# Incremental Speech Production for Polite and Natural Personal-Space Intrusion

Timo Baumann[1] and Felix Lindner[2] *

[1] Natural Language Systems Division
[2] Knowledge and Language Processing Group
Department of Informatics
University of Hamburg
Vogt-Kölln-Straße 30, 22527 Hamburg
{baumann, lindner}@informatik.uni-hamburg.de

**Abstract.** We propose to use a model of personal space to initiate communication while passing a human thereby acknowledging that humans are not just a special kind of obstacle to be avoided but potential interaction partners. As a simple form of interaction, our system communicates an apology while closely passing a human. To this end, we present a software architecture that integrates a social-spaces knowledge base and a component for incremental speech production. Incrementality ensures that the robot's utterance can be adapted to fit the developing situation in a natural way. Observer ratings show that personal-space intrusion is perceived as both natural and polite if the robot has the capability to utter and adapt an apology in an incremental way whereas it is perceived as unfriendly if the robot intrudes personal space without saying anything. Moreover, the robot is perceived as less natural if it does not adapt.

## 1 Introduction

When robots and humans act in common spaces they inevitably encounter each other regularly. Therefore, social robots need to solve the task of passing humans in a socially appropriate manner. Pioneering work on the research question of how robots should pass humans can be attributed to the early studies presented in [15] and [23].

In more recent work the capability to socially pass a human has been modeled using the notion of personal space. Authors from the social sciences like Hall [8] and Sommer [21] use the concept of personal space to explain the various phenomena related to how humans spatially behave towards other humans with particular focus on the distances they maintain to each other. Computational models of personal space have mainly been applied to human-aware robot navigation to avoid personal-space intrusion [5,9,16,18,19,22]. In effect, these approaches result in robots taking detours in accordance with personal-space theory. In fact, there seems to exist a common ground that models of personal space should

* Authors ordered alphabetically.

keep the robot away from humans in the first place. As a result, comparatively few approaches take personal space as a basis for specifying how a robot should behave if it intrudes personal space. Lam and colleagues [10] present a two-stage policy with respect to personal-space usage. As with the other approaches, the robots should avoid personal-space intrusion. However, if it accidently happens that the robot intrudes personal space, the robot will stop moving until it is not within personal space anymore.

All in all, the main line of the reviewed work is that personal spaces should not be entered by robots passing a human. Instead, the robot should take detours and, if the robot finds itself within personal space accidently, it should freeze. According to the available literature on human-aware robot navigation, the title of the paper at hand seems to be inherently contradictory, because it claims that there is a way to intrude personal spaces in a polite and natural manner. In earlier work [13], we already suggest to add social signals to navigation plans to gain permission to enter regions of personal space, thus, to intrude personal space in a planful manner. In this work, we extend this idea and propose to use a model of personal space that acknowledges that humans are not just obstacles to be avoided but potential interaction partners. As a simple form of interaction, our system communicates an apology while closely passing a human. We present a software architecture that integrates a social-spaces knowledge base and a component for incremental speech production (see Sect. 2). Incremental speech production allows a system to start outputting speech based on partial speech plans that can later be extended [20] or even altered to reflect changes of the underlying plan [3]. Incremental speech synthesis is able to continuously render speech with a natural and continuous prosody and at almost the quality of systems that require the full and unchangeable utterance specification in advance [1], even though requiring only a few words of future context.

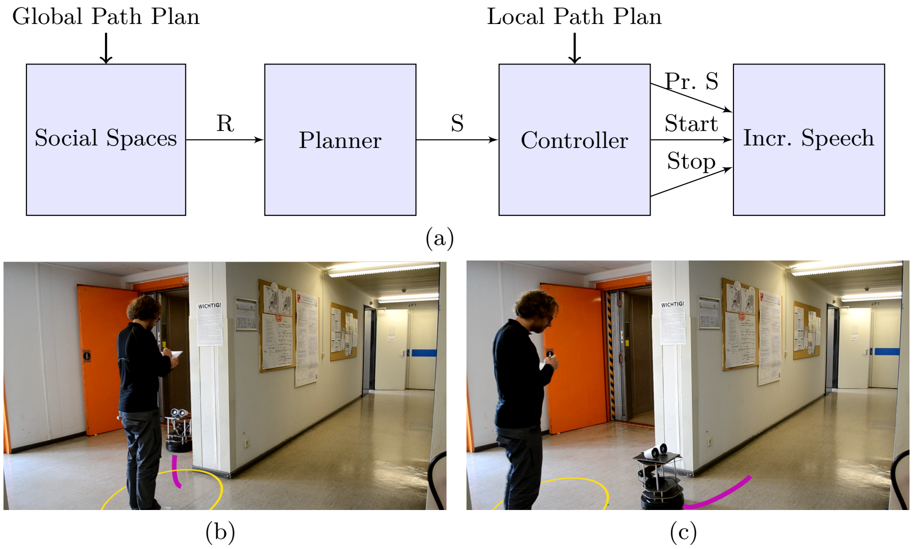
To evaluate our system we conducted an observation study. In particular we tested two main hypotheses:

**Hypothesis A** A robot passing through a personal space is perceived as more polite if it utters an apology rather than saying nothing,

**Hypothesis B** A robot passing through a personal space is perceived as more natural if it has the capability to adapt its speech incrementally as the situation evolves.

A comparable study [7] could not comfirm an effect on the perceived politeness of a robot that signals its intention to pass by making beep sounds as compared to making no sounds at all. This result should discourage our belief in hypothesis A. However, a later study, which investigates the effect of social framing on the reactions of people towards a robot that signals its intention [6], reveals that subjects perceive a speaking robot as more friendly than a beeping robot.

Hypothesis B is grounded in the fact that humans' speech production is inherently incremental [11]. Humans can adapt their utterances while speaking with ease and do so as the situation or interaction requires [4]. Therefore, we expect that a robot with this capability is perceived as more natural than a robot



(a)

(b)

(c)

**Fig. 1.** (a) Architecture integrating a knowledge base about social spaces and a component for incremental speech production. (b) As soon as the local path plan (pink) overlaps the personal space (yellow) the robot starts to say “Excuse me, I need to pass urgently to rescue a patient in the other corridor – thank you.” (c) However, as the person steps aside the robot leaves personal space before the whole explanation was uttered resulting in “Excuse me, I need to pass urgently – thank you.”

that ‘balistically’ utters its whole pre-planned utterance without considering situational changes.

Confirming Hypothesis A, the observation study presented in Sect. 3 shows that personal-space intrusion is perceived as both natural and polite if the robot has the capability to utter and adapt an apology in an incremental way whereas it is perceived as unfriendly if the robot intrudes personal space without saying anything. Confirming Hypothesis B, we found that it is perceived as unnatural if the robot does not adapt its utterance plan incrementally. We find no effects on the control questions regarding the robot’s route, which indicates that observers differentiate between the various aspects of multi-modal robot behaviour.

## 2 A Software Architecture Integrating Social Spaces and Incremental Speech Synthesis

To enable a social robot to planfully intrude personal space while passing a human, we propose the architecture shown in Fig. 1(a). The software architecture integrates the capability to reason about social spaces (i.e., personal spaces among others) and the capability to incrementally utter natural language. An example use case is shown in Fig.s 1(b) and 1(c): Personal space intrusion is accompanied by a verbal explanation, which is adapted as a reaction to the human clearing the way for the robot. The architecture’s components are described below.

### 2.1 Social Spaces

The concept of social spaces subsumes several socio-spatial phenomena among which personal space is the most popular one (cf. [13]). Social spaces can be characterized as socio-spatial entities that are produced by other entities that provide reasons for action to social agents. Particularly, a personal space is produced by a (single) human and the human provides reasons for action to other social entities (e.g., robots). Our reason-driven view is inspired by contemporary work in practical philosophy (e.g., [17]) and motivated by the fact that reasons can be used both for deliberate decision making and for generating justifications or apologies social agents owe to others.

In the example depicted in Figures 1(b) and 1(c) the human produces a personal space. Within the symbolic knowledge base of the robot the human is represented as an individual which provides the robot with a reason against driving along the planned route.[3] Additionally, we assume that there is a patient in the other corridor which needs to be rescued by the robot. Consequently, the patient provides the robot with a reason in favor of driving along the planned route. Hence, given the navigation action *driving along the global path* represented by the global path plan (see Fig. 1(a)) the knowledge base can be queried for reasons that speak in favor of or against actually executing that particular plan.

The geometrical properties of the personal space are represented by an ellipse centered around the human. The major and minor axes were set to 3m and 2m, respectively. Consequently, as the robot crosses personal space from the left to the right hand side of the human it starts to talk to the human at a distance of roughly 1.5m. According to Hall [8] this corresponds to an interaction distance used by strangers.

### 2.2 Verbal-Planner

In cases where there are several alternative ways of acting, knowledge about reasons can be used to make choices among the available options [14]. In the approach presented here, we use reasons in a different way: They play the role of explanations. In particular, reasons that speak in favor of an action play the role of justifications whereas reasons that speak against an action can be used to formulate regret.

For instance, in the example depicted in Figures 1(b) and 1(c) the social-space component informs the verbal planner that there are two reasons $\rho_1, \rho_2$. Reason $\rho_1$ is the fact that the personal space should not be intruded and reason $\rho_2$ is the fact that some patient has to be rescued in the other corridor. Therefore, $\rho_1$ speaks in favor of executing the given path plan and $\rho_2$ speaks against doing so. Consequently, the verbal planner maps $\rho_2$ to an apology and $\rho_1$ to a justification. As a result the component outputs $S :=$ "Excuse me, I need to pass urgently to rescue a patient in the other corridor. Thank you."

We anticipate that $S$ tends to become quite long the more reasons are at stake and hence we propose to order reasons by importance and to insert additional

[3] See [12] for an in-depth technical explanation of the symbolic personal-space model.

chunking information that the incremental speech production may use to skip parts of the resulting utterance for brevity. Such ordering and chunking can be performed by incremental NLG such as [3]. However, this step was simulated in the experiments reported below.

### 2.3 Controller

The controller is a component that interfaces the verbal planner and the incremental speech synthesis. It is implemented as a finite state machine with states $s_0$, $s_1$, and $s_2$. In state $s_0$ the sentence structure $S$ is sent to the incremental speech component in order to internally prepare the sentence that should be uttered as soon as the robot actually enters the personal space. Being in $s_0$ the robot follows the global path plan without saying anything. When the local path plan significantly overlaps the personal space the state machine transitions from state $s_0$ to state $s_1$. In state $s_1$ the command *Start* is sent to the incremental speech component. Now the sentence structure that was prepared in state $s_0$ is actually uttered while the robot is still moving forward. A transition from $s_1$ to state $s_2$ takes place when the robot exits personal space again. In state $s_2$ the *Stop* command is sent to the incremental speech synthesis component. If at this time the robot is still talking, the incremental speech component will adapt the output, i.e., it will quickly but in a fluid way skip ahead in the utterance plan.

### 2.4 Incremental Speech Production

Given the utterance plan $S$ of the verbal planner, the incremental speech production component prepares an *utterance tree* that provide for the alternatives of the original plan (in our case: skipping parts of the explanation). Speech synthesis is a processing problem on multiple layers (determining sentence-level intonation, prosodic contours, generating vocoding parameters and finally producing the actual speech waveform) which must be coordinated across possible continuations of the utterance to produce continuous and natural speech. This is crucial as any discontinuity (spectral, loudness, prosodic, etc.) in the final speech waveform would sound unnatural. It is hence not possible to simply attach separately synthesized utterance parts.

Our speech synthesizer [2] only requires a limited and local lookahead for vocoding, HMM optimization and state selection, and can hence integrate changes between utterance choices in the synthesis process with very little delay (on the order of 50 ms). In our case the *Stop* command from the controller leads the synthesizer to skip the remaining words of the explanation of why it had to intrude and move forward to thanking the user for allowing the robot to pass by in a natural way.

## 3 Observation Study

We tested our hypothesis that sensible interaction when passing through a personal space is superior in terms of perceived naturalness and politeness of the

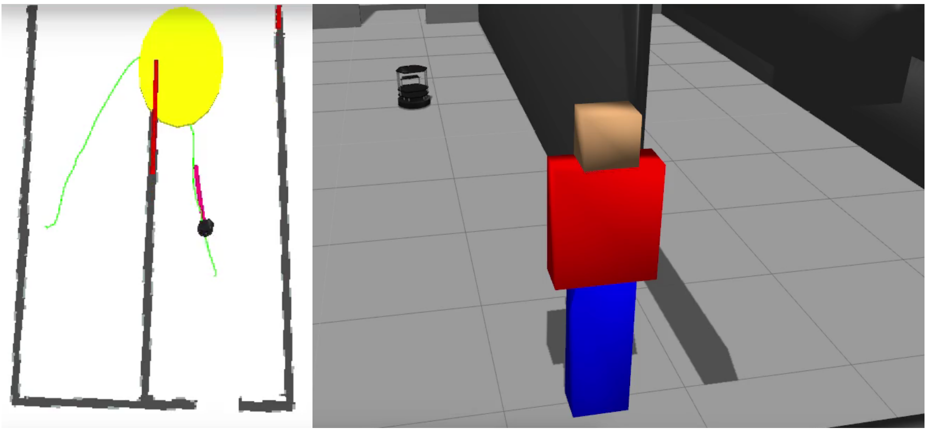


**Fig. 2.** The simulated robot's model (left side) as well as a rendering of the environment (right side) as shown in the observation videos.

robot to other strategies in a highly controlled observer rating experiment. In our conceived test environment, a hospital robot needs to pass by a person that is standing near a narrow passage in order to help a patient in the next corridor. Our test environment is depicted in Fig. 2.

The robot needs to pass through a person's personal space (depicted as a yellow ellipsis in the left part of the figure) in order to reach a target position. The global path plan is depicted as a green line (leading to the target position), the local plan at any time is depicted as a red line.[4] The global path plan was held constant throughout all simulations.

The robot plans upfront that it may want to interact in order to pass through the personal space and generates the utterance plan shown in Fig. 3. The idea of the plan is to gradually escalate the message from a low-profile *excuse me* (which might be sufficient to motivate the human to move away) to a full and thorough explanation of why the robot must violate the human's personal space. The plan finishes off with thanking the human for accepting the intrusion of her personal space.

Of course, the person may move out of the robot's way (and this is actually the robot's intent), however this cannot be relied upon in advance and can only be taken into account locally during speech delivery. To account for the variability of the moment in time at which the robot leaves the personal space, the utterance plan contains several "short-cuts" to seamlessly move ahead to the final *thank you* as indicated by the arrows in Fig. 3. We simulated the robot perception of personal space by directly informing the robot about the position of the person. The geometric properties of the personal space were represented by a polygon defined in the frame of the simulated human.

[4] Simulated laser scans are also shown in red near the walls and should not be confused with the local path plan.

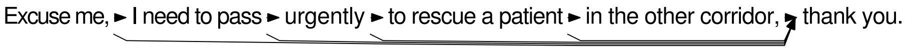


**Fig. 3.** The utterance plan in our example system allows to skip parts of the apology.

### 3.1 Experiment Setup

We screen-recorded the simulated robot's motion along a constant route (cmp. Fig. 2) systematically varying three variables: the speed of the robot (slow or fast), whether the human moves out of the robot's way, and the robot's verbal interaction: whether it delivers the full utterance plan once it enters the personal space, incrementally skips ahead when leaving the personal space, or does not verbally interact at all.

In total there are 12 video stimuli for all combinations of conditions of which 3 show no difference between incremental/non-incremental speech.[5] We played two of the duplicates in the beginning of the experiment and the third in the middle and excluded them from analysis of the verbal interaction variable. All other stimuli were distributed in random order.

We showed the videos to a group of 13 participants[6], who were asked to rate on five-point Likert scales for every video (a) the naturalness of the robot's behaviour (relating to hyp. A), (b) the politeness of the robot (relating to hyp. B), and (c) the appropriateness of the robot's route and speed (as control).

### 3.2 Results

We perform non-parametric paired statistical tests (Wilcoxon signed rank for the two-valued variables *speed* and *human movement*, and Friedman followed by post-hoc Wilcoxon signed rank for the three-valued variable *verbal interaction*) on all three variables and apply Bonferroni correction within the post-hoc tests to control for multiple-hypotheses testing.

We find no significant influence of the robot's *speed* on user ratings ($p = .29$ for naturalness, $p = .83$ for politeness, $p = .60$ for route appropriateness), indicating that there is no general preference for a higher or lower robot speed.

Regarding *human movement*, we find that the robot's behaviour is rated more natural ($p < .0001$) with a median difference of 2 points and the route more appropriate ($p < .01$) with a median difference of 1 point if the human moves aside rather than standing in place when the robot closely passes by. There is no significant effect on politeness ($p = .16$) indicating that the 'tension' of the situation is attributed to the simulated human rather than the robot in this case.

[5] *Being able* to skip does not necessarily imply that the robot *actually does* skip; the time at which the robot leaves personal space depends on the robot's speed and on whether the human steps aside. Thus, incrementality is unobserable in three stimuli (when the robot is slow and the human does not move aside).

[6] Bachelor students of computer science with little or no experience in robot navigation and speech technology (but potentially a higher interest in these topics than the general public) aged 20/20/24 years (median/first/third quartile), 11 male / 2 female, and good listening comprehension of English according to own assessment.

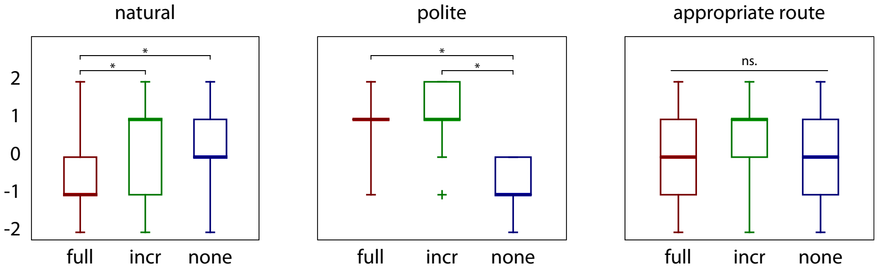


**Fig. 4.** Subjective ratings of naturalness, politeness, and route-appropriateness for the three system configurations. Significantly different ratings between configurations are marked with a star.

The results for our main variable *verbal interaction* are shown in Fig. 4. As can be seen in the figure, the robot is rated as significantly more natural when adapting (or not speaking at all) rather than speaking the full utterance (both $p < .001$), and with median advantages of 2 points (incremental) resp. 1 point (no speech at all). Regarding politeness, both speaking conditions are significantly better than not speaking at all (both $p < .001$), with median advantages of 2 points. We find no significant difference between the speaking conditions on the rated appropriateness of the route and speed, which may serve as an indication that participants successfully distinguish between questions rather than giving highly correlated ratings. Finally, for all three questions the mean rank of the incremental speaking condition is highest, indicating superiority over the other options even where no significant differences are found.

### 3.3 Discussion

A robot is rated as more polite if it verbally apologizes and explains the need to violate the interlocutor's personal space upon entering it. However, a robot is rated as less natural if it continues on this explanation even after leaving the personal space. Thus, in order to act both natural and polite, a robot must adapt its speech output while speaking in order to meet the needs of the evolving situation.

We find that the robot's speed has no overall effect on user ratings, indicating that the robot is free choose a speed that is most suitable. Finally, if the human steps aside to let the robot pass, its route is preferred and its behaviour is rated as more natural than if the human does not move. Of course, human movement is not a variable under the control of the robot. Yet, encouraging the human to move, e.g. by verbally communicating the intent to pass, improves behaviour ratings given by observers.

With respect to the interpretation of our results there are several limitations that should be considered. First, the participants of our study evaluated the behaviour of a simulated robot of a particular kind (Turtlebot) towards a simulated human. Future work will show if our results can be replicated with participants being faced with a real Turtlebot and with another type of robot (as we plan a similar study with a real Care-o-Bot 3). Another limitation is that we did not include a condition in which the robot always utters a short sentence no matter if the human moves away or not. Thus, it may turn out that the incremental condition is perceived as more natural than the non-incremental condition because the sentence uttered in the non-incremental condition is too long. But even if this were the case incrementality serves as a technical solution for producing utterances of adaptable length from arbitrarily long explanations automatically derived from reason-based representations of socio-spatial norms.

## 4 Conclusions

Results show that a comprehensive model of personal space should allow deliberate personal-space intrusion. We model the social norm that personal spaces should be respected as reasons that speak against actions that actually intrude such space. Being reasons, they can be used for decision making but also as pre-verbal representations for natural language generation in case that passing through personal space is weighed as more urgent than avoiding it. We find that adapting a planned utterance is crucial when passing through personal space in order to produce natural and polite behaviour.

We conducted an observation study in order to control for as many aspects as possible by using pre-recorded videos. However, we plan to conduct real-life first-person experiments (rather than third-person observations) in the near future to estimate the influence of speech adaptation in accordance with personal space on perceived naturalness, politeness, and safety of the robot.

Finally, our one-way mode of communication only scratches the surface of a fully interactive, personal space-aware social robot. Such a robot should be able to engage in a full dialogue with the human (or humans) it encounters, either if more elaborate negotiations are necessary for the robot to pass, or by initiative of the human. In such a system, the dialogue management component must be integrated with, or adjoined to local and global behaviour planning and these components need to be able to mutually influence each other.

## References

1. Baumann, T.: Partial representations improve the prosody of incremental speech synthesis. In: Proceedings of Interspeech (2014)
2. Baumann, T., Schlangen, D.: INPRO_iSS: A component for just-in-time incremental speech synthesis. In: Procs. of ACL System Demonstrations. Jeju, Korea (2012)
3. Buschmeier, H., Baumann, T., Dorsch, B., Kopp, S., Schlangen, D.: Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In: Procs. of SigDial. pp. 295–303. Seoul, Korea (2012)
4. Clark, H.H.: Using Language. Cambridge University Press (1996)
5. Dylla, F., Kreutzmann, A., Wolter, D.: A qualitative representation of social conventions for application in robotics. In: Qualitative Representations for Robots: 2014 AAAI Spring Symposium Series. pp. 34–41 (2014)

6. Fischer, K., Soto, B., Pantofaru, C., Takayama, L.: Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance. In: Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE Int. Symposium on. pp. 999–1005 (2014)
7. Fischer, K., Jensen, L., Bodenhagen, L.: To beep or not to beep is not the whole question. In: Beetz, M., Johnston, B., Williams, M.A. (eds.) Social Robotics, Lecture Notes in Computer Science, vol. 8755, pp. 156–165. Springer (2014)
8. Hall, E.T.: The Hidden Dimension, Man's Use of Space in Public and Private. The Bodley Head, London, England (1966)
9. Kirby, R., Simmons, R., Forlizzi, J.: COMPANION: A constraint-optimizing method for person-acceptable navigation. In: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication. pp. 607–612 (2009)
10. Lam, C.P., Chou, C.T., Chiang, K.H., Fu, L.C.: Human-centered robot navigation – towards a harmoniously human-robot coexisting environment. IEEE Transactions on Robotics 27(1), 99–112 (2011)
11. Levelt, W.J.: Speaking: From Intention to Articulation. MIT Press (1989)
12. Lindner, F.: A conceptual model of personal space for human-aware robot activity placement. In: Proceedings of the 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2015). Hamburg, Germany (2015), to appear
13. Lindner, F., Eschenbach, C.: Towards a formalization of social spaces for socially aware robots. In: Egenhofer, M., Giudice, N., Moratz, R., Worboys, M. (eds.) Spatial Information Theory. 10th Int. Conf., COSIT 2011, Belfast, ME, USA, LNCS, vol. 6899, pp. 283–303. Springer: Berlin (2011)
14. Lindner, F., Eschenbach, C.: Affordance-based activity placement in human-robot shared environments. In: Herrmann, G., Pearson, M., Lenz, A., Bremner, P., Spiers, A., Leonards, U. (eds.) Social Robotics – Proceedings of the 5th Int. Conf. (ICSR 2013). LNCS, vol. 8239, pp. 94–103. Springer (2013)
15. Pacchierotti, E., Christensen, H.I., Jensfeld, P.: Human-robot embodied interaction in hallway settings: a pilot user study. In: 2005 IEEE Int. Workshop on Robots and Human Interactive Communication. pp. 164–171 (2005)
16. Pandey, A.K., Alami, R.: A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In: Proceedings of the 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). pp. 5855–5860 (2010)
17. Raz, J.: From Normativity to Responsibility. Oxford University Press (2011)
18. Rios-Martinez, J., Spalanzani, A., Laugier, C.: Understanding human interaction for probabilistic autonomous navigation using risk-rrt approach. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ Int. Conf. on. pp. 2014–2019. IEEE (2011)
19. Sisbot, E.A., Marin-Urias, L.F., Alami, R., Simeon, T.: A human aware mobile robot motion planner. IEEE Transactions on Robotics 23(5), 874–883 (2007)
20. Skantze, G., Hjalmarsson, A.: Towards incremental speech generation in dialogue systems. In: Proceedings of SIGdial. Tokyo, Japan (2010)
21. Sommer, R.: Personal Space: Behavioural Basis of Design. Pentice Hall (1969)
22. Tomari, R., Kobayashi, Y., Kuno, Y.: Empirical framework for autonomous wheelchair systems in human-shared environments. In: Proceedings of the 2012 IEEE Int. Conf. on Mechatronics and Automation. pp. 493–498 (2012)
23. Yoda, M., Shiota, Y.: The mobile robot which passes a man. In: Procs. of the 6th IEEE Int. Workshop on Robot and Human Communication. pp. 112–117 (1997)