A study of elderly people's emotional understanding of prompts given by Virtual Humans

Aarti Malhotra, Jesse Hoey, Alexandra König
University of Waterloo,
Waterloo, Canada
[aarti.malhotra, jhoey,
alexandra.konig]@uwaterloo.ca

Sarel van Vuuren
University of Colorado
Boulder, CO, USA
sarel@colorado.edu

ABSTRACT
This paper presents a study conducted to understand how facial expressions in audio-visual prompts given by virtual humans are understood by elderly people on an emotional level with an aim to design emotionally aligned prompts for persons with cognitive disabilities who often need assistance from a caregiver to complete daily living activities such as washing hands, making food, or getting dressed. Artificially intelligent systems have been developed that can assist in such situations. Our long term aim is to enhance such systems by delivering automated prompts that are emotionally aligned with individuals in order to have better human-system interaction in helping with the tasks. This paper presents a set of prompts of male and female virtual humans with a focus on their facial expressions. A user study with elderly persons was conducted with respect to three basic and important dimensions of emotional experience: Evaluation, Potency, and Activity (EPA). Results show that there is significant consensus on E and P dimensions, and some consensus on the A dimension.

Keywords
Virtual human; Alzheimer; dementia; affective computing

1. INTRODUCTION
People with cognitive disabilities such as Alzheimer's disease and related dementias have trouble completing activities of daily living (AD) and are usually assisted by a human care partner. The use of computerized intelligent cognitive assistants (ICAs) can help reduce a care partner's burden, and can increase feelings of independence and control in a care recipient. These ICAs take the form of automated methods for monitoring a person and inferring their activities, combined with some form of prompting to provide assistance when necessary. However, even when this technology satisfies functional requirements, people often reject it. We believe that a major reason for non-adoption is the lack of an affective (emotional) connection between technology and human.

For example, [14] demonstrates a prompting system called the COACH that can monitor a person with Alzheimer's disease while they are trying to wash their hands, detect when they have lost track of what they are doing, and play a prerecorded assistive prompt. The COACH is effective at monitoring and making decisions about when/what to prompt [14], and works well for some persons, but not as well for others. Considering the heterogeneity in socio-cultural and personal affective identities, a primary reason for lack of effectiveness may be the static, non-adaptive nature of the "canned" (pre-recorded) prompts. While significant efforts have been made to design prompts founded on the methods and styles of human caregivers [23], a simple "one size fits all" style of prompting may be limiting. While one person may find a prompt helpful and motivational, another may find it imperious and impatient. The first person is likely to follow the prompt, to feel respected, valued, and in control, and to adopt and recommend the technology. The second person, on the other hand, may feel confused by the prompt and discontinue the task. However, a different style of prompting (i.e. a more subtle prompt, perhaps with a different tone of voice, or with a different wording), may be much more effective for the second person, but not for the first. Each person comes from a different background, has a different sense of "self", and has different emotional responses to prompts. In this paper, we make use of a sociological theory of identity and the "self" called "Affect Control Theory" (ACT) [6,11]. ACT posits that humans seek emotional consistency in their interactions, and value others who understand and respect their emotional sense of self (their "identity"). Affective identity is believed to be a powerful tool for reasoning about illness in general [10]. Studies of identity in Alzheimer's disease have found that identity changes dramatically over the course of the disease [15], and that persons with AD have more abstract notions of their identity [19].

Our long-term aim is to build technology that will detect and adapt to these differences. Our study took place in two phases. In Phase I, described in [12, 13], we created a set of audio-visual prompts, using a virtual human developed with the USC Virtual Human Toolkit (VHT). We built a set of six audio-visual prompts (for different steps of the handwashing task) with five different emotional deliveries (e.g. "bossy", "motherly" or "bored"). We then did an online survey to measure human responses to these thirty different prompts. Participants were in age groups ranging from 18-54 and 65 & older. We measured responses on three important emotional dimensions of Evaluation (valence), Potency (power/dominance) and Activity, termed as EPA. We analyzed the results primarily in terms of the *consensus* of respondents within each measured dimension (EPA). The respondents agreed (reach *consensus*) on the *evaluation* dimension ("good" vs. "bad"), there was less clear agreement on the potency ("powerful" vs. "powerless") and activity ("active" vs. "asleep") dimensions. The toolkit we used was feature packed but very computationally intensive. The videos were static and not very realistic.

In Phase II of the study, described in this paper, we developed prompts using virtual human characters from the University of Colorado [22], shown in Figure 1. We used their web API to customize for our purpose and achieved better dynamic interaction control. The characters also have a few head and eye movements when in idle state. For this phase, we had a close-up image of the characters with primary focus on their facial expressions as they deliver the prompt. The goal of Phase II of the study is to determine the facial expressions and EPA space mappings of the expressive prompts of the new avatars by conducting a survey with an elderly population without Alzheimer's disease, and by analyzing and comparing results with those from Phase I.

Figure 1: Virtual humans used in the study

2. BACKGROUND

The task of hand washing activity consists of five essential steps: turning water on, using soap, rinsing hands, drying hands, and turning water off. An assistive handwashing system called COACH (Cognitive Orthosis for Assisting with aCtivites in the Home) [14] uses a video camera placed above the wash basin that captures the current activity by tracking hand and towel positions. An artificial intelligence module determines an appropriate action to take: either 'prompt the user' for one of the steps of handwashing, or 'summon the caregiver' or 'continue observing the user activity'. The speaker of the prompt is not currently visualized which limits the interaction.

Affect Control Theory (ACT) arises from work on the sociology of human interaction [6,11]. ACT proposes that social perceptions, behaviors, and emotions are guided by a psychological need to minimize the differences between culturally shared fundamental affective sentiments about social situations and the transient impressions resulting from the interactions between elements within those situations. Fundamental sentiments, **f**, are representations of social objects, such as interactants' identities and behaviors or environmental settings, as vectors in a three-dimensional affective space. The basis vectors of the affective space are called Evaluation/valence, Potency/control, and Activity/arousal (EPA). The EPA space is hypothesized to be a universal organizing principle of human socio-emotional experience, based on the discovery that these dimensions structure the semantic relations of linguistic concepts across languages and cultures [16]. They also emerged from statistical analyses of co-occurrence of a large variety of physiological, facial, gestural, and cognitive features of emotional experience [4], and relate to the universal dimensionality of personality, and social cognition [20].

EPA profiles of concepts can be measured with the semantic differential, a survey technique where respondents rate affective meanings of concepts on numerical scales. In general, within-cultural agreement about EPA meanings of social concepts is high even across subgroups of society, and cultural-average EPA ratings from as little as a few dozen survey participants are extremely stable over extended periods of time [5]. For example, the EPA for the identity of "nurse" is [1.65,0.93,0.34], meaning that nurses are seen as quite good (E), a bit powerful (P), and a bit active (A)[1]. Comparatively a "patient" is seen as [0.9,−0.69,−1.05], less powerful and less active than a "nurse". Social events cause transient impressions, $\boldsymbol{\tau}$, of identities and behaviors that deviate from their corresponding fundamental sentiments, **f**. ACT models this formation of impressions from events with a minimalist grammar of the form agent-behavior-client. Consider, for example, a nurse (agent) who ignores (behavior) a patient (client). Observers agree, and ACT predicts, that this nurse appears (τ) less nice (E), and less potent (P), than the cultural average (f) of a nurse. The Euclidean distance between $\boldsymbol{\tau}$ and **f** is called the deflection (D), and is hypothesized to correspond to an aversive state of mind that humans seek to avoid (the affect control principle). For example, the nurse who "ignores" a patient has a deflection of over 15 (very high), whereas if the nurse "comforts" the patient, the deflection is 1.5 (very low). The affect control principle also allows ACT to compute normative actions for artificial agents: those that minimize deflection. ACT has been shown to be a powerful predictor of human behavior [11].

Recently, a probabilistic and decision theoretic generalization of the ACT model was proposed called BayesAct [7]. BayesAct allows the principles of ACT to be used to guide artificially intelligent systems on an emotional level. It also allows ACT to model more complex affective sentiments, including ones that are multi-modal. A proposal to use BayesAct in the COACH system was presented in [7], and a proof of concept integration was presented in [9]. This combination created an "enhanced" COACH system that could choose an appropriate action with an EPA output that minimized deflection according to ACT principles. In this paper, we develop a set of prompts, using a customized facial expression and operationalized in the same three dimensional space as ACT that could be used by the enhanced COACH system. In this paper, we describe a survey of elderly persons' perceptions of these prompts in terms of EPA dimensions, and the measured EPA values given by this study provide a reference that can be used by the enhanced COACH system from [9] to choose the best (emotionally aligned) prompt for a desired user action step.

3. FACIAL EXPRESSIONS

The facial expressions relating to six universal emotions were studied extensively by Paul Ekman [2,3], indicating how the facial muscles of mouth, brows and eyes contribute towards emotion display on the face. Figure 2 shows mappings based on Ekman's work, categorized according to the three basic emotion dimensions (EPA) of ACT. These mappings are those used in the Interact[2] simulator to generate a set of facial expressions in a simple virtual character (outlined in [17]). Formulas for adjusting lips, brows and eyes are carefully defined based on input EPA profile values so that they appear realistic. Some expressions for basic emotions, along with their EPA values based on two datasets are shown in Figure 3.

4. PROMPTS DESIGN

For this study, we chose to consider just the 'water on' step of the handwashing task. We customized 11 prompts for each male and female virtual character, resulting in a total of 22 prompts. Based on the principles of expressions for EPA profiles given in Figure 2, we customized the virtual human avatars from University of

---

[1] EPA values range from -4.3 to 4.3 by convention

[2] Available download at http://www.indiana.edu/~socpsy/ACT/

Colorado [22], using their web API. Prompt text and expression were designed based on expected EPA in terms of positive, negative and neutral for each dimension. For example, while designing a prompt for EPA + + + i.e. positive on all dimensions, we devised prompt text as ‘Hello.I am so glad to have you here.Turn on the water.’ The corresponding expression had lips curved up for smile, brows arched upwards, lips pulled up higher, eyelids widened, and lower lip dropped.

Facial component	E, P or A impact	E+	E-	P+	P-	A+	A-
mouth (lips)	EPA	curve the lips up	curve the lips down	move lips higher (esp. upper lips)	move lips downwards	drop lower lip and narrow the lips	raise lower lip and draw the lips outward
brows	EP	increase upward arching of brows	reduce upward arching of brows	lower and close brows	raise and separate brows		
eye	A					widen separation between eyelids	reduce separation between eyelids

Figure 2: Facial Expression guidelines

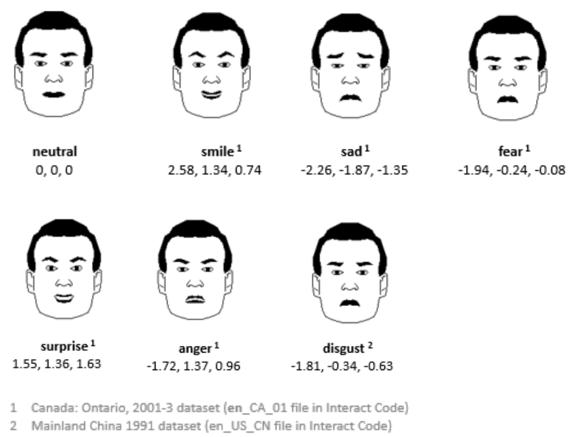

Figure 3: Sample Interact simulator results

5. USER STUDY

An online survey for the newly designed prompts was conducted in which participants were asked to play and watch the 22 prompts and rate them based on Evaluation, Potency, and Activity dimensions (on a scale of -4.3 to +4.3). For each dimension a bi-polar scale called semantic differential was used, labelled as infinitely (-4.3), extremely (-3), quite (-2), slightly (-1), neutral (0), slightly (1), quite (2), extremely (3), infinitely (4.3). We provided meaning of all three dimensions to participants before they started the survey. The survey application was built in-house.

The survey took around 20 minutes to complete. The questions for male and female avatar were provided in an alternate fashion. The participant list for the survey was provided from database of local healthy seniors’ names, contact details and some medical history and who are interested in taking part in research. The participants who confirmed interest were sent the survey link. They also provided their consent online and checked if their audio/video worked fine. They were asked to play each prompt before providing their ratings. They could skip any question or exit the survey at any point in time. At the end of the survey, information on gender, age group, and free-form comments was requested. An example screen shot of a question is shown in Figure 4. Twenty-two respondents (8 males/14 females) completed the survey[3].

[3] Cleared by University of Waterloo Research Ethics Committee

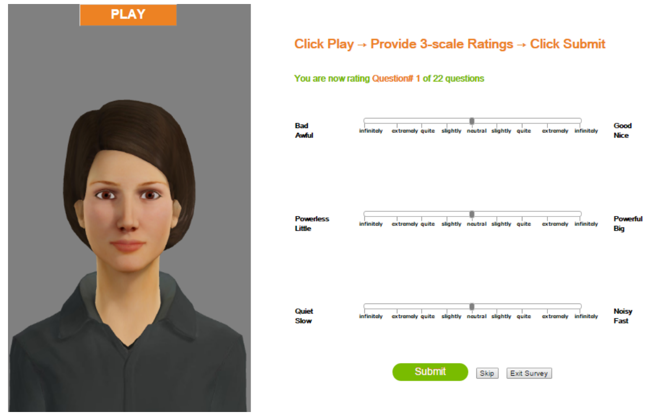

Figure 4: Example Screenshot of the survey showing new virtual human and the three ratings scales

Sub-groups of dataset	Consensus E	Consensus P	Consensus A
All Data	6.3	2.1	1.8
Male Character	3.9	2.9	1.4
Female Character	5.9	5.6	1.9
Male Respondents	4.3	5.5	1.9
Female Respondents	4.7	1.3	1.3
Age Group 65_69 Respondents	11.3	2.7	2.7
Age Group 70 & up Respondents	4.7	1.9	2.4

Figure 5: Consensus Summary.

5.1 Consensus Analysis

To determine consensus amongst participants, we followed the culture-as-consensus model [1,22,5]. This allows measurement of whether respondents homogeneously represent a common culture that explains the similarity in their answers to questions about the cultural norms [18]. The method computes the Eigenvalues of the covariance matrix of all responses for each of E, P, A separately, where the covariance is computed in the space spanned by the participants, with data given by the questions. Thus, we are computing the principal components that indicate the extent to which respondents agree in their ratings across all items. If there is one large first factor (i.e. the first Eigenvalue is notably larger than the second), this reflects cultural commonality in the respondent’s ratings and provides evidence of one dominant factor governing respondent’s judgement [5]. In this paper, if the first-to-second Eigenvalue ratio is equal or greater than 2.0, it is considered as significant, as suggested by [18,5]. We first consider consensus across all the data, followed by more detailed analyses across different sub-groups of the dataset. Figure 5 summarizes result with consensus highlighted in green. Dark green is for values greater than 2.0, while light green is for values between 1.8 and 2.0 which is quite close to consensus. Detailed results in [12].

All Data: There is consensus on E and P dimensions and close to consensus on A.

Gender of virtual character: For male and female character, respondents seem to have consensus on E and P dimension, and for female character, they were close to consensus on A

Gender of respondents: Male respondents seem to have consensus on E and P dimensions and close to consensus on A, whereas female respondents seem to agree more on E dimension only.

Age group of respondents: Respondents in age-group 65-69 yrs. had consensus in E dimension, and respondents in age-group 70 yrs. and up agreed more on P and A dimensions.

For all 22 prompts, we also compared the sign of average rating (positive, negative or neutral) on each E, P and A dimension within groups of respondents with those of expected sign of E, P and A based on design as per Figure 2. The analysis is shown in Table 1. A match in sign is highlighted. Each group has different color for better visualization. Highest match achieved was for E dimension, then A and P dimensions. Respondents in age-group 65-69 yrs. had the highest overall match for E and P signs, while A dimension sign matched highest for female respondents' data.

	All Data			Male Respondent			Female Respondent			Age Group 65_69			Age Group 70_up		
Expected EPA sign	avg E	avg P	avg A	avg E	avg P	avg A	avg E	avg P	avg A	avg E	avg P	avg A	avg E	avg P	avg A
n n n	0.105	0.714	0.086	-0.238	0.963	0.063	0.300	0.571	0.100	0.000	0.629	0.286	0.153	0.753	-0.007
n n n	-1.218	-0.623	-0.295	-0.688	-0.538	-0.125	-1.521	-0.671	-0.393	-0.743	-0.414	0.371	-1.440	-0.720	-0.607
+ + +	0.732	0.645	-0.036	0.350	0.713	-0.075	0.950	0.607	-0.014	0.886	0.571	-0.114	0.660	0.680	0.000
+ + +	-0.418	-0.223	-0.386	-0.025	0.188	-0.288	-0.643	-0.457	-0.443	0.757	0.571	-0.100	-0.967	-0.593	-0.520
+ + +	1.855	1.214	0.245	1.413	1.175	-0.263	2.107	1.236	0.536	2.000	1.429	0.529	1.787	1.113	0.113
+ + +	0.100	0.086	-0.241	-0.125	0.100	-0.713	0.229	0.079	0.029	0.714	0.271	0.143	-0.187	0.000	-0.420
- + +	-2.527	-0.268	0.077	-2.525	-1.063	-0.863	-2.529	0.186	0.614	-2.086	-0.086	0.086	-2.733	-0.353	0.073
- + +	-1.814	-0.773	-0.127	-0.925	-0.913	-0.913	-2.321	-0.693	0.321	-1.457	0.000	0.314	-1.980	-1.133	-0.333
- + -	-2.382	0.114	0.082	-2.438	-1.025	-0.863	-2.350	0.764	0.621	-2.514	0.271	0.114	-2.320	0.040	0.067
- + -	-1.777	-0.700	0.005	-1.588	-0.863	-0.863	-1.886	-0.607	0.500	-1.343	0.086	0.000	-1.980	-1.067	0.007
+ - +	2.059	2.295	0.750	2.488	2.450	0.613	1.814	2.207	0.829	2.243	2.057	0.571	1.973	2.407	0.833
+ - +	1.164	1.073	0.118	1.263	1.113	-0.138	1.107	1.050	0.264	1.671	1.286	0.400	0.927	0.973	-0.013
+ - +	2.050	2.086	0.745	2.075	2.088	0.600	2.036	2.086	0.829	2.086	1.914	0.514	2.033	2.167	0.853
+ - +	1.168	1.077	0.323	0.788	0.988	0.400	1.386	1.129	0.279	1.343	1.429	0.271	1.087	0.913	0.347
+ - -	1.564	1.841	0.845	1.900	1.850	0.688	1.371	1.836	0.936	1.914	1.843	0.357	1.400	1.840	1.073
+ - -	0.632	0.505	0.173	0.963	0.813	0.225	0.443	0.329	0.143	1.386	1.571	0.486	0.280	0.007	0.027
+ - -	1.795	1.764	0.377	1.950	1.938	0.588	1.707	1.664	0.257	2.129	1.786	0.129	1.640	1.753	0.493
+ - -	0.714	0.568	-0.036	0.925	0.988	0.025	0.593	0.329	-0.071	1.529	1.329	-0.014	0.333	0.213	-0.047
+ + +	-0.027	1.218	0.364	0.538	1.200	-0.200	-0.350	1.229	0.686	1.200	1.800	-0.086	-0.600	0.947	0.573
+ + +	-0.032	0.718	0.018	0.388	0.838	-0.400	-0.271	0.650	0.257	0.800	1.014	-0.100	-0.420	0.580	0.073
+ + -	0.409	1.014	-0.005	0.413	0.450	-0.825	0.407	1.336	0.464	1.243	1.329	-0.400	0.020	0.867	0.180
+ + -	0.318	0.091	0.173	1.063	0.850	0.225	-0.107	-0.343	0.143	0.971	0.657	0.357	0.013	-0.173	0.087
total matches	17	8	10	18	8	6	16	8	11	21	10	10	16	6	8

Table 1: Expected EPA sign and actuals in different datasets.

6. CONCLUSION

This paper introduces mapping of facial expressions of a virtual human to the 3 dimensions of emotional space (Evaluation, Potency and Activity), with an aim to improve assistive systems that help persons with cognitive disabilities. In this paper, we present results from a study of the impressions of non-demented elderly persons as a first step towards building emotionally aware cognitive assistants. Results show there is significant consensus on E and P dimensions, and some consensus on the A dimension. In phase I study, only E dimension had maximum consensus.

Future improvements to the study include animations of a real person, or speech variations in the character. A study is planned with the target population (elders with cognitive disabilities). We expect differences will be found in the responses of participants, mainly due to the wide range, quickly changing, and more easily suggestible affective identities that are held by persons with dementia [10,15,19]. We also expect more usability challenges with this population. We are proceeding first with the current study (with healthy elderly people), and with a second ongoing study of identity in Alzheimer's disease [8]. Future studies may also evaluate the prompts in different cultures or languages.

7. ACKNOWLEDGMENTS

This work was supported by AGE-WELL NCE Inc., a member of the Networks of Centres of Excellence program and Alzheimer's Association grant ETAC-14-321494. University of Colorado contributions in part by Nattawut Ngampatipatpong and Robert Bowen. Animation prompting platform supported in part by grant 90RE5019 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR), a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS).

8. REFERENCES

[1] Borgatti, S. P., & Halgin, D. S. Consensus analysis. A companion to cognitive anthropology, 171-190, 2010.

[2] Ekman, P. & Friesen, W. V., Unmasking the face, Prentice-Hall, Oxford, 1975.

[3] Ekman, P.: Emotion in the human face (2nd ed.). New York: Cambridge University Press, 1982.

[4] Fontaine, J.R.J., Scherer, K.R., Roesch, E.B., Ellsworth, P.C.: The world of emotions is not two-dimensional. Psychological Science 18, 1050 – 1057 (2007)

[5] Heise, D. R., Surveying Cultures: Discovering Shared Conceptions & Sentiments, Wiley 2010

[6] Heise, D.R., Expressive Order: Confirming Sentiments in Social Actions. Springer, 2007.

[7] Hoey, J., Schröder, T., and Alhothali, A.: "Affect control processes: Intelligent affective interaction using a partially observable Markov decision process." Artificial Intelligence, 230, 2016.

[8] König, A., Francis, LE, Hoey, J. Defining affective Identities in elderly Nursing Home residents for the design of an emotionally intelligent cognitive assistant. Pervasive Computing Technologies for Healthcare. 2016, Mexico

[9] Lin, L., Czarnuch, S., Malhotra, A., Yu, L., Schröder, T., Hoey, J. Affectively Aligned Cognitive Assistance using Bayesian Affect Control Theory. In Proc. IWAAL, 2014.

[10] Lively, K.J., Smith, C.L.: Identity and illness. In: Handbook of the Sociology of Health, Illness, and Healing, pp. 505–525. Springer New York (2011)

[11] MacKinnnon, N.J., Robinson, D.T.: 25 years of research in affect control theory. Adv. in Group Processing 31 (2014)

[12] Malhotra, A. Design and Study of Emotions in Virtual Humans for Assistive Technologies. Master's Thesis, University of Waterloo, 2016

[13] Malhotra, A,,Yu C., Schröder,T., Hoey, J.An exploratory study into the use of an emotionally aware cognitive assistant. AAAI Workshop on artificial intelligence applied to assistive technologies and smart environments, 2015.

[14] Mihailidis, A., Boger, J., Candido, M., and Hoey, J. The COACH system to assist older adults with dementia through handwashing: an efficacy study. BMC Geriatrics,8(28), 2008

[15] Orona, C.J.: Temporality and identity loss due to Alzheimer's disease. Social Science & Medicine 30(11), 1247 – 1256 (1990)

[16] Osgood, Charles E., W.H. May and M.S. Miron. "Cross-Cultural Universals of Affective Meaning. University of Illinois Press, 1975.

[17] Payr, I.S. and Trappl, R., Enculturating Agents With Expressive Role Behavior.In Sabine Payr and Robert Trappl (Eds.), Agent Culture: Human-Agent Interaction in a Multicultural World. Lawrence Erlbaum Associates. 2004

[18] Romney, A. Kimball, Susan C. Weller, and William H. Batchelder. "Culture as consensus: A theory of culture and informant accuracy." American Anthropologist 88.2 (1986).

[19] Rose Addis, D., Tippett, L.: Memory of myself: Autobiographical memory and identity in Alzheimer's disease. Memory 12(1), 56–74 (2004)

[20] Scholl, W.: The socio-emotional basis of human interaction and communication: How we construct our social world. Social Science Information 52, 3 – 33 (2013)

[21] Schröder, T., Rogers, K. B., Ike, S., Mell, J. N., & Scholl, W.: Affective meanings of stereotyped social groups in cross-cultural comparison. Group Processes & Intergroup Relations, 16(6), 717-733, 2013.

[22] Van Vuuren, S. & Cherney, L. R. (2014). A Virtual Therapist for Speech and Language Therapy. In IVA 2014. LNAI 8637, pp. 438-448. Springer.

[23] Wilson, R., et al.: Quantitative analysis of formal caregivers' use of communication strategies while assisting individuals with moderate and severe Alzheimer's disease during oral care. J. of Communication Disorders 46(3), 249–263 (2013)