Audio-visual synthesized attitudes presented by the German speaking robot SMiRAE

Angelika Hönemann[1,2], *Casey Bennett*[3], *Petra Wagner*[2], *Selma Sabanovic*[4]

[1] Beuth University of Applied Science, Computer Science and Media, Berlin, Germany
[2] Bielefeld University, Linguistics and Literary Studies, Bielefeld, Germany
[3] DePaul University, Computer Science, Chicago, IL, USA
[4] Indiana University, Informatics and Cognitive Science, Bloomington, IN, USA

ahoenemann@beuth-hochschule.de, cabennet@indiana.edu,
petra.wagner@uni-bielefeld.de, selmas@indiana.edu

Abstract

This paper presents the acoustic and visual modeling of nine attitudinal expressions that were realized by the German speaking robot SMiRAE which is a speech-enabled version of the non-speaking robotic face MiRAE previously developed at Indiana University. The parameter-oriented acoustic model is based on the German Mary TTS which is part of the speech processing system InproTK. Visual realization of expressions is based on five defined basic emotions of the Facial Action Coding System (FACS) developed by Ekman. Both models were additionally modified with respect to results of an audio-visual analysis and evaluation of human portrayals of attitudes recorded in our previous work.

The plausibility of synthesized attitudinal expressions is shown by an association study in which 18 participants described 54 attitudes in a free association. Basis for a 5-cluster classification was the first four dimensions of a correspondence analysis which accounted 78% of variance in participant perception. Significant correlations were seen between 66 normalized participant descriptions and the robot's displayed attitudes. For instance, the attitudes *admiration* and *politeness* were associated with the terms *freundlich* and *gluecklich,* the interrogative attitudes *surprise* and *doubt* with the terms *fragend, verwundert* and *skeptisch,* the expression *uncertatinty* was perceived with *traurig* and *besorgt*.

Index Terms: speech production and perception, attitudinal speech, acoustic and visual synthesis, human-robot interaction

1. Introduction

Assistive Technologies (AT) are taking more and more of an important role in daily life, for instance, robots which support household work, assistant system for car driving, and providing for interaction for elderly people [7, 9].

Especially in times of need for care in people's homes, in hospitals and in retirement homes, robots are increasingly used to relieve skilled staff during daily care. Social robots such as Paro developed in Japan [22] also provide therapeutic support, for instance, in the care of people with dementia [6] or as a learning tool for children with developmental disabilities [8]. For older people who want to maintain their autonomy through their own households, intelligent systems or robots can provide assistance and care services, such as monitoring health status, reminding for on-time appointments, or assisting with everyday tasks. In the research area of smart homes, ongoing work focuses on creating intelligent systems that enable independent living even in old age [6]. For instance, researchers at the Cluster of Excellence Cognitive Interaction Technology (CITEC) of the University of Bielefeld are developing an intelligent apartment in which a robot and a virtual agent are integrated to support the residents in every-day life, but also communicate interactively with the residents [17].

The social component in an intelligent system is becoming increasingly important in order to convey a sense of security and trust. A robot capable of demonstrating attitude-specific behavior brings with it an ability to create an authentic social experience. Human-like behavior in a robot can create a bond between the user and the robot, which increases acceptance and facilitates ease-of-use. For instance, a robot may express itself with an uncertain or doubtful expression if it does not understand a user's request or is surprised by a user's behavior, or else expresses his admiration (praise) towards the user, which can be strengthen self-confidence especially for older people. Also situations are conceivable in which the robot has to react authoritatively, e.g. refusing medication. Studies by CITEC have shown that an emotional robot is more accepted by humans as an assistant and contact person [17].

The work of this paper based on a series of investigations of human attitudes [18]. Human attitude-specific behavior expresses itself through the voice and facial expressions, thus attitudes expand a social conversation to a paralinguistic component, which means important information in addition to the linguistic context is conveyed to the listener.

The audio-visual synthesizing of attitudes is a significant challenge because attitudes are very complex in contrast to emotions, as they are built on a mental valuation concept of situations or objects [1]. They are also strongly dependent on the context of the social situation, on the relation of the conversation partners to the speakers, and on the modality of presentation. Different attitudes also sometimes can be very similar in the acoustic and visual cues which makes the recognition difficult. Thus recognition needs additional information at a further level, e.g. the context to separate the attitudinal expression [18].

2. The minimalistic robot

MiRAE is a minimalist robotic face, capable of performing an array of facial expressions and neck motions, and equipped with flexible lips that can simulate speaking movements. In

10.21437/AVSP.2019-16

previous studies, MiRAE has been shown to produce facial expressions that can be identified by human participants at rates similar to more complex robots (~85%), such as Kismet and BERT [2]. MiRAE has also been utilized in several studies of the effects of culture and context on perceptions of robotic facial expressions across Asia and North America [3]. The robotic face can be assembled for under $150, using easily accessible parts, with a construction manual and programming code available online (downloadable from here http://www.caseybennett.com/research.html). The overarching goal of MiRAE is to provide a reproducible, well-documented research platform for studies of human-robot interaction and human perceptions of emotion.

For an association study, the robot SMiRAE (Speech-based Minimalistic Robot for Affective Expressions) has been developed based on the construction of the robot MiRAE. Like the original robot, the face movements are controlled by ten simple servo motors to create 10 degrees-of-freedom (DOF), two for the eye motions (tilt left/right), four for the eyebrow motions (up/down, tilt left/right) and four for the mouth movements (moving the mouth corner in/out, up/down). Combined actuation of these simple DOFs could simulate complex motion, such as the parting of lips to open the mouth (see Figure 6 in [2] for examples). For the left/right and up/down rotation of the head two stronger motors (Hitech HS-485HB) were installed. The eyes and the eyebrows have been printed out by a 3D printer; however the lips were realized with simple pipe cleaners, since they have proven to be flexible, yet stable for realization of lip movements in speaking. Unlike the non-speaking original robot, SMiRAE includes a speaker (disguised as a nose) for voice output, thus an acoustic output through the mouth could be simulated.

The already implemented software of the original robot was used but modified. This included the C++/Python RobotFace library, which provides functions for controlling the motions of the robot using the Arduino Uno R3 micro-controller. Additionally, an implementation of a Python interface for the communication between the control software and the speech synthesis software was necessary. Furthermore, in addition to the existing visual emotional presentation a synchronous lip control was integrated. Figure 1 shows SMiRAE (neutral expression) from the left, front and back view.

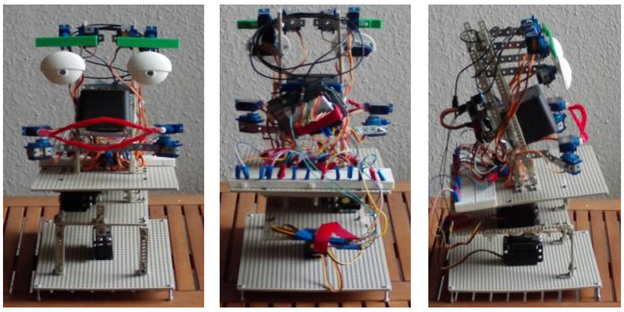

Figure 1: SMiRAE from the front, back and left view, neutral expression

3. Parameter-oriented speech synthesis

In previous work sixteen attitudes performed by sixteen speakers were recorded, evaluated and analyzed. The acoustic speech parameters based on the audio-visual analysis of the human recorded portrayals [18]. The visual modeling took into account five emotions (*anger, fear, surprise, sadness, joy*) described by the Facial Action Coding System (FACS) developed by Ekman [10].

The audio-visual speech synthesis concentrates on nine attitudes: the neutral interrogative attitudes QUES and declarative statement DECL, the interrogative attitudes SURP and DOUB, as well as the declarative attitudes AUTH, UNCE, ADMI, IRRI and POLI. (cf. Table 1). Since the acoustic speech synthesis is already published in [15], we will reduce to a summary and represent the modifications in the current paper.

Table 1: Attitudes description and the short terms using in this paper

Short term	Attitude
ADMI	Admiration
AUTH	Authority
DECL	Neutral Statement
DOUB	Doubt
IRRI	Irritation
QUES	Neutral Question
POLI	Politeness
SURP	Surprise
UNCE	Uncertainty

3.1 Acoustic modeling

For the acoustic speech synthesis, the incremental dialog system InproTK [4, 5] was used which includes the MaryTTS synthesis system [21]. InproTK offers a dynamic speech adaptation while synthesizing which can be used to switch between the attitudinal expressions. However, an extension of the InproTK to create an adaptable speech component was necessary. This entailed an adaptation of prosodic speech parameter, such as the phone duration, F0, intensity and the voice quality parameter jitter and shimmer. These parameters can affect the perception of different attitudes [11, 12, 13].

The initialization of the adaptation process includes the detection of different phone types such as consonants, vowels and long vowels and the determining of the position of them in the current utterance. A distinction is made between the first, middle and end position of a phone in randomly chosen words of the utterance. This allows a specific handling of stressed parts in the utterance. The synthesis process starts with the setting of each speech parameter on the appropriate average human specific value. The parameters were set with respect to the phone type and position in the word. Additionally a random factor between zero and the standard deviation of each speech parameter was added to the respective speech feature. The random-based approach of synthesis is used here to simulate different speaker-specific speech styles.

The dependencies between the parameters have also be take into account, therefore the correlation coefficient as a factor for the significant correlated features where added to the corresponding parameter. The phone duration, for instance, correlates significantly with the range of the F0 and the range of the intensity. Therefore the range of F0 is multiplied by the correlation coefficient of 0.47 and the range of intensity is multiplied with the correlation coefficient of 0.49, both results were then added to the previously calculated duration. The effect of this approach is that in the case of a positive correlation the target parameter increases with an increase of the correlated parameter, and vice versa in the case of a negative correlation.

After all criteria for the synthesis have been taken into account and the value of each parameter was calculated, the parameters were finally reinforced by using a factor of 2.5 in order to give the individual synthesized speech features more expression.

The last step of the synthesis process is the determination of the F0 contour and the intensity contour. For each frame a factor is computed using the previously determined jitter (for the F0) and shimmer (for the intensity) derived from the human data as a multiplier to compute three sine waves, which are then added to each F0 respectively intensity value [16].

3.2 Visual modeling

The visual presentation of attitudes is based on the five basic emotions *anger, fear, surprise, sadness and joy*. For the determination of the relevant facial motions, the attitudes have to be mapped on these emotions. In previous work, the dominance, activation, and valence of each attitude were set based on the results of an evaluation study [18]. Afterwards the attitudes were classified into a 3D emotional model [19, 29] which allows the assignment to appropriate emotions (cf. Figure 2). An exception is the attitude *surprise* because a mapping is not necessary because *surprise* is already defined by Ekman as an emotion [10].

Table 2: Property dominance (dom), activation (act) and valence (val) of the attitudes (+- neutral, +++ high positive, ++ average positive, + low positive, --- high negative -- average negative, - low negative) and the assigned emotions

Attitudes	Property	Emotion
DECL QUES	-val, +-act, +-dom	neutral
SURP	+-val, ++act, --dom	(surprise)
DOUB	--val, +++act, ---dom	anger
IRRI	---val, +++act, +++dom	anger
AUTH	--val, +act, +++dom	anger
UNCE	--val, +-act. ---dom	sad
ADMI	+++val, +++act, +-dom	happy
POLI	+-val, +act, +dom	(contentment) happy

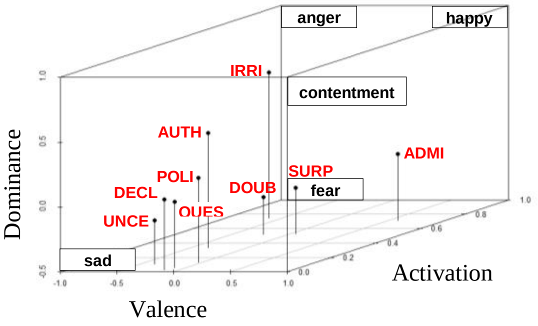

Figure 2: Assignment of the attitudes to the emotions of the 3D emotional model

As mentioned before, the visual representation is based on the FACS but additionally on the visual analysis of human attitudinal expressions [14], therefore the following visual cues could be derived for the individual attitude-specific expressions synthesized by the robot SMiRAE (cf. Figure 2):

- DECL/QUES: Neutral face, looking straight ahead
- DOUB: Raising right outer eyebrow, opening the mouth slightly, raising the corners of the mouth, looking straight ahead, head motion forward
- SURP: Raising eyebrows, opening mouth widely, look straight ahead
- POLI: Lowering the corners of the mouth to a smile, looking straight ahead
- AUTH: raising eyebrows, opening and spreading of the corners of the mouth, slightly lifting, outward movement slight, head movement slightly upwards, thus looking slightly upwards
- ADMI: Raising of eyebrows, lowering the corners of the mouth to a smile, head motions upwards to the right, looking up to the right
- IRRI: Lowering of eyebrows, raising the outer eyebrows, spreading the mouth with lifting and strong outer motion of the corners of the mouth
- UNCE: Raising inner eyebrows which also causes a lowering of the outer eyebrows, opening of the mouth slightly, with raising the corners of the mouth, head movement slightly down to the right, thus looking slightly down to the right

4. Association study

The aim of the association study was the evaluation of the acoustic and visual synthesis presented by the robot SMiRAE, that means the plausibility of the expressions were judged by description entered by participants. The study was carried out with 18 participants (10m, 8w, 20-53 years, mean: 33 years), – 14 students from the Beuth University of applied science and the University of Bielefeld, and 4 participants without an academic background.

The experiment design of the current study was already used in several studies (in different languages) in which attitudinal expressions portrayals by humans were described by participants [11,12,18]. Unlike our previous ‘human study’ for the current study only audio-visual stimuli were included. In addition, the study was conducted via a website, but the participants carried it out in the presence of the experimenter at the university.

4.1 Data corpus and procedure

The nine synthesized attitudes QUES, DECL, SURP, DOUB, AUTH, UNCE, ADMI, IRRI and POLI were presented. For the speech synthesis, the German male MaryTTS voice *dfki-pavoque-neutral-hsmm* was modified, whereby each attitude was synthesized in six speaker variations to represent different speech styles (cf. section 3.1). The data corpus included only one short utterance *Marie tanzte durch den Tag* (engl. Marie was dancing through the day) which was synthesized and spoken by the robot SMiRAE. In total 54 stimuli were finally described by each participant.

After participants called up the website, instructions for the experiment were given and an introductory video appeared in which SMiRAE introduced itself and explained the task for the participants. Relevant movements of the robot such as the mouth, eyes, eyebrows and head orientation were also shown to familiarize the participants with the motions.

The study initiated when the subjects clicked on the 'Start'-button after entering personal data such as initials, age, gender and nationality.

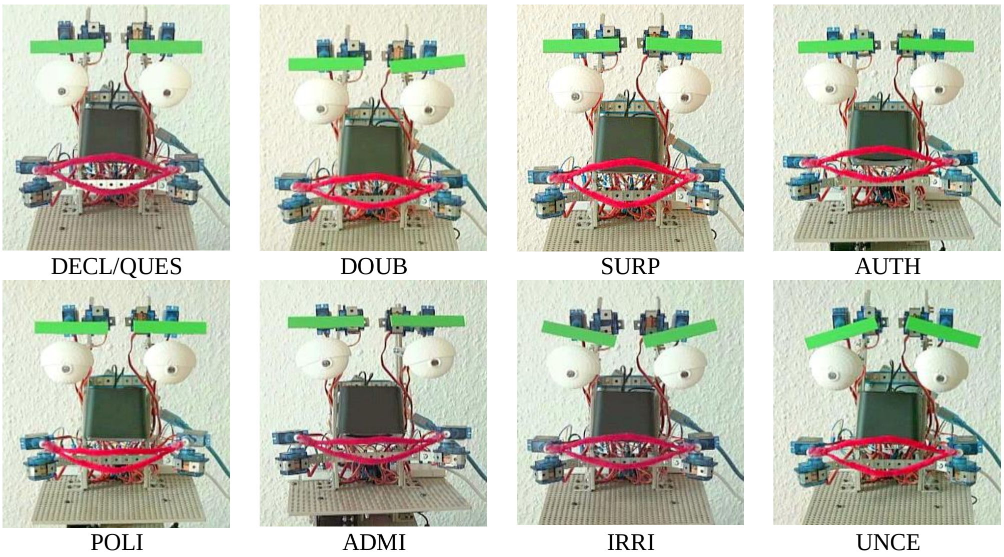

Figure 2: Visual presentation of the nine attitudinal expressions

After each attitudinal expression was presented the participants were asked to enter the perceived property in the form of an adjective or descriptive noun in an input field next to the stimulus. They then proceed to the next stimulus after clicking on a 'Next'-button.

The participants were not limited in time and in case of uncertainty there was the possibility to watch the video repeatedly. The stimuli were presented successively in random order across speaker style and attitude. After completion of the last expression, a request is made to save the result file, which is then sent via email to the experimenter.

4.2 Normalization of terms

In total, 970 German terms (cf. section 6 for translation into English) were collected which need to first be normalized. The normalization process included the correction of typos, the collapsing of similar words, for instance, *wuetend* and *Wut*, as well as of semantic equal words, for instance, *gleichgültig*, *uninterressiert* and *teilnahmslos* onto the more frequent one *teilnahmslos*, or *bedrückt*, *bekümmert* and *besorgt* onto the more frequent one *besorgt*.

After normalization 66 terms were remained for the analysis. The top twenty most frequent terms, such as *neutral*, *gelangweilt*, *veraegert* and *freundlich* their occurrence counts are listed in Table 3.

Table 3: The top twenty normalized terms and occurrence count

Normalized terms	N	Normalized terms	N
neutral	158	erfreut	26
gelangweilt	79	erklaerend	25
veraergert	60	gluecklich	23
freundlich	57	skeptisch	19
traurig	48	aufgeregt	17
fragend	42	froehlich	16
verwundert	39	verwirrt	15
besorgt	31	muede	13
genervt	30	ueberheblich	11
interessenlos	28	eindringlich	10

4.3 Correspondence analysis and clustering

The 66 normalized terms and the 9 attitudes were ordered into a contingency table (9 rows * 66 columns), thus the table summarized the number of times each description occurred for each attitude. These were then analyzed by a correspondence analysis (CA). From the eight dimensions yielded by the CA, a variance of about 78% could be explained by the first four dimensions.

A hierarchical clustering algorithm (HCPC) was applied based on the dimensions yield by the CA in order to group the rows of the matrix with respect to their distribution in the space calculated from the normalized terms.

The classification can be optimally grouped into 5 clusters. The clusters obtained by the HCPC reflect the associated perception of the participants to the attitudinal expressions. The five clusters and contained attitudes with the associated terms are listed in Table 4. Additionally the internal and global percentage share of each terms as well as the result of the v-test of the HCPC is shown. Only terms with a greater internal frequency than the global frequency are reported.

The clusters #1 and #2 only include one attitude these are IRRI (#1) and UNCE (#2). Cluster #1 contained the attitude IRRI which is characterized by a high negative valence as well as a high dominance and activation. The term *veraergert* yielded the greatest frequency (internal share of 53%), but other similar terms such as *boese* (10% internal share) and *genervt* (9% internal share) were perceived by the participants. All these terms align closely with the meaning of that attitude. Cluster #2 mainly consisted of the terms *traurig (*with the greatest internal share of 34%), and *besorgt* (26% internal share) which reflects a negative emotional mood. The attitude UNCE indicates a negative valence as well. Additionally, it is indicated by high negative dominance and neutral activation. The perception of the property *sehnsuechtig* suggests that participants sometimes misinterpret this attitude.

Table 4: Five clusters with contained attitudes, associated terms, the internal and global percentage share of each term, result of the v-test

Cluster Attitude	Normalized Terms	Intern %	Global %	p-value	v-test
#1	veraergert	52.778	6.186	.000	15.990
IRRI	boese	10.185	1.134	.000	6.601
	genervt	9.259	3.093	.002	3.136
	eindringlich	3.704	1.031	.036	2.098
#2	traurig	34.259	4.948	.000	11.065
UNCE	besorgt	25.926	3.196	.000	10.344
	resigniert	2.778	0.309	.003	3.001
	sehnsuechtig	1.852	0.206	.025	2.248
#3	fragend	13.889	4.330	.000	6.826
SURP	gelangweilt	20.370	8.144	.000	6.674
DOUB	skeptisch	7.407	1.959	.000	5.584
	verwundert	9.722	4.021	.000	4.244
	zweifelnd	4.167	1.340	.001	3.384
	muede	4.167	1.340	.001	3.384
#4	neutral	35.714	16.289	.000	11.140
DECL	aufgeregt	4.348	1.753	.000	3.957
QUES	erklaerend	5.280	2.577	.001	3.405
AUTH	hektisch	2.484	0.928	.002	3.125
	gestresst	2.484	0.928	.002	3.125
	aufgeweckt	2.174	0.825	.005	2.817
	aussagekraeftig	1.242	0.412	.024	2.257
	euphorisch	1.553	0.619	.034	2.115
#5	gluecklich	10.648	2.371	.000	8.059
ADMI	freundlich	17.593	5.876	.000	7.313
POLI	erfreut	9.722	2.680	.000	6.253
	vertraeumt	4.630	1.134	.000	4.587
	schwaermend	2.315	0.515	.001	3.275
	zufrieden	1.852	0.412	.005	2.819
	froehlich	4.167	1.649	.006	2.746
	abwesend	1.852	0.515	.020	2.329

For the interrogative attitudes SURP and DOUB which pooled together in cluster #3, the participants chose the terms *fragend* (14% internal share) and *gelangweilt* (20% internal share) with the greatest frequency. Both attitudes are characterized by an approximately neutral valence. The term *gelangweil*t as well as the less frequently used term *muede* reflect a contrasting meaning because the attitudes SURP and DOUB indicate an interest in something but at the same time a question or measure of uncertainty. The property *verwundert, skeptisch* and zweifelnd were also perceived by the participants for those attitudes, which reflect the contrasting meaning of these attitudes much better.

Cluster #4 includes the neutral attitudes QUES and DECL as well as the attitude AUTH. The attitude AUTH shows a high dominance and an average negative valence unlike the attitudes DECL and QUES which exhibit an approximately neutral valence and dominance. The commonality of these three attitudes is their neutral activation. These attitudes were mainly associated with the term *neutral* with an internal share of 36%, but also terms such as *erklaerend, hektisch, aufgeweck* and *aussagekraeftig* were used to describe user perceptions of these robot attitudes. The term *euphorisch* shows a highly positive emotional mood, thus appears to be a bit of an outlier to the other terms.

The last cluster #5 grouped together the attitudes ADMI and POLI which show a positive valence and a neutral dominance of the speaker to the listener. These attitudes were mainly associated by the participants with the terms *freundlich, gluecklich, erfreut* and *vertraeumt* which indicate a positive emotional mood as well. The term *gluecklich* obtained the greatest frequency (internal share of about 11%).

5. Discussion and conclusions

The paper describes the setup of the robot SMiRAE which presents nine synthesized attitudinal expressions via voice and facial movement. The acoustic synthesis was summarized and the facial motions determined for each attitude are presented. Finally an association study is described with the analysis and results shown.

The results of the association study suggest that the participants could recognize the meaning of the attitudes most of the time, e.g. the valence of an attitude. Cluster #1 contains the attitude IRRI, where the motion of the outer eyebrows downwards and the strong voice of the robot reflect a negative and dominant expression. Terms such as the *veraergert, genervt* and *eindringlich* were perceived by participants. Less frequent interpretations, for instance *gelangweilt, traurig* or *beschwingt* mismatch the meaning of the attitude and are difficult to explain. The same can be seen for the attitude UNCE (#2) which is described with the attributes *traurig* and *besorgt* most frequently. The motion of the inner eyebrows downwards and the slowing and quieting speaking of SMiRAE appears to be interpreted correctly by human subjects.

The interrogative attitudes SURP and DOUB were clustered together in #3. The questioning and surprising characteristic is recognized for both attitudes frequently but also the terms *gelangweilt, muede* or *interessenlos* were entered. Obviously leads the slow speaking of the robot to this assumption. Descriptions such as *skeptisch, traurig* and *ueberheblich* show the negative effect of the attitude DOUB on the participants.

The neutral characteristic of the attitudes DECL and QUES were confirmed by the participants' descriptions. In the most cases the term *neutral* was entered for these attitudes. The term *erklaerend* was often indicated for the attitude DECL and the term *fragend* for the attitude QUES which reflects, respectively, the declarative and questioning speech type. The term *hektisch* mostly entered for the attitude DECL can be attributed to the fast speaking. The attitude AUTH was perceived with terms such as *aufgeweckt* and *genervt* more frequently which match the meaning of this attitude. The attitude AUTH points out a negative valence and an average dominance which is reflected by the strength mouth motion and the raising of the eyebrows and head motion upwards. But the movements were obviously misinterpreted by the participants as *erschrocken, freundlich* or *euphorisch*.

The positive attributes such as *freundlich, gluecklich* and *erfreut* were perceived mainly for the attitudes POLI and ADMI (#5), which were visually accompanied with a smile of the robot. The dreaming and sentimental characteristics of the attitude ADMI were recognizable, signaled through the raising of the head and moving of the eyes of SMiRAE upwardly with a lateral movement. Misinterpretations in this case, for instance *ueberheblich* or *genervt,* might lead back to these motions. Participants also identify a neutral mood mainly for the attitude POLI which could be related to the lack of motion of the face and the less expressive voice.

The above results suggests SMiRAE reflects a sufficient presentation of attitudinal expressions via a robotic face, but some misinterpretations occurred which cannot be explained satisfactorily. Our assumption is that some participants may have been more oriented to the visual cues of SMiRAE and others to the auditory speech cues. For instance, the appearance of the eyes on the robot conveyed a sleepy expression (as several participants remarked during the experiment). On that account the term *gelangweilt* is more frequently attributed independent of the attitude. This suggests that the design of the robotic face, and critically the design of the 3D-printed facial components, might impact human perceptions of attitude in ways yet to be understood.

On the other hand, the attitudes were presented without any context, but context gives additional information which helps to interpret an expression if the acoustic and visual cues are very similar [18].

The clustering carried out on the current study shows similar results in comparison to the study carried out using attitude portrayals with human face. Like the current study, the interrogative attitudes SURP and DOUB were pooled together in one cluster, however, the interrogative attitude QUES and the attitude UNCE appeared in different clusters here, unlike the human portrayal studies. The declarative attitudes AUTH and DECL were clustered together which like the previous studies. However, unlike the current study, that cluster also included the attitude POLI. The attitude IRRI (together with the negative attitudes contempt and obviousness) built one cluster in the 'human study', indicating that IRRI could also be clearly separated in human presentations [18].

6. Translation of terms: German – English

German	English	German	English
abwesend	absent	gluecklich	happy
aufgeregt	excited	gleichgültig	indifferent
aufgeweckt	alert	hektisch	hectic
aussagekraeftig	meaningful	ueberheblich	arrogance
besorgt	concerned	muede	tired
boese	angry	neutral	neutral
eindringlich	urgently	sauer	angry
erfreut	pleased	schwaermend	swarming
erklaerend	explanatory	sehnsuechtig	iongingly
erschrocken	scared	skeptisch	skeptical
euphorisch	euphoric	teilnahmslos	indifferent
fragend	questioning	traurig	sad
freundlich	friendly	veraergert	upset
froehlich	cheerfully	vertraeumt	dreamy
gelangweilt	bored	verwundert	bewildered
genervt	annoyed	verwirrt	confused

7. Acknowledgements

Thanks to the students of the Beuth University of Applied Science and the University of Bielefeld as well as to the private persons for the interest and participation at the study. Thanks also to Sebastian Pose for helping to record the videos.

References

[1] Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), A handbook of social psychology (pp. 789–994). Worcester, MA: Clark University Press.

[2] Bennett, C.C., Sabanovic, S. Deriving Minimal Features for Human-Like Facial Expressions in Robotic Faces, International Journal of Social Robotics, Aug. 2014

[3] Bennett, C. C., Sabanovic, S. (2015). The effects of culture and context on perceptions of robotic facial expressions. Interaction Studies, 16(2), 272-302

[4] Baumann. T.. & Schlangen. D. The InproTK 2012 Release. In Proceedings of NAACL-HLT. 2012

[5] Baumann T. Schlangen D. INPRO iSS: A Component for Just-In-Time Incremental Speech Synthesis. In: Proceedings of the ACL 2012 System Demonstrations. ACL: 103–108.. 2012

[6] Chu, M., Khosla, R., Khaksar, S. and Nguyen, K. (2017). Service innovation through social robot engagement to improve dementia care quality. Assistive Technology 29, 8-18.

[7] Cocco,J., Note, Smart Home Technology for the Elderly and the Need for Regulation, 6 J. ENVTL. & PUB. HEALTH L. 85, 92–95, 2011

[8] Conti, D., Di Nuovo, S., Buono, S. and Di Nuovo, S., Robots in education and care of children with developmental disabilities: a study on acceptance by experienced and future professionals. International Journal of Social Robotics 8, 1-12. ,2016

[9] Düring, Michael (2017) Fahrzeugübergreifendes kooperatives Fahrerassistenz- und Sicherheitssystem für automatische Fahrzeuge. Dissertation. sonstiger Bericht.

[10] Ekman P. and Friesen W. V. Manual for the Facial Action Coding System. Palo Alto: Consulting Psychologists Press, 1977

[11] Guerry, M., Shochi, T., Rilliard, A., and Erickson, D. ”Perception of prosodic social attitudes affects in French: A freelabeling study Proceedings of ICPhS 2015, Glasgow, Scotland.

[12] Guerry, A. Rilliard, D. Erickson, T. Shochi, “Perception of prosodic social affects in Japanese: a free-labeling study”, In Proc. Speech Prosody 2016, Boston, 811-815, 2016.

[13] Gobl C., Chasaide A.N., The role of voice quality in communicating emotion, mood and attitude. Speech Communication 40, p. 189–212, 2003

[14] Hönemann, A., Wagner, P. Facial activity of attitudinal speech in German, 14th International Conference on Auditory-Visual Speech Processing, Stockholm, Sweden, 2017

[15] Hönemann, A., Wagner, P. Synthesizing Attitudes in German. SST 2016, Parramatta City, New South Wales, Australia, 2016.

[16] Klatt. D. & Klatt. L. Anaysis. synthesis. and perception of voice quality variations among female and male talkers. J. Acoust. Soc. America 87(2), 820-857, 1990

[17] Meyer zu Borgsen S, Bernotat J, Wachsmuth S, Hand in Hand with Robots: Differences between Experienced and Naive Users in Human-Robot Handover Scenarios, In: Social Robotics. 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017

[18] Mixdorff, H., Hönemann, A., Rilliard, A., Lee, T., Ma, M.K.H., Audio-visual expressions of attitude: How many different attitudes can perceivers decode? Speech Communication 95 (2017) 114–126, 2017

[19] http://tschroeder.eu/weblog/?page_id=2, accessed on 5 February 2016.

[20] Schauenburg, G., Ambrasat, J., Schröder, T., von Scheve, C., & Conrad, M., Emotional connotations of words related to authority and community. Behavior Research Methods, 47: 720-735, 2015

[21] Schröder. M. & Trouvain. J. The German Text-to-Speech Synthesis System MARY: A Tool for Research. Development and Teaching. International Journal of Speech Technology. 6.pp. 365-377. 2003

[22] Shibata T, Wada K. Robot therapy: A new approach for mental healthcare of the elderly. A mini-review. Gerontology 2011