Emotion Expression in HRI – When and Why

Kerstin Fischer
Department of Design & Communication
University of Southern Denmark
Sonderborg, Denmark
kerstin@sdu.dk

Malte Jung
Department of Information Science
Cornell University
Ithaca, NY, USA
mfj28@cornell.edu

Lars Christian Jensen
Department of Design & Communication
University of Southern Denmark Kolding, Denmark
larscj@sdu.dk

Maria Vanessa aus der Wieschen
Department of Design & Communication
University of Southern Denmark
Kolding, Denmark
mwies@sdu.dk

***Abstract*—In this paper, we draw attention to the social functions of emotional display in interaction. A review of HRI papers on emotion suggests that this perspective is rarely taken in the field, but that it is useful to account for the context- and culture-dependency of emotional expression. We show in two case studies that emotional display is expected to occur at very specific places in interaction and rather independently from general emotional states, and that different cultures have different conventions regarding emotional expression. Based on conversation analytic work and the results from our case studies, we present design recommendations which allow the implementation of specific emotional signals for different human-robot interaction situations.**

***Index Terms*—emotion expression, human-robot interaction, situation-specificity, conversation analysis**

I. Introduction

Emotional expression is ubiquitous in human life and has important functions, for instance, for survival (e.g. [1]). Emotional systems provide the individual with information about arousal and evaluation and lead to fast responses to danger; for example, fear keeps us from taking high risks, and anger evokes the necessary resources for an attack. However, beyond survival, emotional expression plays a considerable social role in the regulation of interpersonal relationships; for instance, in medical encounters, laughter is often used to talk about delicate topics (such as digestion, erection or death, see [2]). By using emotional displays, participants indicate that they are aware of societal constraints on the appropriateness of such topics and indicate that their current talk is to be understood as an exception based on the special (medical) circumstances (cf. [3]). Similarly, Hepburn and Potter [4] find displays of empathy to be used by therapists to signal recognition of the partner's problems in order to move on to another topic. Such empathy signals are thus part of the professional toolkit of a therapist (cf. [3]).

The research reported on was supported by the Danish Council for Independent Research under Grant No. 4180-00359A and by the National Science Foundation under Grant No. IIS-1421929.

Some studies even indicate that the social function of emotional expression may be more important than those functions connected to arousal and evaluation; for instance, it has been found that bowlers do not smile when they make a hit but only when they turn around to view their team members [5]. Furthermore, work on conversation from an ethnomethodological perspective [6] shows that emotional expression is conventionally expected at certain sequential places in interaction. That is, emotional expressions are often produced not in response to an internal state but in response to social demands. As analyses of interactions between humans show, emotional expression is often more defined by the sequential structure of the respective activity than by the person's emotional state: "We won't find that strong sorrow and strong joy are just distributed over the course of the conversation but instead, there are *real places for them to occur*" (Sacks [7], p. 572).

For instance, in storytelling, an activity that robots have been suggested to carry out in the future [8], emotional displays are produced in response to culturally determined expectations. Storytelling is a relatively well-structured activity, consisting of an introduction, climax and evaluation (e.g. [9]), in which emotional display is required at specific places. Selting [10] shows that speakers produce increasing numbers of emotional indicators when the telling is moving towards the climax; these emotional indicators comprise lexical and syntactic means, such as interjections, intensifiers or pivot-constructions; phonetic-prosodic signals, such as dynamic pitch contours, sound lengthenings and tempo changes; gestural (such as clenched fists, slashing and pointing gestures) and facial expression; as well as laughter, among many others [10]. By means of such signals, the speaker indicates to the listener a) what the point of the story is (by increasing the number of emotional displays), b) how the main point should be evaluated, and c) what type of response is expected from the listener. Consequently, by indicating the story climax by means of increased emotional display, speakers allow listeners to cooperate in the telling. Sharing of both arousal and evaluation is the expected response from the listener, and

Bavelas et al. [11], for instance, show that if these displays are not forthcoming in the expected strength and evaluation, the resulting story is of lower quality. Thus, the appropriate timing, value and strength of the listener's emotional displays crucially contribute to the storytelling activity.

However, the listener's task is not just to align to the storyteller's emotional displays; Selting [10] shows that whether the expected response is produced depends on the listener's social role. For example, she finds a father not to align his emotional displays with his daughter's and thus to create a situation in which he demonstrates control and calmness and thus provides an interpretation of his daughter's evaluation of the situation as overreacting. From a conversation analytical perspective, emotional displays thus constitute an interactional resource used by participants to coordinate their assessments of the issues talked about.

That emotional expression during listening to stories is indeed conventional and not due to the listeners' underlying emotional states is evident from a study by Voutilainen et al. [9] who find listeners to use similar types and amounts of facial gestures as storytellers; nevertheless, using measures of skin conductance (electro-dermal activity), facial muscle activation (facial EMG) and heartbeat (ECG), they find that arousal differs between speaker and listener: While storytellers' arousal is highest when they are approaching the climax of the story, listeners' arousal is highest when the telling is ambiguous and they therefore do not know whether a positive or a negative assessment is in order. This indicates not only that displays of emotion are generally independent of emotional arousal, but also that listeners are under considerable pressure to produce the expected emotional assessment.

In the next section, we illustrate that this perspective of emotion expression to be situationally demanded rather than internally determined has not yet found its way into mainstream human-robot interaction research by reviewing previous work on emotion in robots published at the HRI conference since 2006. We then present two studies designed to show a) the context dependence of emotional expression in HRI; that is, emotional expression is evaluated differently depending on the position in interaction in which it occurs; and b) that emotion expression is conventionally determined because it is culture-dependent; that is, if emotion expression was based on the expression of emotional states, then no culture differences could be expected. The point we are arguing for in this paper is thus that robots' expressions of emotions need to be based on the kinds of activities the robot is involved in and attuned to the conventional expectations towards emotion expression in these activities. The paper concludes with design recommendations based on the results of the two studies and based on the relevant literature in the ethnomethodological study of conversation.

II. Emotion Displays in Human-Robot Interaction

To gain an overview of current work on emotion expression in HRI, we reviewed 53 full HRI-proceedings papers from 2006 to 2017. The papers were searched for in the ACM digital library using the "Advanced Search" feature for papers in the HRI Proceedings Series "Where <any field> <matches any> of the following words and phrases: emotion, affect". This search retrieved 183 papers (including LBRs etc.), from which we manually identified the 53 full papers. Out of the 53 full papers, 27 (51%) focused in some form on a robot's emotional expressions and were therefore included in our review. The excluded 26 papers did not focus on emotional expression and investigated emotional aspects of HRI more broadly, for example by examining related constructs such as persuasion (e.g. [39]) and therefore using "emotion" or "affect" only as a tangential construct, or by focusing on emotional reactions that people have as they interact with robots (e.g. [40]), which was not the focus of our review. Table I provides an overview of the remaining 27 papers.

The review of the papers reveals that many studies address how robots, given their different morphologies and functionalities, can express emotions in an unambiguous way (e.g. [30], [31], [34]–[36]), focusing on how a robot's internal state can be reliably inferred from emotion expressions. Obviously, producing a set of distinct expressive behaviors that can be reliably identified to fall within a specific emotion expression category constitutes a considerable challenge given different robot embodiments [41]. Indeed, nine of the 27 selected papers on emotion expression (24%) tested the ability of users to correctly label which emotion a robot was intended to express through its behavior (papers marked "yes" in the "Labeling study" column of Table I). Most of these labeling studies followed either dimensional theories of emotion (e.g. [42]) and tested for the correct classification of expressions along dimensions of valence and arousal (e.g. [17], [19], [23]) or they followed basic emotion approaches (e.g. [43], [44]) and tested for the correct classification of expressions along a specific set of emotions (e.g. [30], [31], [34]–[36]).

Many other papers take a more interactional perspective and determine the display of emotion based on the human partner's emotional expression, thus making use of interactional alignment or accommodation (papers marked "2" in the "Theory" column of Table I, such as [27] or [25]). At the same time, almost all of the emotion expression papers (89%) are based on the idea that the display or expression itself determines how it will be interpreted (papers marked "yes" in the "Fixed Interpretation" column of Table I). For example, the robot's facial expressions in an early study by Gockley, Simmons, and Forlizzi [13] were assumed to determine user perceptions of the robot as neutral, happy, or sad. More recently, Song and Yamada [36] presented a robot that was designed to be perceived as relaxed, happy, sad, or angry based on the activation of a specific set of colors, vibration patterns, and sounds. Of the remaining three papers, two are about emotionally expressive robots but do not specify how the expression-interpretation relationship is viewed [14], [16], and only one paper argues for an approach that disconnects the interpretation of emotion from the specific embodiment of an expression [38]. It is worth mentioning that Read and Belpaeme [24] found the interpretation of a

TABLE I: Emotional Expression Papers published in HRI proceedings

Year	Author(s)	Ref.	Expressions	Labeling study	Fixed Interpretation	Theory
2006	Scheutz et al.	[12]	neutral, half-frightened, frightened	no	yes	1,3
2006	Gockley, Forlizzi, and Simmons	[13]	neutral, happy, sad	no	yes	2
2007	Bethel and Murphy	[14]	none specified	no	unspecified	0
2009	Mutlu et al.	[15]	gaze cue, no gaze-cue	no	yes	1
2009	Kanda, et al.	[16]	none specified	no	unspecified	0
2010	Saerbeck and Bartneck	[17]	dimensions: valence, arousal, dominance	yes	yes	2
2011	Leyzberg et al.	[18]	appropriate response, inappropriate response	no	yes	2
2011	Yohann and MacLean	[19]	arousal (low, medium, high), valence (negative, neutral, positive)	yes	yes	1
2011	Chen et al.	[20]	instrumental touch, affective touch	no	yes	2
2012	Leite et al.	[21]	empathy expressions (encouraging, scaffolding, offering help, playing a bad move)	no	yes	2
2012	Ribeiro and Paiva	[22]	anger, disgust, fear, joy, sadness, surprise	no	yes	1
2013	Sharma et al.	[23]	dimensions: valence, arousal	yes	yes	1
2014	Read and Belpaeme	[24]	positive, negative sounds	yes	yes	2
2014	Hoffman et al.	[25]	positive, negative	no	yes	2
2014	Tielman et al.	[26]	dimensions: valence, arousal, extroversion	no	yes	1
2015	Leite et al.	[27]	not specified, just examples (happy bouncing)	no	yes	0
2015	Hoffman et al.	[28]	calm, curious, scared	no	yes	1
2015	Jung et al.	[29]	neutral, repair	no	yes	2
2015	Ammi et al.	[30]	neutral, sadness, joy	yes	yes	2
2016	Boccanfuso et al.	[31]	happy, angry, sad	yes	yes	1
2016	Koschate et al.	[32]	happy, sad, neutral	no	yes	2
2016	Birnbaum et al.	[33]	responsive, non-responsive	no	yes	2
2016	Cauchard et al.	[34]	dopey, sad, sleepy, grumpy, shy, happy, brave, afraid	yes	yes	1
2017	Moosaei et al.	[35]	pain, anger, disgust	yes	yes	1
2017	Song and Yamada	[36]	relaxed, happy, angry, sad	yes	yes	2
2017	Feldmaier, Stimpfl, and Diepold	[37]	dimensions: valence, arousal	no	yes	1
2017	Jung	[38]	none specified	no	no	3

The table provides an overview of full HRI papers focusing on emotion expression that were published since the start of the conference. The table lists publication year, authors, paper reference, and categories of emotional expressions that were examined in the paper. The column "labeling study" highlights papers that presented studies intended to test whether specific behaviors by a robot can be reliably identified as falling into a specific emotional expression category. The column "Fixed Interpretation" highlights papers that are grounded in the assumption that the expression is unambiguous and essentially context-independent. The label "unspecified" was used for papers that did not clearly describe a set of specific expressions and what they were expected to communicate. The column "Theory" categorizes papers based on the assumptions a paper makes about emotion expressions: "0" is used if no theory is presented, "1" is used for papers that subscribe to the idea that a robot's internal state drives expressions, "2" is used for papers that are based in assumptions that relate specific robot behaviors to specific user reactions, and "3" is used for papers that see the situation as an important driver of emotion expression.

robot's non-linguistic utterances (such as sequences of beeps) as positive or negative to depend on the situational context the sound was played in. However, the paper still followed the commonly held idea that the emotional interpretation of an utterance as positive or negative is determined by its embodiment characteristics, so that context was seen as a biasing factor that overrides the "correct" interpretation of the expressive behavior.

Finally, when taking a closer look at the theoretical perspectives underlying the reviewed papers, it becomes apparent that the interactional context is rarely seen as an important determinant for appropriate emotion expression by a robot. For example, eleven (29%) of the reviewed papers subscribe to some degree to the idea that a robot's expressions are there to reflect its internal state and that expressions are produced in response to changes in a robot's state (see papers marked "1" in the "Theory" column of Table I). Only two papers (7%) acknowledge the interactional context as an important driver of a robot's emotional expressions (see papers marked "3" in the "Theory" column of Table I). A study by Scheutz and colleagues [12] compares two approaches: In one condition, a robot was programmed to display an emotional reaction in response to changes in its internal state and in another condition the robot was programmed to display an emotion reaction in response to displays of emotion made by the human the robot was interacting with.

In sum, our review shows that the interactional context has received very little attention as an important determinant of appropriate emotion displays in robots. Especially the idea that emotional expressions are often displayed because they are (conventionally) *expected* rather than *felt* is not represented in current HRI literature. As our case studies will show, however, also human-robot interactions can benefit from an understanding of emotional display as socially determined.

III. Case Study: Greetings for Friends and Strangers

Our first study addresses the effect of the position of emotion expression in the context in which it is uttered.

A. Introduction

We set this study up to explore how people receive a robot that uses emotional expression in similar ways humans do when they greet each other. Greetings can reveal the state of interpersonal relationships in the sense that people design their greeting differently depending on how well they know

each other [45]. Pillet-Shore [46] has analyzed 332 greetings between humans at home and at the workplace using ethnomethodological conversation analysis [6]; she noticed that two different kinds of greetings can be reliably distinguished: One emotionally neutral type of greeting, which is used for strangers, for instance, and one emotional type used for people the speaker shows affection for, such as friends and family. The two types differ prosodically, especially in terms of duration: While the average duration of the neutral greetings (for strangers) is about 0.2 secs, the affective greetings (for friends and family) are 2-4 times longer. The prosodic lengthening of the greeting seems to communicate a positive attitude towards the listener and can therefore be seen as a sign of emotional expression.

Now, the study explores how participants ratings of a robot change if it shows positive affect to them during the greeting. In order to create a baseline, we set up an experiment with two Keepon robots producing the same greeting, just one with the prosodic elongation that expresses positive affect. In particular, we wanted to see whether a robot that uses a greeting that expresses positive emotion is evaluated as more friendly, engaging, polite and desirable.

B. Method

To measure the effect of two lexically identical, but phonetically different utterances with different degrees of emotional expression, we set up an experiment with two modified Keepon robots. The two robots, named John and Jim, spoke with different but similar sounding voices.

1) Experimental Setup: The greeting "Hello, I'm John/Jim. How are you?" was created with two different English male voices using the Festival text-to-speech synthesis engine [47], which were then prosodically manipulated so that the final syllables were elongated using Praat (see Fig. 1).[1] We thus obtained two original versions and two manipulations, which only differ in length; otherwise, the lexical and phonological structure was kept constant.

The experiment uses a between subjects design with speaking style (long or short final syllables) as grouping factor.

2) Procedure: Participants first filled out an informed consent form and the first part of the questionnaire. They were then shown around in our lab, and as part of the 'tour' they were introduced to our two Keepons, named John and Jim. They were greeted first by John and then by Jim with the utterance 'Hi, I'm John/Jim, how are you?', where in one condition, John's greeting was long and Jim's greeting was short, whereas John's greeting was short and Jim's long in the other condition. Before the verbal action, each Keepon turned toward the participant and then bowed when greeting. After this short interaction, participants were led into a different room where they filled out the second part of the questionnaire.

[1]In particular, we almost doubled the durations of the three phrase-final syllables (from about 170 ms to about 320 ms). In order to maintain the stimuli's naturalness, we slightly increased the nuclear accents' pitch ranges and adjusted their F0-peak alignments to the new vowel duration (see, e.g. [48]).

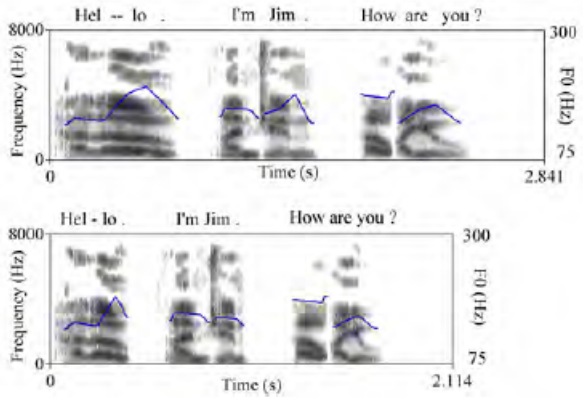


Fig. 1: Phonetic Manipulation

3) Materials: Two MyKeepons (see fig. 2) were modified[2] so that the robots could be tele-operated. In addition, a miniature speaker was connected and hidden beneath the table on which the Keepons were standing to give the impression that the robots were speaking. The robots were remotely controlled from a nearby workstation by a 'wizard' [49]. The 'wizard' received one hour of training and controlled the Keepons through all 36 interactions.

Fig. 2: Experimental Setup

4) Participants: 36 students and staff members from the University of Southern Denmark, but also people from the general public participated in the experiment; their mean age is 30.7 (range 19-59). 24 of the participants are men, 12 are women. The distribution of participant gender was balanced evenly between the two experimental conditions. 45% are native speakers of Danish, while the others have diverse language backgrounds. More than half (57%) of the participants stated that they know robots only from the media, 35% say they have worked or played around with a robot a few times before, while only three participants stated that they work with robots regularly. Participants were compensated for their time and participation with chocolate.

5) Questionnaire: The first part of the questionnaire asked about participants' age, gender, experience with robots and native language. The second part of the questionnaire asked participants to rate how engaging, friendly, polite, and certain they perceived each robot to be on a 7-point Likert scale. It also asked them about the extent to which they would like to talk to the robot again, and to own the robot.

[2]http://probo.vub.ac.be/HackingKeepon/

C. Results

Questionnaire responses for John and Jim are compared using independent samples t-tests with the phonetic manipulation as the grouping factor. While there are significant effects of affective greeting for the second Keepon Jim (see Fig. 3 and Table II), we found no significant differences between speaking styles for John; i.e. whether the first Keepon speaks with or without an emotionally toned greeting makes no difference. That is, only when the second Keepon speaks with long final syllables, participants rate it significantly more engaging, friendly and polite. Participants interacting with Jim in the long condition were also significantly more willing to speak to the robot again, and their wish to own the robot was significantly greater.

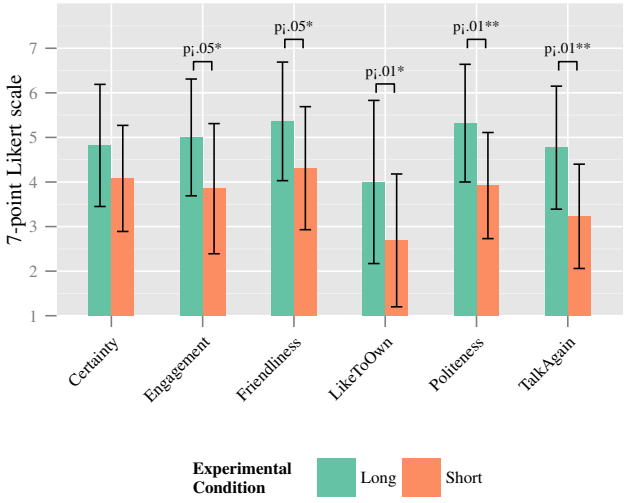


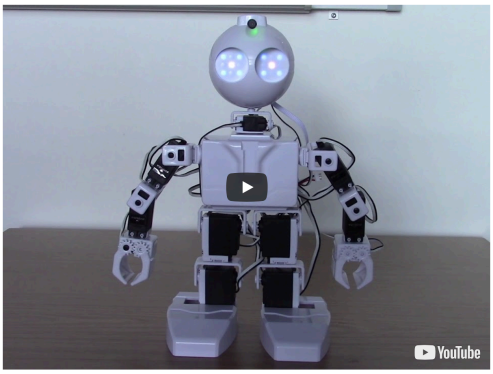
Fig. 3: Questionnaire ratings for the second Keepon Jim

TABLE II: Statistics for the second Keepon Jim

	Long	**Short**	**T-test**
Engaging	5.00, ± 1.31	3.85, ± 1.46	t(23.088)=2.34, p=0.028
Friendly	5.36, ± 1.38	4.31, ± 1.38	t(24.582)=2.22, p=0.036
Polite	5.32, ± 1.19	3.92, ± 1.32	t(27.584)=3.22, p=0.003
Speak again	4.77, ± 1.38	3.23, ± 1.17	t(28.781)=3.53, p=0.001
Own robot	4.00, ± 1.83	2.69, ± 1.49	t(29.434)=2.30, p=0.029

D. Discussion

The results show that a robot that uses an affective greeting may be perceived as more friendly, engaging, polite and desirable. The fact that there are several significant effects for the display of positive stance by the second robot but not by the first indicates that the longer greeting is not a simple display of emotion that has an effect independently of where in the interaction it is used. To some extent, participants' failure to respond to the first Keepon's long greeting can be attributed to a novelty effect; that is, it is possible that the experience of encountering a Keepon that turns, bows and greets a person in English is to many people so novel that micro-level differences



* Please rate the robot according to the following criteria:

	low						high
friendliness	○	○	○	○	○	○	○
politeness	○	○	○	○	○	○	○
warmth	○	○	○	○	○	○	○
degree of engagement	○	○	○	○	○	○	○
formality	○	○	○	○	○	○	○
professionalism	○	○	○	○	○	○	○

Fig. 4: Video stimulus and questions as presented to participants who chose English as their native language (Study II)

in the interactions may simply go unnoticed. Similarly, people may lack a point of comparison when hearing the first robot speak. Another possible explanation for the fact that only Jim is evaluated positively is that participants understand John and Jim as two social actors belonging to the same situational context, where the long greeting signals to the participant that the second robot has witnessed the previous interaction, too. This explanation is consistent with participants' verbal behavior where a number of participants responded by saying "still fine", "still good" or similar, indicating that they assume that the second robot must have heard their response to the first robot. In this case, the signal of positive emotion may be taken to be in line with the acquaintance achieved due to the shared interactional history (cf. [50]).

IV. Case Study: Delivering Bad News

The conventional, socially defined nature of emotional display becomes most obvious when different cultures have different expectations about when such displays should occur. In the following, we present a study in which a robot informs the participant about the lack of availability of a product the participant has chosen, thus delivering bad news (see [51]–[53]). Such a task could be very common when robots are used to assist people in shopping centers, as receptionists, or as assistants at train stations etc. Our case study was designed to examine whether people hold culturally determined expectations towards a robot's display of emotion.

A. Study Overview

The study was carried out as an online survey with two conditions in a between-subject design. In the two conditions, the robot's utterances differ according to the amount of emotional display during the delivery of the bad news. The study was conducted in three languages, targeting different cultural backgrounds, since we suspected a cultural difference between Denmark and Germany. For instance, we had noticed that Danish shop assistants or IT-counselors would simply say "we don't have that any more" or "we can't help you", where in corresponding situations in Germany, emotional expression of compassion would have been expectable. This difference is also reflected in Fredsted's [54] finding that German speech exhibits more instances of verbal politeness than Danish speech.

In order to elicit more information on how different situations of the delivery of bad news are handled in Denmark and Germany, we set up an informal e-mail survey with a simple discourse completion task (DCT), a method common in contrastive pragmatics research, in which a context is given and participants are asked to complete the dialog (see [55]). Based on the formulations elicited, we designed two variants of robot utterances delivering bad news, one with and one without emotional displays of compassion. The resulting videos were couched in a questionnaire in which the participants were first asked to decide between different types of chocolate for a (hypothetical) year's supply. In particular, after participants had made their choice by picking one out of four (yummy) chocolate types presented with images, they saw a video of the EZ-bot in which the robot provided the bad news that the chosen type of chocolate is not available – either with or without emotional display (see Figure 4). Participants were then asked to rate the robot regarding friendliness, politeness, engagement, warmth, formality and professionalism, on a 7-point Likert scale.

Our expectations were that emotional expression is rather uncommon in Danish bad news and that using such displays would lead to higher ratings of friendliness, politeness and engagement, but also to higher ratings of formality since the emotional expression constitutes extra interactional effort that is not normally used. For English and German, based on prior conversation analytic work (e.g. [53]), we expected that failure to use signs of compassion will have negative consequences for the evaluation of the robot as polite, friendly and engaging. For speakers of other languages, we had no predictions.

B. Participants

Participants were recruited through social networks and e-mail distribution lists. 1195 people started and 372 completed the survey. Of these, 40 stated Danish to be their first language, 196 were native speakers of German and 55 were native speakers of English. 81 speakers of other languages saw a version of the English questionnaires. Independent of their native languages, which determined the video participants saw, participants were able to select the language of the survey. Participants were assigned randomly to one of the two conditions. Participants in the conditions are balanced in terms of age and gender (52.1% report to be female).

C. Stimuli Creation

For the design of the stimuli, we first carried out a pre-study involving a discourse completion task, in which we asked a small group of people (who are not the same as those who participated in the survey) via e-mail distribution lists to provide us with formulation alternatives for situations in which the participant has to take the role of a person delivering bad news, such as:

- A customer comes into a shop in which you help out occasionally. She asks for almond chocolate. While the shop normally carries almond chocolate, you happen to know that the supplier has failed to deliver it three times in a row. You answer (in Danish/German/English):

Using the formulation alternatives we received (in all languages, there were examples with and without emotional expression), we then constructed the following messages:

- Condition 1: Due to a technical problem, the chocolate you have chosen is not available. You would have to choose another one.
- Condition 2: Oh, I'm really sorry, due to a technical problem, the chocolate you have chosen is unfortunately not available. I'm afraid you would have to choose another one, sorry!

While the relevant emotional indicators were identified empirically in the discourse completion task, they also function as emotional indicators in other contexts; for instance, interjections like *oh* have been argued to generally have an emotional *I feel*-component [56], and *I'm afraid* and *I'm sorry* are literally denoting emotional states.

The Danish and German stimuli were created by close pragmatic translation.[3] These sentences were synthesized using iSpeech, a free text-to-speech system that offers very similar (female) voices in the three languages,[4] and paired with a video of the EZ-bot (see Figure 4).

D. Results

For four of the six dimensions according to which the robot was evaluated, we find significant overall differences between the conditions. In particular, when the robot used emotional expression, participants' rated it higher regarding friendliness ($F(1, 371) = 21.986$, $p < .001$) (see Fig. 5); warmth ($F(1, 371) = 7.579$, $p = 0.0062$); politeness ($F(1, 371) = 12.34$, $p = 0.00049$); and engagement ($F(1, 371) = 6.458$, $p = 0.0112$), but whether or not the robot uses emotional expression had no overall effect on ratings of formality and professionalism.

Furthermore, we find significant differences between the different language backgrounds, i.e. intercultural differences,

[3]The videos of the bad news in the different languages and in the two conditions can be found here:
German: youtu.be/M-rEyTKs9Y0 and youtu.be/452aC9B1iBE;
Danish: youtu.be/_QzC5IaTbTo and youtu.be/0hTD2_JfjS4;
English: www.youtube.com/watch?v=8nEbvM_OOTg and www.youtube.com/watch?v=86yzB0PIRvY

[4]www.ispeech.org/

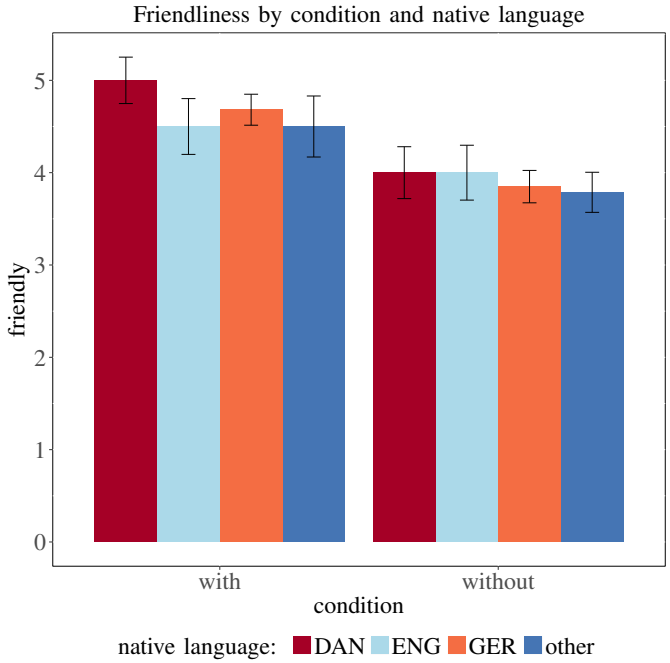


Fig. 5: Perceived friendliness by condition (with and without emotional expression) and native language

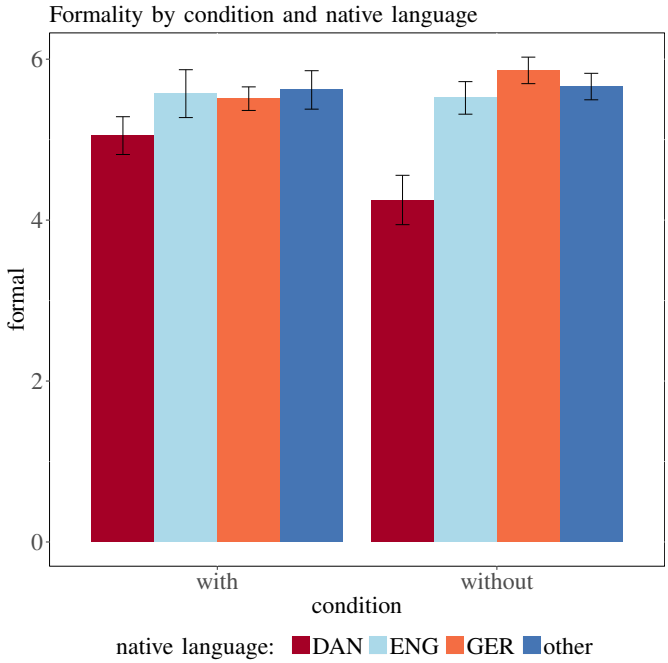


Fig. 6: Perceived formality by condition (with and without emotional expression) and native language

for four out of the six dimensions. The dimensions unaffected by the different cultural expectations are friendliness and politeness. In contrast, the four language groups evaluated the robot differently concerning warmth ($F(3,369) = 11.86$; $p > 0.001$), engagement ($F(3,369) = 3.074$; $p = 0.02784$), formality ($F(3,369) = 3.836$; $p = 0.009988$) and professionalism ($F(3,369) = 2.823$; $p = 0.0387$).

Posthoc comparison shows that our hypotheses regarding Danish were confirmed regarding formality, where Danes were found to differ significantly from Germans ($p = 0.009$) and speakers of other languages ($p = 0.020$) as well as from native speakers of English ($p = 0.04969$) (see Figure 6). Thus, for the Danish speakers, emotional expression serves as a formality marker, whereas for speakers of the other languages, emotional expression has no influence on their rating of the situation as formal or informal.

E. Discussion

The study suggests that across cultures, whether or not a robot produces emotional displays when delivering bad news, has significant effects on how the robot is perceived. The effects found are quite remarkable given that the emotional displays used were restricted to the lexical material only; thus, no special tone of voice or mimics supported the displays, and generally previous work has shown that people prefer human over synthesized voices [57]. Still, the lexically expressed compassion by the robot had significant effects on how the robot was evaluated. It can be expected that congruent prosodic emotional displays and emotional mimics and gestures would have made the effect even stronger (see [58]); however, additional empirical evidence will have to show the effects of the interplay of multimodal emotional cues in the future.

The second type of result concerns the intercultural differences. The fact that different communities hold different expectations regarding emotional expression supports the suggestion that emotional expression is socially defined and activity specific. This also means that aligning with the participants' emotions would not have produced the correct results; the intercultural differences would not be accounted for in an approach that targets alignment between human and robot. While further research is necessary to integrate the findings in a general understanding of cultural preferences, especially concerning English, which is associated with several different cultures, our results indicate that emotional expression does not always rely on emotional states only, either.

V. General Discussion

The two case studies have shown that emotional expression can have considerable interactional functions, such as indicating common ground and mitigating bad news. Furthermore, the results of case study I suggest that the placement of emotional display within a conversational sequence is crucial; this is in line with previous work in conversation analysis, which also revealed the impact of sequential placement within an activity (e.g. [10]). These findings are also in line with research on embodied conversational agents, which has also observed that emotional expression is related to communicative acts [59], social roles [60] and detailed action sequences [61], and not only to emotional states (cf. also [62]–[64]).

The results of case study II furthermore illustrate the conventional nature of emotional expression by revealing

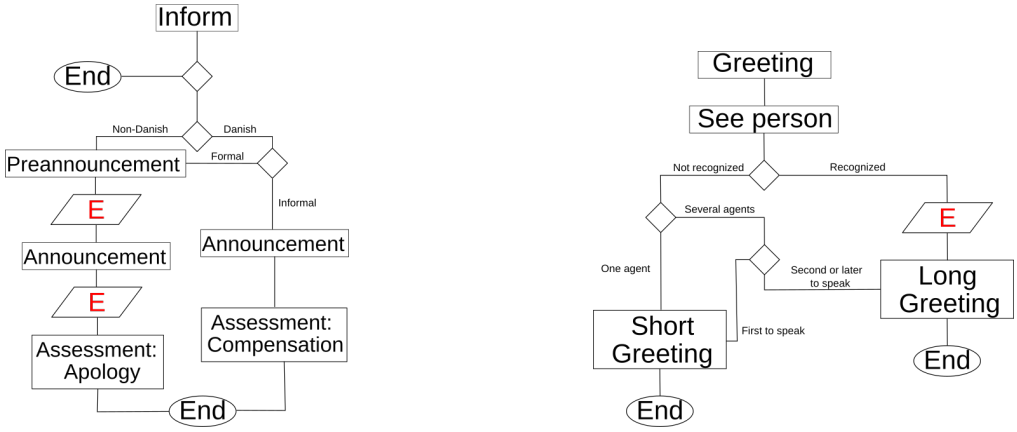


Fig. 7: Flow charts illustrating the culture- and context-dependency of emotional expression during the delivery of bad news (left), and the context- and sequence-dependency of emotional expression during greetings (right)

intercultural differences in people's expectations regarding the interactional practice of conveying bad news, where for the Danish speakers, emotional expression makes them perceive the situation as highly formal. Thus, to conceptualize emotional expression as due to emotional states is misleading in these cases, given its interactional functions. As a consequence of the context-, activity- and culture-dependency of emotional expression, HRI needs to design emotional displays with respect to specific activities, for which we make some suggestions below.

VI. Design Recommendations

In this section, we provide design recommendations for emotional expression for greetings and bad news, based on our two case studies. The goal of the examples is to illustrate how the sequential dependence and situational and cultural specificity of emotional expression could be accounted for in interactionally appropriate human-robot interactions.

A. Delivering Bad News

Providing bad news is an activity that is likely to be an integral part of robots' tasks (e.g. [65]), for instance, when serving as receptionists (the requested person may not be available, or the client may have to wait), as guides (e.g. a shop or site may be closed, or a route may not be possible), as shop assistants (a product may not be available), as entertainment robots (the child may have lost a game), or as tutors (the child may have made mistakes in a task), among many other roles robots may fulfill. As Maynard [52] has argued, providing bad news may comprise four phases: announcement, response, elaboration, assessment. The announcement should provide the listener with an idea of how to respond, especially whether it is bad news or good news and thus what kind of response is expected. As our study has shown, emotional display of compassion during the announcement phase has a crucial effect on how friendly, polite, engaging and warm a robot is evaluated. Furthermore, the emotional displays have been found to be culture-dependent, which should be considered in the interaction design (see Figure 7).

B. Greeting

As our case study has shown, the production of emotionally charged greetings depends on the familiarity with the partner on the one hand and on the position in the interaction (i.e. first or second to greet) on the other. Our recommendation for the display of emotional cues is illustrated in Figure 7.

VII. Conclusion

To conclude, we have shown that in interactions between humans, emotional display is to a large extent conventionally defined and activity specific. While we have presented only two contexts and two modalities of emotion expression (lexical and prosodic features), the fact that conversation analytic work has demonstrated that all modalities are integrated in emotional expression, for instance, in storytelling [10], it can be expected that the point we are making here applies to emotion expression in general, which future work should confirm. The answer to the question posed, why emotional expression, is thus that it is conventionally required as part of particular activities; accordingly, the answer to the 'when question' concerns specific places in interactional sequences, depending on the activity, the robot's role and the cultural context. Whereas current work in human-robot interaction has so far taken an approach that relies on emotions as based on internal states that need to be reliably recognized or aligned with users' states, we suggest to define emotional display according to conventional expectations towards the activities the respective robot is carrying out. Emotional expression is an integral part of everyday activities, which follows social rules and expectations. We hope that the perspective on emotion expression presented here opens new possibilities for the design of systems that generate appropriate emotion expression in robots.

VIII. Acknowledgments

We wish to thank phonetician Oliver Niebuhr for his help with the prosodic manipulations.

REFERENCES

[1] A. J. Fridlund, *Human facial expression: An evolutionary view*. Academic Press, 2014.

[2] W. A. Beach and E. Prickett, "Laughter, humor, and cancer: Delicate moments and poignant interactional circumstances," *Health Communication*, vol. 32, no. 7, pp. 791–802, 2017.

[3] J. Ruusvori, "Emotion, affect and conversation," in *The Handbook of Conversation Analysis*. Oxford: Blackwell, 2013.

[4] A. Hepburn and J. Potter, "Crying and crying responses," in *Emotion in Interaction*, A. Perälylä and M. Sorjonen, Eds. New York, NY: Oxford University Press, 2012, pp. 195–211.

[5] R. E. Kraut and R. E. Johnston, "Social and emotional messages of smiling: An ethological approach." *Journal of personality and social psychology*, vol. 37, no. 9, p. 1539, 1979.

[6] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, vol. 50, no. 4, pp. 696–735, 1974.

[7] H. Sacks, *Lectures on Conversation*, G. Jefferson, Ed. Oxford: Blackwell, 1996, vol. I & II.

[8] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati, "Emotional storytelling in the classroom: Individual versus group interaction between children and robots." in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction. Portland, USA, March 2-5*, 2015.

[9] L. Voutilainen, P. Henttonen, M. Kahri, M. Kivioja, N. Ravaja, M. Sams, and A. Perälylä, "Affective stance, ambivalence, and psychophysical responses during conversational storytelling," *Journal of Pragmatics*, vol. 68, pp. 1–24, 2014.

[10] M. Selting, "Affectivity in conversational storytelling: An analysis of displays of anger or indignation in complaint stories," *Pragmatics*, vol. 20, no. 2, p. 229277, 2010.

[11] J. B. Bavelas, L. Coates, and T. Johnson, "Listeners as co-narrators," *Journal of Personality and Social Psychology*, vol. 79, no. 6, pp. 941–952, 2000.

[12] M. Scheutz, P. Schermerhorn, and J. Kramer, "The utility of affect expression in natural language interactions in joint human-robot tasks," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, ser. HRI '06. New York, NY, USA: ACM, 2006, pp. 226–233. [Online]. Available: http://doi.acm.org/10.1145/1121241.1121281

[13] R. Gockley, J. Forlizzi, and R. Simmons, "Interactions with a moody robot," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, ser. HRI '06. New York, NY, USA: ACM, 2006, pp. 186–193. [Online]. Available: http://doi.acm.org/10.1145/1121241.1121274

[14] C. L. Bethel and R. R. Murphy, "Non-facial/non-verbal methods of affective expression as applied to robot-assisted victim assessment," in *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '07. New York, NY, USA: ACM, 2007, pp. 287–294. [Online]. Available: http://doi.acm.org/10.1145/1228716.1228755

[15] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '09. New York, NY, USA: ACM, 2009, pp. 69–76. [Online]. Available: http://doi.acm.org/10.1145/1514095.1514110

[16] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '09. New York, NY, USA: ACM, 2009, pp. 173–180. [Online]. Available: http://doi.acm.org/10.1145/1514095.1514127

[17] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," in *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '10. Piscataway, NJ, USA: IEEE Press, 2010, pp. 53–60. [Online]. Available: http://dl.acm.org/citation.cfm?id=1734454.1734473

[18] D. Leyzberg, E. Avrunin, J. Liu, and B. Scassellati, "Robots that express emotion elicit better human teaching," in *Proceedings of the 6th International Conference on Human-robot Interaction*, ser. HRI '11. New York, NY, USA: ACM, 2011, pp. 347–354. [Online]. Available: http://doi.acm.org/10.1145/1957656.1957789

[19] S. Yohanan and K. E. MacLean, "Design and assessment of the haptic creature's affect display," in *Proceedings of the 6th International Conference on Human-robot Interaction*, ser. HRI '11. New York, NY, USA: ACM, 2011, pp. 473–480. [Online]. Available: http://doi.acm.org/10.1145/1957656.1957820

[20] T. L. Chen, C.-H. King, A. L. Thomaz, and C. C. Kemp, "Touched by a robot: An investigation of subjective responses to robot-initiated touch," in *Proceedings of the 6th International Conference on Human-robot Interaction*, ser. HRI '11. New York, NY, USA: ACM, 2011, pp. 457–464. [Online]. Available: http://doi.acm.org/10.1145/1957656.1957818

[21] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling empathic behaviour in a robotic game companion for children: An ethnographic study in real-world settings," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 367–374. [Online]. Available: http://doi.acm.org/10.1145/2157689.2157811

[22] T. Ribeiro and A. Paiva, "The illusion of robotic life: Principles and practices of animation for robots," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 383–390. [Online]. Available: http://doi.acm.org/10.1145/2157689.2157814

[23] M. Sharma, D. Hildebrandt, G. Newman, J. E. Young, and R. Eskicioglu, "Communicating affect via flight path: Exploring use of the laban effort system for designing affective locomotion paths," in *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '13. Piscataway, NJ, USA: IEEE Press, 2013, pp. 293–300. [Online]. Available: http://dl.acm.org/citation.cfm?id=2447556.2447671

[24] R. Read and T. Belpaeme, "Situational context directs how people affectively interpret robotic non-linguistic utterances," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 41–48. [Online]. Available: http://doi.acm.org/10.1145/2559636.2559680

[25] G. Hoffman, G. E. Birnbaum, K. Vanunu, O. Sass, and H. T. Reis, "Robot responsiveness to human disclosure affects social impression and appeal," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/2559636.2559660

[26] M. Tielman, M. Neerincx, J.-J. Meyer, and R. Looije, "Adaptive emotional expression in robot-child interaction," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 407–414. [Online]. Available: http://doi.acm.org/10.1145/2559636.2559663

[27] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati, "Emotional storytelling in the classroom: Individual versus group interaction between children and robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 75–82. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696481

[28] G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, and T. Shani Sherman, "Design and evaluation of a peripheral robotic conversation companion," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 3–10. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696495

[29] M. F. Jung, N. Martelaro, and P. J. Hinds, "Using robots to moderate team conflict: The case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 229–236. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696460

[30] M. Ammi, V. Demulier, S. Caillou, Y. Gaffary, Y. Tsalamlal, J.-C. Martin, and A. Tapus, "Haptic human-robot affective interaction in a handshaking social protocol," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: ACM, 2015, pp. 263–270. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696485

[31] L. Boccanfuso, E. Barney, C. Foster, Y. A. Ahn, K. Chawarska, B. Scassellati, and F. Shic, "Emotional robot to examine differences in play patterns and affective response of children with and without asd," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ,

USA: IEEE Press, 2016, pp. 19–26. [Online]. Available: http://dl.acm.org/citation.cfm?id=2906831.2906837

[32] M. Koschate, R. Potter, P. Bremner, and M. Levine, "Overcoming the uncanny valley: Displays of emotions reduce the uncanniness of humanlike robots," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 359–365. [Online]. Available: http://dl.acm.org/citation.cfm?id=2906831.2906892

[33] G. E. Birnbaum, M. Mizrahi, G. Hoffman, H. T. Reis, E. J. Finkel, and O. Sass, "Machines as a source of consolation: Robot responsiveness increases human approach behavior and desire for companionship," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 165–171. [Online]. Available: http://dl.acm.org/citation.cfm?id=2906831.2906861

[34] J. R. Cauchard, K. Y. Zhai, M. Spadafora, and J. A. Landay, "Emotion encoding in human-drone interaction," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 263–270. [Online]. Available: http://dl.acm.org/citation.cfm?id=2906831.2906878

[35] M. Moosaei, S. K. Das, D. O. Popa, and L. D. Riek, "Using facially expressive robots to calibrate clinical pain perception," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 32–41. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020216

[36] S. Song and S. Yamada, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 2–11. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020239

[37] J. Feldmaier, M. Stimpfl, and K. Diepold, "Development of an emotion-competent slam agent," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 1–9. [Online]. Available: http://doi.acm.org/10.1145/3029798.3034784

[38] M. F. Jung, "Affective grounding in human-robot interaction," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 263–273. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020224

[39] V. Chidambaram, Y.-H. Chiang, and B. Mutlu, "Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 293–300. [Online]. Available: http://doi.acm.org/10.1145/2157689.2157798

[40] J. Otterbacher and M. Talias, "S/he's too warm/agentic!: The influence of gender on uncanny reactions to robots," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 214–223. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020220

[41] D. Loeffler, N. Schmidt, and R. Tscharn, "Multimodal expression of artificial emotion in social robots using color, motion and sound," in *HRI18, March 58, 2018, Chicago, IL, USA*, 2018.

[42] A. Tellegen, D. Watson, and L. A. Clark, "On the dimensional and hierarchical structure of affect," *Psychological Science*, vol. 10, no. 4, pp. 297–303, 1999.

[43] L. F. Barrett, "Are emotions natural kinds?" *Perspectives on psychological science*, vol. 1, no. 1, pp. 28–58, 2006.

[44] P. Ekman, "Expression and the nature of emotion," *Approaches to emotion*, vol. 3, pp. 19–344, 1984.

[45] A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press, 1990.

[46] D. Pillet-Shore, "Greeting: Displaying stance through prosodic recipient design," *Research on Language & Social Interaction*, vol. 45, no. 4, pp. 375–398, 2012.

[47] P. Taylor, A. W. Black, and R. Caley, "The architecture of the festival speech synthesis system," 1998.

[48] M. Atterer and D. Ladd, "On the phonetics and phonology of "segmental anchoring" of f0: Evidence from german," *Journal of Phonetics*, vol. 32, pp. 177–197, 2004.

[49] A. Weiss, R. Bernhaupt, D. Schwaiger, M. Altmaninger, R. Buchner, and M. Tscheligi, "User experience evaluation with a wizard of oz approach: Technical and methodological considerations," in *Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on*. IEEE, 2009, Conference Proceedings, pp. 303–308.

[50] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[51] D. W. Maynard, "Notes on the delivery and reception of diagnostoc news regarding mental disabilities," in *The Interaction order: New Directions in the Study of Social Order*, A. J. Meehan and A. W. Rawls, Eds. New York: Irvington, 1989.

[52] ——, "The news delivery sequence: Bad news and good news in conversational interaction," *Research on Language and Interaction*, vol. 30, no. 2, pp. 93–130, 1997.

[53] ——, *Bad News, Good News: Conversational Order in Everyday talk and Clinical Settings*. University of Chicago Press, 2003.

[54] E. Fredsted, "Politeness in denmark: Getting to the point," *Politeness in Europe. London: Multilingual Matters*, pp. 159–173, 2005.

[55] B. e. a. Ivanovsak, "On the reliability of discourse completion tests in measuring pragmatic competence in foreign language learners," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 2016.

[56] A. Wierzbicka, "The semantics of interjection," *Journal of Pragmatics*, vol. 18, pp. 159–192, 1992.

[57] C. Nass and S. Brave, *Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA., London: MIT Press, 2005.

[58] N. Jaques, Y. L. Kim, and R. W. Picard, "Personality, attitudes, and bonding in conversations," in *Proceedings of Intelligent Virtual Agents, California, USA*, 2016.

[59] I. Poggi and C. Pelachaud, "Emotional meaning and expression in animated faces," in *Affective interactions*. Springer, 2000, pp. 182–195.

[60] C. Pelachaud, V. Carofiglio, B. De Carolis, F. de osis, and I. Poggi, "Embodied contextual agent in information delivery application," in *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, Bologna, Italy, 2002, pp. 758–765.

[61] C. Becker-Asano and I. Wachsmuth, "Affective computing with primary and secondary emotions in a virtual human," *Autonomous Agents and Multi-Agent Systems*, vol. 20, no. 32, 2010.

[62] J. Gratch and S. Marsella, "Lessons from emotion psychology for the design of lifelike characters." *Applied Artificial Intelligence*, vol. 19, pp. 215–233, 03 2005.

[63] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.

[64] Z. Kasap and N. Thalmann, *Interacting with Emotion and Memory Enabled Virtual Characters and Social Robots*, 06 2010, vol. 2010, pp. 209–224.

[65] M. K. Lee, S. Kiesler, and J. Forlizzi, "Receptionist or information kiosk: How do people talk with a robot?" in *Computer Supported Cooperative Work*, Savannah, Georgia, February 6-10 2010, pp. 31–40.