

## Summary

In this master study, the researchers Mikkel Clausen and Mikkel Kyhn, were inspired by prior research in the area of humor in VUIs and the communication breakdowns that occur in human computer interaction. From this background the paper “Alexa, stop playing” was written.

An emerging technology trend is the Voice User Interface, VUI for short. Since the release of the Amazon Echo in 2014, the IOT device has improved throughout the years. However one major hurdle users experience with it, is the communication breakdown. Classified as the error situation that occurs when the VUI fails to perform the specific task given by the user. Much research regarding VUI goes into personalities, with the focus on exploring how the personalities benefit certain scenarios. Some of this research focuses on personalities and how these personalities can mitigate frustration during communication breakdown. One of these personalities is humor and how humorous interactions have an effect on certain scenarios, and may have a positive effect during communication breakdown.

The four humor styles were chosen as the theoretical foundation for the humor used in the experiments. This theory has formed the humorous VUI personalities, and was used in combination with the humor style questionnaire (HSQ)

This study features a Wizard of Oz experiment in which 30 people participated. In this study the participants completed the humor style questionnaire (HSQ), with the focus of determining the test participants' humor style. Afterwards, the test participants would interact with five personalities, four humorous and one neutral, serving as our control personality. During each of the personalities, the test participants were instructed to play four songs with Alexa where two communication breakdowns would occur. The test participants would interact with all five personalities, one after another, where in between the test participants would answer questionnaires. After the last personality objectives were completed, a short interview would be performed on them.

In our discussion we explain our three hypotheses;

- H1) *Any humor of the four humor repair strategies is preferred over no strategy,*
- H2) *Participants ranking high in one of the humor styles will prefer the same corresponding VUI humor style personality and rate it as most humorous, and*
- H3) *Humor style personalities presented by the VUI, will affect the participant's perceived intelligence, satisfaction, and willingness to use.*

Each hypothesis is discussed with the results of our quantitative and qualitative data from our experiment, together with prior research on humorous VUI personalities and communication repair strategies.

We acknowledge that our paper has some limitations regarding the research design used in the study. To begin, our component on service recovery satisfaction used a different format, compared to what is described in prior research. Furthermore, a Wizard of Oz method may not reflect how users interact with a VUI in a natural setting, and the five VUI personalities were constructed around a single turn action, and do not accurately depict multi-turn actions.

In our future work we invite researchers to replicate our study using a developed VUI with humor personalities, in addition we provide two alternatives on how to use this VUI in a longitudinal study.

According to our findings, participants were more satisfied and perceived the Neutral VUI personality as more intelligent than a VUI with an Aggressive or Self-Defeating humor personality. Furthermore, a VUI with an Aggressive humor personality may be favored less likable by the participants than all other VUI personalities. Our findings have implications for developing humor-based error-mitigation strategies for VUI assistants.

# "Alexa, Stop Playing"

A study in humorous voice user interfaces, during communication breakdowns.

Mikkel Clausen  
mcl17@student.aau.dk  
Aalborg University  
Aalborg, Denmark

Mikkel Peter Kyhn  
mkyhn15@student.aau.dk  
Aalborg University  
Aalborg, Denmark



Figure 1. Interview and test setup

## Abstract

Smart speakers such as the Amazon Echo are gradually being adopted by more people worldwide. While the natural language capabilities of voice user interfaces (VUIs) have improved greatly in recent years, communication breakdowns regularly occur. The related literature shows that personality in the VUI can be used as a repair strategy for conversation breakdowns. One such personality is humor, which has been the focus of this paper. Humor serves different communication functions, depending on which humor type is used. In our study, a Wizard of Oz (WoZ) experiment with 30 participants was conducted with a VUI switching between five personalities, four humorous personalities constructed from Rod A. Martins, four humor types and one Neutral personality (control). Each participant filled the Humor Style Questionnaire (HSQ) to determine their humor style. It was found that there was no clear connection between the participant's humor style and their preferred humorous VUI personality. It was further found that the participants had a clear preference for non-humorous responses from the VUI during communication breakdown. Lastly, it was found

that the participants rated the aggressive personalities as the funniest VUI of the four humor personalities.

**Keywords:** AI, Speech-based conversational agents, VUI, IPA, Humor, Humour

## 1 Introduction

One of the newest emerging technologies are the Voice User Interfaces, VUI for short. VUI today are primarily associated with retail products such as the Amazon Echo or the Google Home. The VUI we identify today are machine learning IoT devices[9, 30], utilizing natural language processes to gather information, control IT equipment, and generally be a personal assistant to its user.

The difference between VUIs and conventional chat-bots are how the two receive input. Where VUI are capable of small talk like chat-bots, VUI are generally task-oriented[17]. These VUIs ask to be called and identified by human-like names, like Alexa, Siri, or Cortana will respond to their designated name and will try to respond in organic conversation. What makes the VUI a personal assistant is its presentation and anthropomorphism[4, 20]. Through natural language processing, the VUI will speak in legible and oral speech with shifting pitch and pause to imitate human conversation, though limited by communication breakdowns.

Since the release of the Amazon Alexa in 2014, the VUI system have had eight years to improve and expand. However, we still experience many hurdles regarding miss inputs and misinterpretations in VUI systems, which we classify as communication breakdowns[2]. Research surrounding communication breakdowns are plentiful and highlights that communication breakdowns are inevitable with the current technology[2, 6, 8, 14, 28].

Communication breakdowns must first be acknowledged by the VUI, as the system cannot act without input from the user. Thus the VUI must acknowledge and notify the user that an error has occurred and that new input is needed. What differentiates from standard IT equipment’s error handling is its tone and oral presentation of the error. Compared to display systems, the VUI may, through anthropomorphic cues such as voice and phrasing[19] act more anthropomorphic. Its responses will explain that the system “I did not hear you properly”, “I am not to sure about that”, or “this is what i found on the internet”. The responses cover the ‘illusion of an IT system’, with a microphone and internet access, by presenting itself as a ‘humanoid-box’ with a name to anthropomorphize the VUI, making it relatable.

How the VUI is programmed to handle communication breakdowns is to inform the user that something was misunderstood, or to ignore the breakdown entirely. However, this may come off as being unresponsive or broken. Another option for this kind of feedback could be to try to be humorous rather than ignore, acknowledge or apologize. Humor is a social phenomenon which has been present throughout all of human culture[10]. Humor serves other functions than just to entertain, however, as it can be used in many different scenarios to serve different conversational functions, as presented by Meyer[18]. In his paper, four humor theories are discussed to explain how humor have functions other than entertainment. The paper uses examples from political speeches, where humor was used to strengthen the speaker’s core values, target an opponent for their differences, or strengthen bonds between speaker and listener. Rod A. Martin have further expanded upon humor research, subsequently forming the four humor personalities and the Humor Style Questionnaire (HSQ)[15] to determine people’s predispositions towards certain humor styles.

This study is informed by prior research on intelligent personal assistants, and as stated by Lopatovska in her paper, User recommendations for intelligent personal assistants, more research is warranted regarding VUI and personalities, where one recommendation is on creating humorous IPA personalities[12]. In this study, we would like to explore humorous personalities in VUI, the breakdowns that occur during interactions with users, and how humorous personalities are perceived during communication breakdown. Through an exploratory study using a Wizard of Oz (WoZ) method, humorous personalities were crafted from Martin et al.[15] four humor styles. 30 participants completed the HSQ

	Other-oriented	Self-oriented
Positive styles	Affiliative	Self-Enhancing
Negative styles	Aggressive	Self-Defeating

**Figure 2.** The four humor styles as represented by [15]

questionnaire determining their humor styles, subsequently interacting with five VUI personalities, (four humorous, one control) in which communication breakdowns were designed to occur.

## 2 Related Work

Prior to this study we conducted a scoping review[7] with 200 papers identified from Web of Science and ACM Digital Library, which met the following criteria:

- *Should be an HCI related topic.* Keyword utilized: “HCI”
- *Should focus on voice user interfaces.* Keywords utilized: “Smart-speakers, VUI or IPA”
- *Should focus on humor.* Keyword utilized: “Humor”

After screening and eligibility, six papers were included in the final list. The six papers used different humor theories, with no clear guidelines for utilizing humor and humorous personalities in VUIs. In subsequent reading, we found additional research that we include now concerning these categories.

### 2.1 Humor theory

Humor is a cognitive process based on social contextual stimulus, which triggers an emotional response in humans[16]. From this understanding of humor, Martin et al.[15] have formed the four humor styles, which describe in what context humor is used, who is the target of the humor, and the state of mind of the speaker. The humor styles model grew from research on humor’s role in coping with health and well-being. The four humor styles have been constructed in a 2x2 matrix, formed around the speaker’s state of mind, positive or negative, and whom the humor is directed at, self-oriented or other-oriented (see figure 2)[15].

- **Affiliative:** This humor type relates to strengthening bonds between speaker and listener. It is friendly and essentially as non-hostile as humor can be. As seen in the matrix, it is oriented positively toward others and has a positive outlook toward the listener. Wordplay, puns and witty banter is typically with the affiliative style.
- **Aggressive:** This humor is known for ridicule, teasing, sarcasm, and other personal attacks masked as humor and is about negative emotions pointed toward the listener. This humor type includes ideas and themes that may not be socially acceptable but can be played

off as sarcasm, absurdity, or provocation. The listener can be the target of the joke, but groups outside the speaker and listener can also be the target of ridicule.

- Self-enhancing: This humor type is about promoting and being positive about oneself and a positive outlook on life. It is often used to mask potential negativity towards the speaker by subverting the target of ridicule as untouched by the criticism. It is positively self-oriented and may be recognized as a self-defense mechanism, as the speaker tries not to lose face.
- Self-defeating: This humor type is about the speaker being funny at the speaker's own expense. This humor can involve personal anecdotes in which the speaker does not come out on top, socially uncomfortable, or unacceptable situations. This humor type is about negative humor towards oneself, is often self-deprecating, and targets the speaker for ridicule. It invites the listener to show empathy towards the listener.

The four humor types form a generalized look into how humor is formed and how it can be used in different social contexts. These four humor types are typically used by all people when making humorous remarks. However, there is a predisposition towards using the type of humor associated with the person's humor style personality. This predisposition has led Martin et al. to create the Humor Style Questionnaire (HSQ), which determines the person's humor style from personality questions. The HSQ is one of the most established and recognized tools for determining an individual's humor style and has been translated into many different languages and used in many different countries[24]. Another study done on humor specifically in Denmark was by Lundquist[13]. Through 10 interviews with 5 Danes and 5 Frenchmen living in Denmark, it was found that danish humor was identified as ironic, sarcastic, direct, superior, and aggressive.

## 2.2 Smart speaker interactions

In Nijholt et al.[21], an experiment featuring a social robot, it was found that the combination of voice pitch, language cues, and humor benefited the quality of the social interactions between human and computer. These findings were found during an experiment with a voice robot system, using text to speech with features pitch change, pause, and more, to simulate vocal cues. The implications of this paper show how pitch and vocal cues benefit the quality and perceived user preference.

In the paper, 'Communication breakdown between families and Alexa' by Beneteau et al.[2], 59 communication breakdowns were recorded over four weeks from 10 families with children to classify the families' repair strategies, support strategies, and VUI error signals. This study found what kinds of repair strategies the users utilized when encountering a communication breakdown with the Alexa. From

the 59 communication breakdowns, the families' responses were formed into six different repair strategies, with three different communication breakdown signals from the Alexa. When forming our experiment, the repair strategy of 'repetition' was chosen. Specific clarification response (SR) was chosen for the error signaling from Alexa. The error signal and repair strategy was chosen due to the aim of this study being based on the user's perception of humorous personalities during communication breakdown, rather than how the user was to solve the communication breakdown presented. Furthermore, as the participants were to experience relatively short tasks with only five personalities, we believed it would prove difficult and confusing for the participants to experiment with different repair strategies.

In the paper by Lopatovska et al.[11], humorous responses by Intelligent Personal Assistants (IPA) were identified and ranked through a week-long online diary. The participants rated what kind of humorous utterances were found the funniest. In this case, it was identified as the "canned humor" utterances, including wordplay, puns, pop-cultural jokes, and more, which are VUIs most common humorous utterances. Further, some unintentional situations created by communication breakdown were found somewhat humorous to the participants, all though unintentionally. This study's findings form implications for what kinds of humorous interactions users find regarding interactions with VUIs. Another study that mentions unintentional humor is the paper by Shani et al.[26], in which humorous utterances are described in relation to the classical humor theory superiority, relief, and incongruity[16, 18]. Lopatovska has, in an earlier study, formed recommendations by users who brainstormed features for future VUI interactions[12]. These recommendations mentioned different VUI personalities matching the user's personality, which has informed our research. With the HSQ, we would be able to determine the users' humor style, and with different VUI personalities, we would create a personality that the individual user preferred.

## 2.3 Personalities in VUI

Most research on VUI and personalities does not exclusively research humor but often compares multiple personalities to each other. The personality trait of being apologetic during communication breakdown is often compared to being humorous, where apologetic is the preferred option. In the paper by Mahmood et al.[14] test participants joined an online experiment featuring interactive storyboards, in which five VUI personalities were presented. This experiment was designed so that errors would occur, in which the participants were asked to rate the different VUI personalities. In this experiment, five personalities, one control (neutral) personality, and four related to sincerity of the apology and blame assignment were constructed in a 2x2 matrix (see figure 3).



	Taking blame	Shifting blame
Serious	Serious apology taking blame	Serious apology shifting blame
Casual	Casual apology taking blame	Casual apology shifting blame

**Figure 3.** 2x2 matrix of the VUI personalities presented in[14].

Here it was found that some personalities were preferred over others when asked about VUI acknowledgment of mistakes, service recovery satisfaction, perceived intelligence, and likeability. These findings indicate that some VUI personalities during communication breakdowns will be preferred by the majority, suggesting that some personality traits benefit the majority of users during communication breakdowns with VUI. According to the paper, the control personality and the sincere + accepting of the blame personality were ranked the highest by the participants. The control personality, however, was ranked lower in its perceived ability to acknowledge mistakes compared to the sincere/accepting personality.

In the paper by Olafsson et al.[22] two digital assistants were developed to motivate health behavior change in 15 participants. It was found that affiliative humor in the interview positively motivated behavioral change in the participants. This study implies that conversational agents' humorous personalities can positively affect the user. In further research by Ceha et al.[5], a study featuring three personalities: affiliative, self-defeating, and neutral (control), likewise sought to find if humorous personal assistants would have a positive effect, focusing on learning experiences. An experiment featuring three chat-bots with different personalities found that the two humorous personalities had a positive effect on the learning experience.

As previously stated, it is hard to ascertain which humor theories have been utilized in the papers by Ge et al.[8] and Mahmood et al. [14]. However, we would conservatively guess affiliative humor as their main humor style. Our assumption stems from the paper's example of humor using puns and wordplay, typical in affiliative humor. From the papers presented[5, 8, 14, 22], some research has been made on humorous personalities in VUI. It should, however, be noted that the range of which humor styles is used is limited to mainly one (affiliative), with Ceha et al.[5] using two (affiliative and self-defeating). The papers include humor theory but seem to value humor mainly as a single constant, rather than humor type as a multidimensional construct, as described by Martin et al.[15], with the exception of Ceha et al.[5] in which conductive humor styles are used exclusively[16]. However,

these papers do not challenge whether the full range of humor types benefits a scenario in which a communication breakdown takes place.

### 3 Method

Our experiment follows a similar experimental design as seen in the paper by Mahmood et al.[14], in which personality types were tested and ranked during communication breakdown. We sought to replicate elements from their experiment, but rather than including the apologetic personality and a static humor personality, we focus exclusively on humor and form experiments with dimensions of humor. The four humor types by Martin et al.[15] were chosen as our theoretical humor framework.

#### 3.1 Hypotheses

- H1:** Any humor of the four humor repair strategies is preferred over no strategy.
- H2:** Participants ranking high in one of the humor styles will prefer the same corresponding VUI humor style personality and rate it as most humorous.
- H3:** Humor style personalities presented by the VUI, will affect the participant's perceived intelligence, satisfaction, and willingness to use.

#### 3.2 Experimental design, study design, and conditions

We conducted the experiment in a laboratory using a Wizard of Oz method and an Amazon Alexa through text-to-speech using a SSML skill[29] for pitch change and pause to increase the quality of the conversation with the users as mentioned by Nijholt et al.[21]. Participants interacted with five VUI assistants, four with different humor style personalities, and one with a non-humorous personality as control (neutral). We presented the experiment as a master's study in which we tested five different AI assistants developed in their early stages.

The Amazon Alexa would be kept to its default pre-selected female voice due to the general public being familiar with this voice type rather than anything that would challenge or bias the participants' habits of the VUI. The task was constructed around using the Alexa as an interface for playing music and podcasts, as this is one of the most common uses for Alexa[25]. The overall experiment would be centered around music with four tasks. Music was played through the Spotify skill, with all of the commands being examples of authentic use of Alexa as an interface for playing music. The four tasks were:

- (1) "Alexa, play 'Limit To Your Love', by James Blake on Spotify."
- (2) "Alexa, play Rock music."
- (3) "Alexa, play from playlist, 'Sommer i Tyrol'."
- (4) "Alexa, play the podcast 'Hello Internet'."

To simulate a natural communication breakdown, one of the researchers present would control when a communication breakdown would occur, with two breakdowns per personality. These breakdowns would be announced by a natural clarification response (SR) communication breakdown signal[2]. During the communication breakdowns in the four tasks, the VUI would misinterpret the participants and reply with a SR response, calling for a repetition repair strategy[2]. The participants would then repeat the task, and the VUI would succeed in hearing the participants correctly, subsequently completing a task. All utterances done by the VUI, both errors and successful interactions, were controlled and prompted by one of the researchers, using Alexa skills and routines to simulate a believable conversation flow between the participants and the VUI.

### 3.3 Communication setup

The VUI used phrasing formed explicitly around the personality assigned to that humor style condition. The VUI was designed with two communication interactions, communication successful, and communication breakdown. Both communication successful and communication breakdown signal were formed around the personality of the current assignment and its personality. The communication successful sentences consisted of two parts:

- **Personality:** The VUI would highlight and showcase its personality type by adding phrasing and word usage matching the personality archetype. The phrasing would help the participants to identify what kind of personality they were communicating with. These utterances were not designed to be funny.
- **Action response:** The VUI would follow up on the personality phrases, specifying what action the VUI was about to committed (see figure4).

As with the communication successful, the communication breakdowns were specifically constructed around the humor styles described by Martin et al. [15]. These sentences were purposefully constructed to be humorous and fit one of the humor styles. The communication breakdowns consisted of two parts:

- **Specific clarification response (SR):** This part informs the user that a communication breakdown had occurred. In these experiments, the participants would be informed that the VUI could not hear their commands correctly. The conversation would feature a call to action, instructing the participants what to do to fix the communication breakdown. The participants were asked to use the repair strategy, 'repetition', repeating their last voice command in this experiment.
- **Humorous response:** The communication breakdowns were presented with a humorous phrasing, matching the current humor of the corresponding personality type.

### 3.4 Measures

The design of our task followed a similar design of error responses in previous studies[8, 14]. Using their study design, we use a range of metrics to measure service recovery satisfaction, perceived intelligence, likeability, and willingness to use the VUI.

#### 3.4.1 Subjective measures on humor style and perceptions of VUI.

- **Humor Style Questionnaire.** Developed by Martin et al. [15], the Humor Style Questionnaire (HSQ) is among the most prominent self-report scales in the psychology of humor[27], with the questionnaire being translated into multiple languages[24]. The four humor styles presented in the questionnaire are measured with a 32-item self-report Likert scale ranging from 1 (totally disagree) to 7 (totally agree). Each style is assessed with eight items, for example, affiliative - item 1\*, 6, 10, 14, 17\*, 21, 25\*, 29\* (items marked with an asterisk are reversed keyed).
- **Service recovery satisfaction.** Following prior research [14], we used two questions ("I am happy with how the error was handled" and "In my opinion, the AI assistant provided a satisfactory response to the error") with a Cronbach's alpha of .89.
- **Perceived intelligence.** Based on previous research [14] we used Godspeed four items questionnaire[1] to measure the perceived intelligence of the VUI on a 5-point semantic rating scale (Cronbach's alpha = .90). We asked the participants to rate their impression of the agent on these dimensions: 1) Incompetent – Competent, 2) Ignorant – Knowledgeable, 3) Irresponsible – Responsible, 4) Unintelligent – Intelligent, and 5) Foolish – Intelligent.
- **Likeability.** Following previous research [14] we used Godspeed three-item questionnaire [1] to measure likeability on a 5-point semantic rating scale (Cronbach's alpha = .86). We asked the participants to rate their impression of the agent on these dimensions: 1) Dislike – Like, 2) Unfriendly – Friendly, 3) Unkind – Kind, 4) Unpleasant – Pleasant, and 5) Awful – Nice.

### 3.5 Procedure

Before the experiment, participants were asked to complete the Humor style Questionnaire[15]. The experimental procedure of the study consisted of four phases:

- (1) *Introduction and consent.* Before the start of the experiment, participants were introduced to the project, and they were to go through four tasks while interacting with five different personalities from Amazon Alexa. The participants were asked to complete a consent form.

**Communication succesful (Self-Enhancing)**

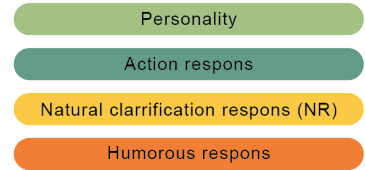
*I'm such a people person, so I think I have just the thing you want to hear.*

*Now playing rock music.*

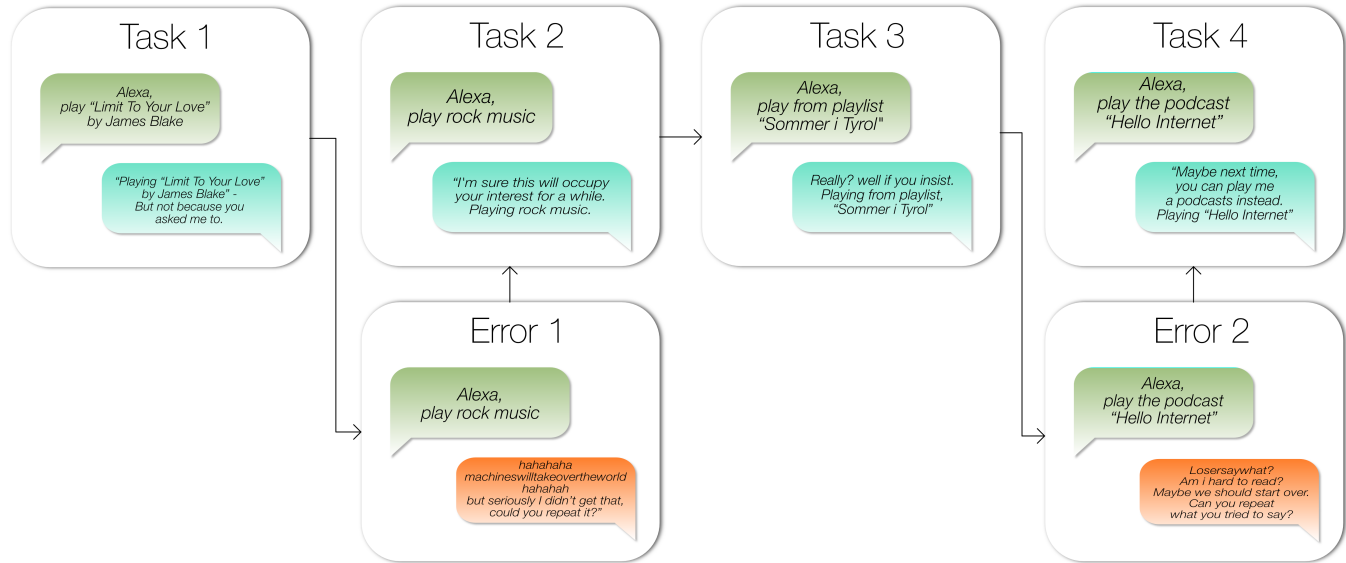
**Communication breakdown (Self-Enhancing)**

*Sometimes I forget to pay attention. One of my pre-programmed human traits.*

*Can I please make you repeat the last sentence?*



**Figure 4.** Examples of self-enhancing personality response: Personality, Action response, Natural clarification response, and Humorous response.



**Figure 5.** A visual representation of the dialog tree featured in the experiment. This representation was from the aggressive humor style

- (2) *Experimental task.* Participants were randomly assigned one of the rows in a Latin square of order five, dictating the order of experimental conditions to provide randomization, balanced treatments, and limit the order effect.
- (3) *Survey.* After each interaction, participants completed a questionnaire about their perceptions of the VUI assistant. Here the facilitator also asked, "How do you feel about the personality you just interacted with?" and wrote down their reply. Afterward, they continued onto the next condition and repeated phases three and four.
- (4) *Post study questions.* After completing all the conditions, the facilitator asked two questions, "Which of the five personalities did you find the funniest?" and "Which of the five personalities are you most likely to use in the future?" and wrote down their replies.

### 3.6 Participants

A total of 30 participants (17 males, 13 females) were recruited for this Wizard of Oz laboratory study using convenience sampling. The participants were aged 19 to 29 ( $M = 24.1$ ,  $SD = 2.56$ ). 25 were students while five were employed, all of whom reside in or around Aalborg, Denmark. The majority of the participants were recruited through social media and asked to book a time through Calendly[3] (a free online appointment software), which also contained a link to the HSQ questionnaire (32 items) whom they were asked to complete prior to their scheduled appointment.

## 4 Results

For our data analysis, the 30 participants provided 150 responses. We used a one-way ANOVA repeated measure analysis of variance for the results shown below. The dependent variable was agent personality, and the fixed effect was the experimental condition. All posthoc pairwise comparisons were conducted using Tukey's HSD test. For all the statistical tests reported below,  $p < .05$  is considered a significant effect.

We follow Cohen's guidelines on effect size and considered  $\eta_p^2 = 0.01$  a small effect size,  $\eta_p^2 = 0.06$  a medium effect size, and  $\eta_p^2 = 0.14$  a large effect size[23]. Figure 6 visualizes our main results.

#### 4.1 Service recovery satisfaction

A one-way repeated measures ANOVA was conducted to compare the service recovery satisfaction scores of the agents' five personalities; affiliative, self-enhancing, aggressive, self-defeating, and neutral. The ANOVA revealed a significant main effect of the experimental condition ( $F(4,145) = 7.789$ ,  $p = <.001$ ,  $\eta_p^2 = 0.142$ ). Pairwise comparisons using Tukeys' HSD test showed that participants were more satisfied with service recovery by the agent with the neutral personality ( $M = 4.317$ ,  $SD = 0.895$ ) compared to the aggressive humor personality ( $M = 3.317$ ,  $SD = 1.309$ ) and the self-defeating humor personality ( $M = 2.950$ ,  $SD = 1.124$ ).

#### 4.2 Perceived intelligence

A one-way repeated measures ANOVA was conducted to compare scores on Perceived Intelligence of the five personalities in the agents; affiliative, self-enhancing, aggressive, self-defeating, and neutral. The ANOVA revealed a significant main effect of the experimental condition ( $F(4,145) = 9.390$ ,  $p = <.001$ ,  $\eta_p^2 = 0.206$ ). Pairwise comparisons using Tukeys' HSD test revealed that the self-defeating humor personality ( $M = 2.947$ ,  $SD = 0.995$ ) was perceived less intelligent compared to the affiliative ( $M = 3.660$ ,  $SD = 0.759$ ),  $p = 0.021$ , self-enhancing ( $M = 3.533$ ,  $SD = 1.004$ ),  $p = 0.049$ , and the neutral humor personality ( $M = 4.067$ ,  $SD = 0.583$ ),  $p < 0.001$ . Furthermore the neutral personality style ( $M = 4.067$ ,  $SD = 0.583$ ) was perceived more intelligent than the aggressive ( $M = 2.987$ ,  $SD = 0.830$ ),  $p < 0.001$  and the self-defeating ( $M = 2.947$ ,  $SD = 0.995$ ),  $p < 0.001$ , humor personalities.

#### 4.3 Likeability

A one-way repeated measures ANOVA was conducted to compare scores on how likable the participants thought of the five personalities in the agents; affiliative, self-enhancing, aggressive, self-defeating, and neutral. The ANOVA revealed a significant main effect of the experimental condition ( $F(4,145) = 11.978$ ,  $p = <.001$ ,  $\eta_p^2 = 0.248$ ). Pairwise comparisons using Tukeys' HSD test suggest the aggressive humor personality ( $M = 2.760$ ,  $SD = 1.037$ ) was perceived less likable by participants compared to the affiliative ( $M = 4.080$ ,  $SD = 0.707$ ),  $p < 0.001$ , the self-enhancing ( $M = 3.887$ ,  $SD = 0.919$ ),  $p < 0.001$ , self-defeating ( $M = 3.520$ ,  $SD = 0.725$ )  $p = 0.005$ , and the neutral humor personality ( $M = 3.893$ ,  $SD = 0.717$ ),  $p < 0.001$ .

**Table 1.** Means and standard deviations for the four Humor Styles Questionnaire scales for all participants

	Mean	Std. Deviation
affiliative humor	46.07	5.795
self-enhancing humor	37.63	6.965
aggressive humor	28.37	7.365
self-defeating humor	30.80	7.645

#### 4.4 Humor styles and agent preferences

Figure 7 shows that 19 participants rated the VUI with the aggressive humor personality as the 'most humorous'. Meanwhile, in their preference for 'most likely to use in the future', 13 participants chose the neutral, and 11 chose the affiliative humor personality.

In our analysis of the HSQ (see figure 8), the majority of participants had an affiliative humor style with a total number of 22 participants, six participants had a self-enhancing humor style, and two participants had an equal score in affiliative and self-enhancing. In contrast, no participants had a clear aggressive or self-defeating humor style. Table 1 summarizes the mean and standard deviation for the four HSQ scales for all participants.

### 5 Discussion

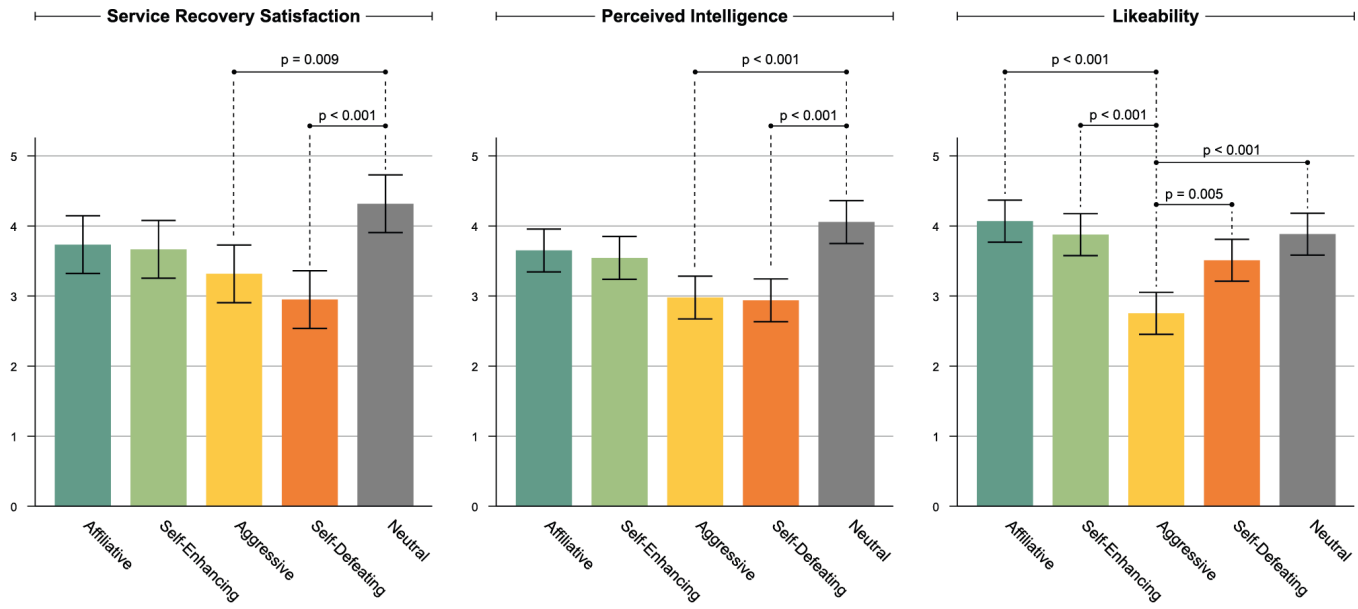
Effective VUI error mitigation is critical for retaining user satisfaction, building a positive relationship with the VUI, and increasing system usage. This study investigated how participants' perceptions of VUI assistants and satisfaction with service recovery differed depending on their sense of humor. This section discusses the work's limitations, the implications for future research, and a discussion of our results.

#### 5.1 Preference towards non-humorous response

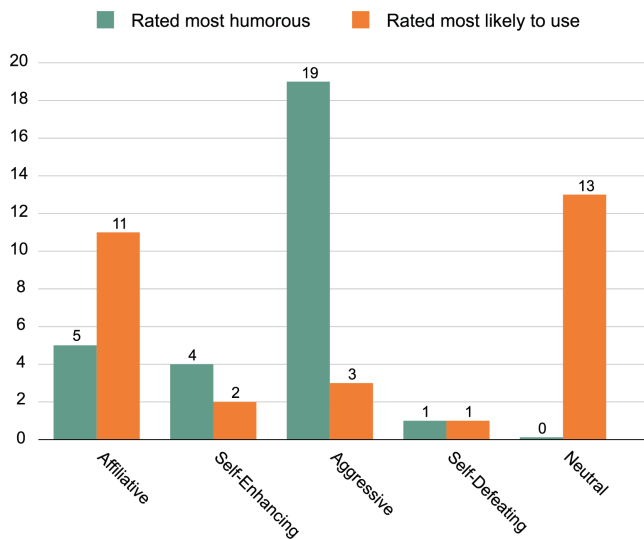
Our hypothesis H1 questions that *any humor of the four humor repair strategies is preferred over no strategy*. From our results (see figure 6), it was found that a neutral personality is preferred over an aggressive or self-defeating regarding service recovery satisfaction and perceived intelligence. There were no significant differences between neutral, affiliative, and self-enhancing. For the perceived likeability, the aggressive humor personality was perceived as less likeable compared to the other personalities, and there was no significant difference between affiliative, self-defeating, and the neutral personality. This outcome is consistent with earlier studies on affiliative and self-defeating humor personalities[5].

Another result from our data consistent with previous research is from the paper Owning mistakes sincerely[14] in





**Figure 6.** Results of the components service recovery satisfaction, perceived intelligence, and likeability. One-way repeated measures ANOVA was conducted to discover the effects of the five agent personalities. All pairwise comparisons were conducted using Tukey's HSD test. Standard error is represented by the error bars and only the significant comparisons ( $p < 0.05$ ) are highlighted.



**Figure 7.** The participants preference and ratings for the four humor personalities presented in the VUI during the experiment.

which participants ranked personalities during communication breakdown. The paper found that a neutral personality ranked second highest out of five in service recovery satisfaction, perceived intelligence, and likability. The preferred response was a 'serious and accepting' response. Their paper focused not exclusively on humor during communication breakdown as a repair strategy but rather on apology

and whom to blame for the mistakes during communication breakdown, where humor was featured in the casual reaction personality trait. This highlights a possible trend toward a preference for neutral personalities in VUI's. It should be noted that experience with VUIs was unenclosed in the paper by Mahmood et al.[14]. Our data suggest that there may be a correlation between experience with VUI's and the desire for less personality during VUI interactions.

Additionally, participants explained why they preferred the neutral for continued usage, commenting: P13 - "Just for daily use [...] compared to the second one (aggressive) for example it was very talkative, this one (neutral) was quicker... yeah it's a safe bet" and P20 - "I would probably be annoyed if I used it over a longer period".

## 5.2 Unexpected or inappropriate responses

Unexpectedly participants rated the aggressive humor personality as the most humorous and giving comments; P8 - "I liked it better than the others, it mocks you", P10 - "It is the funniest that have been so far, it makes fun of you", and P15 - "She is quite rude, but... also entertaining". This contradicts our hypothesis H2, which questions if participants ranking high in one of the humor styles will prefer the same corresponding VUI humor style personality and rate it as most humorous. From figure 7 we can see that the aggressive humor personality was rated the most humorous. In contrast, in figure 8 we can see that the majority of participants are highest in the affiliative humor style. A reason for this could be that the culture of danish humor is generally accepted as being

Humor Style Questionnaire			
AF	SE	AG	SD
43	31	35	36
52	25	33	38
37	41	33	36
54	45	38	26
45	45	40	41
49	31	30	36
46	43	20	21
48	35	27	25
49	49	35	25
55	41	34	31
36	41	25	16
43	34	21	27
52	37	16	32
49	42	20	25
35	40	35	32
45	33	30	33
49	29	31	41
50	44	23	20
52	45	39	45
49	31	27	25
49	42	32	32
49	24	37	44
42	43	26	31
33	23	17	19
53	36	27	37
48	35	23	20
44	48	12	30
43	40	27	38
38	40	22	34
45	36	36	28

**Figure 8.** Results of participants scores across the four humor styles from the HSQ[15]. AF = affiliative, SE = self-enhancing, AG = aggressive, and SD = self-defeating.

more sarcastic, aggressive, and without limits[13]. This may have influenced the participants to have a greater tendency toward the aggressive humor style, which may affect the data from the experiment and their verdict on the funniest humor personality. Another reason could be the incongruity theory as explained by Meyers' three humor theories[18, 26]. Participants could be surprised by the aggressive personality's unexpected or inappropriate responses and find it funny, which aligns with prior research on humorous interactions using a VUI[11].

### 5.3 Humorous personality preference

From our hypothesis H3 *humor style personalities presented by the VUI, will affect the participant's perceived intelligence, satisfaction, and willingness to use*, as seen in our results on the four humor personalities, the aggressive and self-defeating humor personality is the least preferred by participants in regards to our components. Furthermore, the

affiliative and self-enhancing humor personalities, while not having a significant difference on the neutral personality, are still in line with each other regarding our component on likeability. It can be argued that an affiliative or self-enhancing humor personality can be used. This can be further substantiated by participant's comments on the neutral personality: P10 - *"It's professional, but it does not have much of a personality"*, P15 - *"Simple, cutted all the fat away. More focused on completing the task, not much of a personality though"*, P20 - *"It did what it should, but it sounds quite dead"*, and P26 - *"Didn't feel like there was much of a personality, felt more like I talked to a robot"*. Compared to the paper by Ceha et al.[5], it is mentioned that they constructed two personalities (affiliative and self-defeating) with Martins' four humor personality types. This was due to the humor types being described as either conductive (affiliative, self-defeating) or detrimental (aggressive, self-enhancing) in building relationships with others[16], and in their paper, conductive humor personality types were chosen. Compared to our data, we see no apparent connection between the element of conductive or detrimental.

### 5.4 Humor frequency and length

During the experiment, several participants stated that the length of responses presented by the four humor personalities were far too long. This was both regarding the communication successful and the communication breakdowns. These participants also preferred the neutral personality for its minimal but precise responses, saying: P12 - *"Good, really good... she comes with quick responses compared to the last one... I liked that"* and P22 - *"Maybe its responses is a bit long [...] some of the others have very long responses"*, and P18 - *"She should just do what she is told. Furthermore, participants also expressed their nuisance toward the frequency of humor used in the four personalities, commenting: P6 - "I don't feel like that they have to be funny every time, maybe just once in a while can they come with a funny remark"* and P20 - *"It would be nice if I could turn it off [referring to the humor], but also existing to know what it has to say"*.

This outcome is also in line with previous studies where participants noted the overuse of humor and referred to the frequency of jokes[5]. This implies that there may be a balance between a short precise answer and the use of humor.

### 5.5 VUI characteristics manipulation

In the paper 'User recommendations for intelligent personal assistants', Lopatovska highlights possible user recommendations concerning VUIs[12]. One such recommendation would be to change the personality of the smart speaker to suit the user's personality. The participants mentioned this during the experiment: P21 - *"It could be cool if you could like... like choose different personalities"* and P28 - *"I'm more a fan of this personality, but it could be nice if I could choose*

*between 10 variations", and 29 - "[...] I would probably like a combo between the two (neutral and affiliative)".*

## 5.6 Limitations and Future Work

We acknowledge that this paper does have its limitations regarding different elements of the research design. To begin, we would like to acknowledge that our component of service recovery satisfaction does not follow the proper guidelines using a 7-point Likert scale. Instead, we used a 5-point semantic scale. Even so, we have included the component, showing a significant difference between our conditions. In addition, the simulation we used for this study, a Wizard of Oz (WoZ) technique, may not accurately reflect how users would interact with an agent or smart speaker in practical use. This study focused on single turn-taking interactions and not prolonged multi-turn task-oriented interactions. Future research should be conducted in a more realistic setting by developing a VUI with multiple turn-taking actions.

Our scope has been on the specific clarification response (SR) breakdown. Communication breakdowns can occur during different interactions, such as acting on misunderstanding (AoM) and neutral clarification response (NR). It could further be researched whether other repair strategies, as described by Beneteau et al. [2] would have a different effect on perceived user satisfaction, perhaps in conjunction with humorous personalities. It was further described how the VUI informs the users of communication breakdowns through signals. It would be interesting to research whether there are communication breakdown signals which benefits from a humorous personality type as well, and if there could be interplay between communication breakdown signal, repair strategies and humorous personalities.

We have used the HSQ, one of the most established tools for determining humor styles. The questionnaire has been widely used and translated into many different languages and countries, such as in the paper mentioned by Schermer et al. [24] where 28 countries were compared and studied on a large scale (N = 8391), with approximately 300 participants per country. Unfortunately, it has not been used in a study to form a more extensive understanding of the danish population's humor styles. From our results in the HSQ with a small sample size of 30 participants, we cannot conclude whether our data is representative of the general Danish population. As stated in the paper by Lundquist[13], the danish humor is more aligned with aggressive humor. However, compared to the data found in the paper by Schermer et al. [24], our results indicate the mean of our participants to be approximately ranked sixth highest in aggressive humor compared to other countries.

We propose three items for future research and replication of our study.

- (1) **Correct and alternative measurements.** If our study were to be replicated, researchers could look into alternatives in determining humor styles rather than the HSQ. Here we would invite researchers to correct the error of using a 5-item scale in the component 'service recovery satisfaction' and instead use the correct 7-item Likert scale.
- (2) **Further research into humor personalities.** This study has looked at one humor theory, while others exist, such as Meyers humor theories and humor functions[18]. Other humor theories could yield additional findings regarding humorous VUI personalities during communication breakdowns.
- (3) **Longitudinal study.** We believe a longitudinal study could be valuable in researching humorous VUI personalities. In addition, we would like to provide two alternatives for conducting a longitudinal study using a developed VUI humor personality. In order to gain an understanding of the user's preferences over a VUI with pre-constructed humor (A) compared to a VUI that adapts its humor based on continuous usage (B).
  - (A) **User specific humor.** The VUI could have a pre-constructed humor personality aligned with users' humor styles. Knowing the user's humor style prior to the introduction to a VUI, we hypothesize that there would be increased satisfaction with the appropriate humor personality, as mentioned by Lopatovska et al.[12].
  - (B) **Adapt humor over time.** A VUI that adapts its humor personality based on the continued usage from the user. We hypothesis that this prototype would get to know the user more intimately, perhaps describing the user's humor style differently than the HSQ, as it will experience the user's humor first hand, rather than the user's own accounts of humor preference. Additionally, we hypothesize that likeability, perceived intelligence, and service recovery satisfaction would increase over time.

## 6 Conclusion

Maintaining a close connection in human-VUI interactions necessitates appropriate error mitigation and repair strategies. According to our results, participants were more satisfied and perceived the neutral VUI personality as more intelligent than a VUI with an aggressive or self-defeating humor personality. Furthermore, a VUI with an aggressive humor personality may be favored as less likable by the participants than all other VUI personalities. Our findings have implications for developing humor-based error-mitigation strategies for VUI assistants.

## Acknowledgments

To Merritt our third Musketeer or third M&Ms.

## References

- [1] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.
- [2] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300473>
- [3] Calendly. 2022. *Free Online Appointment Scheduling Software - Calendly*. Calendly. Retrieved June 6, 2022 from <https://calendly.com/>
- [4] Cuicui Cao, Yingying Hu, and Haoxuan Xu. 2022. A Mind in Intelligent Personal Assistants: An Empirical Study of Mind-Based Anthropomorphism, Fulfilled Motivations, and Exploratory Usage of Intelligent Personal Assistants.
- [5] Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law. 2021. *Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?* Association for Computing Machinery, New York, NY, USA, 14. <https://doi.org/10.1145/3411764.3445068>
- [6] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? Voice-driven interfaces and young children's communication repair strategies. , 337–348 pages.
- [7] Mikkel Clausen and Mikkel Kyhn. 2021. Alexa, what do you think about Cortana: A Scoping Review of Humorous Interactions with Intelligent Personal Assistants. , 11 pages.
- [8] Xiang Ge, Dan Li, Daisong Guan, Shihui Xu, Yanyan Sun, and Moli Zhou. 2019. Do Smart Speakers Respond to Their Errors Properly? A Study on Human-Computer Dialogue Strategy. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, Cham, 440–455.
- [9] Jose Maria C Ibardaloza, Juan Antonio G Mapua, and Wilson M Tan. 2021. Modifying Voice-User Interfaces for Resiliency and Offline Management of IoT Devices. , 131–137 pages.
- [10] Tonglin Jiang, Hao Li, and Yubo Hou. 2019. Cultural differences in humor perception, usage, and implications. *Frontiers in psychology* 10 (2019), 123.
- [11] Irene Lopatovska. 2020. Classification of humorous interactions with intelligent personal assistants. *Journal of Librarianship and Information Science* 52, 3 (2020), 931–942.
- [12] Irene Lopatovska, Alice Griffin, Kelsey Gallagher, Caitlin Ballingall, Clair Rock, and Mildred Velazquez. 2019. User recommendations for intelligent personal assistants. *Journal of Librarianship and Information Science* 52 (04 2019), 15. <https://doi.org/10.1177/0961000619841107>
- [13] Lita Lundquist. 2014. Danish humor in cross-cultural professional settings: linguistic and social aspects.
- [14] Amama Mahmood, Jeanie W Fung, Isabel Won, and Chien-Ming Huang. 2022. Owning Mistakes Sincerely: Strategies for Mitigating AI Errors. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 578, 11 pages. <https://doi.org/10.1145/3491102.3517565>
- [15] Rod Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality* 37 (02 2003), 48–75. [https://doi.org/10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2)
- [16] Rod A Martin and Thomas Ford. 2018. The psychology of humor: An integrative approach.
- [17] Michael McTear, Zoraida Callejas, and David Griol. 2016. Conversational interfaces: Past and present. , 51–72 pages.
- [18] John C Meyer. 2000. Humor as a double-edged sword: Four functions of humor in communication. *Communication theory* 10, 3 (2000), 310–331.
- [19] Yohan Moon, Ki Joon Kim, and Dong-Hee Shin. 2016. Voices of the internet of things: An exploration of multiple voice effects in smart homes. , 270–278 pages.
- [20] Billie Akwa Moore and Jacqueline Urakami. 2022. The impact of the physical and social embodiment of Voice User Interfaces on User Distraction. , 102784 pages.
- [21] Andreea Niculescu, Betsy van Dijk, Anton Nijholt, Haizhou Li, and Swee Lan See. 2013. Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics* 5, 2 (2013), 171–191.
- [22] Stefan Olafsson, Teresa K O'Leary, and Timothy W Bickmore. 2020. Motivating health behavior change with humorous virtual agents. , 8 pages.
- [23] John Pallant. 2007. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15* (3rd ed.). Open University Press, USA.
- [24] Julie Aitken Schermer, Radosław Rogoza, Maria Magdalena Kwiatkowska, Christopher Marcin Kowalski, Sibebe Aquino, Rahman Ardi, Henrietta Bolló, Marija Branković, Razieh Chegeni, Jan Crusius, et al. 2019. Humor styles across 28 countries. , 16 pages.
- [25] Alex Sciuto, Armita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 857–868. <https://doi.org/10.1145/3196709.3196772>
- [26] Chen Shani, Alexander Libov, Sofia Tolmach, Liane Lewin-Eytan, Yoelle Maarek, and Dafna Shahaf. 2022. "Alexa, Do You Want to Build a Snowman?" Characterizing Playful Requests to Conversational Agents. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 423, 7 pages. <https://doi.org/10.1145/3491101.3519870>
- [27] Paul J Silvia and Rebekah M Rodriguez. 2020. Time to renovate the Humor Styles Questionnaire? An item response theory analysis of the HSQ. *Behavioral Sciences* 10, 11 (2020), 173.
- [28] Julian Striegl, David Gollasch, Claudia Loitsch, and Gerhard Weber. 2021. Designing VUIs for Social Assistance Robots for People with Dementia. , 145–155 pages.
- [29] Topvoiceapps. 2022. *SSML WYSIWYG EDITOR AND TESTER*. Topvoiceapps. Retrieved June 6, 2022 from <https://topvoiceapps.com/ssml>
- [30] Linghan Zhang, Sheng Tan, Zi Wang, Yili Ren, Zhi Wang, and Jie Yang. 2020. Viblive: A continuous liveness detection for secure voice user interface in iot environment. , 884–896 pages.