Evaluation of unimodal and multimodal communication cues for attracting attention in human–robot interaction

Citation for published version (APA):

Torta, E., van Heumen, J., Piunti, F., Romeo, L., & Cuijpers, R. H. (2015). Evaluation of unimodal and multimodal communication cues for attracting attention in human–robot interaction. *International Journal of Social Robotics*, 7(1), 89-96. <https://doi.org/10.1007/s12369-014-0271-x>

DOI:

[10.1007/s12369-014-0271-x](https://doi.org/10.1007/s12369-014-0271-x)

Document status and date:

Published: 01/01/2015

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Evaluation of Unimodal and Multimodal Communication Cues for Attracting Attention in Human–Robot Interaction

Elena Torta · Jim van Heumen · Francesco Piunti ·
Luca Romeo · Raymond Cuijpers

Accepted: 13 November 2014 / Published online: 31 December 2014
© Springer Science+Business Media Dordrecht 2014

Abstract One of the most common tasks of a robot companion in the home is communication. In order to initiate an information exchange with its human partner, the robot needs to attract the attention of the human. This paper presents results of two user studies ($N = 12$) to evaluate the effectiveness of unimodal and multimodal communication cues for attracting attention. Results showed that unimodal communication cues which involve sound generate the fastest reaction times. Contrary to expectations, multimodal communication cues resulted in longer reaction times with respect to the unimodal communication cue that produced the shortest reaction time.

Keywords Human–robot interaction · Attention · Socially-assistive robots · Psychometric study

1 Introduction

During a verbal exchange, people tend to focus their attention to the object of the conversation or to their interlocutor. The same should hold for human–robot communication: the

person and the robot focus their attention to the object of the conversation or to one another. The act of sharing attention to the same object during communication is commonly referred to as joint attention.

A robot can use one or multiple communication cues for attracting attention to itself or to the conversational object. The choice of which cues to use depends on the context and whether the task is joint attention or one-to-one communication.

The most common non-verbal interaction cues used in human–robot joint attention are pointing gestures [12–14, 17] and gaze movements [2, 9, 10, 15]. Pointing gestures and gaze movements are both used to attract a person's attention towards the conversational object and, in the majority of joint attention tasks, they are used to disambiguate the content of speech. As an example, a humanoid robot can use pointing gestures to attract the attention of the person to *that* particular object which the robot is describing [17]. In the same way the robot can look at the conversational object to make interlocutors attention switch to *that* object [10].

A robot can also use multiple non-verbal communication cues to attract a person's attention towards itself. This is of importance when the robot needs to convey messages to a person; a very common task for a socially-assistive robot (especially important when the robot cares for elderly people). Besides uttering reference terms, a robot can direct its gaze towards the person [1, 7], start moving towards the person [4], or use gestures or facial expressions to attract people's attention towards itself. For instance, Bruce et al. [1], study how much people feel engaged with the robot when a robot shows facial expressions and tracks them with its gaze. The experiment was carried out in a public space at Carnegie Mellon University and interaction consisted of people answering a questionnaire after being prompted by the robot. Results do not show a significant effect of people tracking on engagement

E. Torta (✉) · J. van Heumen · R. Cuijpers
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: elena.torta@gmail.com

J. van Heumen
e-mail: j.v.heumen@student.tue.nl

R. Cuijpers
e-mail: r.h.cuijpers@tue.nl

F. Piunti · L. Romeo
Universita Politecnica delle Marche, Ancona, Italy
e-mail: francescopiunti89@gmail.com

L. Romeo
e-mail: l.romeo@univpm.it

but results might have been affected by low performance of people recognition system or timing of tracking behavior. Even though some work has already tackled issues related to how the robot can use its multimodal communication channels to change the focus of human attention during conversation [1,4,6,7], little attention has been directed to studying how the different communication channels can be used by the robot to attract people's attention before initiating conversation, especially when the user is already focused on something other than the robot like watching TV. The present paper focuses on two related research questions reporting results from two experiments.

2 Related Work

Information processing is often studied in one single sensory modality with reaction time tasks. In most experiments, the influence of context or other cues on reaction times is minimized. However, previous research and everyday experience suggest that there is interaction between multiple stimuli' modalities and context which affects reaction time. Presenting concurrent stimuli of different modalities can decrease reaction times with respect to presenting one stimulus alone. The latter is referred to as "intersensory facilitation" [3]. The earliest and simplest account of intersensory facilitation is given by the so-called "race model" [3,8]. The model assumes that multimodal sensory stimuli generate activation processes that are concurrent. The winner determines the response time. If we consider processing time as a random variable we can see that expected reaction to multiple stimuli is, on average, equal or shorter than expected reaction time of single stimuli.

There is evidence that regions of the human brain respond to stimuli in more than one modality [5,16,19]. In particular, multimodal stimuli produce greater activation of these regions than unimodal stimuli [5].

Previous perception studies have also shown that concurrent auditory and visual stimuli can enhance the efficiency in detecting subthreshold masked visual stimuli [5,11].

2.1 Research Questions

In a home setting, a socially-assistive robot needs to attract a person's attention before conveying a message. Attention can be gained by using different types of stimuli such as utterances, gestures or facial expressions. The latter stimuli are primarily of a visual (gestures and facial expressions) or an auditory (utterances) nature.

Although the importance of different communication cues is evident, the relative contribution in terms of efficacy is less clear. Also, previous studies have shown that a combination of cues of different nature can enhance the effectiveness when

Fig. 1 The humanoid NAO robot is 52 cm tall and was used during the experiment to attract people's attention

attracting attention [5,16,19], however, it remains less clear how the contribution of different cues change effectiveness in attracting attention when the stimuli are provided by an embodied agent such as a robot.

The present paper investigates the effectiveness, in terms of reaction time, of unimodal and multimodal communication cues a robot can use for attracting attention. In the first study (experiment A) we investigated the effectiveness, in terms of reaction time, of unimodal stimuli presented by the robot (i.e. speech, gestures, change in facial expression, and gaze) to attract a person's attention when he/she is engaged in another activity and thus is not focused on the robot. Results of experiment A were previously introduced in [18].

A robot could attract attention more effectively when using actions that combine cues with different nature. In the second study (experiment B) we investigated the effectiveness of presenting multimodal stimuli for attracting attention compared to unimodal stimuli. Since visual cues are more salient in peripheral vision, in experiment B we also investigated if robot's position influences reaction time to unimodal and multimodal cues.

In order to collect comparable data, experiments A and B took place in the same mimicked living room. Also, the robotic platform used during both experiments was the NAO (see Fig. 1).

3 Experiment A

During experiment A we manipulated the type of action the robot Nao could perform for attracting attention. Each action is representative of a different communication channel: the actions used were (1) attempting to establish eye-contact by looking at the user, (2) changing the robot's facial appearance by blinking its eye LEDs, (3) making gestures by waving the arms and (4) uttering the word "Hello". Apart from action 4, which is verbal and of an auditory nature, all actions are non-verbal gestures of a visual nature. As actions 1 and 3 involve movements of the robot's body parts, the head for action 1 and the arms for action 3, they are more salient in peripheral vision. The waving gesture (action 3) also involves sound of

the robot's motors. We also manipulated the type of information presented to participants before the robot attempted to attract their attention. Information consisted of short video-clips taken from a Dutch local TV channel. The video clips were presented via a TV positioned in front of them. Video clips were arbitrarily marked as either important or not important. Important video clips were presented with a red box in the upper side of the image. The main task for participants was to respond as soon as they noticed the action performed by the robot by pressing a button on a keyboard positioned in front of them. As a side task, participants were instructed to remember the content of the video clips marked as important because they would be questioned about them at the end of the experiment, even though they were in fact questioned about both important and not important news items. Thus, the experiment had a 4 (actions) $\times$ 2 (importance of news) within-subject design with 3 repetitions. In total every participants experienced 24 trials. During every trial we measured the reaction time between the start of the robotic action for attracting attention and the moment the participant pressed the button.

3.1 Participants

Twelve Dutch elderly people took part in each experiment A (age 62–70). People with reduced vision, hearing problems and/or using hearing aids were excluded from participation. Participants received monetary compensation for taking part in the experiment.

3.2 Set-up

The experiment took place in a mimicked living room. Participants were seated on a couch and had a table and a TV screen in front of them. The set-up mimicked a typical home situation in which the elderly person is sitting on the couch and watching TV. The television played news items retrieved from a Dutch news website and was connected to a computer in order to trigger the start of the different video clips. A keyboard was placed on the coffee table in front of the couch where participants were seated. The robot was standing beside the participants and was remotely connected to the computer in order to trigger the different actions at random moments during the broadcast of the TV news. The experiment was supervised by an operator seated in the control room next to the living room. The experiment set-up is visible in Fig. 2.

3.3 Procedure

Participants were welcomed and given a general explanation of the experiment. They then signed an informed consent form. The participants were seated on the couch in front of

Fig. 2 Experiment Set up. *A* TV screen position, *B* participant position *C* keyboard position and *D* robot position

the television with the keyboard placed in front of them on the coffee table. The experimenter showed participants the various actions the robot could perform to attract their attention and explained how the different importance levels of the video clips were indicated. Participants were instructed to remember the important video clips as they would be questioned about them at the end of the experiment. Then the experiment began with video clips appearing on the TV screen in front of the participants. At random moments during the broadcast of the news the robot performed an action to attract participants' attention. Participants were instructed to press the spacebar of the keyboard as soon as possible after noticing the attempt of the robot to attract their attention. After responding, they were instructed to press a button on the keyboard to categorize the action they saw. The time between the start of the action to attract attention and the moment participants pressed the spacebar was recorded. The experiment had a total of 24 trials with 12 video clips marked as important and 12 unmarked. Marked and unmarked video clips were presented in random order to every participant during the experiment. When all the trials were over, participants filled in additional questionnaires and received monetary compensation for their participation.

3.4 Data Analysis

To test the effects of our manipulations on the reaction time of participants we used a factorial repeated-measures

Fig. 3 Mean reaction time in [s] for every action regardless of the importance of notice. (Adjusted values with 95 % confidence interval)

analysis of variance (ANOVA) with reaction time as the dependent variable and both manipulations (importance of news and type of action) as independent variables. To assure that reaction times had a Gaussian distribution, the analysis was performed on data transformed with a logarithmic function.

3.5 Results

To assess the efficiency of the different ways to attract attention we compared the reaction times. In Fig. 3 the reaction times are plotted for each of the robot's actions.

There was a significant main effect of the type of action on reaction time $F(3, 33) = 140.00$, $p < 0.001$. Contrasts analysis revealed that there was a significant difference, in terms of reaction time, between all the types of actions. In particular, reaction time in the case of eye contact was longer than in the case of blinking the eye LEDs, $F(1, 11) = 140.84$, $p < 0.001$. Reaction time was longer when the robot blinked its eyes compared to waving $F(1, 11) = 8.67$, $p = 0.013$. And finally, reaction time was longer when the robot waved its arms compared to speech $F(1, 11) = 26.65$, $p < 0.001$. The manipulation of news importance had no significant effect on reaction time and no interaction effect between the two manipulations was found.

3.6 Discussion

In terms of reaction time speech was most salient as the average response was fastest, followed by the waving gesture and eye LED blinking gesture. Trying to establish eye-contact was least salient. It is interesting to note that speaking and waving both involve an auditory component either in the form of speech utterance or in the form of noise produced by the robot's actuators, whereas flashing LEDs and eye contact are of a visual nature. Thus, it seems that sound, which is an omnidirectional cue, is more salient than the visual channel.

This makes sense because the visual attention was directed towards the TV. The high acuity foveal area spans only about 5° , so that the robot moved in the participant's periphery. These facts can explain why actions which involve sound production cause a quicker response than actions that involved only visual stimuli. In the same line of reasoning the making eye contact action involves less motion (because the robot had to turn its head only of a few degrees to look at participants) than waving and flashing eye LEDs. It is well-known that peripheral vision is very sensitive to motion explaining why making eye contact is less salient than the other visual cues for attracting attention.

4 Experiment B

Results of experiment A showed that different cues generates different reaction times and that sound resulted in the shortest reaction time. They also showed that the manipulation importance of news appeared not to have an effect on reaction times. A combination of cues might lower reaction times and the position of the robot might have an influence on reaction times related to cues of a visual nature. Therefore, in experiment B, we manipulated the type of action the robot Nao could perform including multimodal cues and the angular position of the robot with respect to the person. The actions used were (a) changing the robot's facial appearance by blinking its eye LEDs, (b) changing the robot's facial appearance by blinking its eye LEDs and uttering the word "Hallo" (c) waving the robot's arms (d) waving the robot's arms and uttering the word "Hallo" and (e) uttering the word "Hello". Actions were of a visual, auditory or visual and auditory nature. We also manipulated the angular position of the robot with respect to participants. The distance between the robot and the person was 2 m and the angular locations were 30° , 45° and 60° from the vertical line that connected the TV to the participant. Experiment B was carried out in the same mimicked living room as experiment A (see Fig. 4). As for experiment A, participants were instructed to press the spacebar of a keyboard as soon as they were noticing the action performed by the robot. The moment between the start of the action and the button press was recorded as reaction time. Thus, experiment B had a 5 (actions) $\times$ 3 (angular positions) within-subject design with three repetitions. In total every participants experienced 45 trials. During every trial we measured the reaction time between the start of the robotic action for attracting attention and the moment the participant pressed the button.

4.1 Participants

Twelve people took part at the experiment. People with reduced vision, hearing problems and/or using hearing aids were excluded from participation.

Fig. 4 Layout of experiment B. The robot was positioned at a distance of 2 m from participants and could be either be at 30°, 45° or 60° from the vertical line that connect the TV set to the participant

4.2 Data Analysis

To test the effects of our manipulations on the reaction time of participants we used a factorial repeated-measures analysis of variance (ANOVA) with reaction time as the dependent variable and both manipulations (type of action and robot's position) as independent variables. As in experiment A, data were transformed with a logarithmic function to assure that reaction times had a Gaussian distribution. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of action, $\chi^2(9) = 18.9$, $p = 0.028$. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\epsilon = 0.574$).

4.3 Results

Data analysis shows a significant main effect of the type of action on reaction time $F(2.296, 25.25) = 8.77$, $p = 0.001$. Contrasts analysis revealed that there was a significant difference between reaction time in the case of a single cue of visual nature and the same cue presented with speech. As one would expect, participants reacted faster to blinking the eye-LED while uttering the word hallo than to blinking alone ($F(1, 11) = 12.75$, $p = 0.002$). On the same line participants reacted faster to waving and saying “Hallo” with respect to waving only ($F(1, 11) = 15.36$, $p = 0.002$). Contrary to expectations, adding a visual cue to sound resulted in longer reaction times compared to sound only, even though the contrast is significant only between waving while uttering the word “Hallo” and uttering the word “Hallo” only ($F(1, 11) = 7.66$, $p = 0.018$). Average reaction times per action are shown in Fig. 5.

Contrary to expectations, a change in robot's position did not influence reaction times significantly.

4.4 Discussion

As one would expect from the results of experiment A, sound generated the fastest reaction time followed by waving and blinking the eye-LED. Contrary to expectations, adding a visual cue to sound did not improve reaction time. On the contrary, reaction times were longer when adding waving or blinking to speech. The difference between results reported in the current paper and the ones presented in [5, 16, 19] could be explained by looking at the embodiment of the source of visual and auditory stimuli. In our experiment the source of the stimuli was the robot that remained constantly visible during the experiment, participant might have been distracted rather than stimulated by a combination of visual and auditory stimuli and, as a result, their reaction time was in between reaction time when only the less salient cue was presented (waving or blinking) and only the most salient cue (speaking) was presented.

A change in robot's position did not affect reaction time significantly. A robot approaching a seated person that watches TV is likely to approach from the side. This poses constraints on the possible robot's position in the environment for attracting attention. In our study we positioned the robot in places it is likely to end up when approaching a person who is watching TV. Results of our study suggest that there is no difference, in terms of effectiveness of reaction, when the robot performs the actions from a range between 30° and 60° from the vertical line connecting the TV to the participant 4.

Fig. 5 Mean reaction time in [s] for every action regardless of the importance of notice. (Adjusted values with 95 % confidence interval)

5 General Discussion

The present section compares results of experiments A and B. The average reaction times per experiment per action are reported in Table 1. As already reported reaction times for cues of visual nature are longer than for unimodal or multimodal cues with an auditory component. During experiment B participants reacted faster to sound with respect to experiment A.

We can also notice that participants took the same time to react to waving in both experiments A and B whereas they reacted faster to blinking in experiment B than in experiment A. The explanation for the observed difference with respect to blinking can be linked to the level of ambient luminance. During experiment B we decreased the level of ambient luminance by turning off some lights in the mimicked living room environment. This increased the contrast between the phases of eye-LED off and on. As a result, the blinking of the eye-LED was more noticeable in experiment B than in experiment A which, we argue, has caused the decrease in reaction time observed in experiment B for the action “blinking”.

We can also observe that reaction times are similar for experiment A and B. In both experiment participants were requested to watch TV while waiting for the action of the robot to attract attention. During experiment A participants were also requested to remember the content of the short video clips which were shown on TV. During experiment B we did not ask participants to remember the content of the TV program explicitly, also the program displayed on the TV was different than the news clips used in experiment A. Since in both experiments, reaction times were similar, we can argue that the content of the TV program, does not influence reaction times significantly.

6 Conclusion

The present paper reported experimental results related to the evaluation of the effectiveness of unimodal and multimodal communication cues a robot can use for attracting a

Table 1 Average reaction times per action per experiment

Cue	Experiment	Average RT [s]	SE [s]
Eye contact	A	4.8	1.025
Blinking		2.7	1.038
Waving		2.3	1.027
Speaking		1.9	1.026
Blinking	B	2.3	1.017
Blinking and speaking		1.8	1.008
Waving		2.3	1.010
Waving and speaking		2.0	1.010
Speaking		1.7	1.016

person’s attention in a naturalistic home scenario. Effectiveness was measured with a reaction time task. Participants were instructed to press a button as soon as they noticed the action of the robot for attracting attention.

Results of experiment A suggest that utterances, which are communication cues with an auditory nature, result in the fastest reaction times compared to cues of a visual nature. The saliency of visual cues depends on environmental conditions such as ambient luminance. In particular, a comparison between reaction times in experiment A and B shows that reaction times to the action blinking decrease if the contrast between the off and on cycle of the LED is increased. We increased the contrast by lowering the environment’s luminance level of the mimicked room. The saliency of visual cues appear not to be influenced by the position of the robot in the range (30°–60°) from the vertical line connecting the participant to the TV set (see Fig. 4).

Interestingly, we did not find a significant difference in terms of reaction time when participants were presented with important and not-important news items for experiment A. Furthermore, reaction times per cue in experiment A were similar to reaction times per cue in experiment B. During experiment B we did not request participants to focus on the TV show. So it seems that, apart from a spatial effect, being more focused does not influence the participant’s ability to notice the actions performed by the robot.

Average reaction times for the auditory cue in experiment A (1.9 s) and in experiment B (1.7 s) are similar. In both experiments the sound came from an embodied source visible to participants, i.e. the robot. Therefore, results of both experiments do not allow to quantify differences in reaction time if participants would not see the source of the sound.

We have also investigated if multimodal cues resulting from the combination of auditory and visual stimuli are more effective when attracting attention than unimodal cues. Results from experiment B suggest that this is not the case and that reaction time, when multiple cues are superimposed, is longer than the average reaction time of the more salient cue. Previous research [11] reported that an auditory stimulus can be perceived as louder if it is presented with visual stimuli in the form of lights. However our results show that this is not necessarily the case for all visual stimuli. In particular, if visual stimuli are presented by an embodied source in the environment such as a robot. In conclusion our results suggest that the robot can effectively attract attention of people using auditory, visual, verbal and non-verbal actions. Attention is obtained faster when using speech followed by gestures. Contrary to our expectations, using multimodal cues for attracting attention appears not to decrease reaction time. The robot's positions tested in our experiment also appear not to influence reaction times.

Acknowledgments The research leading to these results was partially supported by the KSERA project (<http://www.ksera-project.eu>) and has received funding from the European Commission under the 7th Framework Programme (FP7) for Research and Technological Development under grant agreement no 2010-248085. The authors gratefully acknowledge the help received from Martin Boschman for the realization of the experiment.

References

- Allison B, Nourbakhsh I, Simmons R (2002) The role of expressiveness and attention in human–robot interaction. In: Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on, vol 4. pp 4138–4142
- Chaminade T, Okka MM (2013) Comparing the effect of humanoid and human face for the spatial orientation of attention. *Front Neurobot* 7:12
- Diederich A (1995) Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *J Math Psychol* 39(2):197–215
- Finke M, Koay KL, Dautenhahn K, Nehaniv CL, Walters ML, Saunders J (2005) Hey, i'm over here-how can a robot attract people's attention? In: Robot and human interactive communication, 2005. ROMAN 2005. IEEE international workshop on. IEEE. pp 7–12
- Frassinetti F, Bolognini N, Làdavas E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res* 147(3):332–343
- Hoque MM, Onuki T, Kobayashi Y, Kuno Y (2011) Controlling human attention through robot's gaze behaviors. In: Human system interactions (HSI). 2011 4th International Conference on. IEEE, pp 195–202
- Hoque MM, Deb K, Das D, Kobayashi Y, Kuno Y (2013) An intelligent human–robot interaction framework to control the human attention. In: Informatics, electronics & vision (ICIEV). 2013 International Conference on. IEEE, pp 1–6
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn psychol* 14(2):247–279
- Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human–robot conversations: how robots might shape participant roles using gaze cues. In: HRI'09 Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp 61–68
- Nagai Y, Hosoda K, Morita A, Asada M (2003) A constructive model for the development of joint attention. *Conn Sci* 15(4):211–229
- Odgaard E, Ariele Y, Marks L (2004) Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cogn Affect Behav Neurosci* 4:127–132
- Salem M, Eyssel F, Rohlfing K, Kopp S, Joubin F (2011) Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In: Mutlu B, Bartneck C, Ham J, Evers V, Kanda T (eds) Social robotics, vol 7072. Lecture notes in computer science Springer, Berlin, pp 31–41
- Schauerte B, Fink GA (2010) Focusing computational visual attention in multi-modal human-robot interaction. In: International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction. ACM, New York, p 6
- Clair AS, Mead R, Mataric MJ (2011) Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In: RO-MAN, 2011. IEEE, pp 210–216
- Staudte M, Crocker MW (2011) Investigating joint attention mechanisms through spoken humanrobot interaction. *Cognition* 120(2):268–291
- Stein BE, Meredith MA (1993) The merging of the senses. The MIT Press, Cambridge
- Sugiyama O, Kanda T, Imai M, Ishiguro H, Hagita N, Anzai Y (2006) Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection science* 18(4):379–402
- Torta E, Heumen J, Cuijpers RH, Juola JF (2012) How can a robot attract the attention of its human partner? a comparative study over different modalities for attracting attention. In: Ge S, Khatib O, Cabibihan J-J, Simmons R, Williams M-A (eds) Social robotics, vol 7621. Lecture notes in computer science Springer, Berlin, pp 288–297
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91(3):484–488

Elena Torta received her M.S. in Industrial Automation Engineering from Marche Polytechnic University (IT), in 2009 and her doctoral degree from the Eindhoven University of Technology (NL) in 2014 in the context of the European FP7 project Knowledgeable Service Robots for Aging (KSERA). Her main research interests include control system design, ambient assisted living and human–technology interaction.

Jim van Heumen is a master student at the department of Human–Technology interaction at the Eindhoven University of Technology (NL). His main research interests include human–robot interaction and innovation sciences.

Francesco Piuñti received his M.S. in Electronic Engineering from Marche Polytechnic University (IT) in 2014. He is currently research engineer at Loccioni Industry, Ancona (IT). His main research interests

include control system design, smart mobility and human–technology interaction.

Luca Romeo received his M.S. in Electronic Engineering from Marche Polytechnic University (IT) in 2014. He is currently PhD student at the same university. His main research interests include control system design, ambient assisted living and human–technology interaction.

Raymond Cuijpers graduated in Applied Physics at the Eindhoven University of Technology (NL) in 1996. He received his Ph.D. degree in Physics of Man from Utrecht University in 2000. Since 2014 he is associate professor at the Eindhoven University of Technology (NL) at the human–technology interaction group. He was project coordinator of the European FP7 project Knowledgeable Service Robots for Aging (KSERA). In this context he studied how robots can autonomously interact with humans for helping older people.