Integrated Speech-based Perception System for User Adaptive Robot Motion Planning in Assistive Bath Scenarios

Athanasios C. Dometios, Antigoni Tsiami, Antonis Arvanitakis, Panagiotis Giannoulis,
Xanthi S. Papageorgiou, Costas S. Tzafestas, and Petros Maragos
School of Electrical and Computer Engineering, National Technical University of Athens, Greece
{athdom,paniotis,xpapag}@mail.ntua.gr, {antsiami,ktzaf,maragos}@cs.ntua.gr

***Abstract*— Elderly people have augmented needs in performing bathing activities, since these tasks require body flexibility. Our aim is to build an assistive robotic bath system, in order to increase the independence and safety of this procedure. Towards this end, the expertise of professional carers for bathing sequences and appropriate motions have to be adopted, in order to achieve natural, physical human - robot interaction. The integration of the communication and verbal interaction between the user and the robot during the bathing tasks is a key issue for such a challenging assistive robotic application. In this paper, we tackle this challenge by developing a novel integrated real-time speech-based perception system, which will provide the necessary assistance to the frail senior citizens. This system can be suitable for installation and use in conventional home or hospital bathroom space. We employ both a speech recognition system with sub-modules to achieve a smooth and robust human-system communication and a low cost depth camera for end-effector motion planning. With a variety of spoken commands, the system can be adapted to the user's needs and preferences. The instructed by the user washing commands are executed by a robotic manipulator, demonstrating the progress of each task. The smooth integration of all subsystems is accomplished by a modular and hierarchical decision architecture organized as a Behavior Tree. The system was experimentally tested by successful execution of scenarios from different users with different preferences.**

I. INTRODUCTION

Most advanced countries tend to be aging societies, with the percentage of people with special needs for nursing attention being already significant and due to grow. Health care experts are called to support these people during the performance of Personal Care Activities such as showering, dressing and eating [1], [2], inducing great financial burden both to the families and the insurance systems. During the last years the health care robotic technology is developing towards more flexible and adaptable robotic systems designed for both in-house and clinical environments, aiming at supporting disabled and elderly people with special needs.

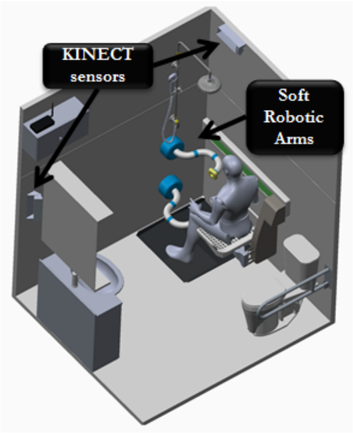

Fig. 1: CAD design of the robotic bath system installed in a pilot study shower room.

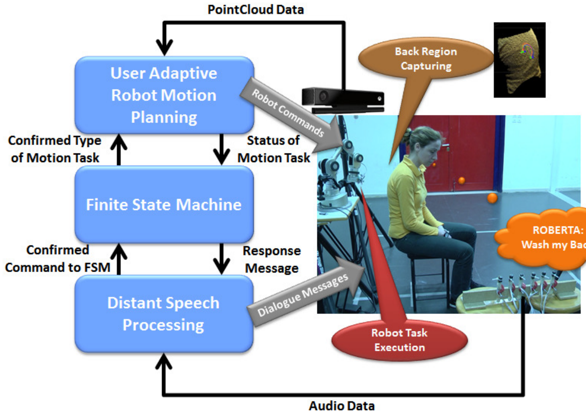

Fig. 2: An overview of the proposed integrated intelligent assistive robotic system.

There have been very interesting developments in this field, with either static physical interaction [3]–[5], or mobile solutions [6]–[8]. Most of these focus exclusively on a body part e.g. the head, and support people on performing personal care activities with rigid manipulators.

However, body care (showering or bathing) is among the first daily life activities which incommode an elderly's life [1], since it is a demanding procedure in terms of effort and body flexibility. On the other hand, soft robotic arm technologies have already presented new ways of thinking robot operation [9]–[11], and can be considered safe for

---

The first four authors contributed equally to this work.

This research work has received funding from the European Union's Horizon 2020 research and innovation programme under the project "I-SUPPORT" with grant agreement No 643666 (http://www.i-support-project.eu/).

direct physical interaction with frail elderly people. Also, applications with physical contact with human (such as showering) are very demanding, since we have to deal with curved and deformable human body parts and unexpected body-part motion may occur during the robot's operation. Therefore safety and comfort issues during the washing process should be taken into account. In addition, control architectures of a hyper-redundant soft arm [11]–[13], in an environment with human motion not only can give important advantages in terms of safety but also is a challenging task, since several control aspects should be considered such as position, motion and path planning [14], stiffness [15], shape [16], and force/impedance control [17].

Another important aspect of these assistive robotic systems is the incorporation of intuitive human-robot interaction strategies, that involve both proper and simple communication and perception of the user and modular interconnection of several subsystems constituting the robotic system. Additionally, Human-Robot trust is a major issue for people especially in contexts of physical interaction. As these people experience this new environment they have difficulty to adapt to the operational requirements. Trust is difficult to be established if the Robot is not dependable, predictable and its actions are perceived as transparent [18]. So the challenge for the system to work safely and intuitively is to enhance to basic properties of the system robustness and predictability.

In particular, we aim to establish the communication of the elderly people with the robotic bathing system via spoken commands. Since speech is the most natural way of human communication, it is the most comfortable and natural way also in human-system communication, and generally the most intuitive way to ask for help. Also, the bathroom scenario as well as the system used by elderly people raise some challenges. Firstly, the necessary distance between the speaker and the microphones imposes distant speech recognition. Secondly, the nature of the environment, i.e. high reverberation (bathroom walls) and noise (coming mainly from water) brings up the need for specific components like voice activity detection, denoising, etc., that should be adapted to the environment. Third, speech recognition for elderly people is a challenging problem, let alone distant speech recognition, due to their difficulty of speaking, tremor, and corruption of speech characteristics. Last, due to the fact that the system is destined to elderly people, it should perform robustly and also confirm the elderly's choices in order to avoid undesired situations.

In this paper an integrated intelligent assistive robotic system designed for bathing tasks is presented. The proposed system consists of two real time perception systems, one for audio instructions and communication and one for user visual capturing, as depicted in Fig. 2. For this reason we employ a speech recognition system with sub-modules to achieve a smooth and robust human-system communication. The system can be adapted to the user's specific room/space, needs and preferences, since it is independent of the microphone positions and the user can change the set of interaction commands according to his needs. The components of this recognition system are a Voice Activity Detection module, a beamforming component for enhancing and denoising, a distant speech recognition engine and a speech synthesis module providing audio feedback to the user. The commanded washing tasks are executed by a robotic manipulator, demonstrating the progress of each task. The smooth integration of all subsystems is accomplished by a modular and hierarchical decision architecture organized as a Behavior Tree.

II. SYSTEM DESCRIPTION

The robotic shower system, which is currently under development Fig. 1, will support elderly and people with mobility disabilities during showering activities, i.e. pouring water, soaping, body part scrubbing, etc. The degree of automation will vary according to the user's preferences and disability level. In Fig. 1, a CAD design of the configuration of the system's basic parts is presented. The robotic system provides elderly showering abilities enhancement, whereas the Kinect sensors are used for user perception and HRI applications.

The robotic arms is constructed with soft materials (rubber, silicon etc.) and is actuated with the aid of tendons and pneumatic chambers, providing the required motion to the three sections of the robotic arm. This configuration makes the arms safer and more friendly for the user, since they will generate little resistance to compressive forces, [19], [20]. Moreover, the combination of these actuation techniques increases the dexterity of the soft arms and allows for adjustable stiffness in each section of the robot. The end-effector section, which will interact physically with the user, will exhibit low stiffness achieving smoother contact, while the base sections supporting the robotic structure will exhibit higher stiffness values.

Visual information of the user is obtained from Kinect depth cameras. These cameras will be mounted on the wall of the shower room in a proper configuration in order to provide full body view of the user, Fig. 1. It is important to mention that information exclusively from depth measurements will be used to protect the user's personal information. Accurate interpretation of the visual information is a prerequisite for human perception algorithms (e.g. accurate body part recognition and segmentation [21]). Furthermore, depth information will be used as a feedback for robot control algorithms closing the loop and defining the operational space of the robotic devices.

Finally, human-system interaction algorithms will include distant speech recognition, [22], employing microphone arrays in order to achieve robustness. This augmented cognition system will understand the user's interactions, providing the necessary assistance to the frail senior citizens.

III. HRI DECISION CONTROL SYSTEM

A. Behavior Tree Formulation

The Decision Control System (**DCS**) consists of distinct modules organized as a Behavior Tree (**BT**). The basic goal of such a system is to properly integrate and orchestrate the sub-modules of the system, in order to achieve both optimal performance of the system and smooth Human Robot Interaction. The BT has two type of nodes, the **Control-Flow** nodes and the **Leaf** nodes. The control-flow nodes are: First, the **"Selector"** which executes its children until one returns *SUCCESS* and can be interpreted as sequential *OR*s. Second is the **"Sequence"** node which executes its children until they all return *SUCCESS* and can be interpreted as *AND*. The Leaf nodes are the ones that are exchanging information with the external components (i.e. outside the scope of the DCS) and are capable of changing the configuration of the system.

The formulation of a BT is described in detail in [23]. Additional nodes are the **"Loop"** node that repeats part of the tree and the **"Parallel"** node which allows for concurrency as in [24].

B. Robot Platform Decision Support Characteristics

Decision modules based on BT have shown to be fault tolerant [24] because they are modular and capable of describing complex behavior as a hierarchical collection of elementary building blocks [25], [26]. These blocks allow easier testing, separation of concerns per module, reusability and the ability to compose almost arbitrary behaviors [25]. Thus a system is built that can be easily audited. Safety auditing is required for medical devices.

The **BT** consists of the following submodules. First, is the auditory module that handles the input from the speech recognition system which enables the communication of the user's commands to the system. Second is the text to speech feedback module which communicates with the user, acknowledges the user's commands and can request for confirmation. Third is the robotic arm module which according with the execution stage of the task and the capabilities of the robot initiates, handles the changes in the user's requests dynamically. The robotic arm can wash the user's back with various patterns. For example during scrubbing in a cyclic pattern, the user can request at any point from the system verbally to change pattern or to stop completely. Except for the described capabilities the user also can request an immediate stop of the whole process for safety reasons by issuing a "Halt" command. The user can also increase and decrease the robot's washing speed, and will be able to perform additional actions as the washing system's capability is expanded.

C. Design of the Automation System

The system consists of modules that are easily testable by themselves separate from the rest of the system, and functionality is replicated throughout the tree due to the design's modularity. The pattern used is that the robot requests from the user for a task using the speech recognition system, gives feedback through the speaker, requests confirmation and proceeds to execute the task based on the behavior tree. If we need to adapt the system with new hardware then the current functionality will be retained as long as any new hardware conforms to the behavior tree's protocol.

IV. DISTANT SPEECH PROCESSING SYSTEM

Our setup involves the use of multiple sensors for audio capture. This is essential for the use cases we have employed, because we expect far-field speech. For this reason, we employ microphone arrays aiming to also address other factors that deteriorate performance, such as noise, reverberation and overlap of speech with other acoustic events.

The distant speech processing system with its subcomponents is depicted in Fig. 3. After audio capture, a Voice Activity Detection (VAD) module detects the segments that contain human voice and it forwards this information to a beamforming component that exploits the known speaker's location (on the chair) to enhance and denoise target speech. Subsequently a distant speech recognition engine transforms speech to text in the form of a command to the system. A speech synthesis module is also used to feed back to the user the recognized command in order to confirm it or not, and if it is confirmed, the correct system prompt is spoken out to the user, employing the same Text-To-Speech (**TTS**) engine. In the next sub-sections each module is described in more detail.

A. Voice Activity Detection

The VAD module constitutes the first step of the distant speech processing system. Given a fast and robust VAD module, the speech processing system can benefit in several ways: (a) better silence modeling (less false alarms), (b) operation of the Automatic Speech Recognition (**ASR**) module in windows of variable length, (c) less computations needed from the ASR module (operation only inside VAD segments).

The proposed VAD system implements a multichannel approach to determining the temporal boundaries of speech activity in the room. The sequence of audio events (speech or non-speech) inside a given time interval, is estimated via the Viterbi algorithm on combined likelihood scores coming from single-channel event models for the entire observation sequence. These combined scores are estimated as averages of scores based on channel-specific event models ("sum of log-likelihoods").

For the modeling of each event at each channel the well-known Hidden Markov Models (HMMs) (single-state) with Gaussian Mixture Models (**GMM**s) for the probability distributions are employed. The features used are baseline Mel Frequency Cepstral Coefficients (MFCCs) with Δ's and $\Delta\Delta$'s. The Viterbi algorithm allows the identification of the optimal sequence of audio events for a given time interval, while the incorporation of the multichannel score essentially leads to a robust decision that is informed by all the microphones in the room.

B. Beamforming

Signals recorded from a microphone array can be exploited for speech enhancement through beamforming algorithms. The beamformed signals can then enhance speech recognition performance due to noise and reverberation reduction. In our specific case, noise mainly comes from water pouring,

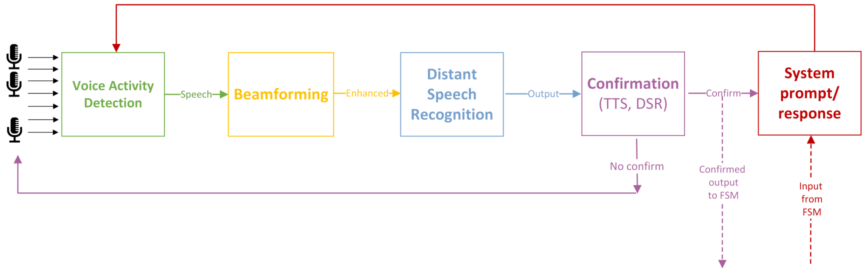

Fig. 3: Distant speech processing system with its sub-components.

and reverberation is high, a usual case in bathroom environments. The offline system offers two beamforming choices: The simple but fast Delay-and-Sum Beamformer (**DSB**) and the more powerful Minimum Variance Distortionless Response (**MVDR**) as implemented in [27]. The online version of the system uses only DSB, because it is faster, thus more suitable for real-time systems than MVDR.

C. *Distant speech recognition*

Several factors mentioned before such as noise, reverberation and distance between system/robot and speaker [28] render speech recognition a challenging task. Thus, we employ Distant Speech Recognition (**DSR**) [29], [30] with microphone arrays.

The DSR module is always-listening, namely it is able to detect and recognize the utterances spoken by the user at any time, among other speech and non-speech events. It is grammar-based, enabling the speaker to communicate with the system via a set of utterances adopted for the context of each specific use case. For the particular use case that will be described later, we have employed English language. In order to improve recognition in noisy and reverberant environments, we use DSB beamforming using the available microphone channels.

Regarding acoustic models, initially we trained 3-state, cross-word triphones, with 8 Gaussians per state, on standard MFCC-plus-derivatives features on Wall street Journal (WSJ) database [31], which is a close-talk database in English. However, mismatch between train and test conditions/environment deteriorates the DSR performance. Maximum Likelihood Linear Regression adaptation has been employed to transform the means of the Gaussians on the states of the models, using data recorded in this particular setup.

D. *Speech synthesis (TTS)*

A speech synthesis module is also part of the system, and has two particular uses: First, after the DSR module produces a result, the TTS module gives feedback to the user, in the form of synthesized speech, about the recognition result asking confirmation. Second, if confirmation is given by the user (recognized by the DSR module), the recognition result is published to the decision support system, which in its turn selects the appropriate system prompt to return to the speech synthesis node, in order to be played back to the user. This way we have a two-step verification process in order to avoid as much as possible mis-recognized commands. The employed speech synthesis module uses the widely available Festival Speech Synthesis System [32].

V. USER ADAPTIVE ROBOT MOTION PLANNING

The on-line end-effector path adaptation problem, of a robotic manipulator performing washing actions over the user's body part (e.g. back region), in a workspace including depth-camera, is considered. The basis of this task lies on the calculation of the appropriate reference end-effector pose, in order for the robotic manipulator to be able to execute predefined washing tasks (such as pouring water on the back region of the user) in a compliant way with the motion and the curvature of each body part.

In addition, this approach takes into consideration the adaptability of the system to different users. There is a great variability to the size of each body part among different users and each user has different needs during the bathing sequence. Therefore, in order to compensate with different operational conditions, the planning of the end-effector path begins on a normalized (in terms of spatial dimensions) 2D "Canonical" space.

Properly designed bathing paths can be followed and fitted on the surface of the user's body part, with use of our motion planning algorithm, satisfying both the time and the spacial constraints of the motion. The resulting point of the motion path at each time segment should be transformed from the fixed 2D "Canonical" space to the 3D actual operational space of the robot (Task space). In particular, we use two bijective transformations for this fitting procedure, Fig. 4(b),(c). The first bijection is an affine transformation (anisotropic scaling and rotation) from the "Canonical" space to the Image space, whereas the second one is the camera projection transformation from the Image space to the 3D operational space, which is included in the camera's field of view (Task space). These bijections ensure that the path will be followed within the body part limits and will be adaptable to the body motions and deformations. Finally, this motion planning algorithm takes into account the emergency situations that might raise on an integrated system and plans the motion of the robot from the operational position to a safe position (away from the user) incorporating obstacle avoidance techniques.

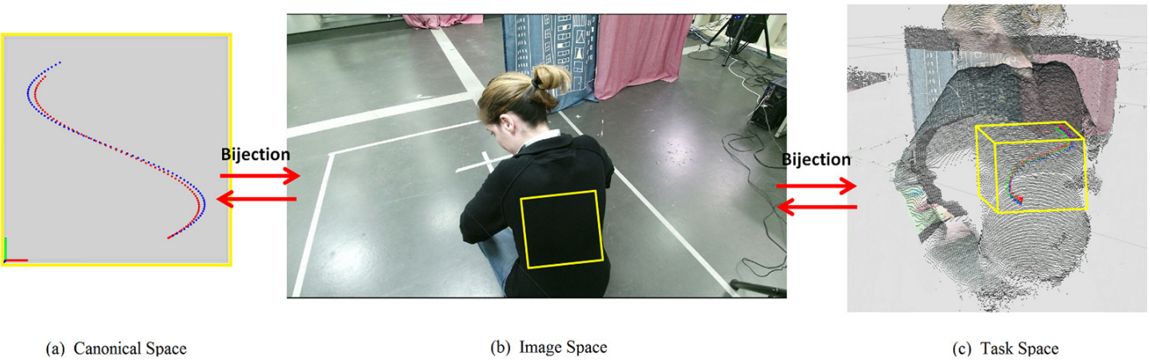

Fig. 4: Three different spaces described in the proposed methodology. (a) 2D space normalized in x,y dimensions notated as "Canonical" space. The sinusoidal path from the Canonical space is transformed the rectangular area on the Image space, using scaling and rotation. (b) 2D Image space is the actual image (of size 512 x 424 pixels) obtained from Kinect sensor. The yellow rectangular area marks the user's back region, as a result of a segmentation algorithm. The transformed path is fitted on the surface of a female subject's back region, using the camera projection transformation. (c) 3D Task Space is the operational space of the robot. The yellow box represents a Cartesian filter, including the points on which the robot will operate.

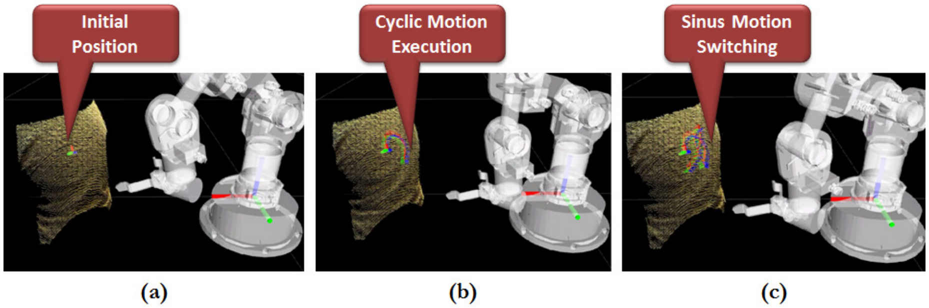

Fig. 5: Point Cloud representation of the user's back region during an experimental execution. (a) Procedure Initialization when the user ask from the system to wash its back with a cyclic motion pattern. (b) Cyclic motion execution until another command came. (c) The robot changed its motion primitive in order to follow the user's will to change the motion pattern to the sinus one.

VI. EXPERIMENTS

A. Setup Description

In order to test and analyze the performance of the integrated system, an experimental setup is used that includes a Kinect-v2 Camera providing depth data for the back region of a subject, as shown in Fig. 4 (c), with accuracy analyzed in [35]. The segmentation of the subjects' back region is implemented, for the purposes of this experiment, by simply applying a Cartesian filter to the Point-Cloud data. The setup also includes a 5 DOF Katana arm by Neuronics for basic demonstration of a bathing scenario. In this study healthy subjects took part in the experiments for validation purposes. The coordination and inter-connection of the subsystems constituting the robotic bathing system is implemented in ROS environment.

For this specific scenario, the speech processing system setup involves a MEMS microphone array [36], which is portable and has been found to perform equally well with other microphone arrays, placed opposite to the user. Regarding distant speech recognition task, a grammar with 11 different commands has been employed. All these commands begin with the word "Roberta", as the keyword that activates the system. This is justified due to the system being always-listening: If there is no keyword, users may be confused when addressing someone else and not the system. The set of commands include action prompts, i.e. "Roberta wash my back", "Roberta scrub my back", start/stop prompts, i.e. "Roberta initiate", "Roberta stop", "Roberta halt", type of motion commands, i.e. "Roberta circular", "Roberta sinus", velocity commands, i.e. "Roberta faster", "Roberta slower" and confirmation commands, i.e. "Roberta yes", "Roberta no".

B. Testing Strategy

The experiments conducted include a scenario in which the user can decide:

- the washing action from a repertoire of tasks, i.e. wash or scrub and the region of actions, i.e. back or legs;
- the confirmation or the rejection of on action based on dialogue in any step is necessary;

- the type of motion that the user may prefer, i.e. cyclic or sinus motion primitive;
- to pause the action, to execute it faster or slower, to change the executed primitive, or to emergency stop the whole system.

Based on these options the subject is free to decide its command in a non-linear way of execution. The robot under the subject's instruction is able to follow a simple sinusoidal or a cyclic path (representing the washing and the scrubbing action respectively), keeping simultaneously a constant distance and perpendicular relative orientation to the surface of the user's respective body region. These paths are chosen for clarity reasons for the presentation of the experimental results.

C. Results and Discussion

Our goal is to experimentally validate that the integrated system can handle correctly the user's information, preferences and instructions, while at the same time it is capable to communicate and interact successfully with the robot that perform the bathing task.

In Fig. 5 the Point Cloud data of the user's back region during a successfully executed experiment is presented. After the initialization procedure the user asks from the system to wash its back with a cyclic motion pattern (i.e. scrubbing or washing action). The robot took an initial position in a constant distance from its back, as depicted in Fig. 5 (a), and therefore the cyclic motion was executed until another command came, Fig. 5 (b). The next instruction was to change the motion pattern to the sinus one, and the robot changed its motion primitive in order to follow the user's will, as it is shown in Fig. 5 (c). The overall performance of the system was satisfying since a full scenario was successfully executed from different users with different preferences.

VII. CONCLUSIONS AND FUTURE WORK

This paper presents an integrated intelligent assistive robotic system designed for bathing tasks. The integration of the communication and verbal interaction between the user and the robot during the bathing tasks is a key issue for such a challenging assistive robotic application. We tackle this challenge by developing the described augmented cognition system, which will interpret the user's preferences and intentions, providing the necessary assistance to the frail senior citizens. The proposed system consists of two real time perception systems, one for audio instructions and communication and one for user capturing. A decision control system integrates all the subsystems in a non-linear way with a modular and hierarchical decision architecture organized as a Behavior Tree. All the instructed washing tasks are executed by a robotic manipulator, demonstrating the progress of each task. The overall performance of the system was tested and a full scenario was successfully executed from different users with different preferences.

For further research, we aim to ameliorate this methodology in order to meet any soft robotic manipulation special requirements. Furthermore, the initial predefined path will be learned by demonstration of professional carers, with the aid of DMP approach, in order to make the bathing sequence more human-like. The speech processing system can be enhanced via training with data more matched with the usecase scenario. Additionally, more microphone arrays distributed in space can be used in order to better enhance speech, remove noise coming from water and mitigate reverberation which is very high in such environments. The system is designed to work locally but it can be adapted to work as a teleoperation platform. Additional functionality can be incorporated and adaptability to users with more specific impairments can be introduced. Users that have back injuries, wounds or other skin abnormalities can be detected by the camera, by the carer or the user himself, towards a decision system with the ability to produce a personalized plans. There are planned tests with elderly patients and the feedback can be incorporated enhancing the systems usability.

ACKNOWLEDGMENT

The authors would like to acknowledge Rafa Lopez, Robotnik Automation, Valencia, Spain, for the CAD representation of the robotic bath system installed in a shower room. The authors also wish thank Petros Koutras, Isidoros Rodomagoulakis, Nancy Zlatintsi and Georgia Chalvatzaki members of the NTUA IRAL for helpful discussions and collaborations.

REFERENCES

[1] D. D. Dunlop, S. L. Hughes, and L. M. Manheim, "Disability in activities of daily living: patterns of change and a hierarchy of disability," *American Journal of Public Health*, vol. 87, pp. 378–383, 1997.

[2] S. Katz, A. Ford, R. Moskowitz, B. Jackson, and M. Jaffe, "Studies of illness in the aged: The index of adl: a standardized measure of biological and psychosocial function," *JAMA*, vol. 185, no. 12, pp. 914–919, 1963.

[3] T. Hirose, S. Fujioka, O. Mizuno, and T. Nakamura, "Development of hair-washing robot equipped with scrubbing fingers," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 1970–1975.

[4] Y. Tsumaki, T. Kon, A. Suginuma, K. Imada, A. Sekiguchi, D. Nenchev, H. Nakano, and K. Hanada, "Development of a skincare robot," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, May 2008, pp. 2963–2968.

[5] M. Topping, "An overview of the development of handy 1, a rehabilitation robot to assist the severely disabled," *Artificial Life and Robotics*, vol. 4, no. 4, pp. 188–192.

[6] C. Balaguer, A. Gimenez, A. Huete, A. Sabatini, M. Topping, and G. Bolmsjo, "The mats robot: service climbing robot for personal assistance," *Robotics Automation Magazine, IEEE*, vol. 13, no. 1, pp. 51–58, March 2006.

[7] A. J. Huete, J. G. Victores, S. Martinez, A. Gimenez, and C. Balaguer, "Personal autonomy rehabilitation in home environments by a portable assistive robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 561–570, July 2012.

[8] Z. Bien, M.-J. Chung, P.-H. Chang, D.-S. Kwon, D.-J. Kim, J.-S. Han, J.-H. Kim, D.-H. Kim, H.-S. Park, S.-H. Kang, K. Lee, and S.-C. Lim, "Integration of a rehabilitation robotic system (kares ii) with human-friendly man-machine interaction units," *Autonomous Robots*, vol. 16, no. 2, pp. 165–191, 2004. [Online]. Available: http://dx.doi.org/10.1023/B:AURO.0000016864.12513.77

[9] C. Laschi, M. Cianchetti, B. Mazzolai, L. Margheri, M. Follador, and P. Dario, "Soft robot arm inspired by the octopus," *Advanced Robotics*, vol. 26, no. 7, pp. 709–727, 2012.

[10] C. Laschi, B. Mazzolai, V. Mattoli, M. Cianchetti, and P. Dario, “Design of a biomimetic robotic octopus arm,” *Bioinspiration & Biomimetics*, vol. 4, no. 1, p. 015006, 2009.

[11] I. D. Walker, “Continuous backbone continuum robot manipulators,” *ISRN Robotics*, vol. 2013, 2013.

[12] D. B. Camarillo, C. R. Carlson, and J. K. Salisbury, “Task-space control of continuum manipulators with coupled tendon drive,” in *Experimental Robotics*. Springer, 2009, pp. 271–280.

[13] F. Renda and C. Laschi, “A general mechanical model for tendon-driven continuum manipulators,” in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3813–3818.

[14] J. Xiao and R. Vatcha, “Real-time adaptive motion planning for a continuum manipulator,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 5919–5926.

[15] M. Mahvash and P. E. Dupont, “Stiffness control of a continuum manipulator in contact with a soft environment,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 863–870.

[16] B. Bardou, P. Zanne, F. Nageotte, and M. De Mathelin, “Control of a multiple sections flexible endoscopic system,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 2345–2350.

[17] A. Kapadia and I. D. Walker, “Task-space control of extensible continuum manipulators,” in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 1087–1092.

[18] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, “A meta-analysis of factors affecting trust in human-robot interaction,” vol. 53, no. 5, pp. 517–527.

[19] T. G. Thuruthel, E. Falotico, M. Cianchetti, F. Renda, and C. Laschi, “Learning global inverse statics solution for a redundant soft robot,” in *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics*, 2016, pp. 303–310.

[20] M. Manti, A. Pratesi, E. Falotico, M. Cianchetti, and C. Laschi, “Soft assistive robot for personal care of elderly people,” in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, June 2016, pp. 833–838.

[21] A. Vedaldi, S. Mahendran, S. Tsogkas, S. Maji, R. Girshick, J. Kannala, E. Rahtu, I. Kokkinos, M. Blaschko, D. Weiss, *et al.*, “Understanding objects in detail with fine-grained attributes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3622–3629.

[22] I. Rodomagoulakis, N. Kardaris, V. Pitsikalis, E. Mavroudi, A. Katsamanis, A. Tsiami, and P. Maragos, “Multimodal human action recognition in assistive human-robot interaction,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 2702–2706.

[23] A. Marzinotto, M. Colledanchise, C. Smith, and P. Ogren, “Towards a unified behavior trees framework for robot control.” IEEE, pp. 5420–5427.

[24] M. Colledanchise, A. Marzinotto, D. V. Dimarogonas, and P. Oegren, “The advantages of using behavior trees in mult-robot systems,” in *Proceedings of ISR 2016: 47st International Symposium on Robotics*, pp. 1–8.

[25] A. Klckner, “Interfacing behavior trees with the world using description logic.” American Institute of Aeronautics and Astronautics.

[26] M. Colledanchise and P. gren, “How behavior trees modularize hybrid control systems and generalize sequential behavior compositions, the subsumption architecture, and decision trees,” vol. PP, no. 99, pp. 1–18.

[27] S. Lefkimmiatis and P. Maragos, “A generalized estimation approach for linear and nonlinear microphone array post-filters,” *Speech communication*, vol. 49, no. 7, pp. 657–666, 2007.

[28] C. Ishi, S. Matsuda, T. Kanda, T. Jitsuhiro, H. Ishiguro, S. Nakamura, and N. Hagita, “A robust speech recognition system for communication robots in noisy environments,” *IEEE Trans. Robotics*, vol. 24, no. 3, pp. 759–763, 2008.

[29] M. Wölfel and J. McDonough, *Distant speech recognition*. John Wiley & Sons, 2009.

[30] A. Katsamanis, I. Rodomagoulakis, G. Potamianos, P. Maragos, and A. Tsiami, “Robust far-field spoken command recognition for home automation combining adaptation and multichannel processing,” in *Proc. ICASSP*, 2014.

[31] D. B. Paul and J. M. Baker, “The design for the wall street journal-based csr corpus,” in *Proceedings of the workshop on Speech and Natural Language*. Association for Computational Linguistics, 1992, pp. 357–362.

[32] A. Black, P. Taylor, R. Caley, R. Clark, K. Richmond, S. King, V. Strom, and H. Zen, “The festival speech synthesis system version 1.4.2,” Software, Jun 2001. [Online]. Available: http://www.cstr.ed.ac.uk/projects/festival/

[33] Y. Zhou, M. Do, and T. Asfour, “Learning and force adaptation for interactive actions,” in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE, 2016, pp. 1129–1134.

[34] ——, “Coordinate change dynamic movement primitivesa leader-follower approach,” in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 5481–5488.

[35] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, “Evaluating and improving the depth accuracy of kinect for windows v2,” *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4275–4285, 2015.

[36] Z. Skordilis, A. Tsiami, P. Maragos, G. Potamianos, L. Spelgatti, and R. Sannino, “Multichannel speech enhancement using mems microphones,” in *Proc. ICASSP*, 2015.