Workshop on Multimodal Semantics for Robotic Systems (MuSRobS)
IEEE/RSJ International Conference on Intelligent Robots and Systems 2015

A Multi-modal Approach for Assistive Humanoid Robots

German I. Parisi, Johannes Bauer, Erik Strahl and Stefan Wermter

Abstract— **Mobile assistive robots can enhance elderly people's perception of safety and prevent loss of confidence at home. Therefore, multi-modal systems that allow robots to operate in complex environments represent an enticing milestone for self-care and independent living applications. We present a humanoid robot that assists a person in daily activities and detects situations of danger such as fall events. Our system integrates multiple sensor modalities to enhance the perception of the robot through visual active tracking, sound source localization, and automatic speech recognition. Robot motor control is triggered from the interplay of audio-visual cues conveyed by onboard sensors. We propose a multi-modal controller to modulate sensor-driven behaviour of the humanoid robot Nao and present preliminary results in a home-like environment for a fall detection scenario.**

I. INTRODUCTION

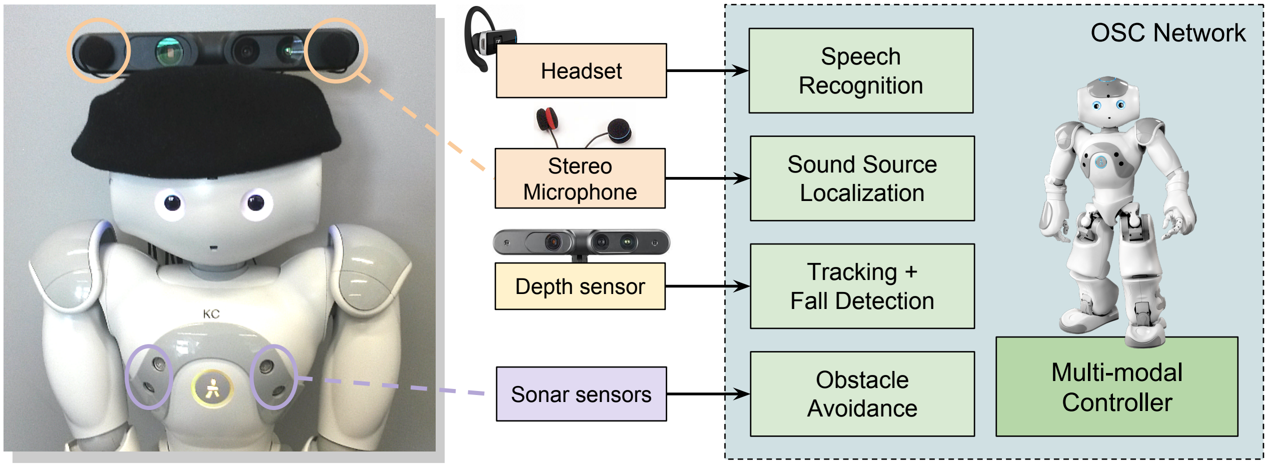
Injuries caused by falling have been identified as the leading cause of loss of independence and premature death in elderly people [1]. Consequently, the development of assistive technologies that detect falls in domestic environments and alert caregivers and relatives has received considerable attention in the health care community in recent years (e.g., [2][3][4]). Mobile robots are a particularly promising area of technology as they are flexible and relatively non-invasive in comparison to larger distributed ambient sensor systems. In addition to watching over people and detecting dangerous events, they can directly undertake actions that benefit the user in everyday situations, thereby enhancing the person's safety perception and preventing the loss of confidence caused by functional disabilities. Moreover, advanced robotic systems may encompass interactive, socially-aware robot companions that not only detect dangerous events, but also enhance the person's experience and well-being through, for instance, flexible and proactive human-robot interaction (HRI) [5][6]. On the other hand, the development of such intelligent systems introduces a vast set of challenges and technical concerns regarding the robot's perception of human activity and the design of sensory-driven robot behaviour.

As humans, our perceptual experience is modulated by an array of sensors that convey different types of information (or modalities), e.g. vision, sound, touch, movement [7]. Similarly, the problem of integrating information conveyed by multiple sensors has been a paramount ingredient of autonomous robots. In particular when operating in natural environments, the robust and efficient processing of multi-modal information plays a key role to perceive human activity. Research efforts have been made towards robots exploiting multi-sensory integration to improve HRI capabilities. For instance, Lacheze et al. [8] used auditory information to recognize objects that were partially occluded and thus difficult to detect through vision only. Sanchez-Riera et al. [9] presented a scenario with a robot companion that performs audio-visual fusion for multi-modal speaker detection. The system targeted multiple speakers in a domestic environment processing information from two microphones and two cameras mounted on a humanoid robot. In the context of assistive robots, Parisi and Wermter [16] presented a humanoid robot with a depth sensor to extract 3D body information and a learning-based system to detect abnormal user behaviour such as fall events. The robot used its head actuators to move the sensor and keep the person in the scene, thereby addressing the limiting field of view (FOV) of the sensor. Martinson [10] introduced a robot with a navigational aid for visually impaired people using a mobile robot platform. The system used depth information to detect other people in the environment and avoid dynamic obstacles. The system communicated to the person the direction of motion to reach the goal destination via a tactile belt around the waist. However, multi-modal systems embedded in mobile robots that remain operative under situations of uncertain sensory information, e.g. temporary unavailability of one of the modalities, represent an enticing milestone for assistive robots and are still to be extensively investigated.

In this paper, we present a humanoid robot that assists a person in daily activities and detects situations of danger such as a fall event. Our system integrates multiple sensor modalities to enhance the perception of the robot through automatic speech recognition (ASR), sound source localization (SSL), visual active tracking and action recognition. In the proposed scenario, the person can communicate with the Nao using speech commands. We enabled Nao to actively track the person using its motor abilities and use the extracted depth information to detect fall events. In the case that the person is out of the FOV of the depth sensor, SSL is used to locate the person and establish visual tracking. For this purpose, we extended Nao with a depth sensor and a stereo microphone system. Information from Nao's sonar sensors is used to avoid obstacles in the environment. When the person asks for assistance or a fall is detected, the humanoid will approach the person and record the scene using the depth's sensor RGB camera. This video recording can then be sent to the person's caregiver or relatives for further human evaluation.

We describe the proposed system with the implementation of ASR, SSL, action recognition, and obstacle avoidance in Sec. II, along with the multi-modal controller to integrate

German I. Parisi, Johannes Bauer, Erik Strahl and Stefan Wermter are with the Department of Informatics, University of Hamburg, Germany. {parisi,bauer,strahl,wermter}@informatik.uni-hamburg.de



(a) (b)

Fig. 1. Overall architecture of our multi-modal system – a) extended Nao with a depth sensor and stereo microphone, b) OSC-based communication network for covey sensor information to the multi-modal controller for robot behaviour.

sensory information. Sect. II includes the evaluation of single modules and the behaviour of the robot driven by the interplay of audio-visual cues. In Sec. III, we present the experimental, assistive scenario in a home-like environment for multi-modal tracking and the detection of fall events. We conclude in Sec. V with a summary and discussion of our approach, as well as open issues and future work directions for our assistive system in terms of technical components and user-centered usability testing.

II. PROPOSED SYSTEM

The goal of our system is to use the available modalities for localizing and tracking the person in the environment with the use of a Nao robot. The behaviour of the humanoid is driven by the integration of audio-visual cues, that include ASR for vocal commands (sentences), vision, to track and detect fall events, and SSL, to detect the position of the person when visual information is not available. For this purpose, we use an array of sensors installed on top of the Nao and a multi-modal controller that integrates the information conveyed by the sensors to modulate Nao's actuators. The overall architecture of the system is shown in Fig. 1.

Nao is a midsize humanoid robot developed by Aldebaran Robotics.[1] We extended the robot Nao with an ASUS Xtion Pro[2] depth sensor installed on top of the head (Fig 1.a). The Xtion has a distance of use between 0.8 and 3.5 meters with a VGA resolution (640x480) at a maximum of 30 fps. In contrast to the Microsoft Kinect, the Xtion has reduced power consumption and weight. For SSL, we use a Soundman OKM II[3] binaural stereo microphone with omni-directional polar pattern and a frequency range of 20Hz–20kHz. We installed the stereo microphone on the Xtion sensor with a distance of 14.5 cm between the right/left channels (Fig. 1.a). We chose the Soundman microphones by comparing the SSL performance also with the stereo microphones embedded in the Nao and the Xtion (see Sec. 2.c). For ASR, we use a bluetooth headset (Sennheiser EZX 80[4]) with an omni-directional microphone that can be comfortably worn by the person and allows more robustness in noisy environments compared to the microphones embedded in the Nao, especially when the robot is moving. We use Nao's sonar sensors to detect obstacles on the way. The sonar sensors have an effective cone of 60° with a resolution of 1 cm and a detection range from 0.25 to 2.55 meters.

A. Active Tracking

The Xtion depth sensor is characterized by a reduced FOV (58° horizontal, 45° vertical, 70° diagonal), limiting its use in expansive environments. This motivates the implementation of an active tracking system, which moves the sensor to keep the person in scene. We use Nao's head to move the sensor and increase the horizontal FOV from 58° to 138° (Fig. 2.a). Nao will then smoothly pan its head by 10° degrees in the required direction, for a maximum pan angle of 40° degrees in each direction. As a strategy for active tracking, we define a bounding box in which the person can act without the sensor being moved (Fig. 2.b). We base the tracking of the person on a 3D skeleton model and consider the point of the upper-body torso as the reference of the person's position. When the torso point lies outside the threshold, the tracking application will compute the operations required to keep the person within the bounding box.

The tracking application is built on top of simple-openni[5],

[1] Aldebaran Robotics: http://www.aldebaran-robotics.com/

[2] ASUS Xtion PRO LIVE: http://www.asus.com/Multimedia/Xtion_PRO_LIVE/

[3] Soundman OKM II Studio: http://www.soundman.de/en/products/okm-ii-studio/

[4] Sennheiser EZX 80: http://en-de.sennheiser.com/bluetooth-headset-smart-phone-headset-mobile-ezx-80

[5] simple-openni – OpenNI library for Processing: https://code.google.com/p/simple-openni/

which wraps the OpenNI–NITE framework[6] for user identification, calibration and estimation of skeletal joints. We use this library with Processing IDE[7] with skeleton tracking provided by OpenNI. In this setting, we obtain the angle of the person with respect to the sensor as follows:

$$\alpha = arctan([x - (x_{max}/2)]/z_{max}), \tag{1}$$

where x is the position of the torso joint w.r.t. the horizontal image plane, $x_{max}/2$ is the center of this plane, and z_{max} is the focal length (max. depth value).

B. Fall Detection

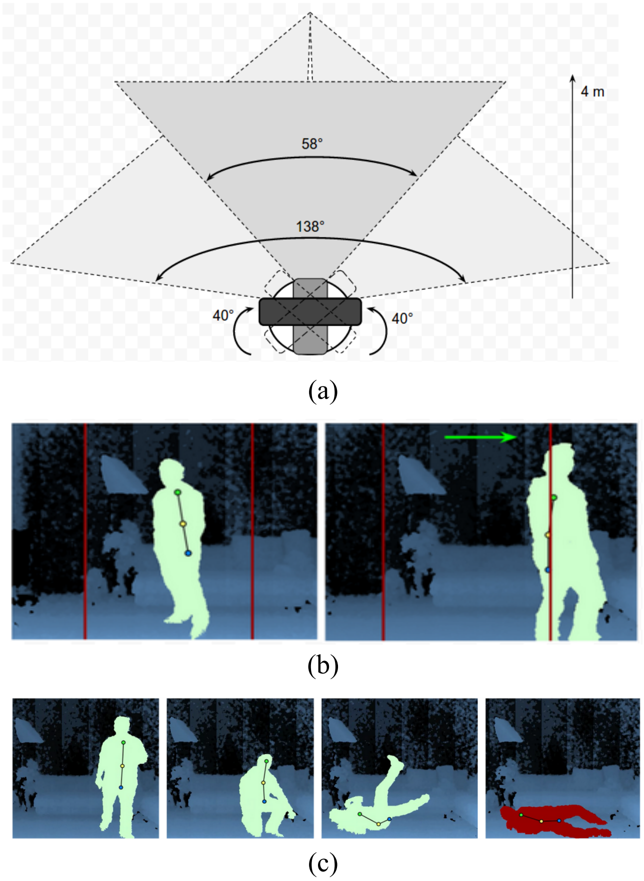
The robust detection of falls in home environments is a major concern in the public health care domain [1]. A combination of computational efficiency and robustness to changes in light conditions in indoor environments have made fall detection systems using depth information increasingly important in the research community (e.g. [11], [12]). However, many approaches do not consider noise-tolerant solutions able to operate with a mobile sensor. In particular, reported experiments with low-cost depth sensors have shown that a moving device has a strong negative impact on the sensor stability, leading to systematic tracking errors and noise.

To detect fall events, we use a learning-based approach (Parisi and Wermter [16]) that reports novel behavioural patterns that were not presented during the training phase. This system trains a neural network architecture on a dataset of 3D body motion from depth map videos comprising normal behaviour, i.e. domestic actions such as walking, sitting, and lying down, and then triggers an alarm when abnormal behavioural patterns are detected, e.g. a fall (Fig. 2.c). To contrast sensor noise and tracking errors, the neural architecture is also responsible for automatically removing noisy samples from the extracted body features during the training and test stage. Experiments in a home-like environment reported that the system detects falls with 96% accuracy [17].

The combination of a depth sensor with the learning-based approach allows us to tailor the robust detection of fall events independently from the background surroundings and changing light conditions. This is especially advantageous in scenarios with a mobile sensor.

C. Sound Source Localization

There are a number of auditory cues that can be used for sound-source localization (SSL). Most of these cues are derived from the spatial separation of sensors. Among these are the difference in the time at which sounds arrive at each microphone (time difference of arrival, TDOA), the difference in intensity (interaural intensity difference, IID), and spectral variations in the signals [14]. Any number of microphones greater than two can be used in principle, but the hardware and computational cost sharply rises with each additional microphone.



(a)

(b)

(c)

Fig. 2. Active tracking and fall detection [17] – (a) Nao with Xtion sensor: extended horizontal field of view from 58 to 138 degrees with a maximum head pan angle of 40 degrees in each direction. sensor (b) Threshold-based active tracking strategy, and (c) Detection of a fall event (red body).

In our scenario, we require fast and reliable SSL. On the other hand, high accuracy is not an issue. We therefore choose a simple but reasonably accurate binaural solution which extracts the TDOA from a stereo signal using the cross-correlation algorithm [15]. This algorithm shifts the signals from the individual microphones with respect to each other and determines the shift producing the greatest cross-correlation. That shift corresponds to the TDOA and thus to the angle of incidence.

It is possible to compute the angle of incidence for a given TDOA from the geometry of the system. However, since the estimate of the TDOA computed by the cross-correlation algorithm can be smeared by the acoustic properties of the environment, the robot body, and the ego-noise it produces, we opted for an empirical approach: We recorded $60\,s$ of recorded speech from 19 directions at 10° intervals between -90° and 90° from the robot. We split each of the recordings into $0.25\,s$ snippets and computed the relative time shift maximizing the cross-correlation between the channels for each snippet. For each occurring time shift, we then selected that angle of incidence for which it occurred most often as

[6] OpenNI/NITE: `http://www.openni.org/software`
[7] Processing IDE: `http://processing.org/`

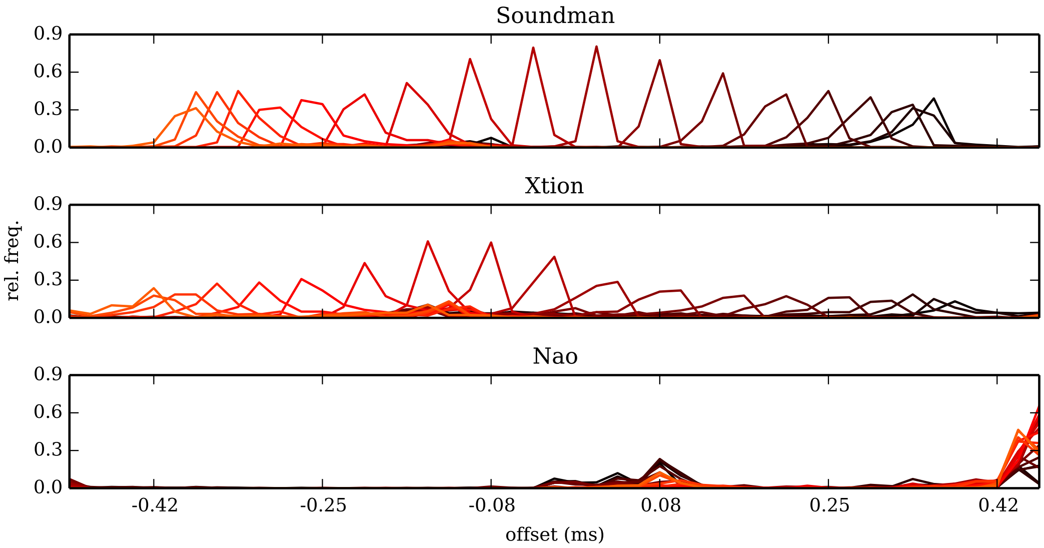


Fig. 3. Results of SSL with cross-correlation using different stereo microphones – Histograms of maximizing time shifts for each angle for the Soundman, Xtion, and Nao microphones. Each shade represents a histogram for one angle.

the most likely angle of incidence. We did this for three different sets of microphones: the Nao's own microphones, those of the Xtion sensor, and the Soundman microphones.

Figure 3 shows histograms of maximizing time shifts for each angle. The TDOA estimated by the cross-correlation algorithm was strongly correlated to the angle of incidence for all stereo microphones, as expected. However, the degree of correlation, measured by Spearman's rank correlation coefficient, differed drastically (Nao: $\rho = 0.506$, Xtion: $\rho = -0.714$, Soundman: $\rho = -0.930$; $p << 0.0001$ for all microphones). We therefore chose the Soundman microphone for SSL.

D. Automatic Speech Recognition

For ASR, we used the approach proposed by Twiefel et al. [13]. This system improves Google's cloud-based speech recognition with domain-dependent post-processing. The post-processor translates each sentence in the list of candidate sentences returned by Google's service into a string of phonemes. To be able to exploit the quality of the well-trained acoustic models employed by Google's service, the ASR hypothesis is converted to a phonemic representation employing the SequiturG2P grapheme-to-phoneme converter. Then, the sentence from a list of in-domain sentences is selected as the most likely sentence, which has the least Levenshtein distance to any of the candidate phoneme strings. For our implementation, we used the 10 top results and the target sentences.

An advantage of this approach is the hard constraints of the results, as each possible result can be mapped to an expected sentence. Experiments reported in [13] showed that the sentence list approach obtained the best performance for in-domain recognition with respect to other approaches such as Sphinx-4 [18] on the TIMIT speech corpus[8] with a *sentence-error-rate* of 0.521.

The sentences that we use for our scenario are: "*Look at me*", "*Come to me*", "*Turn around*", "*Turn to me*", "*Help me*", "*Yes, please*", "*No, thank you*", and "*Stop*".

E. Multi-modal Controller

The multi-modal controller modulates the motor behaviour of the humanoid and other operations of the system based on the information conveyed by the different sensors. This module is responsible for estimating the reliability of the modalities in terms of last arrived valid signal from the audio-visual modules.

When the vision-based position is not available or the last tracked position is older than 3 seconds, then SSL will be used. If the last valid SSL angle is older than 3 seconds, then the robot will ask "*Where are you?*" and wait for either audio or visual input. If audio-visual inputs are in conflict, i.e. the user's position estimated by the tracking framework and the SSL are widely discrepant, then more priority will be given to the visual estimation. This is due to the fact that the SSL module is more likely to return unreliable estimations, e.g., in situations with strong background noise.

At any time, the robot can receive vocal commands that have priority over the other modules. For instance, "*Stop*" will abort the current task of the robot. When the robot is moving, the controller uses Nao's sonar sensors to stop before obstacles that can cause the damage to the robot. If after a stop, the robot is not able to estimate the position of the person, it will wait for vocal hints.

A visual example of the interplay of different modalities is shown in Fig. 4.

[8] TIMIT Acoustic-Phonetic Continuous Speech Corpus: `https://catalog.ldc.upenn.edu/LDC93S1`

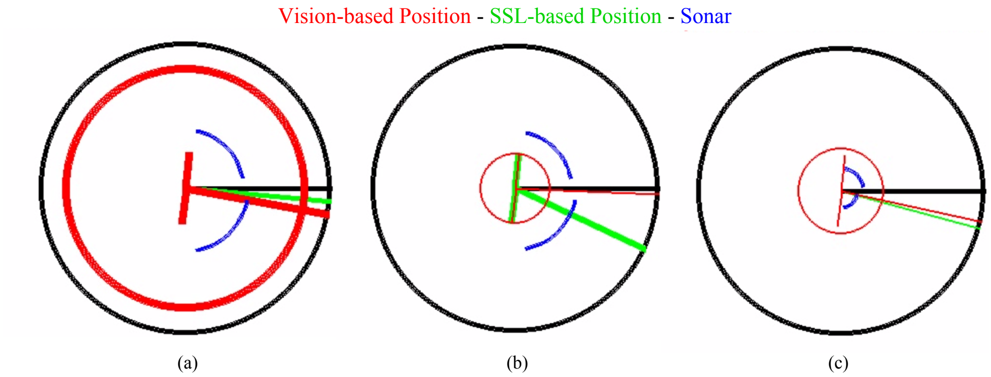


Fig. 4. Visualization of multi-modal robot perception in a ratio of 3 meters - The thickness of the lines represents the reliability of the information sources (the thicker the more reliable). (a) Visual information is used to estimate the position of the person (SSL is also computed but not used); (b) When visual information is not available (e.g. out of date), then SSL is used to estimate the position; (c) The person is too close to the robot (30 cm) so that the depth sensor cannot track the position (out of the operation range) and the sonar sensors detect a possible obstacle.

F. System Interface

All system modules communicate over Open Sound Control (OSC) [22], a message-based protocol developed for communication and data control among multimedia devices. It uses IP/UDP, that makes it very fast and accurate so that is naturally used also in other domains such as robotics [23]. An important advantage of OSC is the compatibility with many programming languages that enables us to connect our modules with a lightweight protocol, in our case using Python, Java, and Processing.

III. SCENARIO

The scenario for fall detection is shown in Fig. 5. The Nao was initially positioned on one side of the room to monitor the scene and connected to the system using wireless communication. The depth sensor and the microphones were connected to a laptop (i5-3320M 2,6 GHz, 4GB RAM) running all system modules through OSC protocol under Linux (Ubuntu desktop 12.04). The bluetooth microphone EZX 80 works up to 10 meters from the laptop (enough to cover a large room).

The person can use a set of vocal commands to interact with the robot that will result in the following behaviours. For *Look at me*, Nao will orient towards the person in the environment using vision and audio. If the position of the person is not known through vision (out of the FOV or occluded), the robot will use SSL. If still the robot is not able to estimate the position of the person, it will ask *”Where are you?”* and wait for vocal hints.

For *Come to me*, the robot approaches the person to a fixed distance of 1 meter using the last estimated position (Eq. 1). When the person is not in the FOV of the robot, the command *Turn to me* is used to rotate the robot (not only the head) towards the person and then establish visual contact.

For *Turn around*, the robot will perform a 180° turn. The command *Stop* will terminate any operation that the robot is performing, for instance if interrupting a turn or stopping the approaching robot at a desired distance.

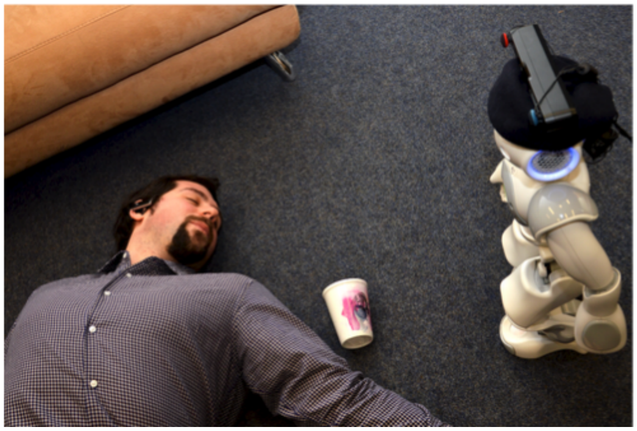


Fig. 5. Fall detection scenario with assistive humanoid robot in a home-like environment.

When the person says *Help me* or when a fall is detected, Nao will approach the person and ask whether assistance is required (e.g., to stand up in case of fall). If the answer is *Yes, please* or no vocal answer is detected, Nao can get in contact with the person’s caregiver or relative for further assessment of the situation. In the case of a fall, the system will store the last 5 seconds of activity before the fall as an RGB video that can be used to evaluate the seriousness of the event.

In the future, we plan to conduct a usability study that allows us to evaluate the system in a real-world setting, for instance by studying the users’ acceptance of the assistive Nao in terms of overall performance, human-robot communication, timing, and task sequence [6].

IV. DISCUSSION

In summary, we have presented a multi-modal system embedded in a humanoid robot for an assistive scenario. The robot integrates multiple sensor modalities to enhance the perception and to detect situations of danger such as fall events. Robot motor control is triggered by the interplay of hearing and vision captured by an array of onboard sensors.

The reported experiments in a home-like environment motivate future work in several directions.

Different from the use of fixed or ambient sensors, mobile robots can use motor capabilities to improve sensory-driven perception and better adapt to complex environments. Multisensory integration (MSI) can be advantageous for a variety of reasons. One is that certain types of information can only be gleaned from some modalities and not from others. This is the case in our scenario for verbal information which is only available in the auditory modality. A second reason why MSI can be useful is that it provides redundancy which can help improve accuracy and disambiguate. In our system, we exploit this aspect of MSI when we integrate segmentation from depth perception and sound cues to estimate the position of a person in the environment. This would be much harder from either modality alone. Finally, it can be useful to employ another type of sensor even if the information gleaned through it could be provided by a different sensor in principle: sometimes one modality just provides information simply in a more appropriate form, as exemplified by our use of the Nao's sonar sensors for obstacle detection which would be possible, at greater computational cost, using just color vision or depth perception.

We plan to improve the reliability of our person localization using more sophisticated and robust biologically-inspired unisensory and multisensory localization modules (based on work by Davila-Chacon et al. [20] and Bauer et al. [21]). At the current state of the system, the depth sensor and the stereo microphone must be wired to an external processing unit. For a better mobility of the robot, these sensors could be wired to an onboard processing unit and then transmit the depth and audio information via WiFi for post-processing in the cloud. From a navigation perspective, the robot does not have any representation about the operational environment. A possible extension is to provide Nao with prior knowledge on the properties of the environment using a ceiling camera [19] or a mechanism for self-localization and mapping such as RatSLAM extended for humanoid robots [24]. This would enhance Nao's navigational capabilities for, e.g., a scenario with multiple rooms in a residential context. Additionally, proxemic behaviours could be explored for socially-acceptable scenarios to navigate safely in a cluttered and dynamically changing domestic environment [25].

ACKNOWLEDGMENT

This work was partly supported by the DAAD German Academic Exchange Service (Cognitive Assistive Systems Project), and by the DFG German Research Foundation (grant #1247) - International Research Training Group CINACS (Cross-modal Interaction in Natural and Artificial Cognitive Systems)

REFERENCES

[1] World Health Organization: Global Report on Falls Prevention in Older Age: http://www.who.int/ageing/publications/Falls_prevention7March.pdf

[2] KSERA: Knowledgable SErvice Robots for Aging: http://ksera.ieis.tue.nl. Cited 15 Feb 2015

[3] ROBOT-ERA: Implementation and integration of advanced Robotic systems and intelligent Environments in real scenarios for the ageing population: http://www.robot-era.eu. Cited 20 Feb 2015

[4] F. Amirabdollahian, S. Bedaf, R. Bormann, H. Draper H., V. Evers, et al. Assistive technology design and development for acceptable robotics companions for ageing years. *Paladyn, Journal of behavioral robotics* 4(2):1–9, 2013.

[5] R. Kachouie, S. Sedighadeli, R. Khosla, and M-T. Chu, M-T. Socially Assistive Robots in Elderly Care: A Mixed-Method Systematic Literature Review. *Int. J. Hum. Comput. Interaction* 30(5):369–393, 2014.

[6] E. Torta, F. Werner, D.O. Johnson, J.F. Juola, R.H. Cuijpers, et al. Evaluation of a Small Socially-Assistive Humanoid Robot in Intelligent Homes for the Care of the Elderly. *J Intell Robot Syst* 76:57–71, 2014.

[7] B.E. Stein, T.R. Stanford, B.A. Rowland. The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hear Res* 258(1–2):4–15, 2009.

[8] L. Lacheze, G. Yan, R. Benosman, B. Gas, and C. Couverture. Audio/video fusion for objects recognition. In: *IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS-09)*, pp. 652–657, St. Louis, MO, 2009.

[9] J. Sanchez-Riera, X. Alameda-Pineda, J. Wienke, and A. Deleforge. Online multimodal speaker detection for humanoid robots. In: *IEEE-RAS International Conference on Humanoid Robots (Humanoids-12)*, pp. 126–133, Osaka, Japan, 2012

[10] E. Martinson. Finding people in home environments with a mobile robot. In: *IEEE Intl. Symposium on Robot and Human Interactive Communication (RO-MAN-14)*, pp. 744–749, Edinburgh, UK, 2014.

[11] R. Planinc and M. Kampel. Introducing the use of depth data for fall detection. *Personal and Ubiquitous Computing* 17:1063–1072, Springer-Verlag, 2012.

[12] G. Mastorakis and D. Makris. Fall detection system using Kinects infrared sensor. *Journal of Real-Time Image Processing*, Springer-Verlag, 2012.

[13] J. Twiefel, T. Baumann, S. Heinrich, and S. Wermter. Improving Domain-independent Cloud-based Speech Recognition with Domain-dependent Phonetic Post-processing. In: *IEEE Conf. on Artificial Intelligence (AAAI-14)*, pp. 1529–1535, Quebec, Canada, 2014.

[14] J. Schnupp, I. Nelken, and A.J. King. Auditory Neuroscience: Making Sense of Sound. 1st ed. Cambridge, MA: MIT Press, 2010.

[15] C.H. Knapp and G. C. Carter. The Generalized Correlation Method for Estimation of Time Delay. In: IEEE Transactions on Acoustics, Speech and Signal Processing 24.4, pp. 320327, 1976.

[16] G.I. Parisi and S. Wermter. Hierarchical SOM-Based Detection of Novel Behavior for 3D Human Tracking. In: *IEEE International Joint Conference on Neural Networks (IJCNN-13)*, pp. 1380–1387, Dallas, US, 2013.

[17] G.I. Parisi, and S. Wermter. Neurocognitive assistive robot for robust fall detection. *Smart Environments*, Springer, in press.

[18] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, J. Sphinx-4: A Flexible Open Source Framework for Speech Recognition. Technical Report SMLI TR2004-0811, Sun Microsystems Inc, 2004.

[19] W. Yan, E. Torta, D. van der Pol, N. Meins, C. Weber, R. H. Cuipers, and S. Wermter. Learning Robot Vision for Assisted Living. *Robotic Vision: Technologies for Machine Leaning and Vision Applications*, ch. 15, pp. 257–280, IGI Global, 2013.

[20] J. Davila-Chacon, S. Magg, J. Liu, and S. Wermter. Neural and statistical processing of spatial cues for sound source localisation. In: *IEEE Intl. Conf. on Neural Networks (IJCNN-13)*, pp. 1–8, Dallas, US, 2013.

[21] J. Bauer, J. Davila-Chacon, and S. Wermter Modeling development of natural multi-sensory integration using neural self-organisation and probabilistic population codes. *Connection Science*, pp. 1-19, 2014.

[22] The Open Sound Control 1.0 Specification: http://opensoundcontrol.org/spec-1_0. Cited 15 Feb 2015.

[23] A. Schmeder, A. Freed, and D. Wessel. Best Practices for Open Sound Control In: *Linux Audio Conference*, Utrecht, NL, 2010.

[24] S. Müller, C. Weber, and S. Wermter. RatSLAM on Humanoids - A Bio-Inspired SLAM Model Adapted to a Humanoid Robot. In: *Intl. Conf. on Artificial Neural Networks (ICANN-14)*, pp. 789–796, Springer Heidelberg, Hamburg, Germany, 2014.

[25] E. Torta, R.H. Cuijpers, J.F. Juola, and D. van der Pol. Design of Robust Robotic Proxemic Behaviour. *Social Robotics* 7072:21–30, 2011.