2017 26th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
Lisbon, Portugal, Aug 28 - Sept 1, 2017.

Speaking Robots: The Challenges of Acceptance by the Ageing Society*

J. Oliveira; G. S. Martins; A. Jegundo; C. Dantas; C. Wings; L. Santos; J. Dias; F. Perdigão

Abstract — **The ability of robots to dialogue with humans appears as one critical Human-Machine Interaction feature when it comes to transferring robots into society. This ability gains additional importance when it comes to elderly people, since they find it more comfortable and natural to interact using voice, due to possible natural physical impairments that hinder the usage of some of the interaction modalities (e.g. touch screens). Challenges like recognition accuracy, distant speech, the idiosyncrasies of elderly voices (fading, muffled pronunciation, etc.), the effects of surrounding environment noise or the expressiveness of the robot when speaking, become highly relevant in the acceptance and usability of service robots by the ageing population. In this paper, we present the results, challenges and solutions developed during a nine-month iterative evaluation process that took place within the GrowMeUp project, with focus on speech recognition and synthesis. The paper concludes with an identification of open scientific and technological problems, based on our interpretation of results, which we identify as critical for the acceptance and usability of robots by an ageing society.**

I. Introduction

This paper analysis the speech system with respect to user acceptance, identifying the limitations of the speech-based human-robot interface and describing the latest developments in the interaction interface of the GrowMeUp system. The objectives within speech interaction are to evaluate the robot's acceptance by the end users (elderly); to identify system limitations and end-user's requests; to test the already-deployed functions of the robot, and to report some improvements with special focus on speech interface.

In recent years, we have been observing a paradigm shift in robotics, with robots progressing from industrial and scientific environments to domestic and social environments. This has been leading to the emergence of new human-robot interaction modalities, striving for an intuitive and human-friendly communication interface. This triggered the emergence of hearing and speaking robots, abilities that are now deemed essential for social interaction with humans. Presenting a social robot to an elderly user instantly triggers an expectation from the person's side: they expect the robot to exhibit human traits, where the ability to dialogue is likely to be the most common.

Speech recognition, synthesis and dialogue models [1] are an active and mature research area, but when deploying these systems in real scenarios we can still observe numerous open problems, both technological and scientific in nature. From a technological perspective, we have the acoustical variability of acquired sound from the distant speaker [2], or the problem of barge-in [1], wherein the human speaker interrupts the system's prompt. From a scientific standpoint, a typical challenge is the problem of identifying who is talking in an environment with multiple persons [3], [4]. Specifically concerning elderly people, the challenges expand into speech impairments related to age - dysarthria [5], for instance. Unforeseen issues like the expressiveness of the robot's face when speaking emerged as critical factors in user acceptance and became unexpectedly crucial during the experimental phase. These problems were out of the scope of the initial version of the GrowMeUp (GMU) system, causing these limitations to persist along the experiment.

In this paper we present an exploratory study conducted within the GMU project, where a social robot was tested for a period of nine months. This study focuses specifically on the speech recognition and synthesis, or Text-To-Speech (TTS), modules.

The work is organized in five sections. Section I has introduced our main problems and goals. Section II reports the methodology of this social experiment and gives some insights about the system itself and the speech components. Section III reports the study's outcomes and the identification of the technological and scientific problems. In Section IV presents improvements already deployed in the system. Finally, Section V presents the conclusions of this study.

II. Materials and Methods

Four elderly persons from two care organization, Cáritas (Portugal) and Zuyderland (The Netherlands), taken part of this study. The elderly were carefully chosen to match the profile of an elderly person in full cognition, without any health issue that could affect their cognitive performance, and who lives autonomously at their own homes.

The elderly tested the different services of GMU robot, summarized in Table 1, in different use-case scenarios, illustrating detailed realistic examples of how users carry out their daily life activities at home. The tests toke place either in Zuyderland (The Netherlands) and Cáritas (Portugal) and lasted for 9 months. The users tested the GMU robot over four 30-min to 1-hour sessions each, with each session taking place approximately one to four weeks from the last. All trials were performed in accordance to legal statement and the disclosure of the results was consented by all participants [6].

* This work was developed in the context of the GrowMeUp project, funded by the European Union's Horizon 2020 Research and Innovation Programme - Societal Challenge 1 (DG CONNECT/H) under grant agreement No 643647.

J. Oliveira (corresponding author), F. Perdigão, G. Martins and L. Santos are with University of Coimbra, e-mail: growmeup@uc.pt). F. Perdigão is also with Instituto de Telecomunicações, Pole of Coimbra. C. Dantas and A. Jegundo are with Cáritas Diocesana de Coimbra, Portugal. C. Wings is with Zuyderland Medisch en Zorgconcern, Sittard-Geleen, The Netherlands. J. Dias is with Khalifa University, Robotics Institute, Abu Dhabi, UAE.

TABLE 1 - TESTED SERVICES LIST

Services	Description
Creating activities and inviting friends	Create new activities in agenda and send an invitation to the user's friends
Objects location (Reminder)	Provide information about a possible location of an object based on static information from database
Behavior analysis and Suggestion	Sugest an activity that will comply with the current user emotional state
Fall detection	Detect whether a person has fallen based on captured images by the camera
Call for help	Contact a caregiver in case of emergency, triggered by the user or by a service (ex: fall detection)
Medication Management	Manage the medication information of elderly, providing some details on medicine including periodicity (day of start and day of end, and time to take the medicines)
Skype Call	Call to a person using Microsoft Skype
Setting Alarm Clock	Set an alarm or a reminder for a task or event on user's agenda
Reminder / Notification Service	Reminder for an agenda event or task
Emotion detection	Interacts with the user to access his current emotional state using dialogue emotion recognition.

Figure 1 - The GrowMeUp robot and an elderly user in trial environment.

Firstly, before the start of the tests, developers ran several laboratory tests to check if the services were ready to be transferred to expert end-users (caregivers, who are closely related to the end-users, are comfortable with technology and have some knowledge about the robot). Secondly, the caregivers tested the services to identify relevant issues that could compromise the elderly trials. With each anomaly found, the caregivers produced a report so that developers could reproduce the issue and solve it before end-user tests took place, as illustrated in Figure 2.

When the caregivers found the prototypes ready for end-user testing, the elderly tested the services by interacting with the robot. Elderly users were given a short verbal brief about how to interact with the robot, namely how to know when the system was listening and what commands does it follows.

After the tests, system evaluation took place with basis on interviews and questionnaires. The evaluation process consisted of a set of questions containing both qualitative and a quantitative information, as reported in [7], and aimed to cover the following topics:

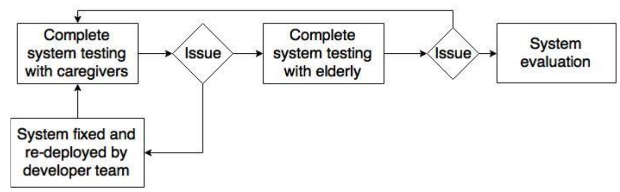

Figure 2 - Iteration process diagram in the GrowMeUp during system development.

- Usability;
- User satisfaction;
- Interaction with the robot;
- Ease of learning;
- Appearance.

During the nine-month period, developers followed the testing process closely to the end-users. Thus, they could perform improvements to the system so the tests could run smoothly. The adjustments made were essentially related to speech synthesis and recognition, namely bad recognition of the command words or bad pronunciations of some words. These issues led the developers to change some of the keywords that triggers the services as well as some of the robot's spoken utterances so that they could be better understood.

A. The GrowMeUp Robot - System Overview

The GrowMeUp robot, illustrated in Figure 1, is a social robot [8] designed to support elderly users in their daily activities at home, improving their everyday life [9]. The robot is equipped with a suite of sensors that enable it to learn the habits, preferences and routines of its user, allowing it to adapt its services over time.

This robot consists of a system based on Robot Operating System (ROS) [10] capable of performing some tasks and routines autonomously. The robot is connected to a cloud network, which allows it to increase its knowledge and share information with other robots, as well as to connect the elderly user to their caregivers, family and friends. The robot is also able to navigate autonomously, to detect human falls, to recognize users' faces [11] and to interact with users. Human-robot interaction happens mainly through speech, as this is the most human natural communication channel [12].

The speech interface (Figure 3) consists of three interconnected ROS nodes: the automatic speech recognizer (ASR), the dialogue manager (DM) and the text-to-speech synthesizer (TTS). The ASR system is based on Chrome's Speech API [13], offloading the recognition effort to remote servers. This method is triggered by the DM for speech recognition. Once DM flags the ASR to acquire audio, it opens the microphone system and starts collecting audio until it detects silence for 0.8 seconds, which is interpreted as end of speech. At this point the ASR stops recording and starts recognizing the audio, sending the audio to the speech API, which recognizes the words in the speech signal as a text string and returns it to the DM.

The DM is the dialogue processing center. It receives information from the recognizer and generates a response according the content of the speech. It can trigger actions, such as a Skype call, or additional verbal interactions with the user. This component also is connected to the cloud, allowing its personalization and update on-the-fly.

The TTS synthesis system that is integrated in GMU system is the Cerevoice® TTS, by Cereproc [14], able to synthesize speech in real time. In the context of these trials, three languages were used: Dutch, Portuguese and English.

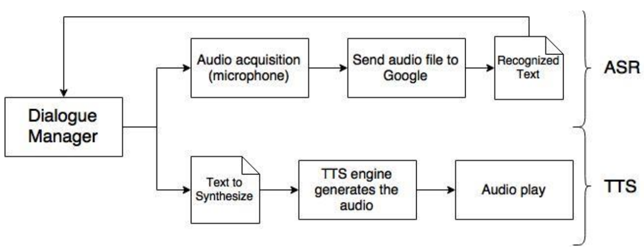

Figure 3 - Speech recognition and synthesis mechanism.

B. *GrowMeUp evaluation methods*

At the end of the tests, participants were asked to carry out a report about the robot (regarding the latest version of the iterative process). The reports consisted in a qualitative and quantitative part, according the following protocols:

- **Interview:** The system usefulness, ease of learning, user satisfaction and ethical and moral aspects were evaluated by a series of questions developed by the care organizations.
- **Think Aloud:** The user experience was tested by a think aloud protocol [15], wherein the elderly were requested to openly express their thoughts, observations, feelings and comments during each trial. These were documented according a pre-defined form.
- **System Usability Scale:** The usability of the system was mapped into the System Usability Scale (SUS), which yields a single measurement representing the overall usability of the system under study. The elderly were requested to answer to a 10-item questionnaire in five-point Likert scales [16].
- **Godspeed Questionnaire:** The effectiveness of the robot as a tool in terms of its appearance and users' perception of it was evaluated through a Godspeed questionnaire, [17], which consists in a validated measurement of the perception of the user in interaction with a robot. This questionnaire consists of five parts, with scores ranging from 1 to 5, that assess i) anthropo-morphism, ii) animacy, iii) likeability, iv) perceived intelligence and v) perceived safety. In the context of GrowMeUp this questionnaire gives insight in any problems with the behavior and/or the appearance of the robot that could not been identified with qualitative tests, once the elderly expect the robot could exhibit human traits

All data from questionnaires and interviews was processed and is reported and discussed in the next section.

III. RESULTS

A. *Quantitative outcomes - questionnaires*

To measure the usability and the impressions of elderly about the GrowMeUp system, two questionnaires were used.

1) *The System Usability Scale*

As described before, the overall SUS score has a range of 0 to 100 (where 0 represents very bad and 100 very good usability). The results of this test are indicated in Table 2.

2) *The Godspeed questionnaire*

In the Godspeed questionnaire the elderly people gave a score on a scale from 1 to 5 for each item. Table 3 shows the scores of each elderly separately and the mean score of all elderly.

From all the results, the animacy, perceived intelligence and likeability revealed to be crucial in the evaluation of the usability, ease of learning, appearance and the user satisfaction on the interaction with the robot.

Generally, the elderly found GMU quite interactive, responsive as well as likeable, friendly, nice and pleasant, as reported previously. However, elderly from Cáritas were more positive about GMU than the elderly from Zuyderland. This may be attributed to the fact that Dutch elderly are better acquainted to technology, since they have participated in several previous technological projects, including ones with robots [18], and are thus less immediately impressed with the system when compared to the Portuguese elderly.

TABLE 2 - SUS SCORES OF THE ELDERLY OF BOTH CARE ORGANIZATIONS

Care Organization	**Elderly**	**SUS Score**
Zuyderland	User 1	72.5
	User 2	50
Cáritas	User 1	80
	User 2	67.5
Mean		67.5
Standard-Deviation		12.75

TABLE 3 - GODSPEED SCORES OF THE ELDERLY OF BOTH HEALTH CARE ORGANIZATIONS.

		Zuyderland		Cáritas			
		User 1	User 2	User 1	User 2	Mean	
Animacy	*Dead*	1	3	4	4	3.00	*Alive*
	Stagnant	2	3	4	4	3.25	*Lively*
	Mechanical	1	1	3	4	2.25	*Organic*
	Artificial	2	3	4	4	3.25	*Lifelike*
	Inert	3	3	4	4	3.50	*Interactive*
	Apathetic	2	3	4	5	3.50	*Responsive*
Likeability	*Dislike*	3	4	4	4	3.75	*Like*
	Unfriendly	3	3	4	4	3.50	*Friendly*
	Unkind	4	3	4	4	3.75	*Kind*
	Unpleasant	3	4	4	5	4.00	*Pleasant*
	Awful	4	3	4	4	3.75	*Nice*
Perceived Intelligence	*Incompetent*	4	3	4	4	3.75	*Competent*
	Ignorant	3	3	3	4	3.25	*Knowledge-able*
	Irresponsible	3	3	4	4	3.50	*Responsible*
	Unintelligent	3	3	4	4	3.50	*Intelligent*
	Foolish	3	3	4	4	3.50	*Sensible*

B. Qualitative outcomes - interviews

The elderly consider that the GMU system can be especially useful for those who are living alone at home. The ability of reminding them of tasks and activities and creating reminders through speech has been pointed out as one of the most valuable features. In general, the users enjoyed the services rendered by the system. The elderly pointed out that the speech interface is one of the most valuated features, however, the quality of the interaction that can be achieved between the robot and the users is still unsatisfactory, due to the bad recognition, wrong pronunciations and machinelike expression of the robot while speaking.

The speech interface softens the learning curve, allowing the elderly to generally learn how to use the GMU system after a short introduction. It was notorious that two sessions after the beginning of the pre-trials, the elderly appeared to be more relaxed and comfortable with the robot and the interaction process. The worst problem reported was a difficult in understanding the robot: sometimes they couldn't hear or understand what GMU was saying. Furthermore, the users had to learn when and how to speak during interaction with the robot. For instance, the elderly users tended to speak in a low tone of voice, and using many idiomatic expressions, which current state-of-the-art speech recognition systems are unable to deal with.

The elderly do not feel that the GMU system interferes negatively with their wishes and will, and feel that their personal information will be safe and treated with confidentiality. However, the system's inability to properly recognize and interpret the user's speech can have a negative impact on these aspects. For instance, if the robot misses a crucial word in a command, for instance a negative, it may be led to execute the wrong service, potentially going against the user's will or exposing their personal information. Thus, it becomes crucial to employ accurate, reliable speech recognition techniques.

C. Discussion: Main Speech-Related Problems

The main feedback of end-users was that the current development state of the robot speech interface still needs some improvements so it can be used by all elderly. However, the capability of communicate and perform tasks through speech was well received. This feedback goes to meet the current development state of the robot prototype, i.e., the tested prototype of the robot had a few development problems, which are in the development pipeline. Due to this development state, some features were not completely developed and some others need improvements.

Although the performance of the speech recognition and synthesis facilities was not directly measured by the questionnaires applied to the users, the robot's speech system became one of the main topics of concern, being described by some of the users and caregivers as one of the main issues with the current prototype.

The interviews conducted have allowed us to isolate several problems that are inherent to this system, which can be split into three basic categories:

- Dialogue
- Speech Synthesis
- Speech Recognition

Dialogue-specific issues, related to the dialogue workflow, are out of the scope of this work. However, some of the dialogue-related problems can be attenuated by improvements on the speech recognition and synthesis systems of the robot. For instance, the users mentioned that it was difficult to follow the timing of the interaction, despite the visual cues used by the robot to signal that it was ready to receive input. This issue could be solved by improving the robot's speech recognition to not require discrete capture intervals.

The users had trouble understanding the robot, revealing problems with the output speech. This is attributed to two main causes: the acoustic qualities of the robot's structure and the too smooth concatenation of speech segments generated by the speech synthesis toolbox, particularly noticed when speaking Portuguese.

The inability to properly understand the robot was a source of frustration and dissatisfaction to both the users and the caregivers, by impairing the user's ability to understand the messages the robot was trying to convey, thus confusing the user as to the state of the interaction and of the service being rendered. This can be seen as one of the reasons for the quantitative measurements obtained. There are issues understanding users. Firstly, speech employed by elderly users exhibits some characteristics that hinder the performance of speech recognition systems. Secondly, a main problem is that the user is far away from the robot's microphones, leading to the distant speech and reverberation problems, which goes beyond the elderly speech scenario [15]. This led the users having to learn how to use the system, adapting their expressiveness to that of the robot, thus speaking slowly, louder, in very discrete windows and making use of very little idiomatic and para-verbal expressions.

Furthermore, the robot was deemed machinelike, mostly because of its inability to move the mouth when speaking, it sounded artificial due to the sound equalization, which was producing a lot of within the robot's housing, resulting in mechanic or unperceivable sound. Also, the turn-taking delays in the dialogue were significant, which hinders the communication. These interpretations emerge from correlating the quantitative and qualitative evaluation methods.

IV. IMPROVEMENTS

From these trials we could understand that one of the most important features of the GrowMeUp robot is the ability to talk clearly and to understand speech. In fact, we observed that the reaction of the users when they start interacting with the robot was empathy and relaxation, due to the inherent naturalness of the interface. However, this interface is still hindered by several problems, as described in III.C, for which we now present a set of recent improvements.

In order to improve the robot's speech synthesis, some adjustments were made: a dynamic sound equalizer was implemented, which adapted the sound to the acoustics of the robot's enclosure, improving the sound quality; a custom user-lexicon of words was developed, improving synthesizer's performance in pronunciation; some prosody contours were applied to sentences. In order to provide a visual feedback of the speech being produced, mouth animations, synchronized with the phonetic events, were added according a phoneme-to-viseme mapping (Figure 4) [19].

Figure 4 - GrowMeUp robot mouth shapes for the word "Hello".

Concerning speech recognition, as described above, the robot did not recognize the user's commands with sufficient accuracy. Furthermore, it was difficult to the users understand what robot was doing due the delay in the response caused by the remote processing of the sound, often confusing users.

Many factors may contribute for these problems. On one hand, due to aging-related problems, the elderly speech may become dysarthric [5], which leads to the utterances not being spoken with clarity. This hinders the performance of the recognition system, since it was not specially trained for elderly or dysarthric speech [20]. On the other hand, the speech acquisition setup has some inherent technical problems, because the user is far from the microphone, giving rise to the well-known distant speech recognition problem [2]. This results in a low-quality audio signal, which may compromise accuracy of the recognition. Although the system is composed by several microphones, no signal enhancement has been made so far on the acquired audio, exposing an interesting avenue of future work, partially due to the computational complexity of the algorithms involved.

To improve the responsiveness of the robot's speech recognition system, a new mechanism for speech-based interaction was developed. The resulting Speech Acquisition System (SAS) comprises a voice activity detection (VAD) mechanism [21], consisting of a low-level method to partition and enhance the Signal-Noise-Ratio (SNR) of the audio before sending it for recognition. It operates even when robot is in standby mode.

This allowed us to change the previous interaction paradigm, wherein speech recognition was performed solely in discrete intervals, in favor of a continuous system: the SAS continuously acquires audio and processes it to detect voice activity. When the SAS detects voice activity, it flags the ASR to start processing audio data continuously until an end of speech condition is reached.

The VAD system consists of a set of circular buffers, containing the audio data from microphones and first-order statistics of the energy of the signal. It is always collecting audio data through the microphones, computing metrics and thresholds, and storing all values in the circular buffers. When consecutive data chunks exhibit energy above a dynamic threshold, the VAD triggers the voice activity flags, activating the recognition module.

The ASR performs speech recognition based on the Google Cloud Speech API [22] (GCSAPI). Since GCSAPI supports real-time streaming, the recognizer module, once flagged by VAD system, starts sending audio chunks to GCSAPI and collecting the recognition results. Thus, when the user stops speaking, the recognizer module has already processed the previous audio, having obtained the text corresponding to the audio and the robot's response becomes virtually instantaneous (Figure 5).

After all, this approach not only fits to the elderly speech, but the general speech, so it can be applied to other scenarios of interaction between human and machines.

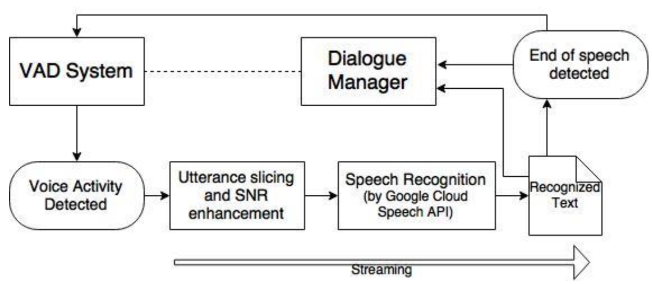

Figure 5 – Mechanism of Speech Acquisition System.

V. CONCLUSIONS

This study allowed us to understand that the execution of the services through speech, in the given context, despite some limitations, makes the GMU robot a useful and desired system by elderly and caregivers. Some emergency services, highly valued by the users, were developed in a way that dialogue failures could not compromise their performance.

The natural speech interface was determined as one of the most important and critical factors for acceptance by the end-users. It also became clear that the interaction through speech allowed elderly to easily learn how to use the robot, softening the associated learning curve.

It was determined that the robot must talk in a natural and understandable way regarding volume, equalization, prosody and pronunciation, visually illustrating its speech with dynamic mouth shapes. Our improvements in these areas have brought the robot's speech a step closer to full naturalness.

Speech recognition is also a very important feature. Therefore, it is important that the speech recognition be accurate and work in real-time, preventing the user from having to repeat their commands and to shorten the delay between the end of users' speech and robot's reaction. Thus, a cloud-based real-time recognition system offers the desired reliability and timing on the recognition, allowing not to overload the system with extremely complex processing, such as the case of local speech recognizers, as well as having access to a wide variety of languages without increasing the system effort. Furthermore, it is not natural to have a short time interval to speak to a machine. A reactive system that avoids timing issues (the implemented voice activity detector), improved the human-robot communication to and turn the system more user-friendly.

This work allowed to expose some speech-related open issues specific to human-robot interaction with elderly users. Some issues like barge-in, distant speech and multiple speakers to interact with the system at the same time are real problems that still need to be answered for this kind of systems.

Meanwhile, GMU will be able to wake up from its standby mode, or react to a help request, even in standby. This feature is a word spotting system, which will work similarly to the speech recognizer, but processing data locally and able to spot a limited set of words. This allows the system to be constantly listening, waking up when a specific command is given, without sending unnecessary information to the cloud.

Furthermore, audio pre-processing techniques are in development, which will allow the acquisition system to simultaneously process the audio from different microphones, reducing the reverberation and the distant-speech problem, allowing the recognition system to increase its accuracy. Since the synthesizer and the recognizer belongs to the same system, the barge-in problem, what may allow to identify the speech that comes from the robot itself or another source.

Lastly, this exploratory study provides a proof of concept for the integration of robots on elderly people lives, exposing some speech communication problems and solutions, tested in real scenarios.

REFERENCES

[1] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[2] K. Kumatani and T. Arakawa, "Microphone array processing for distant speech recognition: Towards real-world deployment," *Proc. APSIPA ASC, …*, 2012.

[3] D. Wang and G. J. Brown, *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley-IEEE Press, 2006.

[4] R. Gomez, K. Inoue, K. Nakamura, T. Mizumoto, and K. Nakadai, "Speech-based human-robot interaction robust to acoustic reflections in real environment," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 1367–1373.

[5] A. B. Kain, J. P. Hosom, X. Niu, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelligibility of dysarthric speech," *Speech Commun.*, vol. 49, no. 9, pp. 743–759, 2007.

[6] GrowMeUp, "GrowMeUp - Public Deliverables." [Online]. Available: http://www.growmeup.eu/index.php/dissemination/deliverables. [Accessed: 01-Mar-2017].

[7] GrowMeUp, "D6.1: Trials Specification and Design," 2015. [Online]. Available: http://www.growmeup.eu/images/Documents/PUDeliverables/D6.1 Trials Specification and Design.pdf. [Accessed: 02-Mar-2017].

[8] S. Taipale, J. Vincent, B. Sapio, G. Lugano, and L. Fortunati, "Introduction: Situating the Human in Social Robots," in *Social Robots from a Human Perspective*, J. Vincent, S. Taipale, B. Sapio, G. Lugano, and L. Fortunati, Eds. Cham: Springer International Publishing, 2015, pp. 1–7.

[9] G. S. Martins, L. Santos, and J. Dias, "The GrowMeUp Project and the Applicability of Action Recognition Techniques," in *Third Workshop on Recognition and Action for Scene Understanding (REACTS). Ruiz de Aloza*, 2015.

[10] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Mg, "ROS: an open-source Robot Operating System," *ICRA*, vol. 3, p. 5, 2009.

[11] G. S. Martins, P. Ferreira, L. Santos, and J. Dias, "A Context-Aware Adaptability Model for Service Robots."

[12] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Rob. Auton. Syst.*, vol. 63, pp. 22–35, 2015.

[13] G. Shires and H. Wennborg, "Web Speech API Specification," *Speech API Community Group, W3C*, 2012. [Online]. Available: https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html.

[14] J. M. Garrido, E. Bofias, Y. Laplaza, M. Marquina, M. Aylett, and C. Pidcock, "The CereVoice speech synthesiser," in *V Jornadas en Tecnología del Habla*, 2008, pp. 126–129.

[15] C. Lewis and J. Rieman, "Task-Centered User Interface Design: A Practical Introduction," *Text*, p. 190, 1993.

[16] J. Nielsen, "Success Rate: The Simplest Usability Metric," *Jakob Nielsen's Alertbox*, vol. 18, pp. 3–5, 2001.

[17] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1. pp. 71–81, 2009.

[18] CaMeLi, "CaMeLi | Care Me for Life." [Online]. Available: http://www.cameli.eu/site/. [Accessed: 13-Jun-2017].

[19] T. Chen, "Audiovisual speech processing," *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 9–21, Jan. 2001.

[20] J. et al Proença, "Characterizing Parkinson's Disease Speech by Acoustic and Phonetic Features," in *Computational Processing of the Portuguese Language: 11th International Conference*, J. Baptista, N. Mamede, S. Candeias, I. Paraboni, T. A. S. Pardo, and M. das G. Volpe Nunes, Eds. Cham: Springer International Publishing, 2014, pp. 24–35.

[21] J. Ramirez, J. Górriz, and J. Segura, "Voice activity detection. fundamentals and speech recognition system robustness," *Robust Speech Recognit. …*, no. June, pp. 1–22, 2007.

[22] Google, "Google Speech API," *Google Cloud Platform*, 2016. [Online]. Available: https://cloud.google.com/speech, 24/05/2016, 17: 28.