Human-Robot Motion: Taking Attention into Account

Thierry Fraichard, Remi Paulin, Patrick Reignier

HAL Id: hal-00965157

https://hal.inria.fr/hal-00965157v2

Submitted on 31 Mar 2014

Human-Robot Motion: Taking Attention into Account

Thierry Fraichard, Rémi Paulin, Patrick Reignier

Human-Robot Motion: Taking Attention into Account

Thierry Fraichard*, Rémi Paulin*, Patrick Reignier†

Project-Team PRIMA

Abstract: *Mobile robot companions* are service robots that are mobile and designed to share our living space. For such robots, mobility is essential and their coexistence with humans adds new aspects to the mobility issue: the first one is to obtain appropriate motion and the second one is interaction through motion. We encapsulate these two aspects in the term *Human-Robot Motion* (HRM) with reference to Human-Robot Interaction. The long-term issue is to design robot companions whose motions, while remaining safe, are deemed appropriate from a human point of view. This is the key to the acceptance of such systems in our daily lives. The purpose of this paper is to explore how the psychological concept of *attention* can be taken into account in HRM. To that end, we build upon an existing model of attention that computes an *attention matrix* that describes how the attention of each person is distributed among the different elements, persons and objects, of the environment. Using the attention matrix, we propose the novel concept of *attention field* that can be viewed as an attention predictor. Using different case studies, we show how the attention matrix and the attention field can be used in HRM.

Key-words: Human-Robot Motion, Human-Robot Interaction, Mobile Robot Companion, Autonomous Navigation.

---

* INRIA, CNRS-LIG, Grenoble University, France.
† Grenoble INP, CNRS-LIG, Grenoble University, France.

Mouvement homme-robot: prise en compte de l'attention

Résumé :   Les *robots compagnons mobiles* sont des robots de service conçus pour partager et se déplacer dans notre espace de vie. Pour de tels robots, la mobilité est essentielle and leur coexistance avec des humans ajoute de nouveaux aspects à ce sujet de recherche: le premier est d'obtenir un mouvement approprié et le second est l'intéraction au travers du mouvement. On regroupe ces deux aspects sous le terme *Human-Robot Motion* (HRM) en référence à Human-Robot Interaction. L'objectif à long terme est la conception de robots compagnons dont le mouvement, tout en restant sans danger, est jugé approprié d'un point de vue human. Ceci est la clé de l'acceptation de tels systèmes dans notre vie quotidienne. L'objectif de ce papier est d'explorer comment le concept psychologique d'*attention* peut être prix en compte dans HRM. A cette fin, nous proposons un concept nouveau de *champ attentionnel* qui peut être vu comme un prédicteur attentionnel. Nos travaux se basent sur un modèlel existant qui quantifie l'attention humaine et fournit une *matrice d'attentionnelle* qui décrit la distribution des ressources attentionnelles de chaque personne entre les différents élements, personnes et objects de son environnement. Le calcul du *champ attentionnel* introduit découle de cette *matrice d'attentionnelle*. En considérant différents scénarios d'étude, on montre comment la matrice et le champ attentionnel(le) peuvent être utilisés en HRM.

Mots-clés :   Robots Compagnons Mobiles, Navigation Autonome, Mouvement Homme-Robot, Interaction Homme-Robot.

1 Introduction
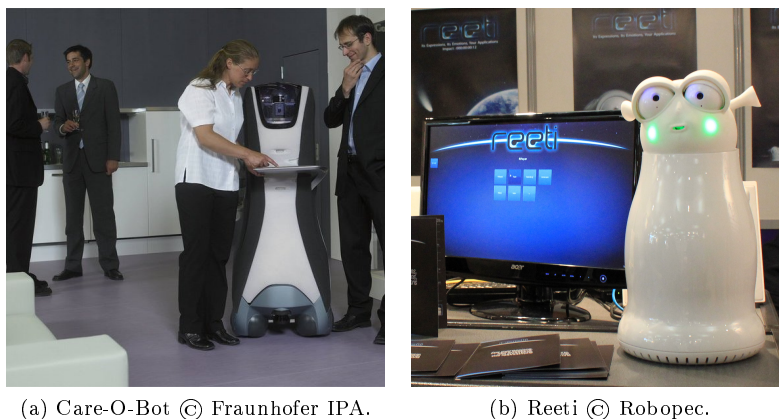
(a) Care-O-Bot © Fraunhofer IPA.   (b) Reeti © Robopec.

Figure 1: Mobile *vs* fixed robot companions.

1.1 Background and Motivations

Manufacturing robots have long dominated the robotics market but recently, we have witnessed the growth of the service robotics sector. Service robots come in a multitude of forms and their application areas are numerous, *e.g.* cleaning, inspection, rescue and security, entertainment, handicap assistance, transportation, logistics. Of particular interest to us are the service robots designed to share our living space (both professional and domestic) and to have varying degrees of interaction with us humans. Henceforth such service robots will be called **robot companions**.

Among the various types of robot companions that exist, we will focus on those similar to Care-O-Bot (Fig. 1a). Such robots are respectable in size (about the size of a person) and are designed to move in environments that can be large. Let's call them **mobile robot companions**. They differ from robot companions like Reeti that are smaller and basically static (Fig. 1b).

For mobile robot companions, **mobility** is an essential problem: they must be able to move freely in their environments. To that end, they have to address all the standard problems pertaining to autonomous motion, *e.g.* world modelling, localization, motion planning, motion control. However, the coexistence in the same environment of robots and humans adds a novel dimension to the mobility issue. For standard mobile robots, mobility usually boils down to computing motions that are both safe and optimal in some sense, *e.g.* traveled distance. However, people are not regular obstacles that can be treated like pieces of furniture. There is a set of social and cultural rules that governs how a person moves around among its peers, *e.g.* Fig. 2. Besides the motion of a person is largely influenced by the set of non-verbal cues, *e.g.* velocity, gaze direction, that can be sent by the other persons (and vice versa). In other words, motion constitutes a form of non-verbal interaction. At the end of the day, two aspects emerge vis-à-vis the mobility of a robot companion: the first one is to obtain **appropriate motion** and the second one is **interaction through motion**. We will encapsulate these two aspects in the term **Human-Robot Motion** (HRM) with reference to Human-Robot Interaction. The long-term issue is to design robot companions whose motions, while remaining safe, are deemed **appropriate** from a human point of view. This is the key to the acceptance of such systems in our daily lives.
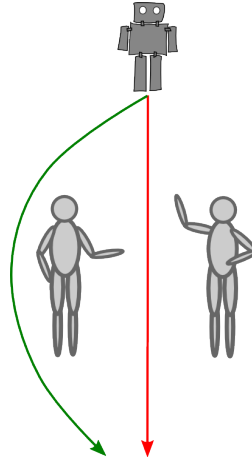
Figure 2: People are not pieces of furniture! Motion in red is definitely shorter but it is not appropriate.

1.2   Related Works and Contributions

Although mobile robots have actually shared the human living space as early as 1997 [1], it is only around 2005 that the appropriate motion issue has surfaced and that human factors have been investigated and explicitly taken into account for motion purposes [2]. The review of the literature on this topic shows that most of the approaches proposed so far rely upon the definition of so-called **social spaces**, *i.e.* regions in the environment that, for different reasons, people consider as psychologically theirs [3]. Such social spaces are primarily characterized using either the position of the person, *e.g.* "Personal space" [4], or the activity (s)he is currently engaged in, *e.g.* "Interaction Space" [5] and "Activity Space" [3]. The interaction/activity spaces are broadly defined by the convex hull of the person at hand and the people or the objects (s)he interacts with. In this framework, human detection and human activity recognition are central in the definition of said social spaces (which explains why activity recognition has grown into an important area of research especially in computer vision [6, 7]). Having characterized the social spaces corresponding to the current situation, the most common approach in HRM is then to define costmaps on such social spaces: the higher the cost, the less desirable it is for the robot to be at the corresponding position. The costmaps are ultimately used for motion planning and navigation purposes, *e.g.* [8, 9, 10].

Such approaches are obviously relevant but we believe that an analysis of the situation based solely on the activities of the persons present may be too limited. Consider for instance the situation depicted in Fig. 3. It involves two persons, one robot companion, a TV set and a phone. Let us focus on person P_1 and assume that she is currently on the phone. Besides the personal space centered around P_1, her current activity would yield an activity space in the form of the convex hull of P_1 and the phone. Both the personal space and the activity space attached to P_1 should be avoided by the others. However it is important to note that the fact that P_1 is currently on the phone does not mean that she is not paying attention to the TV set, the person P_2 or the robot R. Suppose now that the task of R is to deliver a message to P_2. If R relies on social space-based navigation only, it may very well decide to move in a way that hides the TV from P_1. Such a behaviour would not be appropriate should P_1 be actually paying attention to the TV. This simple example is meant to illustrate that there is much more to the cognitive state of a person than what is implied by their current activity and that a person is generally paying
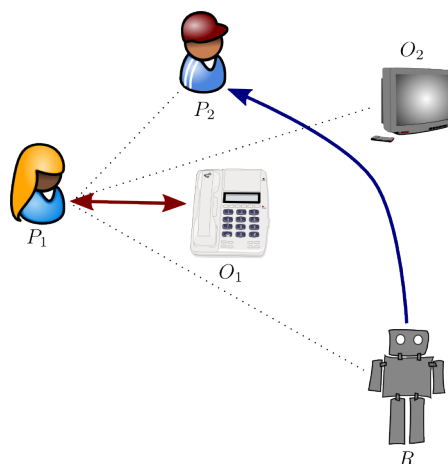
Figure 3: Attention *vs* activity: although P_1's current activity is being on the phone, part of her attention may be directed towards P_2, the TV set or the robot R. Suppose now that R moves towards P_2 in a way that hides the TV from P_1. Such a behaviour would not be appropriate should P_1 be actually paying attention to the TV.

attention to more than one element of their environment (irrespective of their current activity).

The purpose of this paper is precisely to explore how the psychological concept of attention can be taken into account in HRM in order to obtain autonomous navigation schemes for mobile robot companions that yield better appropriate motions. To that end, we build upon a computational model of attention that was earlier proposed in [11]. This attention model was initially developed in the context of ambient applications and pervasive systems. Its purpose was to estimate how attention was shared between the different users and the components of a "smart environment". The output of this model is an *attention matrix* that describes how the attention of each person is distributed among the different elements, persons and objects, of their environment. Using the attention matrix, we propose the novel concept of *attention field* that can be viewed as an attention predictor. Using different case studies, we show how the attention matrix and the attention field can be used in HRM.

1.3 Outline of the Paper

To begin with, the main concepts related to the psychological concept of attention are briefly introduced in section 2. Then, section 3 outlines the attention model originally proposed in [11] and that is used throughout the paper. Section 4 describes the implementation details of the model while section 5 illustrates via three case studies how the attention matrix and the attention field can be used in HRM.

2 Attention

Attention is one of the most intensely studied topics in psychology and cognitive neuroscience and the point of this section is not to cover all the works and theories that have been proposed over the years. Its more modest purpose is to introduce the different psychological concepts that are later used in the paper in relation to the attention model that we use.

Attention, which has also been referred to as the allocation of processing resources [12], is the cognitive process that enables humans to selectively concentrate on one or several aspects of their environment while ignoring others. It can be described as a filter, preventing us from being overwhelmed by all the surrounding sensory data. It selects only few of them to maintain the current activity and resist disruptions. This selection is constantly renegotiated. If the situation changes, attention can shift in response to new events, *e.g.* sudden noise, dimming light. Attention is generally defined as the combination of two components:

- The *endogenous* (*aka* internal) component is considered to be voluntary and based on our current activity. It is the intentional allocation of attentional resources to a predetermined element (object or person) of the environment: that is the *intention*.

- The *exogenous* (*aka* external) component is considered to be reflexive and automatic and is caused by sudden changes in the environment: that is the *distraction*.

The distraction capacity of an element is related to its *salience* (also called saliency). It is the state or quality by which it stands out of its neighbours. Salience and distraction should always be interpreted through one's sensory capabilities. A loud noise for instance will not distract a deaf person.

As mentioned above, attention is the combination of its endogenous and exogenous components. It filters the sensory information coming from the environment. This filtering process can be modeled using two complementary dimensions:

- The *selective* dimension corresponds to the spotlight metaphor [13]. It considers that attention is oriented in a given direction in space. This model promotes the sensory information provided by the elements around this main direction, ignoring partially [14] or totally [15] the elements in the periphery depending on the *attention focus*. Broadly speaking, the attention focus is the size of the region around the attention direction wherein elements can potentially receive attentional resources.

- The *intensive* dimension corresponds to the limited resource metaphor. Attention is a reservoir containing a finite amount of attentional resources [16]. The attention is the energy needed for a cognitive task. Each person, depending of their own skill and the invested effort, will consume more or less attentional resources for each task. This reservoir model captures the ability to perform tasks in parallel (as long as there is enough energy in the reservoir for all of them).

3   Attentional Model

In 2006, [11] proposed a computational model of attention that was later detailed in [17]. As mentioned earlier, this attentional model, henceforth called AM, was initially developed in the context of ambient applications and pervasive systems. Its purpose was to estimate how attention was shared between the different users and the components of a "smart environment". In a situation involving a set of persons and relevant environmental objects, AM can compute the *attention matrix* that characterizes how the attention of each person is distributed among the different elements, persons and objects, of the environment. The purpose of this section is to outline the key principles underlying the definition of AM. A more detailed presentation of AM is found in Section 4. The reader is referred to [17] for a complete presentation of AM.

3.1 AM Principle

The main feature of AM is that it is a *global* attentional model that takes into account both the activity of a person and the influence of their environment. On the one hand, the observation of the activity that a person is currently engaged in (along with other clues whenever available) defines the endogenous factor vis-a-vis the attention of the person. On the other hand, the identification of the salient elements of the environment, persons and objects, defines the exogenous factors vis-a-vis the attention of the person. AM combines these factors in a coherent way. It uses a mathematical model inspired from Newton's gravitation law from physics.
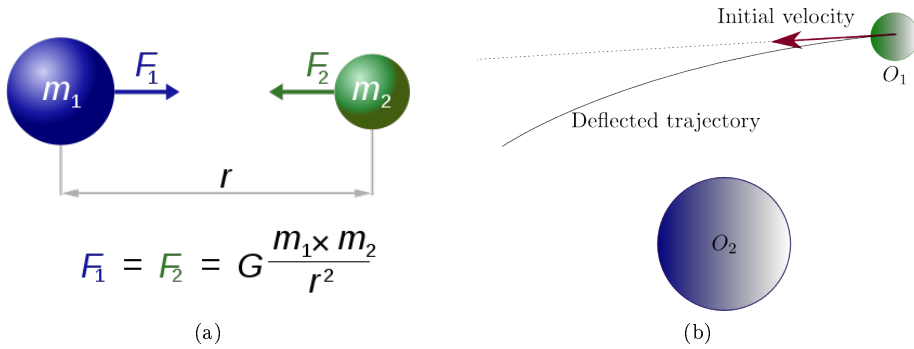

Figure 4: Newton's gravitation law and its effect on motion.

Newton's well known gravitation law states that any two bodies in the universe attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them (Fig. 4a). On the other hand, Newton's second law of motion states that the net force applied to an object is proportional to its acceleration. The consequence of these laws is that the trajectory of a body is deflected in the presence of another object (Fig. 4b). Shifting from physics to attention, the analogy could be formulated as follows:

- An object with an initial velocity and a null net force follows a straight trajectory. It takes a non null acceleration for the object to modify its trajectory and the higher the initial velocity, the more important the acceleration should be in order to significantly modify the initial trajectory. The initial velocity can be interpreted as the resistance to distraction, *i.e.* the endogenous component of attention. In AM, it is modeled as a vector, the *intention vector* henceforth denoted $\vec{I}$.

- The trajectory of an object is modified by its acceleration which is in turn proportional to the sum of the gravitation forces exerted by neighboring objects. Said acceleration can be interpreted as the distraction, *i.e.* the exogenous component of attention where the mass of each neighboring object represents its distraction capability, *i.e.* its salience. In AM, it is modeled as a vector, the *distraction vector* henceforth denoted $\vec{D}$.

3.2 Attention Vector

Each person in the environment has its corresponding intention and distraction vectors. $\vec{I}$ is typically the output of a human activity recognition module [6, 7]. As far as $\vec{D}$ is concerned, it is a function of the distraction of the different elements, other persons and objects, that surround the

person considered. The distraction of an element is a function of its salience that can be estimated using for instance visual salience computation techniques [18, 19]. Both $\vec{I}$ and $\vec{D}$ are two- or three-dimensional vectors depending on the dimensionality of the environment considered. Finally, attention is modeled in AM as a two- or three-dimensional vector, the *attention vector* henceforth denoted $\vec{A}$ that combines both the endogenous and exogenous components of attention. $\vec{A}$ is the result of the competition between *intention* and *distraction*:

$$\vec{A} = f_a(\vec{I}, \vec{D}) \tag{1}$$

3.3 Attentional Resources Allocation

The attention vector $\vec{A}$ is central to AM. Its direction corresponds to the main direction of attention of the person considered, and its magnitude is used to determine both the amount of attentional resources available and the attention focus, *i.e.* the size of the region around the attention direction wherein elements can potentially receive attentional resources. At this point, it remains to allocate the attentional resources of the person considered to the elements within its attention focus. For a given element, person or object, of the environment, this is achieved by taking into account the azimuth of the element with respect to $\vec{A}$, its distance and its salience.

3.4 Attention Matrix

	P_1	P_2	O_1	O_2	O_3
P_1	-	m_{12}	m_{13}	m_{14}	m_{15}
P_2	m_{21}	-	m_{23}	m_{24}	m_{25}

Table 1: Example of attention matrix.

The output of AM is an attention matrix henceforth denoted M. Each column of M is associated with an element, person or object, of the environment while each line of M is associated with a person. Table 1 depicts an example of attention matrix for a situation involving 5 elements: 2 persons (P_1 and P_2), and 3 objects (O_1, O_2 and O_3). The content of each cell, denoted m_{ji}, represents the amount of attention given by the person j to the element i. Note that each line of M represents how the attention of the corresponding person is spread among the different elements of the environment whereas each column represents the attention paid by the different persons to the corresponding element.

4 Attentional Model Implementation

The purpose of this section is to detail how AM is actually implemented. For the sake of simplicity, we have considered two-dimensional environments (the extension to the three-dimensional case is straightforward). Again, the reader is referred to [17] for more details.

4.1 Notations

Let $\mathcal{W}$ denote the environment considered and R the robot at hand. The workspace $\mathcal{W} \subset \mathbb{R}^2$ is populated by a set of e elements $E_i, i = 1 \ldots e$. These elements are either persons, $P_j, j = 1 \ldots p$ or objects $O_k, k = 1 \ldots o$. Unlike objects, persons have senses that provide them with information about their environment. Two sensory modalities have been considered in this work: vision and

hearing. In this respect, every element has a salience, *i.e.* a capacity to distract a person. It will be distinguished between visual and auditory salience. Each element E_i is thus characterized by the following attributes:

- Its pose, *i.e.* its position and orientation in $\mathcal{W}$: $q_i = (x_i, y_i, \theta_i)$.

- Its geometric shape which is denoted by $E_i(q_i)$: it is the closed region of $\mathcal{W}$ which is occupied by E_i when it is in pose q_i.

- Its visual and auditory saliences respectively denoted sal_i^{visual} and sal_i^{audio}. They are both scalar values that represent the intrinsic capacity of an element to distract (irrespective of any sensing capability).

- Its visual and auditory field-of-distractions respectively denoted FoD_i^{visual} and FoD_i^{audio}. FoD_i^{mod} is a mapping from $\mathbb{R}^2 \times S^1 \times \mathcal{W}$ to $\mathbb{R}$. $FoD_i^{mod}(q_i, x, y)$ characterizes the distraction capacity at a given position (x, y) for the sensory modality mod of E_i when it is at pose q_i.

In addition, a person P_j has the following attributes:

- Its visual and auditory field-of-perception respectively denoted FoP_j^{visual} and FoP_j^{audio}. FoP_j^{mod} is a mapping from $\mathbb{R}^2 \times S^1 \times \mathcal{W}$ to $\mathbb{R}$. $FoP_j^{mod}(q_j, x, y)$ characterizes the perception capacity at a given position (x, y) for the sensory modality mod of P_j when it is at pose q_j.

- Its intention $\vec{I}_j$. It is henceforth assumed that $\vec{I}_j$ is constantly estimated thanks to a human activity recognition module [6, 7].

Examples of field-of-distractions and field-of-perceptions are presented in Section 5.1.

4.2 Computing the Distraction

As per (1), the first step in order to compute the attention vector of a person P_j is to compute their distraction vector $\vec{D}_j$. As mentioned earlier, $\vec{D}_j$ is a function of the distraction capacities of the different elements, other persons and objects, that surround P_j. Let $\vec{D}_{ij}$ denote the distraction vector that characterizes the distraction caused by the element E_i on P_j. $\vec{D}_{ij}$ depends on the salience and field-of-distraction of element E_i and the field-of-perception of P_j. In line with the gravitation analogy, $\vec{D}_{ij}$ is a vector pointing from P_j to E_i whose norm is computed as follows:

$$\|\vec{D}_{ij}\| = \sum_{mod} sal_i^{mod} FoD_i^{mod}(q_i, x_j, y_j) FoP_j^{mod}(q_j, x_i, y_i) \tag{2}$$

and $\vec{D}_j$ is the sum of the distractions of each element surrounding P_j:

$$\vec{D}_j = \sum_{i \neq j} \vec{D}_{ij} \tag{3}$$

4.3 Computing the Attention

Both $\vec{I}_j$ and $\vec{D}_j$ are now available for a given person P_j, it is therefore possible to compute its attention $\vec{A}_j$ as follows:

$$\vec{A}_j = f_a(\vec{I}, \vec{D}) = \vec{I}_j + e^{-\|\vec{I}_j\|} \vec{D}_j \tag{4}$$

The exponential term is introduced to weaken the contribution of the distraction when the person is very focused and therefore less sensitive to external distractions (a determined person is hard to distract). The intention vector is scaled accordingly, i.e. $\|\vec{I}_j\| < 1$ for a person who can be easily distracted and $\|\vec{I}_j\| \gg 1$ for a person who is very focused.

4.4   Allocating the Attention

The attention vector $\vec{A}_j$ is used to determine both the amount of attentional resources available and the attention focus, *i.e.* the size of the region around the attention direction wherein elements, persons or objects, can potentially receive attentional resources. In AM, the amount of attentional resources denoted AR_j is computed as follows:

$$AR_j = \tanh(\|\vec{A}_j\|) \tag{5}$$

To capture the fact that the attention is primarily allocated to elements that are within the attention focus of a person, AM uses a function that determines the amount of attention given the incidence angle between the attention vector $\vec{A}_j$ and the line passing through the person and the element at hand: the higher α, the lower the amount of attention received. This function denoted f_j^{focus} is defined as:

$$f_j^{focus}(\alpha) = AR_j e^{-\alpha^2/2\sigma^2} \tag{6}$$

with α the incidence angle and σ the standard deviation of a Gaussian function. It represents the size of the attention focus:

$$\sigma = \pi(1 - AR_j) \tag{7}$$

For a given person P_j, it is now possible to compute the amount of attentional resources (s)he gives to a given element E_i (and therefore to fill in the corresponding cell m_{ji} in the attention matrix M). It is done by ordering the set of elements $\{E_i\}, i \neq j$ by increasing order of incidence angle α_i. Then each element in the order of the list is allocated an amount of attention equal to:

$$m_{ji} = f_j^{focus}(\alpha_i)\|\vec{D}_{ij}\|/\|\vec{A}_i\| \tag{8}$$

This amount is deducted from AR_j, the total amount of attentional resources that P_j has. The process stops when AR_j becomes zero.

5   Attention Model and Human-Robot Motion

The purpose of this section is to demonstrate how attention can be used to address Human-Robot Motion problems. To that end, three case studies aimed at illustrating different ways to use the attention model AM presented earlier are considered. Details concerning the set up that is used in all case studies are given first.

5.1   Set Up Description

The case studies covered in the next sections concern two-dimensional scenarios with one robot, up to two persons and up to two different objects: a TV and a phone. As far as their geometric shape is concerned, they are all modeled as disks for the sake of simplicity. The robot and the persons have both visual and auditory sensory modalities. Let us therefore see how the field-of-distractions and the field-of-perceptions are defined for these elements.

5.1.1   Field-of-Distractions

For the phone and the robot, the auditory field-of-distraction is assumed to be isotropic, centered at the elements' position (the *noise source* so to speak), and decreasing quadratically with the
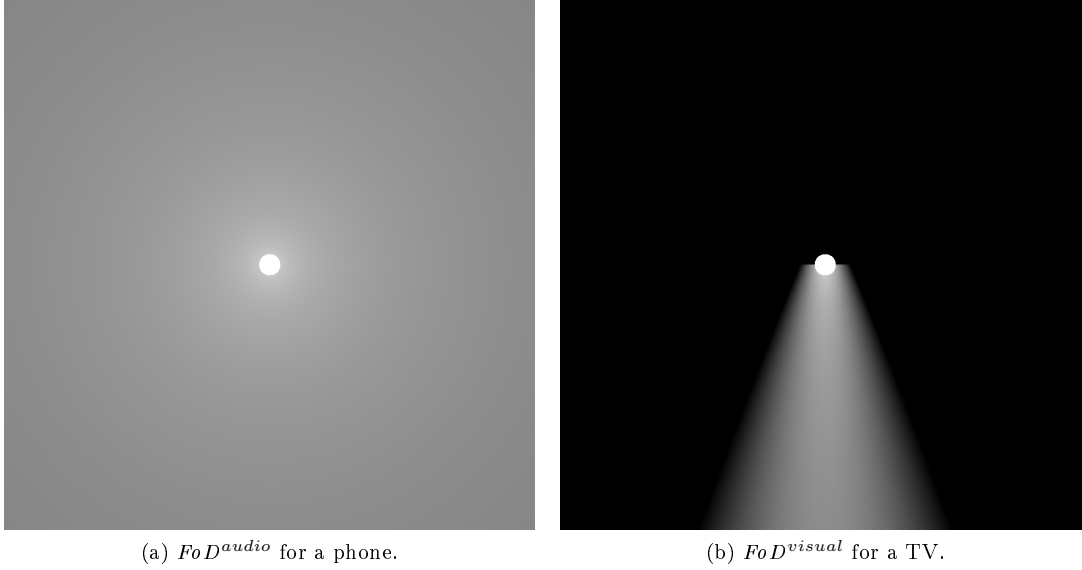
(a) FoD^{audio} for a phone.

(b) FoD^{visual} for a TV.

Figure 5: Isotropic auditory field-of-distraction for a phone (left), and visual field-of-distraction for a TV whose screen is facing downward (right). For the TV, distraction is maximum when standing right in front of the screen. It decreases with the distance to the screen and the incidence angle with the normal to the screen, the rate of decrease obeys an Euler Beta function.

distance to the noise source. FoD_i^{audio}, the auditory field-of-distraction for the element E_i is formally defined as:

$$FoD_i^{audio}(q_i, x, y) = \frac{1}{(x - x_i)^2 + (y - y_i)^2} = \frac{1}{d_i(x, y)^2} \tag{9}$$

The same definition applies to the visual field-of-distraction for the phone and the robot. However things are different for the TV and the persons. In this case, it is assumed that the distraction, both visual and auditory, is maximum when one is directly facing the TV screen/person's face, and that it decreases with the distance to the TV screen/person's face and the incidence angle with the normal to the TV screen/person's face. This can be modeled as follows:

$$FoD_i^{visual}(q_i, x, y) = \frac{1}{d_i(x, y)^2} f_i^{visual}(\alpha) \tag{10}$$

$$FoD_i^{audio}(q_i, x, y) = \frac{1}{d_i(x, y)^2} f_i^{audio}(\alpha) \tag{11}$$

where α is the incidence angle between the normal to the TV screen/person's face at pose q_i and the line passing through (x_i, y_i) and (x, y). f_i^{mod} is the function that weights the amount of distraction as a function of the incidence angle. We choose it based on an Euler Beta function with parameters $(2, 2)$

$$f_i^{visual}(\alpha) = (\alpha - \frac{\pi}{2})^2(\alpha + \frac{\pi}{2})^2 \tag{12}$$

where the incidence angle α is expressed in radians. Other functions could be used, *e.g.* Gaussian. Examples of field-of-distractions for a phone and a TV are depicted in Fig. 5.

5.1.2   Field-of-Perceptions

The persons and the robot have visual and auditory sensing capabilities that are characterized by their field-of-perceptions. As far as hearing is concerned, the auditory field-of-perception FoP_j^{audio} is assumed to be isotropic. As far as vision is concerned, the visual field-of-perception FoP_j^{visual} is defined similarly to (10) (without the quadratic decrease) with an incidence angle now defined with respect to the gaze direction. An example of field-of-perception for a person is depicted in Fig. 6. We chose a Gaussian function for f_i^{mod} with standard deviation $\sigma = \frac{\pi}{4} rad$:

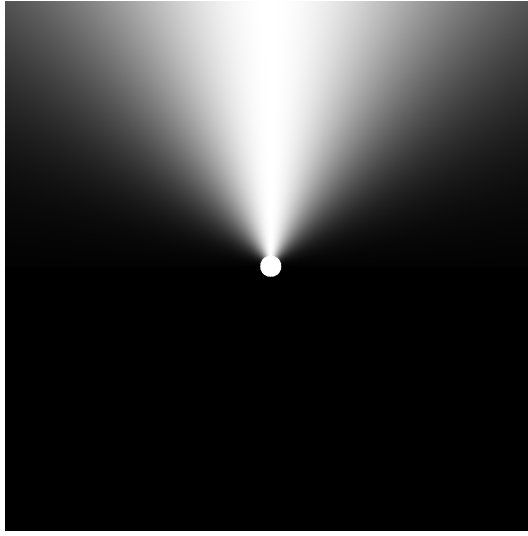$$f_i^{visual}(\alpha) = \exp \frac{-\alpha^2}{2\sigma^2} \qquad (13)$$

Figure 6: Visual field-of-perception of a person whose gaze direction is pointing upwards. Perception is maximum when standing right in the gaze direction. It decreases with the distance to the person and the incidence angle with the gaze direction, the rate of decrease obeys a Gaussian function.

5.2   Case Study #1: Using the Attention Matrix

	R	O_1
sal^{visual}	100	10
sal^{audio}	50	5

Table 2: Visual and auditory saliences for the Person-TV-Robot scenario.

The purpose of this case study is to show how the attention matrix M that is computed at every time instant by AM can be used by the robot R in order to better understand the current situation and plan its next action accordingly. Let us begin by illustrating the type of information that AM brings. Fig. 7 depicts a scenario with a person P_1, a TV O_1 and the robot R (the Person-TV-Robot scenario). It is assumed that the person is currently watching the TV
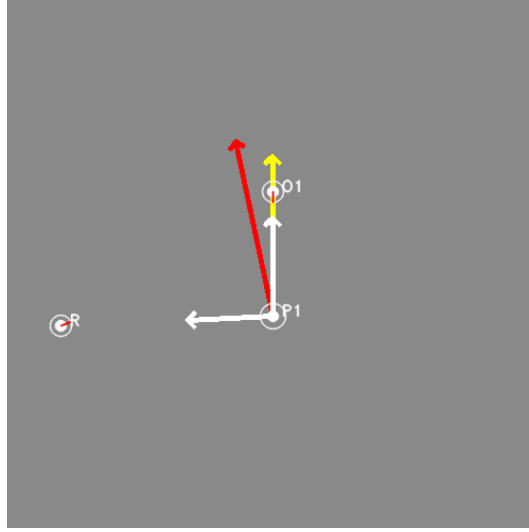
Figure 7: Scenario with a person P_1, a TV O_1 and the robot R (Person-TV-Robot scenario). The person is currently watching the TV. The yellow vector is the intention, white vectors are distractions and the red vector is the attention.

(this is their current activity). Their intention $\vec{I}_1$ is represented by the yellow vector pointing towards O_1 in Fig. 7. The two white vectors represent the distractions $\vec{D}_{i1}$ caused to the person by the robot and the TV respectively (the corresponding visual and auditory saliences are given in Table 2). The red vector is the attention $\vec{A}_1$ that has been computed by AM, the corresponding attention matrix M is given in Table 3. As expected, most of the person's attention is given to the TV.

	P_1	R	O_1
P_1	-	0.02	0.61

Table 3: The attention matrix M for the scenario depicted in Fig. 7.

Let us now extend the previous scenario by adding a phone O_2 in the environment (Fig. 8). The phone is an interesting element whose salience depends on whether it is ringing or not. The corresponding visual and auditory saliences are given in Table 4.

	R	O_1	O_2 silent	O_2 ringing
sal^{visual}	100	10	0	0
sal^{audio}	50	5	0	1000

Table 4: Visual and auditory saliences for the Person-TV-Phone-Robot scenario. The phone salience depends on whether it is ringing or not.

Let us assume now that the task of the robot R is to deliver a message to the person P_1. R is slowly approaching P_1 while constantly monitoring where P_1's attention is. As R gets closer to P_1, the attention that P_1 is paying to R increases: compare for instance the situation corresponding to Figs. 7 and 9a. As per the attention matrix of Table 5 (silent case), it can be
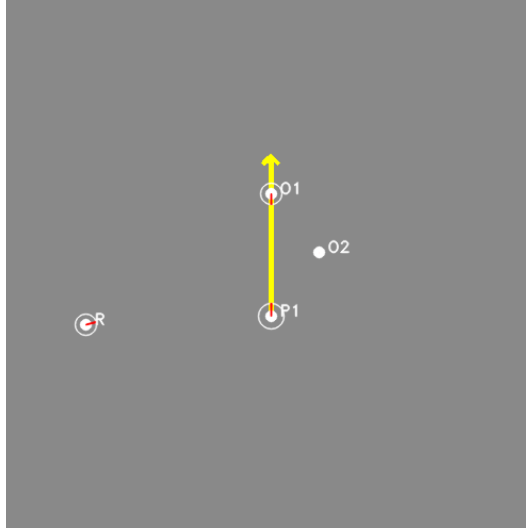
Figure 8: Scenario with a person P_1, a TV O_1, a phone O_2 and the robot R. The person is currently watching the TV (Person-TV-Phone-Robot scenario).

seen that the attention given to R by P_1 is very much increased now that R is close to P_1: it increased from 0.02 to 0.70. All things being equal, if the phone rings, things change drastically and most of the attention of P_1 is now given to the phone: compare Figs. 9a and 9b, and the attention matrices of Table 5. The attention given to R by P_1 drops from 0.70 to 0.09. In such a situation, the appropriate reaction for R is to standby and wait politely until P_1 has picked up the phone. R could later decide to resume its approach as soon as P_1 is once again paying attention to R (when the phone conversation is over). In this respect, it can be seen how taking into account attention allows to anticipate the future activity of the person, *i.e.* picking up the phone, and yields a more appropriate behavior.

	P_1	R	O_1	O_2 silent		P_1	R	O_1	O_2 ringing
P_1	-	0.70	0.17	0	P_1	-	0.09	0.05	0.67

Table 5: The attention matrices M for the situations in Figs. 9.

As simple as these scenarios are, they illustrate how the knowledge of the current global *attentional situation* (embodied in the attention matrix which is constantly updated) can be used to adapt the behaviour of the robot. The next two case studies will illustrate another way to use AM, namely as an *attention predictor*.

5.3   Case Study #2: Introducing the Attention Field

	R	P_1	P_2
sal^{visual}	100	10	10
sal^{audio}	50	5	5

Table 6: Visual and auditory saliences for the Person-Person-Robot scenario.

(a) Phone is silent.

(b) Phone is ringing.

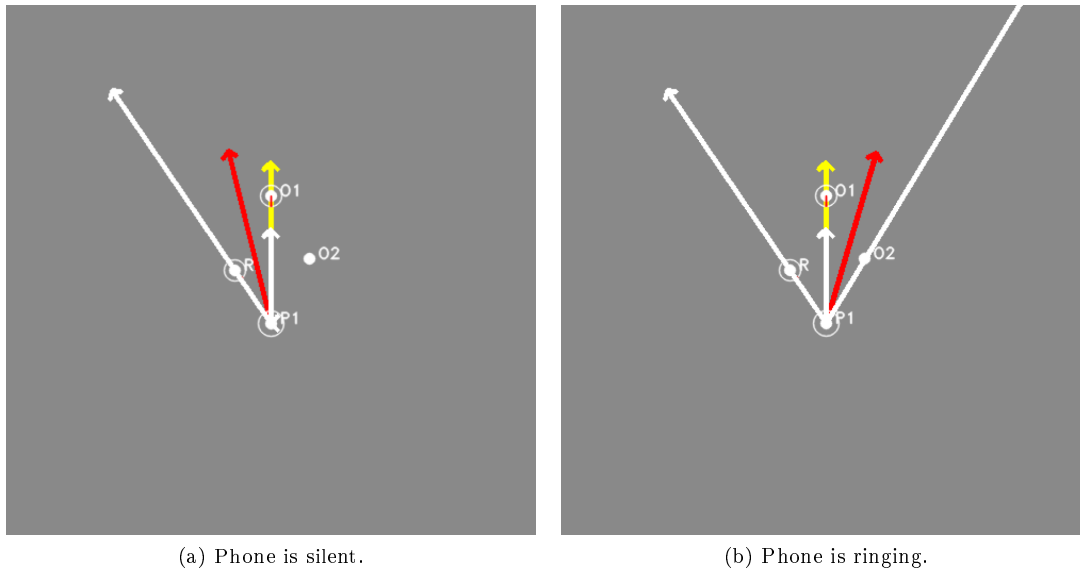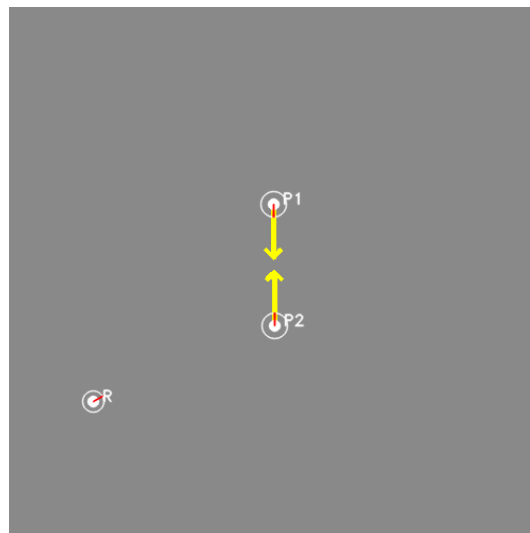Figure 9: Two similar situations for the Person-TV-Phone-Robot scenario.

Figure 10: Scenario with two persons, P_1 and P_2, and a robot R (Person-Person-Robot scenario).

	P_1	P_2	$R_{x,y}$
P_1	-	m_{12}	m_{13}
P_2	m_{21}	-	m_{23}

Table 7: Attention matrix $M_{x,y}$ for the Person-Person-Robot scenario depending on the position of R.

The purpose of this case study is to demonstrate how AM can be used to predict potential attentional situations and to use that knowledge in order to decide what the robot will do depending on its current task. Fig. 10 depicts a scenario featuring two persons, P_1 and P_2, and a robot R. The two persons are currently engaged in a conversation (this is their current activity). Their intentions $\vec{I_1}$ and $\vec{I_2}$ are respectively modeled by the two yellow vectors in Fig. 10, they point at each other. The corresponding visual and auditory saliences are given in Table 6. Let us assume that the task of R is to deliver a message to P_1 while minimizing the disturbance caused to the other persons. To achieve this task, the appropriate behavior for R should be to move to a position where it can (1) attract the attention of P_1, and (2) minimize the distraction caused to P_2. To that end, R must be able to estimate the attention that the persons will pay to it depending on its position (x, y). It is easily done using AM in order to compute the attentional matrix $M_{x,y}$ for every possible positions of R in $\mathcal{W}$. Table 7 gives the general form of $M_{x,y}$. By doing so, it becomes possible to compute a so-called *attention field* for each person. Let F_j denote the attention field for the person P_j, it is mapping from $\mathcal{W}$ to $\mathbb{R}$ that gives the amount of attention that P_j is paying to R when it is at position (x, y). This mapping is readily obtained from $M_{x,y}$, it is the value contained in the cell m_{j3}.

(a) Attention field F_1 for P_1.

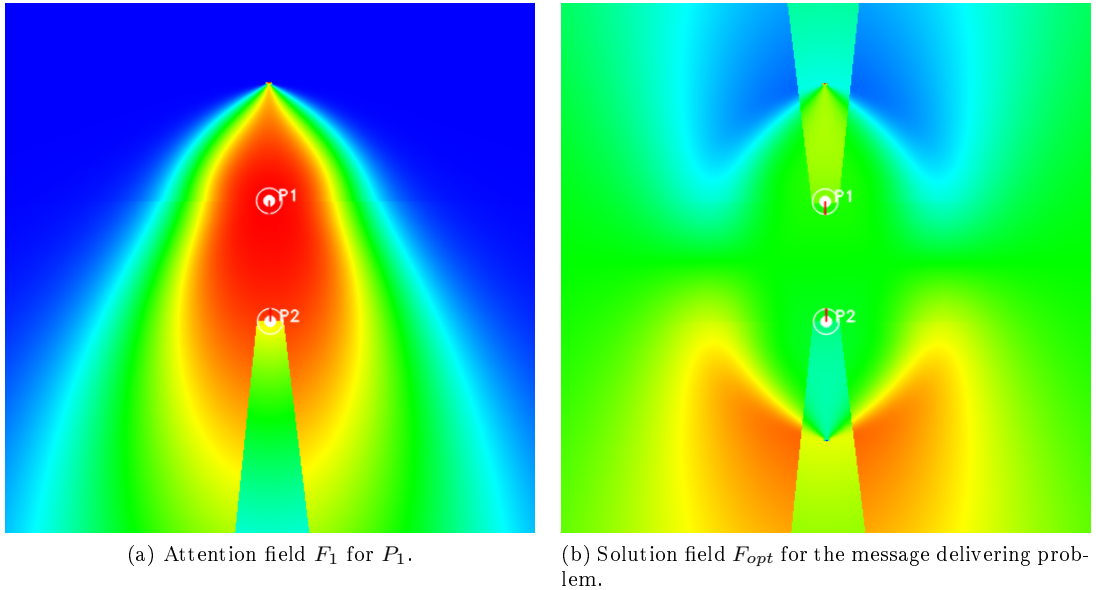(b) Solution field F_{opt} for the message delivering problem.

Figure 11

Fig. 11a depicts F_1, the attention field for P_1 for the Person-Person-Robot scenario. It should be interpreted as follows: the warmer the color, the higher the amount of attention given by P_1 to R. It integrates the visual and auditory perception capabilities of P_1. Note in particular how the occlusion of P_1's field-of-view by P_2 impacts the field. Because of the symmetry in P_1 and

P_2's situations, the attention field F_2 for P_2 would be very similar (F_1 rotated by 180 degrees).

Once the attention fields F_j are available, it becomes possible to use them to address the task at hand: attracting P_1's attention while disturbing P_2 as little as possible. This task can be formulated as a multiobjective optimization problem, *i.e.* maximizing P_1's attention and minimizing P_2's attention. In this case, a simple weighted sum can be used to compute a field F_{opt} defined as:

$$F_{opt} = F_1 - F_2 \tag{14}$$

F_{opt} is depicted in Fig. 11b. It should be interpreted as follows: the warmer the color, the better the solution to the multiobjective optimization problem at hand. F_{opt} confirms the intuition that the best way to attract P_1's attention while disturbing P_2 as little as possible is to be in the red region of Fig. 11b, *i.e.* a region where R is almost facing P_1 while remaining outside the field-of-view of P_2. F_{opt} can readily be used to drive R's behavior (the next case study will show how to actually do that).

5.4 Case Study #3: Motion Planning using Attention Fields

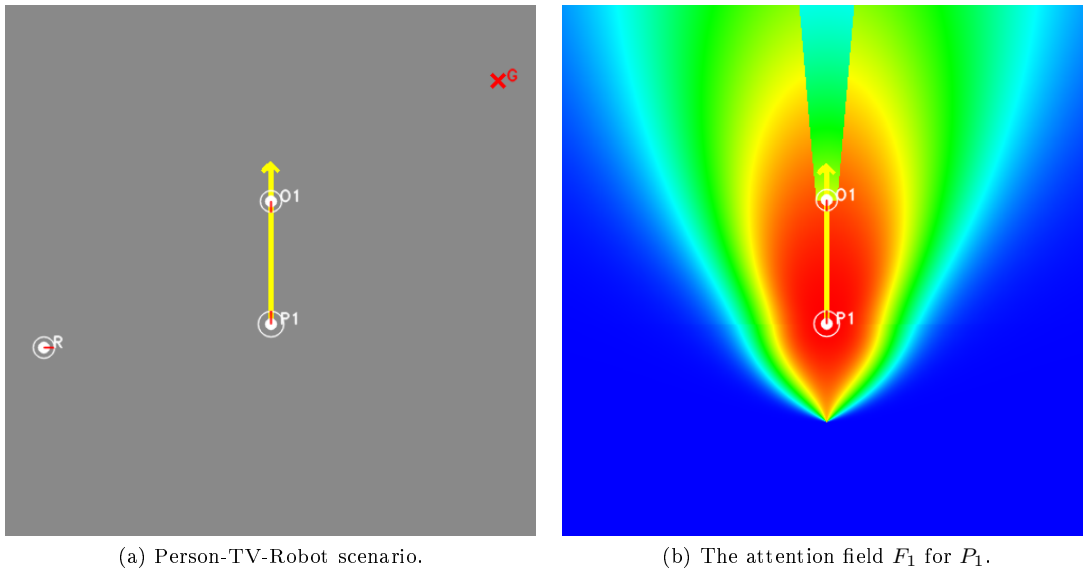(a) Person-TV-Robot scenario. (b) The attention field F_1 for P_1.

Figure 12

The previous case study has shown how the attention field can be used to determine where the robot should go to achieve a given task. The purpose of this case study is to illustrate how the attention field can be used for motion planning purposes, *i.e.* to actually determine how the robot should move. To that end, the Person-TV-Robot scenario introduced in the first case study is used. The person is watching the TV (yellow vector pointing towards O_1 in Fig. 12a). The visual and auditory saliences are given in Table 2).

Two different tasks are assigned to the robot R: the first one is to deliver a message to P_1 (Deliver task). The second one is to reach a goal G while disturbing P_1 as little as possible (Goto task). For both tasks, the attention field F_1 for P_1 is used. F_1 is depicted in Fig. 12a. At this point, it is interesting to look at the relationship between our attention model and the standard social space model. The Person-TV-Robot scenario would yield two social spaces: (1) a personal

space centered around the person, and (2) an activity space in the form of the convex hull of P_1 and the TV. Note how the red region of F_1 encompasses these social spaces.
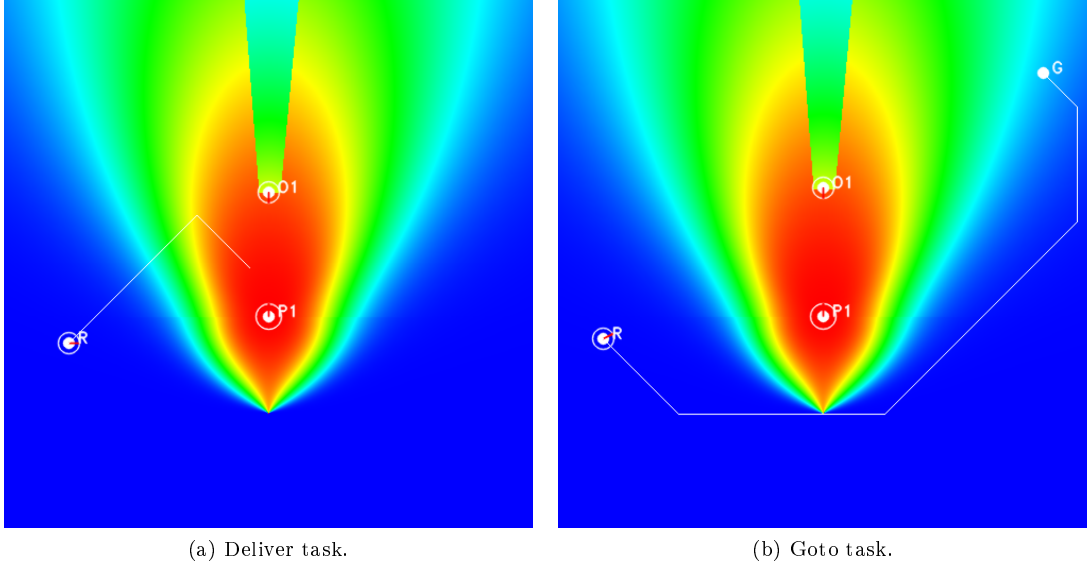

<div align="center">(a) Deliver task.   (b) Goto task.</div>

<div align="center">Figure 13</div>

To address the Deliver task, the behavior of R is driven using a simple gradient descent based on F_1. The idea is to move towards P_1 so as to smoothly increase the attention that P_1 is paying to R (hence the gradient descent). The resulting motion is depicted in Fig. 13a. Note how R ends up approaching P_1 with an incidence angle of about 45 degrees. Such a behavior is coherent with the experimental results presented in [20] that establish that this is the best way to approach a person politely without causing discomfort. This behaviour has been confirmed for arbitrary start poses.

Addressing the Goto task requires motion planning capabilities. A Dijkstra algorithm is used to compute the optimal path between the current position of R and the goal. In our case, the optimality criterion is to minimize the total sum of the attention paid by P_1 to R along the path. A resulting path is depicted in Fig. 13b. Note how R passes behind the person so as not to disturb it (with respect to its visual and auditory modalities). Note also that, had R's behavior been driven based solely on the social spaces, it would have passed above the TV since it would yield the shortest path towards the goal. Doing so, it would have distracted P_1 more. As simple as this example is, it illustrate the interest of taking attention into account.

6   Discussion and Conclusion

We have shown in this paper that the psychological concept of attention can be used to address Human-Robot Motion (HRM) in order to obtain autonomous navigation schemes for mobile robot companions that yield better appropriate motions. We have built upon the computational attention model initially proposed in [11] that computes the attention matrix, *i.e.* a model of how the attention of each person is distributed among the different elements, persons and objects, of their environment. Through several case studies, we have demonstrated different ways to use the

attention matrix in HRM. We have also introduced the novel concept of attention field and used it for navigation purposes. We believe that attention is a concept that is a beneficial complement to the social space concept that is classically used in HRM.

The next step of this work is to investigate this approach further with more complex case studies. Then, although this paper has focused on attention only, we plan to combine attention with social space concepts (in particular our earlier work on social space-based navigation in dynamic environments [10]). Finally, real life experiments on a real robot will be carried out and used to validate the interest of using attention to address HRM.

References

[1] W. Burgard, A. Cremers, D. Fox, D. Hahnel, D. Lakemeyer, D. Schulz, W. Steiner, S. Thrun, The interactive museum tour-guide robot, in: AAAI Nat. Conf. on Artificial Intelligence, Madison (US), 1998, pp. 11–18.

[2] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, H. Christensen, Navigation for human-robot interaction tasks, in: IEEE Int. Conf. on Robotics and Automation, Barcelona (ES), 2004. `doi:10.1109/ROBOT.2004.1308100`.

[3] F. Lindner, C. Eschenbach, Towards a formalization of social spaces for socially aware robots, in: M. Egenhofer, N. Giudice, R. Moratz, M. Worboys (Eds.), Spatial Information Theory, Vol. 6899 of Lecture Notes in Computer Science, Springer, 2011. `doi:10.1007/978-3-642-23196-4`.

[4] E. Hall, A system for the notation of proxemic behaviour, American Anthropologist, New Series 65 (5) (1963) 1003–1026.

[5] A. Kendon, Spacing and orientation in co-present interaction, in: A. Esposito, N. Campbell, C. Vogel, A. Hussain, A. Nijholt (Eds.), Development of Multimodal Interfaces: Active Listening and Synchrony, Vol. 5967 of Lecture Notes in Computer Science, Springer, 2010. `doi:10.1007/978-3-642-12397-9`.

[6] J. Aggarwal, M. Ryoo, Human activity analysis: A review, ACM Comput. Surv. 43 (3) (2011) 16:1–16:43. `doi:10.1145/1922649.1922653`.

[7] A. Bulling, U. Blanke, B. Schiele, A tutorial on human activity recognition using body-worn inertial sensors, ACM Comput. Surv. 46 (3) (2014) 33:1–33:33. `doi:10.1145/2499621`.

[8] E. A. Sisbot, L. F. Marin-Urias, R. Alami, T. Simeon, A human aware mobile robot motion planner, IEEE Trans. Robotics 23 (5) (2007) 874–883. `doi:10.1109/TRO.2007.904911`.

[9] D. Shi, E. Collins, B. Goldiez, A. Donate, X. Liu, D. Dunlap, Human-aware robot motion planning with velocity constraints, in: Int. Symp. on Collaborative Technologies and Systems, Irvine (US), 2008. `doi:10.1109/CTS.2008.4543969`.

[10] L. Scandolo, T. Fraichard, An anthropomorphic navigation scheme for dynamic scenarios, in: IEEE Int. Conf. on Robotics and Automation, Shanghai (CN), 2011. `doi:10.1109/ICRA.2011.5979772`.

[11] J. Maisonnasse, N. Gourier, O. Brdiczka, P. Reignier, Attentional model for perceiving social context in intelligent environments, in: I. Maglogiannis, K. Karpouzis, M. Bramer (Eds.), Artificial Intelligence Applications and Innovations, Vol. 204 of IFIP Int. Federation for Information Processing, Springer, 2006. `doi:10.1007/0-387-34224-9_20`.

[12] J. Anderson, Cognitive Psychology and Its Implications, Worth Publishers, 2005.

[13] W. James, The Principles of Psychology, Holt, 1890.

[14] A. Treisman, Strategies and models of selective attention, Psychological Review 76 (3) (1969) 282–299.

[15] D. Broadbent, Perception and communication, Pergamon Press, 1958.

[16] D. Kahneman, Attention and effort., Prentice-Hall, 1973.

[17] J. Maisonnasse, Estimation des relations attentionnelles dans un environnement intelligent, Phd thesis, University Joseph Fourier, Grenoble (FR), in French (Dec. 2007).

[18] S. Frintrop, E. Rome, H. Christensen, Computational visual attention systems and their cognitive foundations: A survey, ACM Trans. Appl. Percept. 7 (1) (2010) 6:1–6:39. `doi:10.1145/1658349.1658355`.

[19] A. Kimura, R. Yonetani, T. Hirayama, Computational models of human visual attention and their implementations: A survey, IEICE Trans. Inf. and Syst. 96 (3) (2013) 562–578.

[20] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, T. Simeon, How may i serve you? a robot companion approaching a seated person in a helping context, in: ACM Conf. on Human-robot Interaction, Salt Lake City (US), 2006. `doi:10.1145/1121241.1121272`.