 2 | 14 | 50 | 19 | 15 |
| Sound | sound | 9 | 30 | 21 | 27 | 13 | 8 | 29 | 31 | 23 | 9 | 17 | 25 | 32 | 20 | 7 | 19 | 35 | 30 | 15 | 2 | 2 | 13 | 49 | 19 | 18 |
| | no-sound | 6 | 25 | 43 | 19 | 7 | 7 | 25 | 44 | 18 | 5 | 11 | 23 | 51 | 12 | 3 | 10 | 28 | 47 | 11 | 4 | 3 | 21 | 52 | 13 | 11 |

<= 6.3% | <= 10% | < 15% | > 30% | >= 40%

**Figure 3**

*Response distributions in percentages for the likert scale online experiment. Columns represent response levels for each perceptual scale. Rows represent corpus variable values by which the video ratings are grouped to compute the percentage of responses.*

| | Aggressive Gentle | Authoritative Polite | Confident Doubtful | Inspires Conf. Does not | Nice Disagreeable |
|---|---|---|---|---|---|
| Kinematics(2) | 235 *** | 198 *** | 106 *** | n.s. | 19 * |
| Sequence(5) | n.s. | n.s. | 221 *** | 78 *** | 33 * |
| Variant(2) | 47 *** | 33 *** | 246 *** | 178 *** | 37 *** |
| Eyes(2) | 53 *** | 69 *** | 145 *** | 76 *** | 118 *** |
| Head(3) | n.s. | n.s. | 39 *** | 28 ** | 27 ** |
| Base(1) | n.s. | n.s. | 55 *** | 47 *** | n.s. |
| Sound(1) | 68 *** | 60 *** | 107 *** | 76 *** | 25 *** |

| | Sturdy Frail | Strong Weak | Smooth Abrupt | Rigid Supple | Tender Insensitive |
|---|---|---|---|---|---|
| Kinematics(2) | 50 *** | 96 *** | 59 *** | 48 *** | 44 *** |
| Sequence(5) | 102 *** | 90 *** | 49 *** | n.s. | n.s. |
| Variant(2) | 306 *** | 203 *** | 97 *** | n.s. | 24 ** |
| Eyes(2) | 113 *** | 107 *** | 33 *** | n.s. | 75 *** |
| Head(3) | 35 *** | n.s. | n.s. | n.s. | 24 * |
| Base(1) | 104 *** | 53 *** | 12 * | n.s. | n.s. |
| Sound(1) | 115 *** | 46 *** | 86 *** | 82 *** | 47 *** |

**Table 4**

*Chi-square association test results between corpus variables and perceptual scales, for the first online experiment. Reported as chi-square statistic, significance* $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.



underlying cause of the neutral responses could give a better idea of the robustness of the associations. We designed the second perception experiment in order to explore this phenomena.

### 5.2 Second online experiment: binary choice

The second online experiment uses the same stimuli with only slight differences in methodology which we detail in this subsection. In this experiment, the rating scales were presented to participants as simple binary choices between the two adjectives. In perception tests involving subtle differences in stimuli, participants may feel like they are answering at random, hence some could choose to respond with the neutral value on the first experiment. If participants have consistent inclinations towards certain perceptions for given variables, then forcing them to pick a side of the scale would result in higher mass of responses on one end of the scale. However if the perception of participants is truly not impacted by the variable, the responses may become more random, and the mass equally spread to either side of the scale. We chose to use only stimuli with sound since they were more informative, reducing the total number of videos to 450, split into 10 groups.

#### *5.2.1 Binary choice results*

A total of $n = 65$ participants completed the second online perception experiment. Participants of all ages were recruited via university mailing lists, experiment recruiting lists, and social media. When observing the distributions of responses, the trends of the first experiment tended to be confirmed. Distributions with an existing bias towards one side of a scale were shifted further towards that side, and distributions with little to no bias remained similar (see Fig. 4[2]).

As in the first experiment, we used chi-square association tests to study dependencies between corpus variables and perceptual scales. Results are reported in Table 5. The significant dependencies are similar to the first experiment, with just nine out of sixty dependencies being significant in only one of the two experiments. The kinematics, profile, variant and eyes variables have an impact on at least eight of the ten scales, compared to only six and two for the base and head variables respectively. Certain scales are dependent only on a few of the corpus variables such as the Aggressive-Gentle, Rigid-Supple, and Tender-Insensitive scales which only depend on the kinematics, variant, and eyes variables. The three most significant dependencies are Aggressive-Gentle on kinematics, Sturdy-Frail on variant, Authoritative-Polite on kinematics, and Confident-Doubtful on sequence.

In order to establish the relation between each *value* of the corpus variables and response levels on the perceptual scales, we computed the standardized Pearson residuals for each of the chi-square tests. For a given pairing of a variable value and a perceptual scale level, the residual's absolute value indicates whether the difference is statistically significant, and its sign indicates whether there are more or fewer responses than expected. In Table 6 we show only associations between corpus variable values and adjectives having standardized residuals greater than 2, indicating that the association is significant (Sharpe, 2015). The head variable is omitted from the table since there were no significant dependencies between its values and the scales.

#### *5.2.2 Binary choice analysis*

Once again, several participants commented about the difficulty in responding due to videos seeming similar or even identical, and reported having answered at random. Despite this, trends in participant response distributions were confirmed in the second experiment, which suggests that participants mostly had a consistent inclination towards certain sides of the scales when presented with similar videos.

When considering the influence of each corpus variable value (see Table6 ), some variables stand out. The medium kinematics type does not influence any of the

---

[2] This table is available on our project page https://osf.io/5csrg/ which will be updated as the project continues.

| | | Aggressive (-) Gentle (+) | | Authoritative (-) Polite (+) | | Condifent (-) Doubtful (+) | | Inspires conf. (-) Doesn't insp. (+) | | Kind (-) Disagreeable (+) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | - | + | - | + | - | + | - | + | - | + |
| **Kinematics** | high | 59 | 41 | 57 | 43 | 65 | 35 | 42 | 58 | 47 | 53 |
| | medium | 32 | 68 | 34 | 66 | 50 | 50 | 44 | 56 | 56 | 44 |
| | low | 14 | 86 | 18 | 82 | 41 | 59 | 51 | 49 | 68 | 32 |
| **Motion Sequence** | A | 37 | 63 | 36 | 64 | 55 | 45 | 47 | 53 | 53 | 47 |
| | B | 32 | 68 | 28 | 72 | 49 | 51 | 49 | 51 | 61 | 39 |
| | C | 34 | 66 | 36 | 64 | 35 | 65 | 37 | 63 | 56 | 44 |
| | D | 38 | 62 | 35 | 65 | 32 | 68 | 37 | 63 | 57 | 43 |
| | E | 39 | 61 | 46 | 54 | 74 | 26 | 49 | 51 | 53 | 47 |
| | F | 33 | 67 | 39 | 61 | 68 | 32 | 56 | 44 | 61 | 39 |
| **Variant** | none | 34 | 66 | 38 | 62 | 65 | 35 | 54 | 46 | 59 | 41 |
| | saccade | 43 | 57 | 38 | 62 | 25 | 75 | 24 | 76 | 50 | 50 |
| | increment | 33 | 67 | 32 | 68 | 43 | 57 | 44 | 56 | 57 | 43 |
| **Eyes** | none | 37 | 63 | 34 | 66 | 43 | 57 | 39 | 61 | 55 | 45 |
| | round | 26 | 74 | 36 | 64 | 65 | 35 | 60 | 40 | 65 | 35 |
| | squint | 40 | 60 | 44 | 56 | 68 | 32 | 52 | 48 | 55 | 45 |
| **Head** | straight | 36 | 64 | 38 | 63 | 53 | 47 | 47 | 53 | 57 | 43 |
| | side | 37 | 63 | 34 | 66 | 52 | 48 | 44 | 56 | 56 | 44 |
| | side to straight | 34 | 66 | 36 | 64 | 51 | 49 | 45 | 55 | 59 | 41 |
| | straight to side | 35 | 65 | 38 | 62 | 54 | 46 | 47 | 53 | 55 | 45 |
| **Base** | unstable | 36 | 64 | 36 | 64 | 50 | 50 | 44 | 56 | 57 | 43 |
| | stable | 34 | 66 | 40 | 60 | 60 | 40 | 52 | 48 | 58 | 42 |

| | | Sturdy (-) Frail (+) | | Strong (-) Weak (+) | | Smooth (-) Abrupt (+) | | Rigid (-) Supple(+) | | Tender (-) Insensitive (+) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | - | + | - | + | - | + | - | + | - | + |
| **Kinematics** | high | 64 | 36 | 65 | 35 | 55 | 45 | 74 | 26 | 34 | 66 |
| | medium | 52 | 48 | 50 | 50 | 63 | 38 | 66 | 34 | 44 | 56 |
| | low | 47 | 53 | 40 | 60 | 77 | 23 | 55 | 45 | 55 | 45 |
| **Motion Sequence** | A | 58 | 42 | 54 | 46 | 71 | 29 | 63 | 37 | 43 | 57 |
| | B | 49 | 51 | 46 | 54 | 64 | 36 | 60 | 40 | 50 | 50 |
| | C | 47 | 53 | 44 | 56 | 63 | 37 | 64 | 36 | 42 | 58 |
| | D | 41 | 59 | 40 | 60 | 60 | 40 | 69 | 31 | 46 | 54 |
| | E | 69 | 31 | 66 | 34 | 63 | 37 | 68 | 32 | 40 | 60 |
| | F | 63 | 37 | 60 | 40 | 70 | 30 | 64 | 36 | 46 | 54 |
| **Variant** | none | 68 | 32 | 63 | 37 | 71 | 29 | 60 | 40 | 47 | 53 |
| | saccade | 21 | 79 | 26 | 74 | 49 | 51 | 77 | 23 | 36 | 64 |
| | increment | 45 | 55 | 42 | 58 | 63 | 37 | 65 | 35 | 45 | 55 |
| **Eyes** | none | 45 | 55 | 43 | 57 | 61 | 39 | 68 | 32 | 42 | 58 |
| | round | 66 | 34 | 62 | 38 | 76 | 24 | 57 | 43 | 58 | 42 |
| | squint | 71 | 29 | 67 | 33 | 68 | 32 | 64 | 36 | 39 | 61 |
| **Head** | straight | 58 | 42 | 54 | 46 | 67 | 33 | 64 | 36 | 44 | 56 |
| | side | 50 | 50 | 50 | 50 | 63 | 37 | 64 | 36 | 45 | 55 |
| | side to straight | 55 | 45 | 53 | 47 | 66 | 34 | 64 | 36 | 44 | 56 |
| | straight to side | 51 | 49 | 48 | 52 | 63 | 37 | 67 | 33 | 44 | 56 |
| **Base** | unstable | 51 | 49 | 49 | 51 | 64 | 36 | 65 | 35 | 45 | 55 |
| | stable | 70 | 30 | 62 | 38 | 71 | 29 | 64 | 35 | 43 | 57 |

| <= 20% | <= 30% | < 40% | > 60% | >= 70% | >= 80% |
|---|---|---|---|---|---|

**Figure 4**

*Response distributions in percentages for the binary choice online experiment. Columns represent response levels for each perceptual scale. Rows represent corpus variable values by which the video ratings are grouped to compute the percentage of responses.*

| | Aggressive Gentle | Authoritative Polite | Confident Doubtful | Inspires Conf Does not | Nice Disagreeable |
|---|---|---|---|---|---|
| Kinematics(2) | 467 *** | 298 *** | 94 *** | 21 *** | 95 *** |
| Sequence(5) | n.s. | 30 *** | 262 *** | 51 *** | 14 * |
| Variant(2) | 45 *** | 17 *** | 267 *** | 177 *** | 31 *** |
| Eyes(2) | 37 *** | 20 *** | 145 *** | 97 *** | 46 *** |
| Head(3) | n.s. | n.s. | n.s. | n.s. | n.s. |
| Base(1) | n.s. | 5 * | 35 *** | 10 ** | n.s. |

| | Sturdy Frail | Strong Weak | Smooth Abrupt | Rigid Supple | Tender Insensitive |
|---|---|---|---|---|---|
| Kinematics(2) | 66 *** | 138 *** | 106 *** | 83 *** | 72 *** |
| Sequence(5) | 111 *** | 91 *** | 13 * | n.s. | n.s. |
| Variant(2) | 344 *** | 212 *** | 90 *** | 57 *** | 20 *** |
| Eyes(2) | 136 *** | 100 *** | 48 *** | 30 *** | 64 *** |
| Head(3) | 14 ** | 9 * | n.s. | n.s. | n.s. |
| Base(1) | 69 *** | 32 *** | 5 * | n.s. | n.s. |

**Table 5**

*Chi-square association test results between corpus variables and perceptual scales, for the second online experiment. Reported as chi-square statistic, significance* $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

scales, whereas the low and high types are associated with opposite ends of every scale except Inspires-Doesn't inspire confidence. This could suggest that the medium kinematics type acts as a reference point for the other two.

### 5.3 Embodied Experiment

The goal of the in-person experiment was to attempt to replicate the online results in an embodied interaction. Participants were informed of the general goal of the study prior to the experiment. Participants were asked to stand at a fixed position facing the robot as it moved towards them using one of the combinations of variables. The distance at which a mobile robot stops when approaching a person has been investigated, usually to determine what people consider as an acceptable distance (Brandl et al., 2016). In order to control for potential effects of different stopping distances, we ensured that the robot would always stop at a distance of $50cm$ from the person by using the same hard-coded velocity profiles as were used to film the video corpus.

We chose to focus on three variables, using two values for each: the kinematics type (low/high), variant(smooth/saccade), and head position(straight/side). The

| | Aggressive<br>Gentle | Authoritative<br>Polite | Confident<br>Doubtful | Inspires Conf<br>Does not | Nice<br>Disagreeable |
|---|---|---|---|---|---|
| Kin. low | gentle | polite | doubtful | inspires conf | nice |
| Kin. medium | n.s | n.s | n.s | n.s | n.s |
| Kin. high | aggressive | authoritative | confident | n.s | disagreeable |
| Sequence A | n.s | n.s | n.s | n.s | n.s |
| Sequence B | n.s | n.s | n.s | n.s | n.s |
| Sequence C | n.s | n.s | doubtful | n.s | n.s |
| Sequence D | n.s | n.s | doubtful | does not | n.s |
| Sequence E | n.s | authoritative | confident | n.s | n.s |
| Sequence F | n.s | n.s | confident | inspires conf | n.s |
| Var. none | n.s | n.s | confident | inspires conf | n.s |
| Var. saccade | aggressive | authoritative | doubtful | does not | disagreeable |
| Var. increment | n.s | n.s | doubtful | n.s | n.s |
| Eyes none | n.s | n.s | doubtful | does not | disagreeable |
| Eyes round | gentle | n.s | confident | inspires conf | nice |
| Eyes squint | n.s | authoritative | confident | n.s | n.s |
| Unstable | n.s | n.s | n.s | n.s | n.s |
| Stable | n.s | n.s | confident | inspires conf | n.s |
| | Sturdy<br>Frail | Strong<br>Weak | Smooth<br>Abrupt | Rigid<br>Supple | Tender<br>Insensitive |
| Kin. low | frail | weak | smooth | supple | tender |
| Kin. medium | n.s | n.s | n.s | n.s | n.s |
| Kin. high | sturdy | strong | abrupt | rigid | insensitive |
| Sequence A | n.s | n.s | n.s | n.s | n.s |
| Sequence B | frail | n.s | n.s | n.s | n.s |
| Sequence C | frail | weak | n.s | n.s | n.s |
| Sequence D | frail | weak | n.s | n.s | n.s |
| Sequence E | sturdy | strong | n.s | n.s | n.s |
| Sequence F | sturdy | strong | n.s | n.s | n.s |
| Var. none | sturdy | strong | smooth | supple | n.s |
| Var. saccade | frail | weak | abrupt | rigid | insensitive |
| Var. increment | frail | weak | n.s | n.s | n.s |
| Eyes none | frail | weak | abrupt | n.s | n.s |
| Eyes round | sturdy | strong | smooth | supple | tender |
| Eyes squint | sturdy | strong | n.s | n.s | n.s |
| Unstable | frail | n.s | n.s | n.s | n.s |
| Stable | sturdy | strong | n.s | n.s | n.s |

**Table 6**
*Relative influence of corpus variable values on the perceptual scales*

kinematics and variant variables were selected since they had the most influence on participant's responses in the second online experiment. The head position variable was selected since we aimed to test the hypothesis that gaze would be more influential in an embodied interaction. Each participant saw all combinations of values, resulting in 8 stimuli. The base variable was set to unstable, in order to make the saccade variant more visible; eyes were set to round in order to better indicate gaze direction, and the motion sequence was set to one of the shorter sequences (A) since it induced the least variability on the distance to the participant upon stopping ($\pm 5cm$). In addition, sequence A was the only one to have no significant impact on any of the scales in the second online experiment (see Table 6).

### *5.3.1 Embodied experiment results*

A total of $n = 22$ participants completed the embodied experiment, consisting of students recruited from our university, half of which were in the field of computer science and applied mathematics.

As in the online experiments, we performed chi-square association tests, reported in Table 7. Associations were found between most of the scales and the kinematics and variant variables. As in the second online experiment, no association was found between the head variable and any of the scales. Unlike in the second online experiment, no association was found between variant and Authoritative-Polite, nor between kinematics and Sturdy-Frail.

| | Aggressive Gentle | Authoritative Polite | Confident Doubtful | Inspires Conf Does not | Nice Disagreeable |
|---|---|---|---|---|---|
| Kinematics(1) | 87 *** | 74 *** | 18 *** | 8 ** | 47 *** |
| Variant(1) | 4 * | n.s. | 57 *** | 25 *** | 4 * |
| Head(1) | n.s. | n.s. | n.s. | n.s. | n.s. |

| | Sturdy Frail | Strong Weak | Smooth Abrupt | Rigid Supple | Tender Insensitive |
|---|---|---|---|---|---|
| Kinematics(1) | n.s. | 28 *** | 22 *** | 18 *** | 31 *** |
| Variant(1) | 51 *** | 19 *** | 35 *** | 4 * | 4 * |
| Head(1) | n.s. | n.s. | n.s. | n.s. | n.s. |

**Table 7**

*Chi-square association test results between a subset of corpus variables and all perceptual scales for the embodied, in-person experiment. Reported as chi-square statistic, significance* $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

#### *5.3.2 Embodied experiment analysis*

While the results of the embodied experiment are not directly comparable to the online experiments, the general trends seem to be confirmed. The hypothesis that gaze would have an influence in the in-person experiment is not supported.

# 6 Discussion

## 6.1 Limitations

Firstly, a limitation of this preliminary statistical analysis is that interactions between variables were not considered. As such, these results only indicate relative impact of different values for a given corpus variable, not how the robot will be perceived overall given a full specification of the corpus variables.

Secondly, our corpus videos show a robot moving in an empty environment without any interaction with a person, whereas one could include a form of interaction with a person, given we are studying HRI. This decision was taken since our first goal is to isolate the physical navigation primitives from other factors that impact interaction. We hypothesize that in addition to the physical properties of motion, the *relation* between the robot and the person also plays a role in how a robot is perceived, and how we interact with it. If we impose a relation by framing a specific type of interaction in the experiments, it would be difficult to analyze whether a person's reaction to the robot was induced by motion and appearance primitives or by the relation.

## 6.2 Implications

Our results show that the impressionistic adjectives used in our perceptual scales are useful in aiding participants to distinguish and characterize various elements of movement prosody in robot navigation. Furthermore, the typology of variables proposed in the corpus can be used to establish socio-affective traits of expressive navigation which exhibit significant contrasts between one-another. As such, even if a given type of motion was chosen solely on the basis of practical considerations, our work suggests that it will be perceived and interpreted by humans as socio-affective expression, therefore impacting HRI. Thus, it is essential to take into account the fact that navigating intrinsically entails communicating with the human. To achieve this, we must understand and control what types of navigation profiles should be used to generate elements of interaction, whose communicative and ethical effects also require further rigorous study.

One of our perceptual scales is based on the concept of frailty, which has already been found to have significant impacts on the people interacting with a frail robot. In a prior work (Sasa & Aubergé, 2016), isolated elderly people interacted with a small butler robot by giving it voice commands. During the experiment the authors discovered by serendipity that when the robot showed signs of frailty by making mistakes (bumping into a wall while moving), participants became more attached and changed their attitude towards the robot by starting to *take care* of the robot. Similarly, in (Matsumoto, 2021) the authors compared a typical robot to a fragile robot which broke, requiring participants to fix it; finding that participants reported feeling more attached to the fragile robot, as well as finding it more pleasant, less boring, and more interesting. The interest of this is not so much that participants are attached, but rather that because they are attached, they tend to be more active in their interactions with the robot, often helping it, which may have positive effects on the person's physical and mental health when compared to passively receiving care (Takenaka, 2005; Tanaka, 1997). As such, the impact of the impressions generated by robot navigation variables seems to go far beyond the issues of user preference, usability or comfort.

### 6.3 Future work

The next step in our work is to integrate the results of these experiments into the design of a navigation algorithm for mobile robots in human environments that allows navigation tasks to be performed while generating a certain impression based on our perceptual scales. Our study showed that all of our corpus variables which define the velocity profile (kinematics type, motion sequence and variant) had significant impacts on people's perceptions of the robot, hence our navigation algorithm should allow explicit control over these variables. These variables define the strength of the acceleration, maximal velocity, smoothness of the motions, and style of acceleration and deceleration (including hesitations and pauses).

We propose to control these dimensions through time-dependent constraints on the velocity and acceleration values used when generating candidate trajectories for the robot to follow. In our experiments, only straight-line motions were performed by the mobile robot using hard-coded velocity profiles with fixed lengths. In order to handle dynamic navigation in human environments, some assumptions have to be made with respect to how one should modify a profile in order to travel different distances while generating a similar impression. One possibility is to assume that maintaining a constant velocity over different durations does not affect the person's perception, although this would require further experiments to be confirmed. Our design and implementation of such an algorithm is ongoing, and is built as an extension of our prior work (Scales et al., 2020).

Another important future work is to determine the effect of a *combination* of these variables on a person's perception of the robot, which requires further statistical analysis of the *interactions* between variables and perceptual scales. Methods such as linear mixed models (Bates et al., 2015) could be used in order to study the interactions between corpus variables, enabling a full specification of the perceptions generated by a given combination of our corpus variables. Additionally, we are running another embodied experiment where the interaction between humans and the robot takes place within and ecological context with a task to accomplish, as opposed to the basic lab study presented in this paper.

## 7 Conclusion

In this paper, we propose an incremental bottom-up approach to the understanding of how fundamental properties of appearance and navigation impact a person's perception of a mobile robot. We constructed a novel holistic robot motion corpus in order to study the impact of navigation *and* audio-visual cues on people's perceptions of robots, in contrast to the more specialized studies of previous works. The variables contained in the corpus are hypothesized to be involved in what we define as *movement prosody*, a concept we derive by analogy with vocal prosody. The corpus was used in two online perception experiments ($n = 42$, $n = 65$) and one in-person experiment ($n = 22$). Participants rated a robot performing a navigation task along ten perceptual scales opposing adjectives describing physical aspects as well as perceived intentions and attitudes of the robot, such as Frail-Sturdy, Aggressive-Gentle or Confident-Doubtful. A statistical analysis of the dependencies between each variable and scale showed that all scales had significant dependencies on several corpus variables, and most corpus variables impacted several scales. This includes variables related to the robot's navigation such as its maximal velocity, acceleration, smoothness, pauses and hesitations. These initial results show that this experimental methodology can bring some insights into people's perception of mobile robots, and more generally, how humans process cues from various modalities in order to build their perception of an agent.

## Acknowledgment

We thank Jean-Philippe Guilbaud and Djamel Hadji for their assistance in preparing and filming the video corpus. We thank Solange Rossato for the discussions on the statistical analysis.

## References

Augustine, A. C., Ryusuke, M., Liu, C., Ishi, C. T., & Ishiguro, H. (2020). Generation and evaluation of audio-visual anger emotional expression for android robot. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 96–98. https://doi.org/10.1145/3371382.3378282

Barchard, K. A., Lapping-Carr, L., Westfall, R. S., Fink-Armold, A., Banisetty, S. B., & Feil-Seifer, D. (2020). Measuring the perceived social intelligence of robots. *ACM Transactions on Human-Robot Interaction*, *9*(4). https://doi.org/10.1145/3415139

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int J Soc Robot*, *1*, 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brandl, C., Mertens, A., & Schlick, C. M. (2016). Human-Robot Interaction in Assisted Personal Services: Factors Influencing Distances That Humans Will Accept between Themselves and an Approaching Service Robot. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *26*(6), 713–727. https://doi.org/10.1002/hfm.20675

Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 708–713. https://doi.org/10.1109/IROS.2005.1545011

Campbell, N., & Mokhtari, P. (2003). Voice quality: The 4th prosodic dimension. *Proc. 15th Int. Congr. Phonetic Sciences*, pp. 2417–2420.

Carpenter, J. (2013). The Quiet Professional: An investigation of U.S. military Explosive Ordnance Disposal personnel interactions with everyday field robots. *Doctoral dissertation University of Washington*.

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation. *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 254–262.

Carton, D., Olszowy, W., Wollherr, D., & Buss, M. (2017). Socio-Contextual Constraints for Human Approach with a Mobile Robot. *International Journal of Social Robotics*, *9*(2), 309–327. https://doi.org/10.1007/s12369-016-0394-3

Chan, L., Zhang, B. J., & Fitter, N. T. (2021). Designing and validating expressive cozmo behaviors for accurately conveying emotions. *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 1037–1044. https://doi.org/10.1109/RO-MAN50785.2021.9515425

Chen, Y. F., Everett, M., Liu, M., & How, J. P. (2017). Socially aware motion planning with deep reinforcement learning. *IEEE International Conference on Intelligent Robots and Systems*, *2017-Septe*, 1343–1350. https://doi.org/10.1109/IROS.2017.8202312

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3-4), 369–397. https://doi.org/10.1080/11762320903123567

Di Cesare, G., De Stefani, E., Gentilucci, M., & De Marco, D. (2017). Vitality Forms Expressed by Others Modulate Our Own Motor Response: A Kinematic Study. *Frontiers in Human Neuroscience*, *11*, 565. https://doi.org/10.3389/fnhum.2017.00565

Drumm, P. (2012). Köhler, W. In R. W. Rieber (Ed.), *Encyclopedia of the history of psychological theories* (pp. 610–612). Springer US. https://doi.org/10.1007/978-1-4419-0463-8_153

Fischer, K., Jensen, L. C., Suvei, S. D., & Bodenhagen, L. (2016). Between legibility and contact: The role of gaze in robot approach. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*, 646–651. https://doi.org/10.1109/ROMAN.2016.7745186

Gil, Ó., Garrell, A., & Sanfeliu, A. (2021). Social robot navigation tasks: Combining machine learning techniques and social force model. *Sensors*, *21*(21). https://doi.org/10.3390/s21217087

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1-2), 189–212. https://doi.org/10.1016/S0167-6393(02)00082-1

Guillaume, L., Aubergé, V., Magnani, R., Aman, F., Cottier, C., Sasa, Y., Wolf, C., Nebout, F., Neverova, N., Bonnefond, N., Nègre, A., Tsvetanova, L., & Girard-Rivier, M. (2015). Hri in an ecological dynamic experiment: The gee corpus based approach for the emox robot. *2015 IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 1–6. https://doi.org/10.1109/ARSO.2015.7428207

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold, A. R., Durbin, M., Edmonson, M. S., Fischer, J. L., Hymes, D., Kimball, S. T., Barre, W. L., Frank Lynch, S. J., McClellan, J. E., Marshall, D. S., Milner, G. B., Sarles, H. B., Trager, G. L., & Vayda, A. P. (1968). Proxemics [and comments and replies]. *Current Anthropology*, *9*(2/3), 83–108. http://www.jstor.org/stable/2740724

Hebesberger, D., Koertner, T., Gisinger, C., & Pripfl, J. (2017). A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults. *International Journal of Social Robotics*, *9*. https://doi.org/10.1007/s12369-016-0391-6

Honig, S., & Oron-Gilad, T. (2020). Comparing laboratory user studies and video-enhancedweb surveys for eliciting user gestures in human-robot interactions. *ACM/IEEE International Conference on Human-Robot Interaction*, 248–250. https://doi.org/10.1145/3371382.3378325

Honour, A., Banisetty, S. B., & Feil-Seifer, D. (2021). Perceived Social Intelligence as Evaluation of Socially Navigation. *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 519–523. https://doi.org/10.1145/3434074.3447226

Irfan, B., Kennedy, J., Lemaignan, S., Papadopoulos, F., Senft, E., & Belpaeme, T. (2018). Social Psychology and Human-Robot Interaction: An Uneasy Marriage. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 13–20. https://doi.org/10.1145/3173386.3173389

Kamezaki, M., Kobayashi, A., Yokoyama, Y., Yanagawa, H., Shrestha, M., & Sugano, S. (2019). A Preliminary Study of Interactive Navigation Framework with Situation-Adaptive Multimodal Inducement: Pass-By Scenario. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00574-3

Khambhaita, H., & Alami, R. (2020). Viewing Robot Navigation in Human Environment as a Cooperative Activity. Springer, Cham. https://doi.org/10.1007/978-3-030-28619-4_25

Knight, H., Thielstrom, R., & Simmons, R. (2016). Expressive path shape (Swagger): Simple features that illustrate a robot's attitude toward its goal in real time. *IEEE International Conference on Intelligent Robots and Systems*, *2016-Novem*, 1475–1482. https://doi.org/10.1109/IROS.2016.7759240

Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013). Human-Aware Robot Navigation: A Survey. *Robotics and Autonomous Systems*, *61*(12), pp.1726–1743. https://hal.archives-ouvertes.fr/hal-01684295

Magnani, R., Aubergé, V., Bayol, C., & Sasa, Y. (2017). Bases of Empathic Animism Illusion: audio-visual perception of an object devoted to becoming perceived as a subject for HRI. *Proc. of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots – VIHAR 2017.*

Matsumoto, M. (2021). Fragile Robot: The Fragility of Robots Induces User Attachment to Robots. *International Journal of Mechanical Engineering and Robotics Research*, *10*(10), 536–541. https://doi.org/10.18178/ijmerr.10.10.536-541

Mavrogiannis, C., Hutchinson, A. M., MacDonald, J., Alves-Oliveira, P., & Knepper, R. A. (2019). Effects of Distinct Robot Navigation Strategies on Human Behavior in a Crowded Environment. *ACM/IEEE International Conference on Human-Robot Interaction*, *2019-March*(March), 421–430. https://doi.org/10.1109/HRI.2019.8673115

Mavrogiannis, C. I., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., & Oh, J. (2021). Core challenges of social robot navigation: A survey. *CoRR*, *abs/2103.05668*. https://arxiv.org/abs/2103.05668

McGinn, C., & Torre, I. (2019). Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots. *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, 211–221.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Menne, I. M., & Schwab, F. (2018). Faces of Emotion: Investigating Emotional Facial Expressions Towards a Robot. *International Journal of Social Robotics*, *10*(2), 199–209. https://doi.org/10.1007/s12369-017-0447-2

Moon, A., Parker, C. A. C., Croft, E. A., & Van der Loos, H. F. M. (2013). Design and Impact of Hesitation Gestures during Human-Robot Resource Conflicts. *Journal of Human-Robot Interaction*, *2*(3). https://doi.org/10.5898/jhri.2.3.moon

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations. *Proceedings of the 3rd international conference on Human robot interaction - HRI '08*, (May 2014), 287. https://doi.org/10.1145/1349822.1349860

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, *7*(3), 437–454. https://doi.org/https://doi.org/10.1075/is.7.3.14nom

Ramirez, O. A., Khambhaita, H., Chatila, R., Chetouani, M., & Alami, R. (2016). Robots learning how and where to approach people. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*, *1*, 347–353. https://doi.org/10.1109/ROMAN.2016.7745154

Reinhardt, J., Prasch, L., & Bengler, K. (2021). Back-off. *ACM Transactions on Human-Robot Interaction*, *10*(3), 1–25. https://doi.org/10.1145/3418303

Rios-Martinez, J., Spalanzani, A., & Laugier, C. (2015). From Proxemics Theory to Socially-Aware Navigation: A Survey. *International Journal of Social Robotics*, *7*(2), 137–153. https://doi.org/10.1007/s12369-014-0251-1

Robair mobile robot, designed and built by fabmstic, grenoble [Accessed: 2021-07-19]. (2021).

Robinson, F. A., Velonaki, M., & Bown, O. (2021). Smooth operator: Tuning robot perception through artificial movement sound. *ACM/IEEE International Conference on Human-Robot Interaction*, 53–62. https://doi.org/10.1145/3434073.3444658

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, *33*, 201–212. https://doi.org/10.1016/j.chb.2014.01.004

Saerbeck, M., & Bartneck, C. (2010). Perception of affect elicited by robot motion, 53–60. https://doi.org/10.1109/hri.2010.5453269

Saldien, J., Vanderborght, B., Goris, K., Van Damme, M., & Lefeber, D. (2014). A Motion System for Social and Animated Robots. *International Journal of Advanced Robotic Systems*, *11*(5), 72. https://doi.org/10.5772/58402

Sasa, Y., & Aubergé, V. (2016). Perceived isolation and elderly boundaries in eee (emoz elder-ly expressions) corpus: Appeal to communication dynamics with a socio-affectively gluing robot in a smart home. *Gerontechnology*, *15*.

Sasa, Y., & Aubergé, V. (2017). SASI: perspectives for a socio-affectively intelligent HRI dialog system. *1st Workshop on "Behavior, Emotion and Representation: Building Blocks of Interaction"*. https://hal.inria.fr/hal-01615470

Savery, R., Rose, R., & Weinberg, G. (2019). Establishing human-robot trust through music-driven robotic emotion prosody and gesture. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956386

Savery, R., Zahray, L., & Weinberg, G. (2021). Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication. *Paladyn, Journal of Behavioral Robotics*, *12*(1), 454–467. https://doi.org/doi:10.1515/pjbr-2021-0033

Scales, P., Aycard, O., & Aubergé, V. (2020). Studying navigation as a form of interaction: A design approach for social robot navigation methods. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 6965–6972. https://doi.org/10.1109/ICRA40945.2020.9197037

Schulz, T., Holthaus, P., Amirabdollahian, F., Koay, K. L., Torresen, J., & Herstad, J. (2020). Differences of Human Perceptions of a Robot Moving using Linear or Slow in, Slow out Velocity Profiles When Performing a Cleaning Task. *2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019*. https://doi.org/10.1109/RO-MAN46459.2019.8956355

Sharpe, D. (2015). Your chi-square test is statistically significant: Now what? *Practical Assessment, Research and Evaluation*, *20*, 1–10.

Shiomi, M., Zanlungo, F., Hayashi, K., & Kanda, T. (2014). Towards a Socially Acceptable Collision Avoidance for a Mobile Robot Navigating Among Pedestrians Using a Pedestrian Model. *International Journal of Social Robotics*, *6*(3), 443–455. https://doi.org/10.1007/s12369-014-0238-y

Sorrentino, A., Khalid, O., Coviello, L., Cavallo, F., & Fiorini, L. (2021). Modeling human-like robot personalities as a key to foster socially aware navigation. *2021 30th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2021*, 95–101. https://doi.org/10.1109/RO-MAN50785.2021.9515556

Takenaka, H. (2005). Loss experience and rebirth of elderly people. Seitosha Publishing.

Tanaka, K. (1997). Geratology isagoge. Nihon hyoron sha.

Tennent, H., Moore, D., Jung, M., & Ju, W. (2017). Good vibrations: How consequential sounds affect perception of robotic arms. *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, *2017-January*, 928–935. https://doi.org/10.1109/ROMAN.2017.8172414

Torre, I., Linard, A., Steen, A., Tumová, J., & Leite, I. (2021). Should robots chicken? How anthropomorphism and perceived autonomy influence trajectories in a game-theoretic problem. *ACM/IEEE International Conference on Human-Robot Interaction*, 370–379. https://doi.org/10.1145/3434073.3444687

Tsvetanova, L., Aubergé, V., & Sasa, Y. (2017). Multimodal breathiness in interaction : From breathy voice quality to global breathy "body behavior quality". *Proc. of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots – VIHAR 2017*.

Watanabe, K., Greenberg, Y., & Sagisaka, Y. (2014). Sentiment analysis of color attributes derived from vowel sound impression for multimodal expression. *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 1–5. https://doi.org/10.1109/APSIPA.2014.7041586

Zecca, M., Endo, N., Momoki, S., Itoh, K., & Takanishi, A. (2008). Design of the humanoid robot KOBIAN - preliminary analysis of facial and whole body emotion expression capabilities-. *2008 8th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2008*, 487–492. https://doi.org/10.1109/ICHR.2008.4755969

Zhou, A., & Dragan, A. D. (2018). Cost Functions for Robot Motion Style. *IEEE International Conference on Intelligent Robots and Systems*, 3632–3639. https://doi.org/10.1109/IROS.2018.8594433