Natural Listening Robot for AAL Applications

Yasser Mohammad and Toyoaki Nishida
Kyoto University
yasser@ii.ist.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

ABSTRACT

The increase in the proportion of senior citizens to the total population is both a challenge and an opportunity for industrial societies. To meet this challenge an increasing research effort is being allocated to the area of Ambient Assisted Living (AAL). Most of the research in this area is concerned with designing novel interfaces and novel services based on embedded intelligence targeting the elderly, and architectures to integrate heterogeneous sensors and solutions. In this paper we argue that a mobile robot companion can provide benefits to this program that are beyond what is possible using only embedded intelligence techniques. We further argue for the benefits of utilizing the robots as knowledge media paradigm for AAL systems. This paper then describes an ongoing effort to realize a humanoid robot that can use human like nonverbal behavior to give the interacting human a *natural* listening experiment as a first step for realizing a useful knowledge media companion robot for AAL applications. The details of a proof of applicability study of our approach is given and discussed in relation to AAL.

WHY A ROBOT

The direct application of robots in AAL systems is to meet the physical needs of the elderly by manipulating different objects in the home, assist in mobility, etc. [1]. In this paper we are more concerned with the psychological, social, and disaster-prevention role of robots in the AAL home. From this point of view, robots can provide several important additions to any AAL system including:

1. Robots are embodied agents, which means that they live in the physical domain and therefore can be active in manipulating the environment to facilitate the life for the accompanied senior agents. For example, robots can not only detect a fall condition, which is a major concern for many AAL systems, but it can assist in recovering or even preventing it. A robotic nurse can not only remind the accompanied human of the medication time, but can also bring the medicine to her if it was not easily accessible. Another area in which robots can assist is efficient and accurate localization of the accompanied person which needs otherwise a complex and expensive localization system. The main point is that a robotic companion is more like a human caregiver than a complex network of embedded devices.

2. As shown in [2], humans tend to anthropomorphize robots more than other intelligent agents even if they have exactly the same appearance. This ability to act as a personalized companion is a critical need for senior citizens who suffer usually from a loneliness problem. Although robots, with the current state of the are technologies, cannot provide a complete companion *person*, they can in principle provide an interesting interacting companion *agent*.

3. In [3], it was shown that any AAL system should evolve with the senior citizen building on already familiar services. A robot that can accompany the citizen for a long period of time and learn her own style of life can be very beneficial from this point of view. Another potential supporting point of designing companion robots that target AAL applications is the envisioned increase in the number of robots that will be living in our homes in the near future . This means that the senior citizen may be already familiar with dealing with personal assistant robots even before she needs AAL support.

Other than providing physical and psychological benefits for the senior citizens, robots can work as interactive knowledge media through which the senior citizen can transfer her own experiences to other family members and through which she can get useful knowledge when needed which gives a social dimension to the benefits of augmenting the AAL system with a robot. An example of such a use of robots is the Knowledge Medium Robot proposed by the authors in [4]. This robot can use natural means of communication to acquire knowledge from human explanations, and to provide it as needed. Such a robot can have several uses for the senior citizen like automatically generating multimedia content summarizing her life or everyday activities which can be transferred to other family members under her control. Another use of such a robot can be as a form of an embodied mobile external memory that can partially compensate for the memory problems associated with aging. Yet another application for such a robot is to replace boring and complex operating manuals of different *intelligent* artifacts that exist in the home with a life interactive experience that is more fun and easier to learn and manage for senior citizens.

The main obstacles that face the wide utilization of personal robots in the AAL applications are robot prices which is expected to decrease dramatically with wide adoption of personal and service robots in the next years and the need of special technical training for operating current robots. [3] demonstrated that living assistance systems must realize flexibility and adaptability at the algorithmic, architectural and human interface level to an extent unknown in present systems. This intensifies the need for robots that can use human like verbal and nonverbal communication channels in the field of AAL.

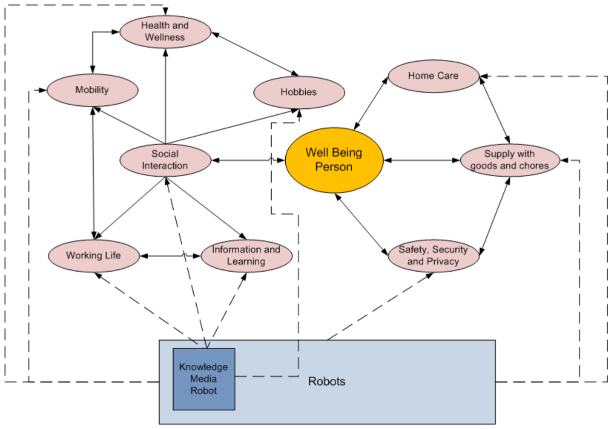


Figure 1. The needs and opportunities of the AAL program as depicted in AAL169 annotated with services that can be provided by robots

As shown in Fig. 1 robots are not intended to replace the intelligent supporting environment envisioned in most current research in AAL but to augment it with new services that are only possible using physically embodied agents.

In the rest of this paper, the author's ongoing effort to realize a robot that can give a natural listening experiment to the human partner using only the nonverbal aspects of the interaction under the umbrella of the *robots as knowledge media* framework advocated in [4] and [5] is detailed with some preliminary results that provide a proof of applicability of the proposed approach.

A ROBOT THAT LISTENS

Natural listening is defined in this paper in a subjective sense by the ability of the robot to use human-like nonverbal behavior to convey its attention to the speaker so that the cognitive load on the speaker is minimum. To realize the provisioned listening robot, a robotic architecture that can easily combine interactivity with autonomy, and that allows for incremental design is needed. The authors' proposed the Embodied Interactive Control Architecture (EICA) architecture to meet this challenge. For details about this architecture refer to [6].

EICA is a massively parallel system and the main active entities in the EICA architecture are:

1. Intentions that represent simple reactive plans that generate a short path control mechanism from sensing to actuation. The action integration mechanism provides the means to integrate the actuation commands generated by all running *intentions* into final actions sent to the actuators of the robot based on the *intentionality* assigned to every *intention*.

2. Processes that provide a high level control over the behavior of the robot by controlling the temporal evolution of the *intentionality* of various *intentions*. The listener robot utilizes only three simple processes but the interaction between them results in the emergence of complex behavior that mimics human nonverbal listening.

3. Reflexes that are the only running processes in EICA that can bypass the action integrator and send direct commands to the actuators. Reflexes provide safety services like collision avoidance during navigation, or safety measures to prevent any possible accidents to the interacting person due to any failure or bugs in higher level modules.

4. Low Level Emotions that summarize the sensed environment and internal state of the robot into a set of *modes* that can be utilized to provide a simple form of pseudo-personality for the robot and as an efficient mechanism to react to environmental changes. Currently the listener robot does not utilize this feature of the EICA system.

The design principles of the EICA architectures are detailed in [6] but it is important for the sake of this discussion to clarify how a massively parallel system like EICA can produce coherent behavior that is perceived as *Intentional* by human partners. Intention is classically defined as *a goal with commitment*. The major point stressed by this definition is that intentional behavior should have a form of inertia that makes the agent committed to the goal of its actions. In EICA this form of inertia can emerge only from the careful timing of the adjustment of the *intentionality* of various base intentions. This reliance on accurate timing may appear as a disadvantage in the first glance, but in actual robot programming situations the correct division of the robot behavior between ground intentions and adjusting processes can dramatically reduce the effort needed to find the required timing parameters. Another option that is not explored in this study but will be explored in later experiments is to use machine learning techniques to select those timing parameters which opens an opportunity for continuous learning of the interaction. In this way the EICA system can adapt to the interaction task and the human partner which is an important feature for AAL systems as argued by [7].

As a proof of concept a pilot study of the interactive behavior of the robot using a minimalist approach was employed. The goal of this study was to check the applicability of the EICA architecture in this domain, and to realize a behavior pattern that can be objectively compared with the human-human known behavior in close encounters. The average time of showing four kinds of behaviors were selected as a measure of the system performance because of the availability of objective human-human data [8],[5], their relative simplicity, and their important role in natural interactions [9], namely:

1. Mutual Attention which is defined here as the behavior of attending (looking) to the same object attended to by the speaker.
2. Gaze Toward Instructor which is defined as looking within the face area of the speaker.
3. Mutual Gaze which is defined here by having the two interacting agents looking within the face area of each other in the same time.
4. Initial Eye Contact which is defined here as the first mutual gaze instant in the interaction.

As a minimal design, only the head of the robot was controlled during this experiment. This decision was based on the hypothesis accepted by many researchers in the nonverbal human interaction community that gaze direction is one of the most important nonverbal behaviors involved in realizing natural listening in human-human close encounters [8].

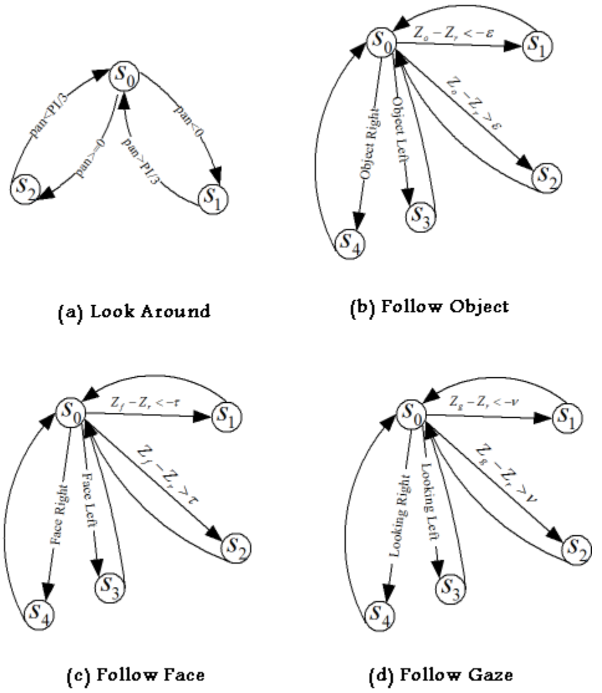
The evaluation data was collected as follows:

1. Six different explanation scenarios were collected in which a person is explaining the procedure of operating a hypothetical machine that involves pressing three different buttons, rotating a knob, and noticing results in an LCD screen in front of a Robovie II robot while pretending that the robot is listening to the explanation. The data was collected using the PhaseSpace Motion Digitizer system [10] by utilizing 18 LED markers attached to various parts of the speaker's body. The data was logged 460 times per second.
2. The logged data were used as the input to the robot simulator and the behavior of the robot's head was analyzed.
3. For every scenario 20 new synthetic scenarios were generated by utilizing 20 different levels of noise. The error level is defined as the percentage of the mean value of the noise term to the mean of the raw signal. The behavior of the simulator was analyzed for every one of the resulting 120 scenarios and compared to the original performance.
4. The same system was used to drive the Robovie II robot and the final behavior was subjectively studied.

It should be noted that in the current setup of the experiment the speaker's behavior cannot be assumed completely normal because the robot is not actually executing the natural listening behavior in real time which breaks the interaction loop. This is the reason that in the current proof of concept experiment only the means of the time of execution of various interaction behaviors are compared to the known human-human means. It was assumed that the effect of the inaccuracies resulted from not using a human in the listener position will have smaller effect on the means compared to its effects on the dynamics of the behavior. In the full scale experiment the listening robot will be actually *listening* to the instructor using the software presented in this paper and the interaction loop will be closed which will make it possible to compare the dynamics of the robot's behavior to the human-human case.

Design

Four reactive intentions were designed that encapsulate the possible interaction actions that the robot can generate, namely, looking around (Fig. 2-a), following the human face (Fig. 2-a), following the salient object in the environment (Fig. 2-c), and looking at the same place the human is looking at (Fig. 2-d). The sufficiency of those intentions was based on the fact that in the current scenario the robot simply have no other place to look, and the necessity was confirmed empirically by the fact that the three behavioral processes needed to adjust the intentionality of all of these intentions.



(a) Look Around

(b) Follow Object

(c) Follow Face

(d) Follow Gaze

Figure 2. Four basic intention implementations used to achieve mutual attention

The analysis of the mutual attention requirements showed the need of three behavioral processes. Two processes to generate an approach-escape mechanism controlling looking toward the human operator which is inspired by the *Approach-Avoidance* mechanism suggested in [8] in managing spacial distance in natural human-human situations. These processes were named Look-At-Human, and Be-Polite. A third process was needed to control the realization of the mutual attention behavior. This process was called Mutual-Intention. The details refer to [4]. A brief description of them is given here:

1. *Look-At-Human*: This process is responsible of generating an attractive virtual force that pulls the robot's head direction to the location of the human face. This process first checks the Gaze-Map's current modes and if their weights are less than a specific threshold for more than 10 seconds, while the human is speaking for more than 4 seconds , it increases the intentionality of the *followFace* intention and decreases the intentionality of the other three reactive intentions based on the difference in angle between the line of sight of the human and the robot and the *Confirming* condition (if the human is confirming the robot should look at him more as was noted in [9]).

2. *Be-Polite*: This process works against the *Look-At-Human* process by decreasing the intentionality of the *followFace* intention in reverse proportion to the angle between the line of sight of the human and the robot depending on the period the human is speaking.

3. *Mutual-Attention*: This process increases the intentionality of the *followObject* or the intentionality of the *followGaze*. The rate of intentionality increase is determined based on the confirmation mode.

Five perception processes were needed to implement the aforementioned behavioral processes and intentions:

1. *Human-Head*, which continuously updates a list containing the position and direction of the human head during the last 30 seconds sampled 50 times per second.

2. *Robot-Head*, which continuously updates a list containing the position and direction of the robot head during the last 30 seconds sampled 50 times per second.

3. *Gaze-Map*, which continuously updates a representation of the distribution of the human gaze both in the spacial and temporal dimensions. The spacial distribution is stored as a mixture-of-Gaussians like structure where the mean μ_i represents the location of an important object and the variance σ_i is a measure of the size of that object. The weight w_i represents the importance of the place according to the gaze of the human. The details of this process will not be given here due to lack of space, refer to [4] for details.

4. *Speaking*, which uses the power of the sound signal to detect the existence of human speech. The current implementation simply assumes there is a human speech whenever the sound signal is not zero. This was acceptable in the simulation but with real world data a more complex algorithm that utilizes fourier analysis will be used.

5. *Confirming*, which specifies whether or not the human is making a confirming action. Currently this value is manually added to the logged data although the algorithm proposed in [9] can be used in future experiments.

Fig. 3 shows the evolution of intentionality of the aforementioned four basic reactive intentions under the control of the three control processes used to implement natural interaction in one case.

In the beginning the robot was scanning the environment for salient features that require attention. The interaction with the human started when the human directed his gaze to the robot for a few seconds. The *Look-At-Human* process increased the intentionality of the *followFace* intention while decreasing the intentionality of the *LookAround* Intention which initialized the eye contact that started the interaction. After a while (21.3 seconds in average) the *Be-Polite* process takes over reducing the intentionality of the *followFace* which along with the increased intentionality of *followGaze* and *followObject* results in breaking the eye contact. As the interaction goes the robot will tend to look to the human for around 77.87% of the time while attending to the shared objects of interest around 53.12% of the time in which the human is attending to those objects.

RESULTS AND DISCUSSION

Table 1. Comparison Between the Simulated and Natural Behavior

Item	Statistic	Simulation	H-H value
Mutual Gaze	Mean Std.Dev.	31.5% 1.94%	30% –
Gaze Toward Instructor	Mean Std.Dev.	77.87% 3.04%	75% –
Mutual Attention	Mean Std.Dev.	53.12% 4.66%	– –
Initial Eye Contact	Mean Std.Dev.	21.3sec 5.47%	15–45sec –

To analyze the applicability of EICA to the natural listening behavior an objective evaluation criteria was selected. For the four behaviors chosen the mean and standard deviation of the time spent doing each of them were calculated and compared with the known values in natural human-human interactions when available.

Some of the results of numerical simulations of the listening behavior of the robot are given in Table 1. The table shows the average value obtained from the simulated robot in comparison to the known values measured in human-human interaction situations. The source of the average time in the human-human case are reported from [8]. As the table shows the behavior of the robot

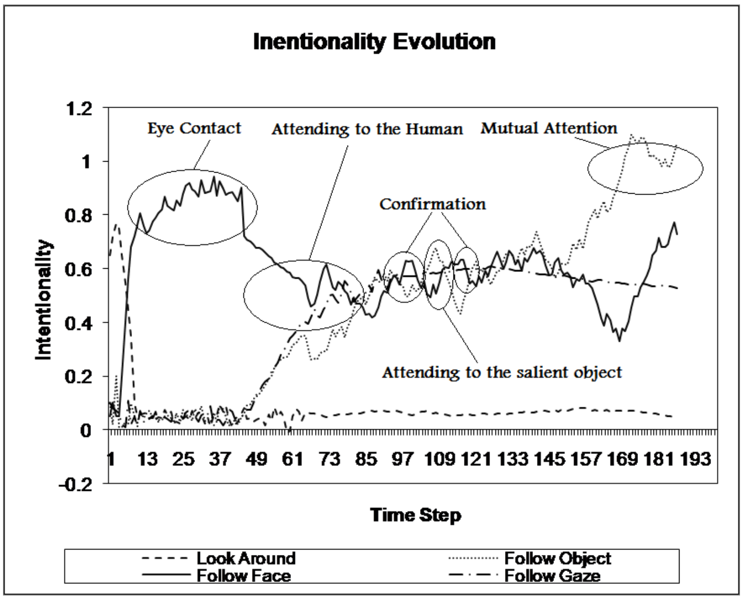


Figure 3. The Evolution of Intentionality of the four Basic Intention Implementing Natural Listening Behavior

is similar to the known average behavior in the human-human case for both mutual gaze and gaze toward instructor behaviors and the standard deviation in both cases is less than 7% of the mean value which predicts robust operation in real world situations. These results suggest that the proposed approach is at least applicable to implement natural listening behavior.

It should be noted that the naturalness of the behavior does not only depend on the averages specified but on the detailed movements of the head and eye during the interaction. As a first step in analyzing the dynamical behavior of the robot, the proposed system was used to drive the Robovie II robot and the resulting video of the robot was aligned to the explanation used to get the data. Some of the lessons learned evaluating this video are listed here.

- The head movement alone is not enough for realizing natural listening as the fine control of the eye direction is also needed.
- The proposed system can sometimes cause more total movements of the robot head than in natural situations which means that driving the parameters of the behavioral processes from a large set of training examples is needed.
- In general, the behavior of the robot looks more or less similar to a human listener and can cause a suspense of disbelief which is a promising result of the proposed system.

In AAL applications the environment in which the robot is to function is expected to have high levels of noise due to the fact that it is not a controlled environment and because of the expected existence of other intelligent embedded electronic devices that generate interference signals. For this reason it is essential for this application to study the effect of noise.

Fig. 4 shows the effect of increasing the error level on the percentage of time mutual gaze, gaze toward instructor, and mutual attention behaviors were recognized in the simulation. As expected the amount of time spent on these interactive behaviors decreases with increased error level although this decrease is not linear but can be well approximated with a quadratic function. Regression Analysis revealed that in the three cases the effect on the mean time spent doing the studied behavior grows with the quadrable of the inverse SNR (Signal to Noise ratio) . This result suggests that the EICA based implementation is robust to low levels of correlated noise and that the system has a graceful degradation of performance with the increase of noise level. As shown in the previous paragraph, this is an appealing feature from the point of view of AAL applications.

LIMITATIONS AND FUTURE DIRECTIONS

The main limitations of the proposed system are the fact that it only controls the head of the robot neglecting the possible effect of other body movements on the final behavior and ignoring the effect of the verbal behavior of the speaker on the nonverbal behavior of the listener. Those limitations were accepted to simplify the problem in this first exploration.

Another problem with the current system is its reliance on hand tuning of various parameters, the small number of interactions used in the design, and ignoring the dynamical aspects of the behavior in the evaluation. All of those problems will be addressed in the near future by

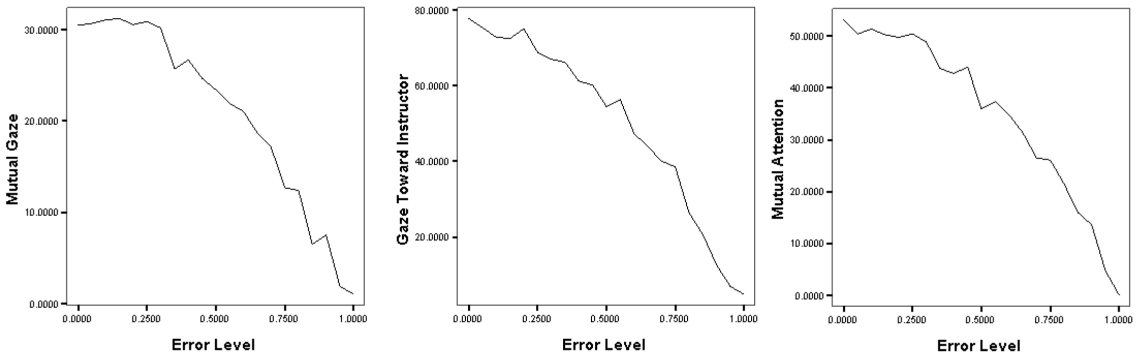


Figure 4. Effect on the error level on the behavior of the robot

using a learning component that is trained using a large scale human-human experiment, and incorporating objective physiological signals, subjective evaluations and dynamics of the behavior in the evaluation.

Another important point to address in future research is to find an objective way to measure the naturalness of the interaction. We are currently investigating using various kinds of physiological signal analysis and external behavior analysis for the sake of finding this measure.

CONCLUSION

This paper argued that a companion robot that uses natural human-like interaction means of communication can be a valuable addition to any AAL system by providing services that otherwise need a human caregiver. The possible services that can be provided by a specific kind of interactive robots called Knowledge Medium Robots was also illustrated. A proof of applicability experiment with a robot that can exhibit *natural* nonverbal listening abilities based on the processing of motion tracking and sound signals of the speaking human partner is also presented showing that the proposed system can provide a listening behavior comparable to human-human nonverbal listening in terms of four basic dimensions of comparison. This feature makes the proposed robot suitable for AAL applications. The error analysis revealed that the proposed system is robust to small levels of signal correlated noise which is another attractive feature from the AAL point of view.

REFERENCES

1. G. Riva, F. Vatalaro, F. Davide, and M. Alcaniz. *Ambient Intelligence: The Evolution Of Technology, Communication And Cognition Towards The Future Of Human-Computer Interaction.* Ios Pr Inc, March 2005.

2. C. D. Kidds and C. Breazeal. Effect of a robot on user perceptions. In *IEEE/RSJ Conference on Intelligent Robots and Systems 2004 (IROS 2004)*, volume 4, pages 3559–3564. IEEE, September 2004.

3. Stinne Aalokke Ballegaard, Jonathan Bunde-Pedersen, and Jakob E. Bardram. Where to, roberta?: reflecting on the role of technology in assisted living. In *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, pages 373–376, New York, NY, USA, 2006. ACM.

4. Yasser F. O. Mohammad, Taku Ohya, Tatsuya Hiramatsu, Yasuyuki Sumi, and Toyoaki Nishida. Embodiment of knowledge into the interaction and physical domains using robots. In *International Conference on Control, Automation and Systems*, October 2007. to appear.

5. Takashi Takima Makoto Hatakeyama Yoshiyasu Ogasawara Yasuyuki Sumi Yong Xu Yasser Mohammad Kateryna Tarasenko Taku Ohya Toyoaki Nishida, Kazunori Terada and Tatsuya Hiramatsu. Toward robots as embodied knowledge media. *IEICA Trans. Inf. and Syst.*, E89-D(6):1768–1780, June 2005.

6. Yasser F. O. Mohammad and Toyoaki Nishida. A new, hri inspired, view of intention. In *AAAI-07 Workshop on Human Implications of Human-Robot Interactions*, pages 21–27, July.

7. Jürgen Nehmer, Martin Becker, Arthur Karshmer, and Rosemarie Lamm. Living assistance systems: an ambient intelligence approach. In *ICSE '06: Proceeding of the 28th international conference on Software engineering*, pages 43–50, New York, NY, USA, 2006. ACM.

8. Michael Argyle. *Bodily Communication.* Routledge; New Ed edition, 2001.

9. Toyoaki Nishida T. Tajima, Y. Xu. Entrainment based human-agent interaction. In *IEEE Conference on Robotics, Automation, ans Mechatronics*, December 2004.

10. http://www.phasespace.com/, 2007.