



Field Evaluation with Cognitively-Impaired Older Adults of Attention Management in the Embodied Conversational Agent Louise

Pierre Wagnier, Giovanni Carletti, Yann Laurent-Corniquet, Samuel Benveniste, Pierre Jouvelot, Anne-Sophie Rigaud

► To cite this version:

Pierre Wagnier, Giovanni Carletti, Yann Laurent-Corniquet, Samuel Benveniste, Pierre Jouvelot, et al.. Field Evaluation with Cognitively-Impaired Older Adults of Attention Management in the Embodied Conversational Agent Louise. 4th International Conference on Serious Games and Applications for Health (IEEE SeGAH 2016) , <http://web.ipca.pt/segah2016/index.htm>, May 2016, Orlando, United States. hal-01266477v1

HAL Id: hal-01266477

<https://minesparis-psl.hal.science/hal-01266477v1>

Submitted on 2 Feb 2016 (v1), last revised 5 Feb 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Field Evaluation with Cognitively-Impaired Older Adults of Attention Management in the Embodied Conversational Agent *Louise*

Pierre Wagnier ^{*†}, Giovanni Carletti ^{‡¶}, Yann Laurent-Corniquet [‡],
Samuel Benveniste ^{*‡}, Pierre Jouvelot ^{*} and Anne-Sophie Rigaud ^{†§}

^{*} MINES ParisTech, PSL Research University, France (name.surname@mines-paristech.fr)

[†] Université Paris Descartes, France (name.surname@parisdescartes.fr)

[‡] Centre d’expertise nationale en stimulation cognitive, France (name.surname@censtimco.org)

[§] Broca Hospital, Assistance Publique - Hôpitaux de Paris, France (name.surname@aphp.fr)

[¶] LIAS - Institut Marcel Mauss, CNRS-EHESS, France (name.surname@ehess.fr)

Abstract—We present the first experiment we conducted to evaluate the attention monitoring performance of Louise, following a *Wizard of Oz* method, during the interactions with a cohort of 8 elderly users in a day hospital environment. Louise is a new, semi-automatic prototype of an Embodied Conversational Agent (ECA), a virtual character interacting with users through social-like communication, adapted to the special needs of older adults with cognitive impairment; it is intended to ultimately provide assistance in their activities of daily living. We recorded and analyzed both videos of the conversation-like interactions and Louise’s tracking data. In our experiment, Louise’s attention estimation algorithm achieved about 80% accuracy; moreover, in almost all cases, the user’s attention was successfully recaptured by Louise after a planned, experimenter-induced distraction. These results are in line with what was observed in previous experiments involving only younger adults, thus suggesting that attention measurement tools embedded in cognitive prostheses will not need to be adapted to elderly patients. Finally, to gain further insights on conversation management and provide evidence-based suggestions for future work, we performed an anthropological analysis of the whole experiment.

Index Terms—Assistive technologies, attention, cognitive impairment, dementia, embodied conversational agent, older adults.

I. INTRODUCTION

Dementia is a general term that designates the symptoms caused by several neurodegenerative diseases that affect older adults. The most prevalent of them is Alzheimer’s disease (AD). The main risk factor to develop such diseases is old age. The aging of the global population yields an increase in the number of people living with dementia and a caregiver shortage. For these reasons, dementia care is now widely recognized as a public health priority. Indeed, the World Health Organization predicted that the number of people living with dementia in the world, will exceed 100 million people by 2050 [1].

To tackle this challenge, assistive technologies have gained momentum in the past decade. Assistive devices are meant to provide patients and their caregivers (whether families or professionals) with means to improve quality of life and

quality of care. In that regard, computer-based technologies may be of great help. However, due to the low computer literacy of this public, introducing this new tool in their environment is challenging in terms of accessibility. The main reason is the great difficulty of older adults with cognitive impairment to acquire new skills. In this context, Embodied Conversational Agents (ECAs), virtual characters interacting with users through social-like communication, could be an effective interaction modality for assistive technologies, relying on verbal and non-verbal communication, as well as visual media, to convey information and take users’ inputs.

The goal of our project is to rely on video game technologies (game engine, 3D animation and Kinect) to create an ECA, called Louise, that is adapted to the special needs of older adults with cognitive impairment. Older adults with dementia have executive and attentional disorders [2], which come in addition to the diminution of attentional capabilities observed in normal aging [3]. Our team, hosted at the Broca Hospital in Paris, in the LUSAGE living lab [4], observed in many assistive technology trials that people with dementia often lose track of what they are doing and their attention needs to be redirected towards the task. By “attention”, we mean the ability one has to focus on a specific stimulus or task; we thus decided to equip Louise with attention management capabilities. The attention estimation method and the ECA prototype were developed and validated with healthy younger adults in a previous work [5].

In this paper, we present the first experiment we conducted to evaluate the attention monitoring performance of Louise, following a *Wizard of Oz* method, during the interactions with a cohort of 8 elderly users in a day hospital environment. We recorded and analyzed both videos of the conversation-like interactions and Louise’s tracking data. Our experiment shows that Louise’s attention estimation algorithm achieved about 80% accuracy; moreover, in almost all cases, the user’s attention was successfully recaptured by Louise after a planned, experimenter-induced distraction. These results are in line with what was observed in previous experiments involving only

younger adults, thus suggesting that attention measurement tools embedded in cognitive prostheses will not need to be adapted to elderly patients. To gain further insights on conversation management and provide evidence-based suggestions for future work, we performed an anthropological analysis of the whole experiment.

The first contribution of this paper is to introduce in an assistive ECA prototype attention management capabilities which, based on the experience of our interdisciplinary team of physicians, neuropsychologists and computer scientists, is an important feature, and evaluate it with older adults with cognitive impairment. Our second contribution is to provide evidence that the analysis of interaction videos with tools from anthropology helps gaining insights on the specifics of conversation management with our target user group and drive future research.

The paper is organized as follows. Section II discusses previous work related to ECAs and conversation management techniques. Section III describes our system and simple attention monitoring algorithm. Its validation with older and cognitively-impaired adults is addressed in Section IV, where we describe the participants, the protocol, the results and compare those with the ones obtained previously with younger adults. To help us define future work, we performed an anthropologically-based analysis of the whole project, which is summarized in Section V. Section VI highlights its key findings and how they will impact future work. We then conclude in Section VII.

II. RELATED WORK

In this section we review the works on the use of ECAs in elderly care. More specifically, we sum up evidences that show that ECAs are a well adapted Human-Machine interaction modalities and review works on specific interaction management for cognitively impaired people.

A. ECA in Cognitive Assistance for Older Adults

In the literature, some authors have worked on showing that ECAs are the most effective interaction modality for older adults with cognitive impairment. The first study was performed by Ortiz *et al.* [6], who compared three prompting modalities: a virtual character, speech with synchronized on-screen text and text alone. The study involved 15 elderly adults, some of whom had Mild Cognitive Impairment (MCI) or AD. They found that most subjects liked the virtual character more than other modalities and performed significantly better in a task when guided by the virtual character. These findings are also supported by other results in [7]. In this study, Morandell *et al.* compared a photograph with animated lips with a speech and text prompt. The experiments involved 10 older adults with cognitive impairment. These authors also found that guidance by the talking face yielded better task performance than with the other modalities, especially for the most cognitively impaired participants. In addition, the study participants reported that the talking face was more pleasant than the disembodied voice. Lastly, the authors observed that

the talking face captured the participants' attention better than the speech and text and was understood better, thanks to co-articulation of the character's lips with the speech. Lastly, Morandell's team findings on task performance were confirmed in a larger-scale study, involving 12 healthy older adults and 12 seniors with MCI [8].

An interesting feature of ECAs for this public is that, being virtual, they can be personalized and configured depending on the users' tastes and cognitive abilities. Indeed, dementia symptoms have high variability across patients. Lapointe *et al.* have proposed comprehensive guidelines for the choice of prompting modalities to use in ambient assisted living (assistive smart homes), depending on the person's remaining abilities, to increase effectiveness. This approach is also supported by Diaz-Orueta *et al.*, who have studied the influence of older adults' cognitive status on their interaction with an ECA [9]. In this study, 20 participants with cognitive impairment ranging from MCI to moderate AD interacted, through a two-buttons remote control with a character displayed on a TV set. They concluded that cognitive screening can help predict the user's performances in the task and that paying attention at specific kinds of deficits can help produce better assistive device designs.

Regarding the user input modalities, Carrasco *et al.* [10] proposed to use a TV remote control with two buttons (yes and no), as this device is familiar for the elderly. They reported that all 21 participants with AD involved in their study successfully interacted with the ECA. More Recently, several authors have proposed to use verbal and non-verbal inputs. Sakai *et al.* built a prototype of a listener agent that can do active listening, through non-verbal speech analysis (power and pitch) [11]. The conversational agent asks a question and performs backchannel (nods and acknowledgment vocalizations) according to the pitch variations of the user's speech. This prototype also performs speech recognition. An ECA for people living with dementia has also been developed by Yasuda *et al.* [12]. The virtual character, a cartoon-like young boy, asks people reminiscence questions about their past and performs active listening while people tell their story. It was evaluated with a group of 8 older adults with AD. The authors found that people uttered only 26% less words, on average, than with a human conversational partner. Lastly, Wilkis *et al.* tested a prototype ECA, which includes speech recognition and emotional state estimation, for smart home applications with a brain-injured veteran; good communication with the test patient was observed [13].

To sum up, firstly, the studies presented here suggest that ECAs are better media to convey information and provide guidance to older adults with cognitive impairment than text and disembodied speech. Secondly, ECAs are found pleasant by most elderly people, attract attention better than other modalities and their use yields better task performance. Thirdly, some authors recommend to tailor the prompting modalities and interaction modalities to the cognitive and sensory abilities of each patient. Lastly, the authors we cited above showed that people with MCI or dementia successfully

interacted with ECAs, whether through remote control button presses or social-like interaction.

B. Conversation Management

We have only found one team that published works on the details of the interaction management between an ECA and a cognitively impaired person [14]. In this study, Yagoubzadeh *et al.* asked 11 younger adults with intellectual disability and 6 healthy older adults to interact with Billie, a virtual assistant that looks like a young boy, to enter appointments in a virtual calendar. There were three information per appointment: date, time and topic. The study was conducted following the *Wizard of Oz* method and specifically focused on error recovery. The goal was to compare two error recovery strategies: after completing a calendar entry, the participant was asked to check the information. In the global condition, all the information of the entry was checked at once whereas, in the local condition, each piece of information was checked separately. The authors reported that all participants but one successfully interacted with the ECA to put their schedule in the virtual calendar. In addition, they observed that the subjects repaired significantly more errors in the local condition than in the global condition; this improvement was much more significant in the cognitively impaired group, who performed just as well as the healthy elderly group in this condition.

III. SYSTEM DESCRIPTION

We conducted a *Wizard of Oz* study with a semi-automated ECA prototype, called Louise, that monitors the user's attention during the interaction. In this section we present the test system, developed in a previous work [5], and describe the attention estimation method we used.

A. The Louise ECA

The Louise ECA is a virtual cartoon-like female character. It is animated and displayed with a neutral background and includes speech synthesis. The layout is presented in Figure 1. A Microsoft Kinect sensor is used to monitor the user during the interaction to perform attention estimation, thanks to the algorithm described in Section III-B. The interaction is managed thanks to a written scenario, consisting in a list of utterances. The utterances are spoken in a predefined order. When the user answers a question Louise asks, a hidden operator has to press a key to move one to the next utterance. For transitions, Louise performs an acknowledgment utterance, randomly selected in a dedicated list. When a loss of attention is detected, Louise automatically sends a prompt, also randomly chosen from a list. Lastly, when the user pays attention again, a transition phrase is spoken, before asking the last unanswered question again.

The prototype is built from three software modules (see Figure 2, where the arrows stand for data exchanges).

- **Attention estimator.** Body and face tracking data is extracted and processed to estimate the user's attention as detailed in Section III-B.



Fig. 1. Louise, our virtual character

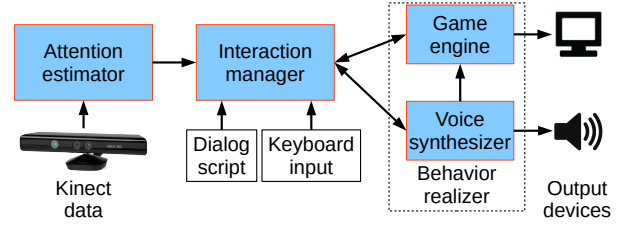


Fig. 2. Functional diagram of the Louise ECA prototype

- **Interaction manager.** A scenario and keyboard presses are used as inputs to perform the conversation. The course of the dialog is automatically interrupted to perform attention prompting when a loss of user's attention is detected.
- **Behavior realizer.** This part of the program animates and displays the character, performs voice synthesis and lip synchronization. The character animation was implemented using the Unreal Engine 4 game engine and voice synthesis uses the Cereproc Cerevoice speech synthesis engine.

B. Attention Monitoring Algorithm

Older adults with cognitive impairment often lose track of what they are doing. We thus implemented an attention estimator to monitor the user's gaze direction throughout the interaction. Our method relies solely on determining, in real time, if the user is gazing towards the screen or away from it. It combines the measures of orientation of the user's shoulders [15] [16] and head pose [17], seen here as proxies for his or her intensity of attention towards Louise.

This method relies on the Kinect's skeleton and face tracking data. Only the 3D positions of the shoulders and the yaw and pitch rotations of the head are used (see Figure 3 for notation). It assumes that the sensor is placed on top and in the middle of the screen displaying the ECA.

The azimuth of the user is defined as $\theta = \arctan(N_x/N_z)$ and the angle of the upper-body α as

$$\alpha = \arccos\left(\frac{N_z - SL_z}{\sqrt{(SL_x - N_x)^2 + (SL_z - N_z)^2}}\right) - 90. \quad (1)$$

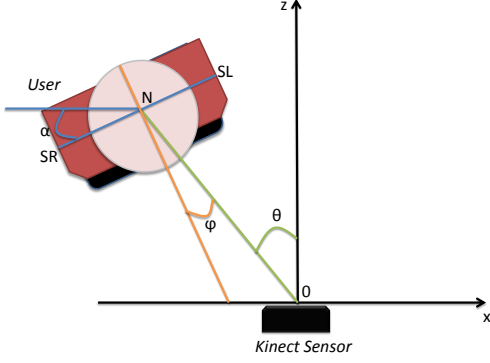


Fig. 3. Estimation of the angle φ between the shoulder line and the Kinect sensor. θ is the angular position of the user, SR the position of the right shoulder, SL the position of the left shoulder, and N the position of the neck (calculated as the center of the vector $[SR, SL]$).

The posture feature f_1 is thus, at each time t :

$$f_1 = \varphi = \alpha - \theta. \quad (2)$$

The Kinect's face tracker outputs, for every sampling time t , the three rotation angles of the tracked head: pitch $\gamma_{pitch}(t)$, yaw $\gamma_{yaw}(t)$ and roll. To make the estimation more stable and more robust to noise and to the failure of the face tracker over a few frames, these angle values are averaged over a period $T = 30$ frames, which roughly corresponds to one second, given the sensor's sample frequency (see Equations 3 and 4). This calculation produces Values f_2 and f_3 for the monitoring algorithm, at each time t :

$$f_2 = yawMean = \frac{1}{T} \sum_{k=t-T+1}^t \gamma_{yaw}(k), \quad (3)$$

$$f_3 = pitchMean = \frac{1}{T} \sum_{k=t-T+1}^t \gamma_{pitch}(k). \quad (4)$$

The three features f_j ($j = 1, 2, 3$) are normalized as $\overline{f_j}$ in the same way:

$$\overline{f_j} = \frac{\cos(f_j) - \cos(Max_j)}{1 - \cos(Max_j)}, \quad (5)$$

where Max_j represents the maximum value for each feature f_j ; these values correspond to the Kinect's tracking limitations (30° for yaw, 20° for pitch and 60° for upper-body pose).

A sum of the $n = 3$ normalized features $\overline{f_j}$, weighted by coefficients ω_j (see Equation 6), is computed to assess the *AttentionLevel* for each sampling instant. The face's horizontal rotation has the heaviest weight to account for the importance of the face's orientation in the attention estimation. This corresponds to using this information as a proxy to the user's gaze direction. For normalization purposes, the sum of the weights is equal to 10:

$$AttentionLevel = \sum_{j=1}^n \omega_j \overline{f_j}. \quad (6)$$

The obtained attention level values range from 0 to 10, 10 being the maximum level, when the user's body and face are directly oriented towards the sensor. These values are then used to decide the user's attentional state, i.e., whether the user is engaged or not, using a hysteresis threshold rule: the user is considered engaged if the attention value is more than 8. Transition from engaged to disengaged is triggered when the attention value decreases below 6. Two more states are used: "user detected" (at the beginning, when the user has been detected and is not engaged yet) and "no user".

IV. EXPERIMENTS WITH OLDER AND COGNITIVELY-IMPAIRED ADULTS

We performed a series of pilot experiments with a small group of older and cognitively-impaired adults to assess the viability of the attention monitoring algorithm presented above for such a population. Our study involved 8 participants, 2 males and 6 females. All participants were older adults from 63 to 91 years old. The average age of the subjects was 78 ± 1 years old (*mean \pm standard deviation*). 6 participants had cognitive impairment (3 MCI and 3 AD). The mini-mental state examination (MMSE) scores of these 6 participants ranged from 17 to 29 (*mean* = 23 ± 3). The two other participants did not have cognitive impairment. Participants were recruited on a voluntary basis, among patients from the Broca geriatric hospital in Paris. All participants signed an informed consent form and a written authorization to film them before their participation to the experiment. The experimental data was analyzed anonymously.

A. Experiment Protocol

Participants were seated in front of the screen on which the ECA was displayed. They were told that the character on the screen was going talk to them and to ask them questions, to which they were instructed to answer.

The interaction consisted of 3 utterances for the introduction, 7 questions, 5 acknowledgment utterances, 5 prompting utterances, a transition utterance and 2 utterances for the conclusion. When necessary, the acknowledgment and prompting utterances are randomly selected from the corresponding list in the scenario; the full list of utterances can be found online [18]. When a user got distracted, Louise automatically stopped and prompted the subject to attract his or her attention. After such an interruption, she always asked if they wanted to continue the interaction, at which point the participant could decide to stop the test.

During the interaction, two distractions were voluntarily introduced at fixed moments. The first one was introduced by the experimenter in the room at the beginning of the third introduction utterance: he asked the participant if the sound was loud enough. The second distraction was introduced by another experimenter, opening the door and asking the participant if everything was fine, during the fifth question.

The data samples produced by the Kinect and used in the attention estimation method were recorded during the interaction, as well as the color image with an overlay showing the tracking information (face keypoints and skeleton). These videos were then annotated by three experts, asked for each recorded instant to judge if the user was paying attention to the ECA (i.e., looking at the screen) or not. Two of the annotators worked independently. The third annotator had to arbitrate when the two annotators did not agree. The annotations were then compared to the decisions taken by the attention estimator, using only the third annotator's data.

To evaluate our attention estimation method, we computed the correct detection rates by comparing the decisions taken by the algorithm during the experiment with the human annotations. This was done for each data sample, on all the data at once and per subject (see Table I). We also computed a receiver operating characteristics (ROC) curve for our classifier, using 21 threshold values from 0 to 10, with a step of 0.5 (see Figure 4). The area under the ROC curve (AURC) was also computed.

In addition, we computed correlations to see which of the three features used in the attention estimation method are the most relevant, using point-biserial correlation coefficients [19] between the ground-truth human annotations of subjects' attention and each of our three features, as well as with the aggregation of all three features in a single attention value. We chose this correlation coefficient because it is adapted to check for correlations between qualitative and numerical data. Lastly, we did a group comparison with the data from the previous experiment that had the same protocol and only differed by the content of the questions. This experiment involved 14 healthy younger adults, and intended to check if there were significant differences in the attention estimator between the two groups. This was done using the Wilcoxon rank sum test for equal medians, which is adapted to small group sizes [20]. Most computations were performed using Mathworks Matlab. Only the per-participant performance scores were computed using Microsoft Excel.

Lastly, we observed the videos of the experiment to evaluate the efficacy of the prompts to attract the participants' attention. The only indicator we used was the number of times the experimenter in the room had to ask the participant to look at the screen and talk to the character.

B. Results

6 out of 8 participants successfully interacted with Louise. One of the participants had a difficult time understanding the voice of the character, because he had very poor hearing. However, when he did understand or was helped, he could interact successfully. The reason why another participant could not interact successfully throughout the scenario is that she lost track of the context and refused to continue after the first distraction; her cognitive impairment was the highest of the cohort (MMSE = 17).

Regarding the evaluation of the attention estimator, all but one tests produced usable data. In that test, the Kinect tracking completely failed (it tracked a chair in the background instead

of the subject). Lastly, during one of the tests, body and face tracking was lost for a few seconds. For this reason, the data showed below is presented with and without the major tracking failures (face tracking was sometimes lost for a few data frames, but we did not remove the corresponding data).

The mean per-participant correct detection rates, presented in the last two rows of Table I, show encouraging results. The individual scores ranged from 24% (the participant for which the tracking completely failed) to 92%. The figures presented in the last row exclude the data of the participant with the lowest score. The difference in performance observed when considering the subjects with cognitive impairment only (PCI, in Rows 3 and 4 of Table I) is mostly due to the fact that all major tracking failures happened with this group.

Regarding the ROC curve, it shows that the algorithm is quite sensitive for most values of the threshold. Unfortunately, the specificity is not as good. However, using a two-threshold hysteresis for decision making allows being more specific to transition of the user's state representation from "paying attention" to "not paying attention".

When computing the point-biserial correlations, the most relevant feature is f_2 ($r_2 = 0.46$), which, considered separately, has higher correlation with the attention annotations than the weighted sum of the features ($r = 0.38$). f_1 and f_3 have lower correlation coefficients: $r_1 = 0.14$ and $r_3 = 0.14$.

Lastly, out of a total of 19 distractions over all the experiment sessions, the experimenter only had to ask the participant to look at the screen once. This shows that the attention-recapture strategy is quite effective. However, most of the time, the subject did not get distracted long enough for the character to say one of the prompting phrases (one or more prompting phrases were triggered only 6 times in 19 distractions), and directly asked the after prompting question. Contrary to what we expected, we did not observe any situation in which a patient with cognitive impairment looked away from the screen because he or she lost track of what he or she was doing. The only self-induced distractions were related to hearing and understanding issues.

C. Comparison with Healthy Younger Adults

In previous work [5], this attention estimation algorithm was evaluated with younger adults, who are assistive technology experts. The method showed over 80% accuracy with this group. The experiment reported in this paper aimed at validating the attention estimator with older adults. Our motivation for this second set of experiments was to assess whether simple attention monitoring algorithms were adapted to an older population with cognitive disabilities or whether specific algorithms were needed to cope with such a population.

The results presented above, together with those of our previous experiment, enable us to perform such a group comparison. The same experiments, when performed with the two groups formed of younger adults and older ones, did not show any significant differences in medians ($p > 0.05$). This means that the validation data of both groups could be considered together. The overall performance of our method,

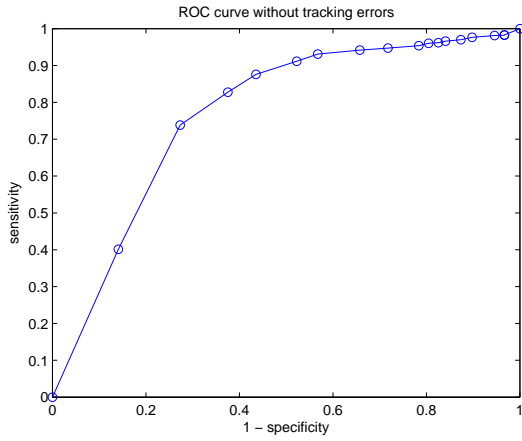


Fig. 4. ROC curve of the attention estimator.

TABLE I
PERFORMANCES OF THE ATTENTION ESTIMATOR IN PERCENTAGE OF
CORRECT DECISIONS.

Data set	Perf.	AURC	samples rem.
All samples	76%	0.64	
No tracking errors	80%	0.77	8.3%
PCI only	72%	0.6	
PCI no tracking errors	77%	0.77	10.9%
Mean per subject	$75 \pm 17\%$		
Mean p/sub. no tr. error	$82 \pm 7\%$		1 subject

taking all data, is 84%, with an AURC of 0.68, or 86% (AURC = 0.74) if the skeleton tracking errors (3.3% of the samples) are removed.

A key insight gained from our experiments is thus that simple attention monitoring technologies embedded within ECAs appear to be usable, without specific tuning, by both types of population. Such a result provides thus a reasonable degree of assurance that off-the-shelf and thus inexpensive technologies for attention monitoring can be used when needed in more advanced cognitive prosthesis developments.

V. ANTHROPOLOGICAL INTERACTION ANALYSIS

To gain insights on the conversation management issues addressed by our experiment and provide evidence-based suggestions for improving the human-computer interaction (HCI) design for Louise, anthropological expertise was recruited. The goal was to analyze the interactions between the participants and Louise during the experiment carried out to evaluate the attention estimator presented in Section IV.

A. Method and Hypotheses

The empirical material consisted in the audio/video recordings of the experimentation (provided by the experimenter). To guide our analysis, take into account the context of the experiment, and follow best practices of ethnographic research [21], [22], the data was complemented with interviews with the experts and a data session with Louise’s leading designers.

The video material was coded by two independent annotators in order to perform a quantitative conversation analysis (CA) using the ELAN software to supplement and test the qualitative analysis [23].

The data and interviews highlighted two targets of general interest: a) to improve the ECA’s overall interactional adaptability to people with dementia (PWD); b) to cope with the structural asymmetry of the interactional setting (programmed “one-off volley” questions/answers vs. improvised conversation; institution vs. individual; enabler vs. enabled). Five hypotheses emerged from our qualitative analysis: H1) PWD proportionally utter more words than healthy older adults, seen as control subjects (CS); H2) PWD develop more topic expansion; H3) silences are longer before answers to open questions than before answers to polar questions; H4) the social setting of the experiment significantly modifies the HCI; H5) PWD speak more with the experimenter than CS.

The following indicators, averaged over participants, were computed to assess the validity of each hypothesis (hypothesis, PWD/CS): word count (H1, 60.7/26.5); number of introductions of new information (H2, 6.0/2.5); silence’s length, in seconds, per question type, open (H3, 1.0/0.5) or polar (H3, 1.4/0.9); and number of utterances shared by the patient and the physician, namely physician turns (H4 and H5, 9.0/7.5) and patient-to-physician (H4 and H5, 6.3/5.5).

B. Results

The following result analysis is based on the ethnographic data indicators and CA. In this small sample, PWD uttered more than the healthy older adults, producing spontaneous topical development (H1 and H2). The average conversational time was around two minutes and a half. Contrary to our expectations, silences were surprisingly shorter when answering open questions compared to polar questions (H3). Silences shorter than 300 ms were considered part of a speech turn. Interestingly, qualitative evidences suggest that the silences and pauses could be related with the stage of the disease. The average silence was shorter for CS, with a mean of 0.83, than for PWD (mean 1.32). Further research on this topic is needed, due to the small number of participants.

Both annotators describe the protocol as a multi-party interaction setting. In fact, there are at least three participants to the conversation: physician, ECA, patient. Sometimes, PWD are accompanied by relatives who spontaneously produce some turns. In this respect, sequential analysis cannot be performed by treating the situation as a general face-to-computer interaction. The physician’s tokens were 24.5 out of 63.5 produced by PWD and 140 for the ECA. Every time the experimenter spoke, people modified their behavior (H4). Lastly, patients addressed more utterances to the physician than CS, but the sample is too small to judge about the significance of this result.

C. Discussion

The linguistic analysis of the tested scenario highlighted that almost all of the questions present a wide focus design.

This could explain the absence of statistically significant differences in silence length before open questions (H3). Wide-focus questions are cognitively heavier than contrasted and narrow questions (e.g., do you like music or theater more? [24]) and are used with preference for contiguity in natural talk [25]. Moreover, wide-focus questions are more open to interpretation and may trigger longer answers, mitigation or contextualization reactions. In natural interaction speech, the sequential embedding of topic shifts has the function of indicating some linguistic context in order to notify the conversation partner(s) of a change of topic [26], [27]. It seems relevant to ask if it is clinically desirable that PWD are put in a position to speak fluently with ECAs.

Qualitative interaction analysis showed that the healthy older adults displayed some typical adaptation behavior towards the ECA (anthropomorphic robot voice imitation). PWD seemed to adapt too, but in a distinctively polite manner. Further comparison and context-aware scenarios may help analysis and positioning in answering the following question: should machines allow and promote this behavior or should they stimulate different, clinically relevant reactions? This question only accounts for one part of the interactional deal, i.e., the stimulating machine. Although the tested conversational scenario followed a principle of equilibrium-keeping, it was built as a series of questions designed to address the experts' issues without focusing on the conversational nature of the experimental task. The numerous superposed utterances show that categorically splitting conversation and interaction is not relevant. In fact words' meaning is strictly dependent on their use in discourse [28].

To take into account this coupling, one could ask if it is preferable to limit the topic shifts and concentrate on fluent topic design. Finally, if machines are needed to engage the user in a focused interaction, it could be useful to embed social features in the protocols' design. Instead of pretending that the equality of conditions in the interaction baseline exists, it could be possible to integrate conversational and interactional analysis findings, in order to compensate asymmetry and account for social variations.

VI. FUTURE WORK

Via anthropological analysis and the data session involved in the process the participants' behaviors and the influence of the context on the experiments' validity can be characterized, providing insights on conversation management and contents.

Regarding the experiments' context, the presence of the experimenter in the room and the fact that he or she addresses the subjects directly, yield to a three-party interaction, whereas the system is intended for an interaction only involving the user and the ECA. Furthermore, our analysis revealed that the subjects emitted few criticisms about the system, contrary to what we expected. Lastly, all participants but the one with the strongest cognitive impairment focused very hard on the task. In our future experiments, we should thus try to move the experimenter out of the way, for the interaction to happen only between the participant and the ECA. We think the

day hospital environment in which the tests were performed yielded what we call a "white-blouse effect", which prevented people from expressing their criticism and made them try harder to focus than they normally would. Future experiments in a more casual context, and with more participants, are thus warranted to increase ecological validity.

Regarding interaction management, our observations show that we should address the issue of "context reminding" after distractions, to help people keep track of what they were doing. At a more fundamental level, following the guidelines obtained after our anthropological analysis, future research should investigate the socio-conversational context of such experimental settings and focus on topic organization design. Lastly, with system automation in mind, asking only narrow or contrasted questions should lead to shorter answers, which is easier for automatic speech recognition (ASR), and reduces cognitive load on patients. However, since PWD tend to provide longer answers than healthy people, a keyword spotting ASR solution should be preferred.

VII. CONCLUSION

We tested the attention management capabilities of the Louise ECA prototype on two aspects: performance of the attention estimator and effectiveness of the attention recapture strategies. The results suggest that our simple attention estimation method is about 80% accurate and group comparison with a previous experiment showed that its performance is independent of the user's cognitive status. Of course, these findings should be confirmed with a larger sample of users. The attention-recapturing strategy was shown to be effective in almost all cases but was not always necessary. The fact that people looked at the screen again after a voluntarily-introduced distraction suggests that the ECA is quite engaging, though it may be partly due to "white-blouse effect". However, we could not observe self-induced distractions due to people losing track of the task.

We then performed an anthropological analysis of the interactions videos and a data session, to guide our future work, yielding the following findings: 1) the context of the experiment should be considered carefully to increase ecological validity; 2) interaction management should include recontextualisation capabilities; 3) socio-conversational and topic organization aspects of the interaction should be carefully designed; 4) narrow or contrasted questions should be privileged.

ACKNOWLEDGMENT

The authors would like to thank Région Île-de-France for funding part of this project. Opinions expressed in this paper do not necessarily reflect those of the Région Île-de-France local government.

REFERENCES

- [1] World Health Organization and Alzheimer's Disease International, *Dementia: a public health priority*, 2012, [Online] available at: http://www.who.int/mental_health/publications/dementia_report_2012/en/.

- [2] J. Lapointe, A. Verreault, B. Bouchard, J. Bouchard, A. Potvin, and A. Bouzouane, "How to maximize the effectiveness of prompts in assistive technologies according to the particular cognitive profile of people with alzheimer?" *International Journal of Smart Home*, vol. 7, no. 5, pp. 19–38, Sep. 2013.
- [3] A. D. Fisk, W. A. Rogers, N. Charness, S. J. Czaja, and J. Sharit, *Designing for Older Adults: Principles and Creative Human Factors Approaches*, 2nd ed., ser. Human Factors & Aging Series. CRC Press, 2009.
- [4] M. Pino, S. Benveniste, R. Picard, and A.-S. Rigaud, "User-driven innovation for dementia care in France: The LUSAGE living lab case study," *Interdisciplinary Studies Journal*, vol. 3, no. 4, pp. 251–268, 2014.
- [5] P. Wagnier, A. Malaise, J. Jacquemot, S. Benveniste, P. Jouvelot, M. Pino, and A.-S. Rigaud, "Towards attention monitoring of older adults with cognitive impairment during interaction with an embodied conversational agent," in *Virtual and Augmented Assistive Technology (VAAT), 2015 3rd IEEE VR International Workshop on*, March 2015, pp. 23–28.
- [6] A. Ortiz, M. del Puy Carretero, D. Oyarzun, J. J. Yanguas, C. Buiza, M. F. Gonzalez, and I. Etxeberria, "Elderly users in ambient intelligence: Does an avatar improve the interaction?" in *Universal Access in Ambient Intelligence Environments*, ser. Lecture Notes in Computer Science, C. Stephanidis and M. Pieper, Eds. Springer Berlin Heidelberg, January 2007, no. 4397, pp. 99–114.
- [7] M. M. Morandell, A. Hochgatterer, S. Fagel, and S. Wassertheurer, "Avatars in assistive homes for the elderly," in *HCI and Usability for Education and Work*, A. Holzinger, Ed. Springer Berlin Heidelberg, Jan. 2008, pp. 391–402.
- [8] M. M. Morandell, A. Hochgatterer, B. Wöckl, S. Dittenberger, and S. Fagel, "Avatars@home," in *HCI and Usability for e-Inclusion*, ser. Lecture Notes in Computer Science, A. Holzinger and K. Miesenberger, Eds. Springer Berlin Heidelberg, Jan. 2009, no. 5889, pp. 353–365.
- [9] U. Diaz-Orueta, A. Etxaniz, M. F. Gonzalez, C. Buiza, E. Urdaneta, and J. Yanguas, "Role of cognitive and functional performance in the interactions between elderly people with cognitive decline and an avatar on TV," *Universal Access in the Information Society*, vol. 13, no. 1, pp. 89–97, Mar. 2014.
- [10] E. Carrasco, G. Epelde, A. Moreno, A. Ortiz, I. Garcia, C. Buiza, E. Urdaneta, A. Etxaniz, M. F. González, and A. Arruti, "Natural interaction between avatars and persons with alzheimer's disease," in *Computers Helping People with Special Needs*, ser. Lecture Notes in Computer Science, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Springer Berlin Heidelberg, Jan. 2008, no. 5105, pp. 38–45.
- [11] Y. Sakai, Y. Nonaka, K. Yasuda, and Y. Nakano, "Listener agent for elderly people with dementia," in *7th IEEE/ACM International Conference on Human-Robot Interaction (HRI)*, Mar. 2012, pp. 199–200.
- [12] K. Yasuda, J.-I. Aoe, and M. Fuketa, "Development of an agent system for conversing with individuals with dementia," in *The 27th Annual Conference of the Japanese Society for Artificial Intelligence*, Jun 2013.
- [13] Y. Wilks, J. M. Jasiewicz, R. Catizone, L. Galescu, K. M. Martinez, and D. Rugs, "CALONIS: An artificial companion within a smart home for the care of cognitively impaired patients," in *Smart Homes and Health Telematics*, ser. Lecture Notes in Computer Science, C. Bodine, S. Helal, T. Gu, and M. Mokhtari, Eds. Springer International Publishing, Jan. 2015, pp. 255–260.
- [14] R. Yaghoubzadeh, M. Kramer, K. Pitsch, and S. Kopp, "Virtual agents as daily assistants for elderly or cognitively impaired people," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds. Springer Berlin Heidelberg, Jan. 2013, no. 8108, pp. 79–91.
- [15] W. Benkaouar and D. Vaufreydaz, "Multi-sensors engagement detection with a robot companion in a home environment," in *Workshop on Assistance and Service robotics in a human environment at IEEE International Conference on Intelligent Robots and Systems (IROS2012)*, Oct 2012, p. 45–52. [Online]. Available: <http://hal.inria.fr/hal-00735150/>
- [16] L. Li, Q. Xu, and Y. K. Tan, "Attention-based addressee selection for service and social robots to interact with multiple persons," in *Proceedings of the Workshop at SIGGRAPH Asia*, ser. WASA '12. New York, NY, USA: ACM, Nov 2012, pp. 131–136.
- [17] R. Stricker, S. Hommel, C. Martin, and H.-M. Gross, "Realtime user attention and emotion estimation on a mobile robot," in *Proceedings of 55th Ilmenau International Scientific Colloquium*, Sep 2010, p. 13–17.
- [18] P. Wagnier. List of utterances. [Online]. Available: <http://cri.enscm.fr/classement/doc/E-385.pdf>
- [19] R. F. Tate, "Correlation between a discrete and a continuous variable. point-biserial correlation," *The Annals of mathematical statistics*, pp. 603–607, 1954.
- [20] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, pp. 80–83, 1945.
- [21] D. Hymes, *Foundations in sociolinguistics: An ethnographic approach*. Psychology Press, 2001.
- [22] K. O'Reilly, *Ethnographic methods*. Routledge, 2012.
- [23] H. Sloetjes and P. Wittenburg, "Annotation by category: Elan and iso dcr," in *LREC*, 2008.
- [24] R. Vipperla, M. Wolters, and S. Renals, "Spoken dialogue interfaces for older people," *Advances in Home Care Technologies*, vol. 1, pp. 118–137, 2012.
- [25] H. Sacks, "On the preferences for agreement and contiguity in sequences in conversation," in *Talk and Social Organisation*, G. Button and J. R. E. Lee, Eds. Multilingual Matters LTD, 1987, pp. 54–69.
- [26] D. W. Maynard and D. H. Zimmerman, "Topical talk, ritual and the social organization of relationships," *Social psychology quarterly*, pp. 301–316, 1984.
- [27] K. Lambrecht, *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge university press, 1996, vol. 71.
- [28] P. Encrevé and M. De Fornel, "Le sens en pratique [construction de la référence et structure sociale de l'interaction dans le couple question/réponse]," *Actes de la recherche en sciences sociales*, vol. 46, no. 1, pp. 3–30, 1983.