

My virtual assistant sounds like my neighbor: how human-like is human-like enough?

An experimental study to examine the effects of text, standard language, and accent used by a virtual assistant between regional and non-regional people on anthropomorphism and appropriateness

Rick Oosterholt
Snr 2049244

Master's thesis
Communication and Information Sciences
Specialization Business Communication and Digital Media

School of Humanities and Digital Sciences
Tilburg University, Tilburg

Supervisor: Dr. T. Lentz
Second reader: Dr. M. Mos

June 2021

Abstract

This study investigated whether (1) speech use (instead of text) affects perceived anthropomorphism, (2) whether perceived appropriateness and anthropomorphism differ between a virtual assistant using a regional accent and a non-regional accent, and (3) whether this difference is influenced by (non-)correspondence of accents between the participant and the VA. By using a 3 (text virtual assistant vs. standard Dutch virtual assistant vs. regional accent virtual assistant) X 2 (regional participants vs. non-regional participants) between-subjects design, participants ($N = 139$) interacted with a virtual assistant from a vacation park to find the most suitable residence. As there were no significant findings, there is no evidence that (1) speech is more anthropomorphic than text and that (2) regional participants consider a corresponding regional accent more appropriate compared to non-regional participants. Furthermore, there was no evidence that (3) non-regional participants consider regional accent more anthropomorphic compared to regional participants. For follow-up research, we suggest experimenting in a controlled environment and with a better-manipulated accent.

Keywords: chatbot, virtual assistant, anthropomorphic speech cue, communication accommodation theory, accent appropriateness, accent correspondence

Introduction

Interacting with a chatbot has become common in recent years, and sometimes it can be unclear whether someone is interacting with a real person or a robot. A chatbot is defined as a computer program developed to hold conversations with users mimicking natural language (Shawar & Atwell, 2007). Araujo (2018) describes a chatbot as a form of a ‘disembodied conversational agent’ designed to perform tasks as simple as sending theater tickets or as complex as giving subscription, travel, or lifestyle advice. The definition of a chatbot is often related to the definition of a virtual assistant. A chatbot becomes a virtual assistant when it performs tasks (Telang et al., 2018). Thus, these definitions suggest that chatbots as virtual assistants are designed to facilitate and simplify everyday tasks. In this study, the term 'virtual assistant' (VA) is used.

Not only do virtual assistants simplify the everyday life of individuals, but they are also efficient and inexpensive within a company's customer service department and streamline and lower the workload within a company (Følstad et al., 2018). This is an explanation for the increasing number of companies implementing virtual assistants (Elsner, 2019). However, VA technology frequently associates with consumer skepticism, and people show a preference for engaging with genuine humans (Elsner, 2019). Machine-like VAs that lack any 'human resemblance' (anthropomorphism; Araujo, 2018) have a detrimental effect on the trust an individual has in the VA in question (Følstad et al., 2018). Since people treat VAs as genuine humans, they apply the same social rules to them (CASA; Nass, Steuer & Tauber, 1994; Reeves & Nass, 1996), causing possible detrimental effects on trust when they lack anthropomorphism. Since individuals show this preference for having a conversation with genuine humans, making the VA more human by letting the VA speak could undermine this phenomenon. Therefore, the following research question is proposed.

RQ1: "Does the use of speech (instead of text) affect the level of perceived anthropomorphism of VAs?"

Although there is little research on the effects of the use of speech by website VAs, Barcelos et al. (2018) investigated the impact of tone of voice adaption in an online (social media) context. Their research is based on the *communication accommodation theory* (CAT; Giles et al., 1991). This theory states that while people interact, they individually tend to accommodate each other's verbal and non-verbal behavior. One can converge by adopting the recipient's behavior or diverge by purposely deploying a communication style different from the other party's communication style. This can be considered adapting the dominant language style (e.g., formal vs. informal) and the use of accents (e.g., regional accent vs. standard language).

Accents are speech characteristics that reveal to a listener from which region a speaker originated (Torre & Maguer, 2020). Firstly, the question is whether a human characteristic like a 'regional accent' is perceived as more 'human-like' than standard language since the regional accent is more sophisticated than standard language. Secondly, as a VA is a standardized installed program that does not know where the interaction person is originating from (nor can it accommodate its communication and accent accordingly), the question therefore is whether people with a non-corresponding accent perceive the different accent use of the VA as equally appropriate as people with a corresponding accent.

RQ2a: " Does the use of a regional accent affect the perceived anthropomorphism of a VA?"

RQ2b: "Does the use of a regional accent affect the perceived appropriateness of a VA?"

People who speak with a regional accent can cope well with the dominant standard language. It is the language spoken on national radio and TV and therefore mirrors the norm (Finegan, 2007). Moreover, people listening to people with a corresponding regional accent perceive the accent as less noticeable than people from an outside region (with a non-corresponding accent) who listen to the people in question (Scales et al., 2006). This assumes that someone from the province of North-Brabant who listens to someone from the same region considers the regional accent from the person in question to be less noticeable than a person from the Randstad (i.e., the West) who listens to someone from the province of North-Brabant. With regards to this phenomenon, this study investigates whether someone outside the accent region and who thus perceives it as more noticeable considers a VA using a specific accent as equally anthropomorphic and appropriate as people with a corresponding accent. Therefore, the third research question is formulated:

RQ3a: "Is there a difference in perceived anthropomorphism between people with corresponding regional accents interacting with a virtual assistant that uses a similar regional accent and people with non-corresponding regional accents¹?"

RQ3b: "Is there a difference in perceived appropriateness between people with corresponding regional accents interacting with a virtual assistant that uses a similar regional accent and people with non-corresponding regional accents?"

Theoretical Framework

Humanizing a chatbot and social presence

Araujo (2018) defines 'anthropomorphism' as 'humanizing' objects by giving it human-like cues. He describes that a VA has anthropomorphism cues when it introduces itself, when it uses informal language, and says 'hello' or 'goodbye' during the interaction. His results indicate that the use of anthropomorphic language or a human name increases the perception of the VA as human-like. Furthermore, he found that emotional connection with the organization was higher with an anthropomorphic VA. Thus, giving VA anthropomorphism cues seems beneficial for brands.

An anthropomorphic presentation increases the feeling of social presence among individuals interacting in a conversation with such VAs (Go & Sundar, 2012). Social presence is described as the level of the salience of the person in the chat (Short et al., 1976), and also the sense of "being with someone else in a mediated environment" (Biocca, Harms, & Burgoon, 2003, p.14). Kim and Sundar (2012) suggest that anthropomorphic cues such as a human avatar increase the salience of the 'other person' because the presence of this avatar enclosed to the VA can suggest the existence of the 'other person' in the chat. Short et al., (1976) suggest that media that score high on social presence are considered more friendly, more sensitive, warmer, more personal, having more favorable attitudes, and constructing more trusting relationships with a brand.

Liebrecht and van der Weegen (2019) used linguistic elements to make a distinction between the human-like (i.e., anthropomorphic) and machine-like VA. The human-like VA's messages included various aspects of the Conversational Human Voice (i.e., CHV; Kelleher, 2009; Kelleher & Miller, 2006; Van Noort et al., 2012). For starters, they used informal language by mimicking sound in written text and used emoticons (e.g., 'yaaay ☺'). Next, they

used personalization of the message by doing a personal salute to the consumer (e.g., 'Hi Rick'). Finally, they used invitational rhetoric by showing sympathy and empathy (e.g., 'nice, have fun!'). The human-like VA included a personal name ('Booky') and a unique avatar. The machine-like VA, however, did not include elements of CHV in its messages, included an impersonal name ('Bookbot'), and the avatar included the brand's logo. The scholars showed that interacting with a human-like VA results in a higher brand attitude and that this effect was mediated by perceived social presence.

Virtual voice assistants

Additionally to VAs consisting of written-out text interactions, there are also VAs where speech is implemented and where both the user and the VA can respond through speech. Examples are Google Home, Apple's Siri, and Samsung's Bixby. Analysts predict that there will be 4.2 billion speech VA's being used worldwide (Young, 2018; Statista, 2021). The exact forecasts suggest that by 2024, the number of these speech VA's will reach 8.4 billion units. In the education field, the virtual voice assistants are successfully making their entry, helping students with complex course-related issues, resulting in a decrease in teachers' workload (Harvey et al., 2016). Accordingly, these assistants are promising for the help they can provide through their using a voice.

This study contributes to the existing scientific literature as there is little research on adding human speech to a VA on a corporate website. Previous research has investigated the effects of adding human characteristics in written-out texts used by the VA (Araujo, 2018; Go & Sundar, 2012; Liebrecht & van der Weegen, 2019). Since a voice is also a human-like cue, which brings a salient addition to the set of common anthropomorphism cues, a speech VA could have a higher level of anthropomorphism than a text-only VA.

Adding a regional accent to the Virtual Assistant

Scott and colleagues (2020) found significant evidence of anthropomorphism, nostalgia, and the emotional connection with regard to text-to-speech voices. Furthermore, human-sounding synthetic voices provoked significant negative feedback. Similar results in the study of Kühne et al. (2020) indicate that synthesized voices score lower on human likeness than human voices and are also considered less trustworthy.

Human voices thus score higher on anthropomorphism and trustworthiness than synthesized voices. However, in addition to that, every genuine human voice has a regional accent. A standard language that differs in speech characteristics is called an 'accent' (Grondelaers & Van Hout, 2010). As described by the dissertation of Van der Harst (2011), vowels of the Randstad region (i.e. 'the West') have certain formants that differ from those of the South region. Analyzing differences in such formants is possible by using the speech analysis tool 'Praat' (Boersma & Weenink, 2007). The tool shows F1 & and F2 data from sound recordings. In phonetics, F1 relates to tongue height, and F2 relates to the forward/backward movement of the tongue. F1 decreases as tongue height increases, so the higher the tongue is placed in the mouth, the lower F1. For instance, for the vowel /a/, the F1 tends to be lower for standard Dutch language (i.e., N-R) compared to a regional accent from the south (i.e., N-S) (see Appendix E) (Van der Harst, 2011). Thus, people who speak standard Dutch position their tongue higher in their mouth when saying Dutch words like 'hallo' and 'kan' than people speaking with a south regional accent.

The Netherlands contains four main accent areas: The West (i.e., the central zone), the North, the South, and a transitional area between the North and the South (Driessen, 2006). Biases emerge in all of these regions, especially in terms of the social cues that speakers of these accents give (Smakman, 2006). For instance, people from the western region, which

reflects the standard Dutch language, are less tolerant to other accents. Someone with an accent from the south of the Netherlands is considered friendly yet also as someone with a lack of refinement. The accent from the West (i.e., the Randstad) is understood to be the most 'prestigious' variant, seen as best suited in formal environments, and the pronunciation from this region is therefore often considered standard Dutch (Grondelaers & Van Hout, 2010; Grondelaers, Van Hout, & Van Gent, 2018).

Dahlbäck and colleagues (2001) investigated the effects of the correspondence of accents between native and non-native speakers of English. They found that when participants match with accents similar to their own, they (1) reveal socially undesirable behavior, (2) find the interviewer more socially rich, and (3) consider the interviewer more enjoyable. These findings are in line with CAT and this, therefore, suggests that accents are appropriate when the counterpart matches the accent. Furthermore, Torre and Maguer (2020) state that the in-groupness that results from individuals hearing each other's similar accent is vital since individuals tend to rather communicate with others who belong to the same group.

Although literature suggests that people would rather communicate with people with corresponding accents (than with people with different accents) (Torre & Maguer, 2020), we expect that people with a non-corresponding accent rate a higher level of anthropomorphism for a VA that uses a regional accent, as for them, the accent is more salient than for people with corresponding accents (Scales et al., 2006). People with corresponding accents are both familiar with the accent and with the standard language. However, for people with non-corresponding accents, it may be more striking that the VA has such a human characteristic.

Virtual assistants and the communication accommodation theory

Barcelos et al. (2018) suggest that the intention of the person reaching out to a company is the moderator for deciding the most suitable response. For instance, if someone has a utilitarian attitude towards the brand (i.e., the person only desires service), it is more beneficial for the brand to stay formal and use a corporate tone of voice. However, if someone has a hedonic attitude towards the brand (i.e., the person wants to build a relationship with the brand), it is more beneficial to accommodate the user's type of language by mimicking a more human voice. Hence, to the present study, it is crucial that participants have a hedonic attitude towards the brand to succeed in the use of human voice (Barcelos et al., 2018). In addition, an unfamiliar company must be developed to ensure that each participant has the same intention when interacting with the VA, without the possibility of bias.

Although Barcelos et al. (2018) focused on language style adaption, mentioning CAT as their explanation for the positive results, the question is whether regional accent use by a VA will result in positive results as well. Whereas standard language can be considered 'formal language,' regional accents can be considered 'informal language.' Again, a VA is a standardized installed program that does not know where the interaction person is originating from (nor can it align its accent accordingly). The question, therefore, is whether people with a non-corresponding accent perceive the different accent use of the VA as equally appropriate as people with a corresponding accent.

People treat VAs as genuine humans since they are apply the same social rules to VAs (CASA; Nass et al., 1994; Reeves & Nass, 1996). Since the VA will be treated as a genuine human but cannot comply with the CAT, the assumption is that the VA that uses a regional accent will be less polite and therefore less appropriate by people with a non-corresponding

regional accent. This phenomenon has been shown in Gretry et al.'s (2017) study by manipulating the communication style. Gretry et al.'s (2017) argumentation is grounded in Role Theory (Sarbin & Allen, 1968). The Role Theory entails that the evaluation and success of interactions depend on the appropriateness of the interaction partner's behavior regarding their social roles. If the interaction partner (i.e., the VA) is a stranger, a formal communication style is considered appropriate. Hence, the assumption is that people with a non-corresponding regional accent prefer a formal communication style and consider that appropriate. As mentioned previously, people with a corresponding regional accent consider the counterparty as more 'sociable' (Dahlbäck et al., 2001), and therefore probably also more 'appropriate.'

Again, people from the western region are less tolerant towards other accents (Smakman, 2006). This assumes that when someone with a non-corresponding regional accent interacts with someone with a regional accent that sticks to the regional accent the complete conversation, the person with the non-corresponding regional accent will perceive that person as a stranger and also as unfriendly and inappropriate (Gretry et al., 2017). Thus, using a regional accent for people with non-corresponding regional accents could be considered less appropriate.

In sum, this study expects that human speech will be more anthropomorphic above all else than text. However, this study then expects that the regional accent is the superlative of the two types of human speech, explaining an interaction effect between people with corresponding and non-corresponding regional accents. This study expects that regional accents are more appropriate for people from the corresponding region (Torre & Maguer, 2020). This study also expects that the regional accent is considered more anthropomorphic for non-regional people.

Hypotheses

H1: "A speech human-like virtual assistant has a higher level of anthropomorphism than a text-only human-like virtual assistant."

H2: "There is a larger positive difference for anthropomorphism between SVA and RVA for non-regional participants than for regional participants" (one-tailed interaction)

H3: "There is a larger negative difference for appropriateness between SVA and RVA for non-regional participants than for regional participants" (one-tailed interaction)

Method

Overview and introduction of the study

The research questions and hypotheses were tested with a 3 (*text virtual assistant* vs. *standard language speech virtual assistant* vs. *regional accent speech virtual assistant*) X 2 (*participant living in accent area* vs. *not living in accent area*) between-subjects design. The virtual assistant was developed to interact with the participant using informal written language or speech, had a human avatar and name (Eline), and the participant was requested to choose one of the three predetermined answer options that the VA showed after she finished her message. As the VA concerned a predetermined script, each participant finished with the same final offer from the VA (see Appendix C). Participants ($N = 142$) were randomly assigned to one of three different groups: (1) *text virtual assistant* (TVA), (2) *standard language speech virtual assistant* (SVA), and (3) *regional accent speech virtual assistant* (RVA).

Materials

Company. Concerning Barcelos et al. (2018), the company was fictional and, therefore, an unknown vacation park. A vacation park is considered 'hedonic' (rather than utilitarian), and therefore a hedonic attitude towards the fictional brand was established. The park was named 'Polderparks' and was located somewhere in the Netherlands. Participants did not know the location, nor was a location given. All participants were introduced to a random type of VA that helped them determine the most suitable residence for them.

Virtual assistants. The *anthropomorphic text virtual assistant* (TVA) was developed to interact with the participant using informal written language, with anthropomorphism cues, and the participant was requested to choose one of the three predetermined answer options

that the VA showed after he finished his message. The name 'Eline' was chosen since it is a commonly known and used female name across the Netherlands and therefore not fixed to a specific area of the country, making it more representative for each participant (Meertens Instituut, n.d.). Furthermore, the TVA contained a human avatar (Go & Sundar, 2019), used personalized message (e.g., using 'we' instead of the name of the brand: 'we from Polderparks...') (Liebrecht & van der Weegen, 2019), and used invitational rhetoric (e.g., showing sympathy and empathy: 'nice choice!') (Van Noort et al., 2012), to make the VA more anthropomorphic.

The *standard language speech virtual assistant* (SVA) was identical to the TVA, except that instead of text, it pronounced messages by using speech. This condition used the standard language speech. The *regional accent speech virtual assistant* (RVA) was identical to the *standard language speech robot* but used an accent from the region of North-Brabant instead of standard language speech while uttering the messages. A female speech therapist who is proficient in speaking with accents voiced both speech VAs.

The VAs only differed in text or type of speech. Besides, the two speech conditions did not include written-out responses like TVA. Finally, the female avatar and the website environment were identical in each condition (see Appendix D). The interaction flow was based on the study of Go and Sundar (2018). The VA waited two seconds before responding to the chosen input from the participant in order to mimic a more human-like response (Araujo, 2018). See Appendix 7 for more details about the development of the virtual assistant.

Pretest. A pretest was conducted to check the quality of the accents. Participants ($N = 6$) were provided an informed consent form to approve before continuing their participation in the pretest. Participants were exposed to a beta version of the audio that is used in the main

study. One type contained standard Dutch language (SVA), and the other type contained an accent from North-Brabant (RVA). Participants were subsequently asked to classify which region the voice messages originated. This question contained six answer options, consisting of all the Dutch regions. Participants could choose between *North, East, Middle, North-West, South-West, and South-East*. The "South-East" answer option was accepted as "correct" for the North-Brabant accent version (RVA). "North-West," "South-West," and "Middle" was accepted for the standard Dutch version (SVA). Participants were shown a country map of The Netherlands indicating the borders of the regions (see Appendix A). Three participants originated from North-Brabant (i.e., regional participants), and three participants did not (i.e., non-regional participants) (see Table 1). Five out of six native speakers of Dutch classified the accent from North-Brabant correctly as an accent from the "Southeast." Participants from the pretest were excluded from the main study to preserve the validity and reliability of the main study.

Praat. In addition to deploying the small survey with both types of speech messages, the differences in vowels were also analyzed. This was done by using the sound-analyzing program Praat (Boersma & Weenink, 2007). The data from Van der Harst (2011) was used to analyze whether his findings would comply with the findings of this pretest. As the accent from the Randstad is considered 'standard Dutch,' (Grondelaers & Van Hout, 2010; Grondelaers, Van Hout, & Van Gent, 2018), the N-R (i.e., Randstad) data from Van der Harst (2011) was interpreted and utilized as 'standard Dutch' in order to find similar differences. Furthermore, since Van der Harst (2011) did not investigate this for the Brabant accent particularly, the N-S data was interpreted and utilized for the Brabant accent as this accent belongs to the south of the Netherlands. For this analysis, the vowels /a/, /u/, and /e/ were analyzed as Van der Harst (2011) suggested that these vowels differ significantly between N-

R and N-S (see Appendix E). As shown in Figure 4, there are constantly similar differences in the chosen vowels between the two types of speech.

Contrary to the findings of Van der Harst (2011) (see Appendix E), the present analysis found that for /a/, the F1 tended to be higher for standard Dutch type (SD) compared to the accent condition of Brabant (North-Brabant). Additionally, there were no notable differences for /u/ and /ɛ/ found between the two types of speech. Nevertheless, the manipulation was considered effective as the results of our pretest survey have shown that participants distinguished and categorized the two types of speech correctly.

Procedure

Participants. Participants ($N = 148$) from the Netherlands were recruited through convenience sampling because social distancing measures were in effect. Participants had to be at least 18 years old. Nine participants were excluded due to inconsistencies in their answers. For instance, these participants gave the wrong answer to the control question and were inconsistent with their origin and their province identification. Participants ($N = 139$) identified themselves as someone from the province of Noord-Brabant (i.e. ‘regional’) ($N = 81$) or as someone from another province in the Netherlands (i.e. non-regional) ($N = 58$). They were randomly assigned to one of the three conditions, resulting in a normally distributed dataset. Their mean age was 30 years ($SD = 12.87$). Participants' demographics (e.g., “*Which province have you lived in the longest?*”) were asked at the end of the survey to ensure that they did not become aware of the purpose of the study and, therefore, might react differently to our main questions. The demographic questions helped answer the research questions and hypotheses because they provided insight into where each participant comes from (see Appendix B).

Main study. Participants were provided an informed consent to approve before continuing with their participation in the main study and were exposed to an attention check. Participants were instructed to interact with a VA as if they were looking for a vacation residence on a vacation park and asked the VA to find out what type of residence would be most suitable for the participant (see Appendix C). Each proposal that the VA suggested came with three clickable answer options from which the participant could choose, resulting in a Q&A format (Go & Sundar, 2019). The participant was able to answer the VA by clicking on one of the buttons. The participant was to interact with the VA for no more than three minutes and continue to the next question irrespective of whether the wrong button was clicked. The three-minute-interaction time was based on the total length of all the audio files combined, which lasted 90 seconds in total. This information was recorded as a control variable in the survey program Qualtrics. As participants who took more than three minutes might have been distracted, their data was discarded. Finally, participants subsequently answered questions considering their perceptions about the VA (see Appendix B) and finalized the survey by answering a set of demographic questions.

Measurements

Anthropomorphism. The variable 'anthropomorphism' was operationalized by asking eight questions about human-likeness (Powers & Kiesler, 2006; Kim & Sunder, 2012; Araujo, 2018). The measurement constructs used by Araujo (2018) were duplicated, *mindful* anthropomorphism and *mindless* anthropomorphism were measured, and both were generalized under the guise of "anthropomorphism." The items of Powers and Kiesler (2006) were used, asking participants about their perception with regards to the VA and its service when it comes to being (1) *human or machine-like*, (2) *natural or unnatural*, and (3) *lifelike or artificial*, along with 7-point semantic differential scales. The items of Kim and Sundar

(2012) were used to analyze to what extent the VA and its service were considered (4) *likable*, (5) *sociable*, (6) *friendly*, (7) *personal*, and (8) *unsociable*. The eighth question was a reversed repetition of the fifth question and was used as a control question to analyze whether the participant was still answering the questions consistently. Participants could indicate this using the 7-point Likert scale (e.g., 1 = strongly unfriendly, 7 = strongly friendly) (see Appendix B).

Appropriateness. Since there is no literature regarding measuring appropriateness in an online VA setting, a new scale was developed. Appropriateness of the tone of voice was measured by six items on the 7-point Likert scale (e.g., 1 = strongly inappropriate, 7 = strongly appropriate), consisting of (1) *appropriateness*, (2) *politeness*, (3) *selfishness*, (4) *arrogance*, (5) *rigid*, and (6) *impoliteness*. These six terms used are associated with studies of communication accommodation theory (Giles et al., 1991). The sixth question was a reversed repetition of the second question and was used as a control question to analyze whether the participant was still answering the questions consistently.

Results

A reliability analysis for ‘anthropomorphism’ (seven items; $\alpha = .878$) and ‘appropriateness’ (five items; $\alpha = .805$) showed that both scales were reliable. To ensure that the manipulations for SVA and RVA succeed sufficiently, it was analyzed whether there is an association between the speech conditions the participants were exposed to (i.e., SVA vs. RVA) and identifying the accent correctly (e.g., ‘The accent I heard was most probably from Noord-Brabant’). Therefore, an χ^2 test of association was performed ($\chi^2 (1) = 24.27, p <.001$). As the TVA was left out of this analysis, 99 participants were analyzed. 37 participants who got exposed to SVA identified the accent as coming from a province outside Brabant, whereas twelve did not. 37 participants who got exposed to RVA identified the accent as an accent from Brabant, whereas twelve did not². The analysis showed that, on average, 74,8% identified the exposed accent exactly correctly. So, participants mostly recognize the accent they heard.

The data of two types of speech conditions (SVA and RVA) were combined to analyze H1. To test whether a speech virtual assistant has a higher level of anthropomorphism than a text-only virtual assistant, an independent sample t -test was performed (IV = type of VA [text, speech], DV = anthropomorphism). The data were normally distributed. Homogeneity can be assumed as Levene's test was not significant $F(1, 137) = 2.347, p = .128$). On average, a speech virtual assistant ($M = 4.69, SD = 1.20$) has a higher level of anthropomorphism than a text-only virtual assistant ($M = 4.48, SD = .97$), but the difference between the speech virtual assistant and the text-only virtual assistant was not significant ($M_{dif} = .21, t(137) = .986, p = .326$). The difference represents an almost irrelevant-sized effect $d = 0.19$. Thus, the hypothesis that a speech virtual assistant has a higher level of anthropomorphism than a text-only virtual assistant was not confirmed by this test.

Since the last two hypotheses concern two different dependent variables (i.e., anthropomorphism and appropriateness), it was analyzed whether both variables correlate with each other. Pearson's correlation test showed a significant positive correlation between anthropomorphism and appropriateness of the speech VA, $r = .659$, $p = <.001$ (see Table 2). Since the dependent variables appear to correlate, there is a possibility that there is a mediation effect. However, first, they were analyzed individually. Any significant results were interpreted with caution, as this study was given two not-independent chances to find a significant result.

To test whether there is a larger positive difference for anthropomorphism between SVA and RVA for non-regional participants than for regional participants, a factorial ANOVA was performed (see Table 3 for the means per cell). The data were normally distributed. The assumption of homogeneity of variances was not met for anthropomorphism. Levene's test was significant: $F(3, 95) = 2.88$, $p = .040$. The factorial ANOVA is fairly robust against violations of these assumptions, but the outcomes may not be completely reliable, and therefore the corrected F-value from the Welch F-test and Brown-Forsythe F-test was used, which were the same ($F(1, 86.36) = .866$, $p = .355$).

The ANOVA showed no significant main effect of participant identity (i.e. regional or not), $F(1, 95) = .415$, $p = .521$, $\eta^2 = <.001$. Furthermore, the ANOVA neither showed a significant main effect of VA type (i.e. exposed to RVA or SVA condition), $F(1, 95) = 1.96$, $p = .165$, $\eta^2 = <.001$. Finally, there was no significant interaction effect, $F(1, 95) = .146$, $p = .352$ (one-tailed), $\eta^2 = <.001$. Figure 2 and Figure 3 show the levels of anthropomorphism per condition.

To test whether there is a larger negative difference for appropriateness between SVA and RVA for non-regional participants than for regional participants, again a factorial ANOVA was performed (see Table 4 for the means per cell). The data for appropriateness was not normally distributed. The appropriateness scale for non-Brabanders was left-skewed (z -score skewness = -2,1). Therefore, bootstrapping was used, and the results need to be interpreted carefully. The assumption of homogeneity of variances was met. Levene's test was not significant: $F(3, 95) = 1.307, p = .277$.

The ANOVA showed no significant main effect of participant identity, $F(1, 95) = .466, p = .496, \eta^2 = <.001$ (95% CI [-.285, .584]). Furthermore, the ANOVA neither showed a significant main effect of VA type, $F(1, 95) = .815, p = .369, \eta^2 = <.001$ (95% CI [-.237, .632]). Finally, there was no significant interaction effect, $F(1, 95) = .031, p = .43$ (one-tailed), $\eta^2 = .000$. Figure 4 and Figure 5 show the levels of appropriateness per condition. From a mediation analysis was refrained, as there was no significant effect on either of the two.

Discussion

This study aimed to analyze whether speech is considered more anthropomorphic than text in a VA and whether a regional accent would influence anthropomorphism and appropriateness among regional and non-regional people. This study did not indicate that a speech VA is considered more anthropomorphic than a text VA. There was no evidence that regional participants consider the RVA as more appropriate compared to non-regional participants. The explanation that accent accommodation (CAT; Giles et al., 1991) results in possible feelings of in-groupness (Torre and Maguer, 2020) and is therefore considered more appropriate is not supported by the data of the present study. Furthermore, there was no evidence that non-regional participants consider the RVA as more anthropomorphic compared to regional participants. The results indicate that it was not significantly striking for non-regional people to hear a different accent, which is not in line with the findings of Scales et al. (2006).

Dahlbäck et al. (2011) suggest that when participants match with accents similar to their own, they (1) reveal socially undesirable behavior, (2) find the interviewer more socially rich, and (3) consider the interviewer more enjoyable. Their findings, however, did not seem to correspond with the results, as there was no significantly larger negative difference for appropriateness between SVA and RVA for non-regional participants than for regional participants. An explanation for this outcome may be because, in the study of Dahlbäck et al. (2011), participants spoke with the agent via voice instead of clicking on predetermined, written buttons. It is likely that only when speaking to each other evolves in feelings appropriateness.

As there are no findings that text VAs are significantly less anthropomorphic, a VA that uses merely text with anthropomorphic cues may be considered anthropomorphic enough for people. People are probably aware that they are interacting with a robot that uses pre-

recorded audio or text, and therefore they could see through it. Also, as this study was done between subjects, participants would not have known any better as there was no other VA type for them to compare.

Implications

This study has addressed the need for further theory development for scientists in the field of webcare. The findings of our study contribute to what yet has little been investigated. Furthermore, they may help scientists in the field understand how (not) to develop a study concerning VAs and the use of an accent.

This study has practical implications for practitioners in the field of webcare and virtual assistance domain by showing how organizations, in this case, the vacation parc industry, could apply specific types of VAs to their business. However, follow-up research before the concepts are put into practice is recommended, as this is one of the first studies regarding the use of virtual assistants with a regional accent.

Limitations and future study

This study has several limitations. For starters, due to Covid-19 restrictions, the study was conducted in an online environment, and participants could participate from wherever they were. Furthermore, participants were gathered by snowball sampling, and therefore the representativeness is not guaranteed. Each participant completed the survey in their preferred environment, which may have increased noise in the experiment. In order to simulate the experience of successful interaction and control for content effects and thus determine the effects of anthropomorphism and appropriateness, this study used a script in which participants were asked to follow a particular chat procedure. Although the use of the script allowed the study to control for unwanted effects that may occur in the online survey, it also introduced low ecological validity to the study.

Next, the theme and setting could have been important factors for our results. In this study, the setting of our experiment was a vacation park. While this decision has been made carefully (Barcelos et al., 2018), the question is whether a utilitarian company environment (e.g., an insurance company or governmental related) would yield similar results, as these types of companies are not considered 'hedonic.' Future research comparing VAs in a hedonic and utilitarian business would be interesting.

Third, this study was given two chances to find something because of the correlation between the two dependent variables. There was less chance of finding something separately since they were correlated. Perhaps the same construct was measured twice, and therefore H2 and H3 cannot be tested independently. This conflict may be due to the experimental setup. Perhaps there is a difference between anthropomorphism and appropriateness, but that does not seem to manifest itself here. Alternatively, it could be that anthropomorphism and appropriateness are just strongly related in this context with a VA. In contrast, if there would have been an effect, it is unknown whether anthropomorphism causes appropriateness or the other way around. However, it could then have been analyzed with a mediation analysis.

Fourth, gender might have played a role in the results. This study used a female speech therapist who provided the voice of the VA in both speech conditions. However, people treat computers with female voices differently than "male" ones (Reeves & Nass, 1996). Follow-up research adding a male condition would be interesting.

Finally, the χ^2 test showed that the accent was probably not sufficiently manipulated that 100% of the participants could recognize the correct type of accent. On average, 25% had given an incorrect answer to the question of which region the VA's voice would originate. Although a pretest was conducted using a small survey and using the speech-analyze program Praat (Boersma & Weenink, 2007), this pretesting proved insufficient. Furthermore, contrary to the findings of Van der Harst (2011), the Praat analysis showed that

for the vowel /a/, the F1 tended to be higher for standard Dutch type (SVA) compared to the accent condition of Brabant (RVA). Therefore, for follow-up research, this study recommends that more focus should be placed on this component by, for example, analyzing and comparing more data points in Praat between two accents and measuring any significance between the two.

Conclusion

In conclusion, the study contributes to the researchers' knowledge that there is still work that needs to be done to find significant results regarding VAs and the use of speech and an accent. As the hypotheses were not significant and therefore cannot be confirmed, it can be presumed that merely text with anthropomorphic cues (Araujo, 2018; Kim & Sundar, 2012; Liebrecht & van der Weegen, 2019) is 'anthropomorphic enough' and that there is possibly a ceiling for perceived anthropomorphism.

This study investigated whether adding speech and an accent to a VA has different effects for non-regional and regional people, but that is not likely. The challenge is that there are various types of users and different aims, and there are multiple different situations and many kinds of preferences. The effect of modest regional accents is at best negligible. Researchers looking into this should make sure that their accent is correctly identified. We suggest running this study differently by doing the experiment in a controlled test environment, by manipulating the accent better, developing scales that do not correlate, and taking our other future research advice as a guideline.

References

- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189.
doi:10.1016/j.chb.2018.03.051
- Barcelos, R. H., Dantas, D. C., & Sénécal, S. (2018). Watch Your Tone: How a Brand's Tone of Voice on Social Media Influences Consumer Responses. *Journal of Interactive Marketing*, 41, 60–80. doi:10.1016/j.intmar.2017.10.001
- Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, 12(5), 456–480. doi:10.1162/105474603322761270
- Boersma, P. and D. Weenink (2007). Praat: doing phonetics by computer (Version 4.5.11), retrieved Januari 2007, from <http://www.praat.org/>.
- Dahlbäck, N., Swamy, S., Nass, C., Arvidsson, F., & Skågeby, J. (2001). Spoken interaction with computers in a native or non-native language-same or different. *Proceedings of INTERACT*, 294-301.
- Driessen, G. (2006). Ontwikkelingen in het gebruik van streektalen en dialecten in de periode 1995-2003. *Toegepaste Taalwetenschap in Artikelen*, 75(1), 103-113.
- Elsner, N. (2019, November 27). *KAYAK Mobile Travel Report - KAYAK UK*. KAYAK Blog UK. <https://www.kayak.co.uk/news/mobile-travel-report-2017/>
- Finegan, E. (2007). Language: its structure and use. Belmont: Thomson Wadsworth.

- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. *Internet Science*, 194–208. doi:10.1007/978-3-030-01437-7_16
- Giles, H., & Baker, S. C. (2008). Communication Accommodation Theory. *The International Encyclopedia of Communication*, 1–4. doi:10.1002/9781405186407.wbiecc067
- Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304–316. doi:10.1016/j.chb.2019.01.020
- Grondelaers, S. & Van Hout, R. (2010). Is Standard Dutch with a Regional Accent Standard or Not? Evidence From Native Speaker's Attitudes. *Language Variation and Change* 22(2), 221-239. doi:10.1017/S0954394510000086
- Grondelaers, S., Van Hout, R., Van Gent, P. (2018). Re-evaluating the prestige of regional accents in Netherlandic Standard Dutch: The role of accent strength and speaker gender. *Journal of Language and Social Psychology* 38(2), 1-22. doi:10.1177/0261927X18810730
- Harvey, P. H., Currie, E., Daryanani, P., & Augusto, J. C. (2016). Enhancing Student Support with a Virtual Assistant. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 101–109. doi:10.1007/978-3-319-28883-3_13
- Kelleher, T. (2009). Conversational Voice, Communicated Commitment, and Public Relations Outcomes in Interactive Online Communication. *Journal of Communication*, 59(1), 172–188. doi:10.1111/j.1460-2466.2008.01410.x

- Kelleher, T., & Miller, B. M. (2006). Organizational Blogs and the Human Voice: Relational Strategies and Relational Outcomes. *Journal of Computer-Mediated Communication*, 11(2), 395–414. doi:10.1111/j.1083-6101.2006.00019.x
- Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1), 241–250. doi:10.1016/j.chb.2011.09.006
- Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. *Frontiers in Neurorobotics*, 14. doi:10.3389/fnbot.2020.593732
- Liebrecht, C., & van der Weegen, E. (2019). Menselijke chatbots: een zegen voor online klantcontact? Het effect van conversational human voice door chatbots op social presence en merkattitude. *Tijdschrift voor Communicatiewetenschap*, 47(3-4), 217–238.
- Meertens Instituut. (n.d.). NVB - verspreiding voornaam Eline. Retrieved March 13, 2021, from <https://www.meertens.knaw.nl/nvb/verspreiding/naam/Eline>
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 72–78.
- Powers, A., & Kiesler, S. (2006). The advisor robot. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, 218–225. doi:10.1145/1121241.1121280
- Sarbin, T. R., & Allen, V. L. (1968). Increasing Participation in a Natural Group Setting: A Preliminary Report. *The Psychological Record*, 18(1), 1–7. doi:10.1007/bf03393733

- Scales, J., Wennerstrom, A., Richard, D., & Wu, S. H. (2006). Language Learners' Perceptions of Accent. *TESOL Quarterly*, 40(4), 715. doi:10.2307/40264305
- Scott, K. M., Ashby, S., & Hanna, J. (2020). "Human, All Too Human": NOAA Weather Radio and the Emotional Impact of Synthetic Voices. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–9. doi:10.1145/3313831.3376338
- Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful?. *Ldv forum* 22(1), 29-49.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Toronto; London; New York: Wiley.
- Smakman, D. (2006). Standard Dutch in the Netherlands. A sociolinguistic and phonetic description. Utrecht: LOT.
- Statista. (2021, January 22). *Number of digital voice assistants in use worldwide 2019-2024*. <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/#statisticContainer>
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge, UK: Cambridge university press.
- Telang, P. R., Kalia, A. K., Vukovic, M., Pandita, R., & Singh, M. P. (2018). A Conceptual Framework for Engineering Chatbots. *IEEE Internet Computing*, 22(6), 54–59. doi:10.1109/mic.2018.2877827

Torre, I., & Maguer, S. L. (2020). Should robots have accents? *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 208–214.
doi:10.1109/ro-man47096.2020.9223599

van der Harst, S. & Landelijke Onderzoekschool Taalwetenschap. (2011). *The Vowel Space Paradox*. LOT.

van Noort, G., & Willemsen, L. M. (2012). Online Damage Control: The Effects of Proactive Versus Reactive Webcare Interventions in Consumer-generated and Brand-generated Platforms. *Journal of Interactive Marketing*, 26(3), 131–140.
doi:10.1016/j.intmar.2011.07.001

Young, W. (2018, August 10). *The voice search explosion and how it will change local search*. Search Engine Land. <https://searchengineland.com/voice-search-explosion-will-change-local-search-251776>

Footnotes

¹ Non-regional accent is defined as an accent outside the region of interest in this paper

² To ensure reliable data, all the analyses were run again excluding each participant that did not identify the accent correctly. Ultimately, this neither yielded any significant results and after a discussion it was decided to stick to the shown data as the participants in question answered the other control questions correctly.

Table 1

Results of the pretest

Participant number	Origin participant	Result fragment 1 (Noord-Brabant)	Result fragment 2 (Standard Dutch)
1	Noord-Brabant	South-east	South-west
2	Noord-Brabant	South-west	Middle
3	Noord-Brabant	South-east	South-west
4	Overijssel	South-east	Middle

Table 2

Grand mean and standard deviation of the scales ‘anthropomorphism’ and ‘appropriateness’

Outcome variables	<i>M</i>	<i>SD</i>
Anthropomorphism	4.69	1.20
Appropriateness	4.93	1.06

Table 3

Anthropomorphism scores per VA type and per participant group

Group	<u>VA type</u>			
	<u>SVA</u>		<u>RVA</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Regional participants	4.63	1.03	4.88	1.27
Non-regional participants	4.73	.89	4.81	1.52

Table 4

Appropriateness scores per VA type and per participant group

Group	<u>VA type</u>			
	<u>SVA</u>		<u>RVA</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Regional participants	5.08	.92	4.92	1.10
Non-regional participants	4.97	.88	4.73	1.33

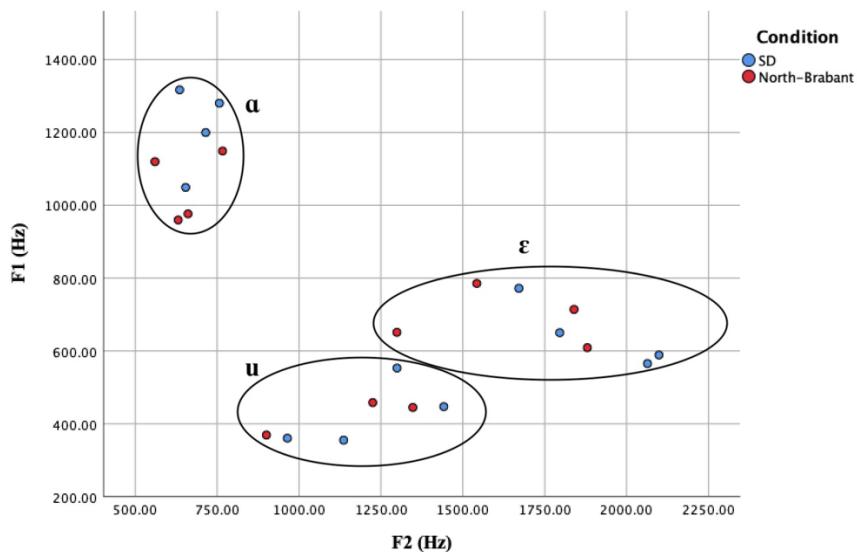


Figure 1. F1 & F2 for three sampled vowels

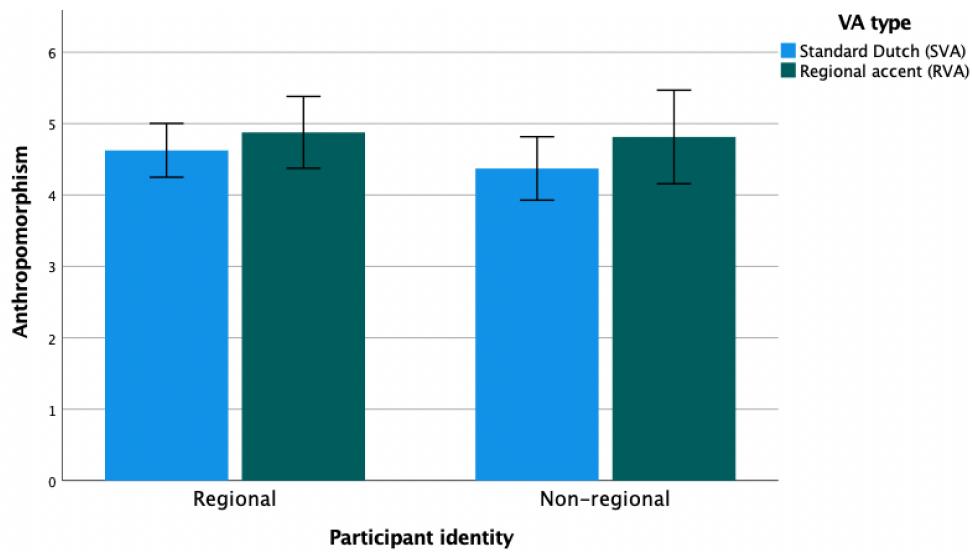


Figure 2. Clustered bar chart of anthropomorphism based on participant identity and the VA type the participant got exposed to.

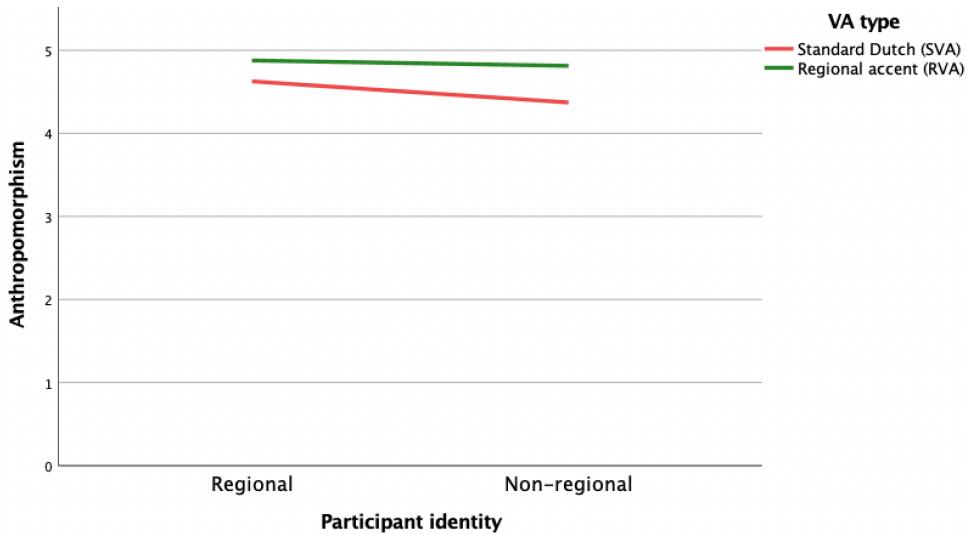


Figure 3. Multiple line graph of anthropomorphism based on participant identity and the VA type the participant got exposed to.

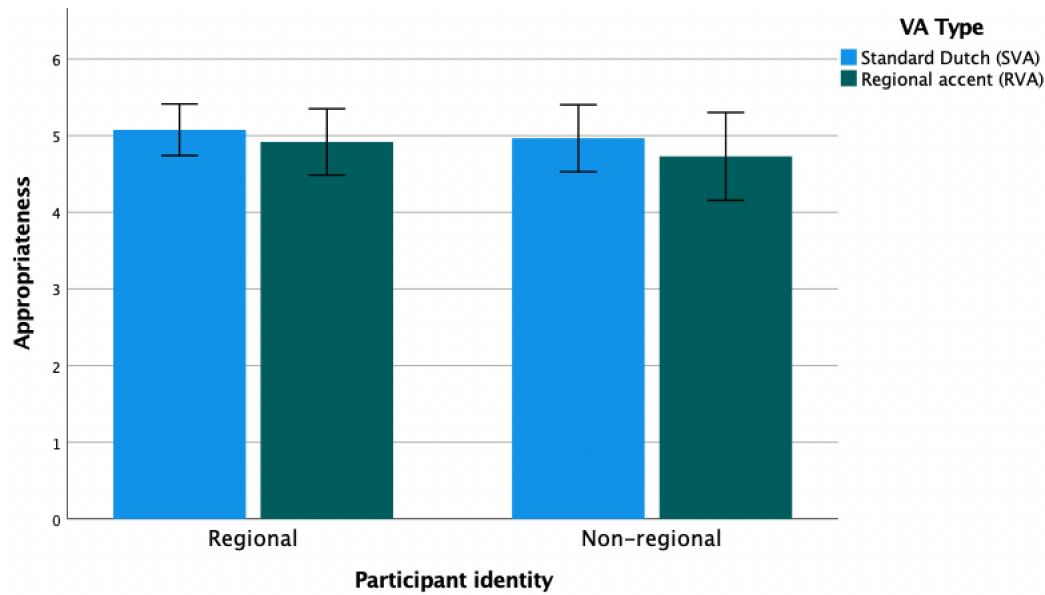


Figure 4. Clustered bar chart of appropriateness based on participant identity and the VA type the participant got exposed to.

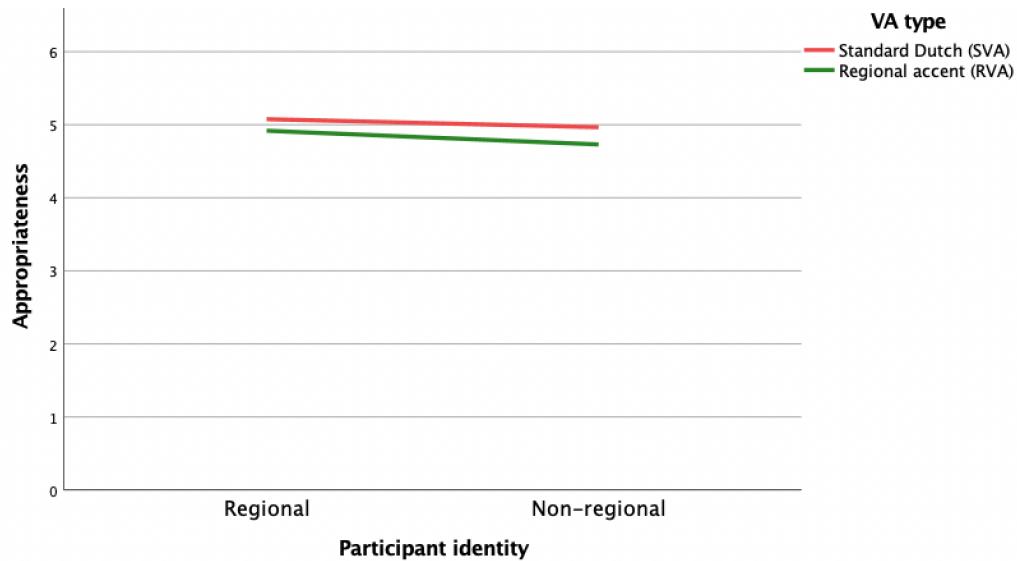


Figure 5. Multiple line graph of appropriateness based on participant identity and the VA type the participant got exposed to.

Appendices

Appendix A

Pretest of the VA prototype

Beste deelnemer,

Wij waarderen het enorm dat u de tijd neemt om deel te nemen aan ons onderzoek. Dit onderzoek wordt uitgevoerd door Rick Oosterholt en Mike Vos van de master Communicatie- en Informatiewetenschappen aan de Universiteit van Tilburg. U wordt blootgesteld aan een experiment, gevolgd door een aantal vragen.

Uw deelname aan dit onderzoek is geheel vrijwillig, en u kunt zich op elk moment terugtrekken zonder uitleg of consequenties. De verzamelde informatie zal vertrouwelijk en anoniem worden behandeld. De resultaten van dit onderzoek zullen worden gebruikt voor wetenschappelijke doeleinden en kunnen worden gedeeld met de Universiteit van Tilburg.

Voordat u aan het onderzoek begint, verzoeken wij om uw geluid aan te zetten. Daarnaast verzoeken wij u ook om een rustige omgeving op te zoeken, vrij van afleiding, en, indien nodig, een koptelefoon te gebruiken.

Als u vragen heeft over het onderzoek, neem dan contact op met
r.oosterholt@tilburguniversity.edu

Door op de onderstaande knop te klikken, geeft u aan dat u:

- Minimaal 18 jaar oud bent.
- De bovenstaande informatie heeft gelezen.
- Vrijwillig instemt met deelname aan dit onderzoek.
- Zich ervan bewust bent dat u uw deelname op elk moment en om welke reden dan ook kunt stopzetten.

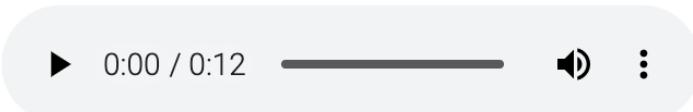
Indien u niet wenst deel te nemen aan de onderzoeksstudie, verzoeken wij u uw deelname te weigeren door te klikken op de niet akkoord knop.

Alvast hartelijk dank!

- Ik ga akkoord
- Ik ga niet akkoord

Vraag 1.

Wij zouden je willen vragen naar het volgende fragment te luisteren. Let er op dat je volume hard genoeg staat om het te verstaan.



Controleer met behulp van deze afbeelding waar de regio's zich in Nederland bevinden om de volgende vraag te beantwoorden.

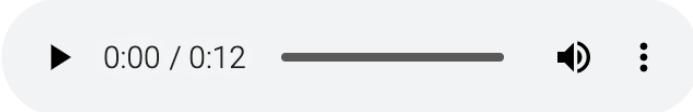


Uit welke regio denkt u dat de betreffende assistent vandaan komt?

- Noorden (N)
 - Oosten (O)
 - Midden (M)
 - Noord-West (NW)
 - Zuid-West (ZW)
 - Zuid-Oost (ZO)

Vraag 2.

Wij zouden je willen vragen naar het volgende fragment te luisteren. Let er op dat je volume hard genoeg staat om het te verstaan.



Controleer met behulp van deze afbeelding waar de regio's zich in Nederland bevinden om de volgende vraag te beantwoorden.



Uit welke regio denkt u dat de betreffende assistent vandaan komt?

- Noorden (N)
- Oosten (O)
- Midden (M)
- Noord-West (NW)
- Zuid-West (ZW)
- Zuid-Oost (ZO)

Vraag 3.

In welke provincie heb je het langst gewoond?

A. Noord-Holland

B. Utrecht

C. Zuid-Holland

D. Zeeland

E. Noord-Brabant

F. Gelderland

G. Overijssel

H. Drenthe

I. Groningen

J. Friesland

K. Flevoland

L. Limburg

VERSTUUR

Appendix B

Survey for the main study

BLOCK 1

Beste deelnemer,

Wij waarderen het enorm dat je de tijd neemt om deel te nemen aan ons onderzoek.

Dit onderzoek wordt uitgevoerd door Rick Oosterholt en Mike Vos van de master Communicatie- en Informatiewetenschappen aan de Universiteit van Tilburg. Wij doen een onderzoek naar de percepties van gebruikers tegenover virtual assistants.

Jouw deelname aan dit onderzoek is geheel vrijwillig en je kan jezelf op ieder moment terugtrekken zonder uitleg of consequenties. De verzamelde informatie zal vertrouwelijk en anoniem worden behandeld. De resultaten van dit onderzoek zullen worden gebruikt voor wetenschappelijke doeleinden en kunnen worden gedeeld met de Universiteit van Tilburg. Gedeelde of gepubliceerde gegevens zullen echter nooit herleidbaar zijn tot personen.

Voordat je aan het onderzoek begint, verzoeken wij je om je **geluid aan te zetten**. Daarnaast verzoeken wij je ook om een rustige omgeving op te zoeken, vrij van afleiding, en, indien nodig, een koptelefoon te gebruiken.

Als je vragen hebt over het onderzoek, neem dan contact op met r.oosterholt@tilburguniversity.edu of m.vos@tilburguniversity.edu

Door op de onderstaande knop te klikken, geef je aan dat je:

- Minimaal 18 jaar oud bent.
- De bovenstaande informatie hebt gelezen.
- Vrijwillig instemt met deelname aan dit onderzoek.
- Je ervan bewust bent dat je jouw deelname op elk moment en om welke reden dan ook kunt stopzetten.

Indien je niet akkoord gaat, of indien je tijdens het onderzoek jezelf wil terugtrekken van deelname, worden je antwoorden na vier uur verwijderd. Door het tabblad weg te klikken, eindigt je deelname. Deelname aan dit onderzoek duurt ongeveer 8 minuten.

Alvast hartelijk dank!

- Ik ga akkoord, ik neem deel aan dit onderzoek
- Ik ga niet akkoord, ik neem geen deel aan dit onderzoek

BLOCK 2 - Randomizer

Let op! Wanneer je op de link klikt opent een nieuw tabblad met daarin het prototype. Het is belangrijk dat je het tabblad van deze enquête open laat staan zodat je straks weer terug kunt gaan. Het prototype geeft zelf aan wanneer je weer terug mag gaan naar deze enquête.

Klik [hier](#) om naar het prototype te gaan

Bij terugkomst kan je op de 'verder'-pijl drukken.

- VA startscherm -

Welkom bij het prototype van de virtual assistant! Deze virtual assistant is van het fictieve bedrijf Polderparks en kan je helpen bij verschillende vragen, zoals reserveringen en klachten. Polderparks is een nationaal bedrijf die een groot aanbod aan vakantiehuisjes aanbiedt in Nederland.

Wij willen je vragen de volgende taak uit te voeren met behulp van de virtual assistant:

- Je bent op zoek naar een vakantiehuisje
- Voor jezelf en drie vrienden, dus 4 personen totaal
- In de periode juli
- De overige opties mag je invullen naar eigen wensen

Nadat de virtual assistant jou een vakantiehuisje heeft geadviseerd, geeft de virtual assistant aan dat je weer terug mag keren naar de enquête. Je kunt de virtual assistant dan verlaten door op het kruisje te klikken en de enquête in het andere tabblad te openen.

- VA flow -

Hallo en welkom! Mijn naam is Eline, en ik ben de virtuele assistent van Polderparks die je graag verder helpt. Wat kan ik vandaag voor je doen?

- A. Ik ben op zoek naar een vakantiehuisje
- B. Ik heb een vraag over een reservering
- C. Ik heb een andere opmerking

Leuk dat je op zoek bent naar een vakantiehuisje! Wij bieden ruime keuze aan in typen vakantiehuisjes in Nederland. Voordat ik je een huisje kan adviseren heb ik nog wel een aantal vragen voor je. Om te beginnen, in welke periode zou je een vakantiehuisje willen reserveren?

- A. Ik wil reserveren ergens in juli
- B. Ik wil reserveren ergens in augustus
- C. Ik wil reserveren ergens in september

Duidelijk! Op onze vakantieparken is er zowel ruimte voor grote als kleine groepen mensen. Met hoeveel personen ben je van plan te verblijven?

- A. 1 tot 3 pers.
- B. 4 tot 6 pers.
- C. 7 tot 12 pers.

Helder! Hoeveel wil je maximaal besteden aan het vakantiehuisje?

- A. Maximaal 100 euro
- B. Maximaal 300 euro
- C. Dat maakt mij niet uit

We bieden vakantiehuisjes aan voor elke type comfort. Van wegdroegen in een luxe chalet tot een lekker knusse bungalow. Welke type comfort heeft je voorkeur?

- A. Basic heeft mijn voorkeur
- B. Comfort heeft mijn voorkeur
- C. Luxe heeft mijn voorkeur

Wij van Polderparks vinden het belangrijk dat je jezelf op je gemak voelt tijdens je verblijf. Daarom richten wij onze huisjes in op jouw behoeften. Wil je kunnen roken in het huisje?

- A. Ja, graag
- B. Nee, nee dat hoeft niet

We zijn er bijna! Op de vakantieparken van Polderparks zijn dieren van harte welkom! Ben je van plan om je huisdier mee te nemen?

- A. Ja, die neem ik graag mee
- B. Nee, die neem ik niet mee
- C. Weet ik nog niet

Hartelijke bedankt voor je antwoorden! Wij denken dat **Wolken verblijf 5** het meest aan jouw wensen voldoet! Wil je een digitale folder ontvangen met meer informatie? In de volgende stap kun je deze folder aanvragen

Mijn antwoord:

“Het gesprek met Eline afsluiten”

TEXT “Dit is het einde van het prototype, je mag weer teruggaan naar de survey en dit tab sluiten.”

BLOCK 3 – Controle vraag

Welk vakantiehuisje heeft de virtual assistant je aangeraden?

- A. Bungalow XL
- B. Wolkenverblijf 5
- C. Chalet 85
- D. De virtual assistant heeft me niets aangeraden

BLOCK 3 – Anthropomorphism & Appropriateness

Je hebt zojuist een interactie gehad met de virtual assistant. Het onderzoek duurt nu nog ongeveer 3 minuten.

De volgende vragen gaan over hoe je de *interactie* met de virtual assistant hebt ervaren.

In hoeverre kwam de interactie met de virtual assistant over als...

Machine-achtig Menselijk

Natuurlijk Onnatuurlijk

Kunstmatig Levensecht

Onsympathiek Sympathiek

Gezellig Ongezellig

Onvriendelijk Vriendelijk

Persoonlijk Onpersoonlijk

De volgende vragen gaan over hoe je de toon van de virtual assistant beschouwde.

In hoeverre kwam de toon van de virtual assistant over als...

Ongepast Gepast

Beleefd Onbeleefd

Egoïstisch Altruïstisch

Arrogant Nederig

Professioneel Onprofessioneel

BLOCK 4 – Afkomst stem VA

De volgende vragen hebben betrekking op jouw ervaring met de virtual assistant.

Uit welke provincie(s) denk je dat de stem van de virtual assistant afkomstig was? (*er zijn meerdere opties mogelijk*)

1. Noord-Holland
2. Zuid-Holland
3. Utrecht
4. Zeeland
5. Flevoland
6. Limburg
7. Gelderland
8. Noord-Brabant
9. Drenthe
10. Groningen
11. Friesland
12. Overijssel

Met betrekking tot je vorige gekozen antwoord(en), wat is de meest waarschijnlijke provincie waar de stem van de virtual assistant vandaan kwam?

1. Noord-Holland
2. Zuid-Holland
3. Utrecht
4. Zeeland
5. Flevoland
6. Limburg
7. Gelderland
8. Noord-Brabant
9. Drenthe
10. Groningen
11. Friesland
12. Overijssel

BLOCK 5 – Demographics & participant identity

Je bent bijna klaar! Hier volgen nog een aantal algemene vragen.

- Wat is uw geslacht?

- A. Man
- B. Vrouw
- C. Non-Binary
- D. Zeg ik liever niet

- Wat is uw leeftijd? (in jaren)

- In welke provincie ben je geboren?

- A. Noord-Holland
- B. Utrecht
- C. Zuid-Holland
- D. Zeeland
- E. Noord-Brabant
- F. Gelderland
- G. Overijssel
- H. Drenthe
- I. Groningen
- J. Friesland
- K. Flevoland
- L. Limburg

- In welke provincie woon je nu?

- A. Noord-Holland
- B. Utrecht
- C. Zuid-Holland
- D. Zeeland
- E. Noord-Brabant
- F. Gelderland
- G. Overijssel

- H. Drenthe
- I. Groningen
- J. Friesland
- K. Flevoland
- L. Limburg

- In welke provincie heb je het langst gewoond?

- A. Noord-Holland
- B. Utrecht
- C. Zuid-Holland
- D. Zeeland
- E. Noord-Brabant
- F. Gelderland
- G. Overijssel
- H. Drenthe
- I. Groningen
- J. Friesland
- K. Flevoland
- L. Limburg

- Met welke provincie identificeer je jezelf?

- A. Noord-Holland
- B. Utrecht
- C. Zuid-Holland
- D. Zeeland
- E. Noord-Brabant
- F. Gelderland
- G. Overijssel
- H. Drenthe
- I. Groningen
- J. Friesland
- K. Flevoland
- L. Limburg

De laatste vragen gaan over je algemene gebruik omtrent virtual assistants.

- In hoeverre maak je gebruik van een virtual assistant? (Bijvoorbeeld: Google Home, Apple Siri, Amazon Alexa)
 - A. Nooit
 - B. Jaarlijks
 - C. Maandelijks
 - D. Wekelijks
 - E. Dagelijks

In hoeverre beschouw je jezelf als ‘technologie-bekwaam’ met betrekking tot nieuwe technologieën?

Ik vind het leuk om iets nieuws te leren over nieuwe technologie

Ik verdiep me gemakkelijk in nieuwe technologie

Ik leer niet gemakkelijk wat betreft nieuwe technologie

Ik ben technologie-bekwaam

Nieuwe technologie interesseert me niet zo

Zeer mee oneens – Mee oneens – Neutraal – Mee eens – Zeer mee eens

BLOCK 6 – The end

Hartelijk dank voor je deelname! Het doel van dit experiment was om te onderzoeken wat het effect is van het accent van een virtual personal assistant op gebruikers. Om dit te onderzoeken hebben wij het prototype gemanipuleerd met 2 verschillende accenten (algemeen Nederlands en Brabants). Daarnaast hadden wij twee controlegroepen in de vorm van tekst en gesynthetiseerde spraak.

Graag willen wij je vragen het doel van het onderzoek nog niet te bespreken met vrienden of kennissen.

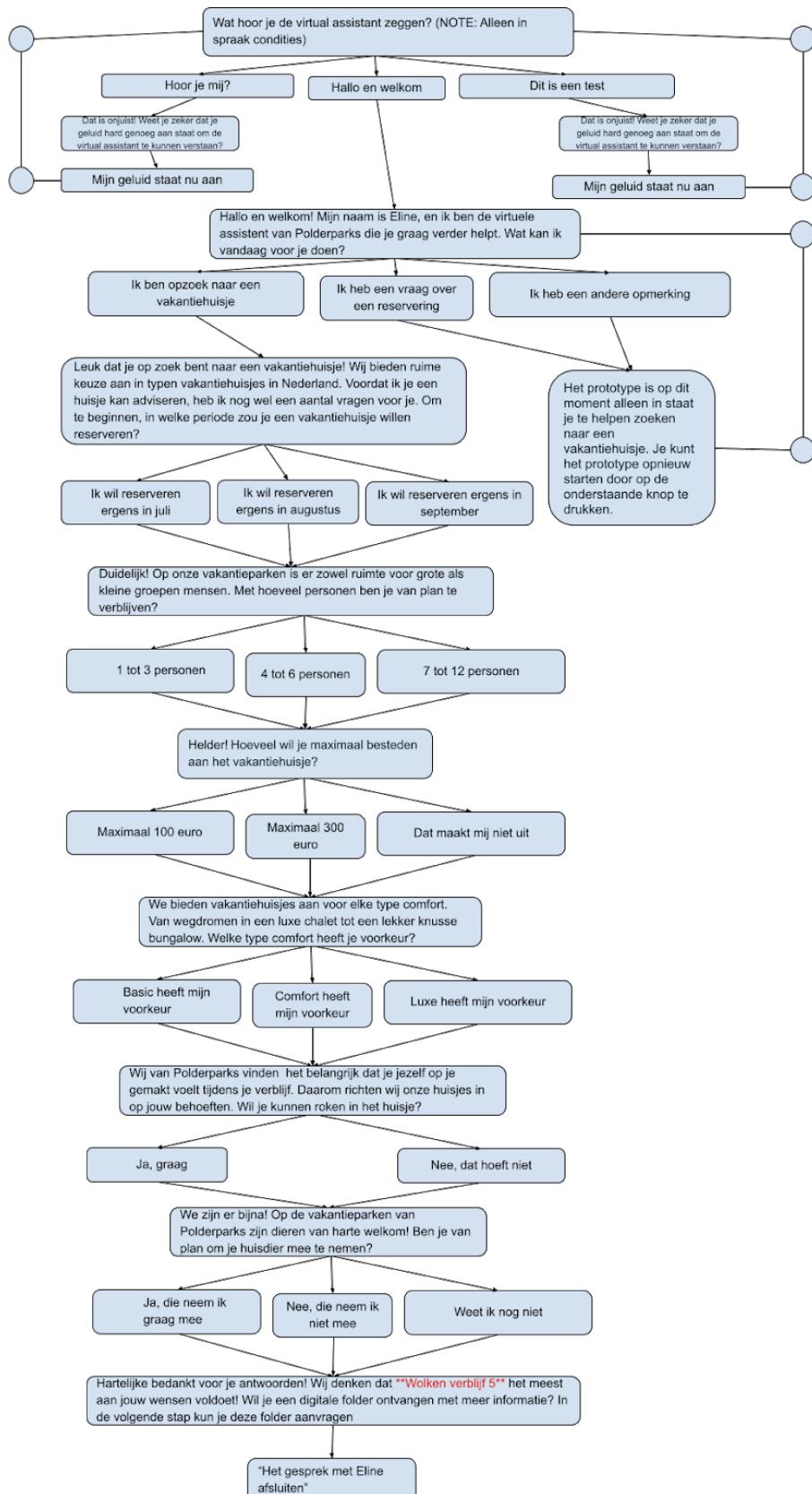
Als je interesse hebt in de uitkomsten van het onderzoek of indien je overige vragen hebt, dan kun je een e-mail sturen naar m.vos@tilburguniversity.edu of naar r.oosterholt@tilburguniversity.edu. Wij zullen dan vervolgens, wanneer het onderzoek voltooid is, een samenvatting sturen van de uitkomsten van dit onderzoek.

Je kan ons nog verder helpen met het onderzoek door de volgende link te delen met vrienden of familie!

bit.ly/helpckenmikeafstuderen

Appendix C

Interaction flow of the VA



Appendix D

The interface of the VA

Eline van Polderparks

Hallo en welkom! Mijn naam is Eline, en ik ben de virtuele assistent van Polderparks die je graag verder helpt. Wat kan ik vandaag voor je doen?

Eline van Polderparks

Leuk dat je op zoek bent naar een vakantiehuisje! Wij bieden ruime keuze aan in typen vakantiehuisjes in Nederland. Voordat ik je een huisje kan adviseren, heb ik nog wel een aantal vragen voor je. Om te beginnen, in welke periode zou je een vakantiehuisje willen reserveren?

Eline van Polderparks

Duidelijk! Op onze vakantieparken is er zowel ruimte voor grote als kleine groepen mensen. Met hoeveel personen ben je van plan te verblijven?

Mijn antwoord:

- Ik ben op zoek naar een vakantiehuisje →
- Ik heb een vraag over een reservering →
- Ik heb een andere opmerking →

- Ik wil reserveren ergens in juli →
- Ik wil reserveren ergens in augustus →
- Ik wil reserveren ergens in september →

- 1 tot 3 personen →
- 4 tot 6 personen →
- 7 tot 12 personen →

Mijn antwoord:

Helder! Hoeveel wil je maximaal besteden aan het vakantiehuisje?

Eline van Polderparks

We bieden vakantiehuisjes aan voor elke type comfort. Van wegdragen in een luxe chalet tot een lekker knusse bungalow. Welke type comfort heeft je voorkeur?

Eline van Polderparks

Hartelijke bedankt voor je antwoorden! Wij denken dat 'Wolken verblijf 5' het meest aan jouw wensen voldoet! Wil je een digitale folder ontvangen met meer informatie? In de volgende stap kun je deze folder aanvragen.

Mijn antwoord:

- Maximaal 100 euro →
- Maximaal 300 euro →
- Dat maakt mij niet uit →

- Basic heeft mijn voorkeur →
- Comfort heeft mijn voorkeur →
- Luxe heeft mijn voorkeur →

Mijn antwoord:

Het gesprek met Eline afsluiten →

The image displays four sequential screenshots of a web-based virtual assistant application, likely a chatbot or survey tool, titled "Eline van Polderparks".

Screenshot 1: A welcome message from "Eline van Polderparks" introducing the prototype of a virtual assistant for Polderparks. It asks if the user wants to perform a task using the virtual assistant.

Screenshot 2: A question "Wat hoor je de virtual assistant zeggen?" followed by a large blue circular button labeled "Geluid". Below the button is a note: "Vind je het goed dat er geluid wordt afgespeeld?" with a "Ja" button.

Screenshot 3: A question "Wat hoor je de virtual assistant zeggen?" followed by a large blue circular button labeled "Geluid". Below the button is a note: "Vind je het goed dat er geluid wordt afgespeeld?" with a "Ja" button.

Screenshot 4: A list of three responses under "Mijn antwoord:":

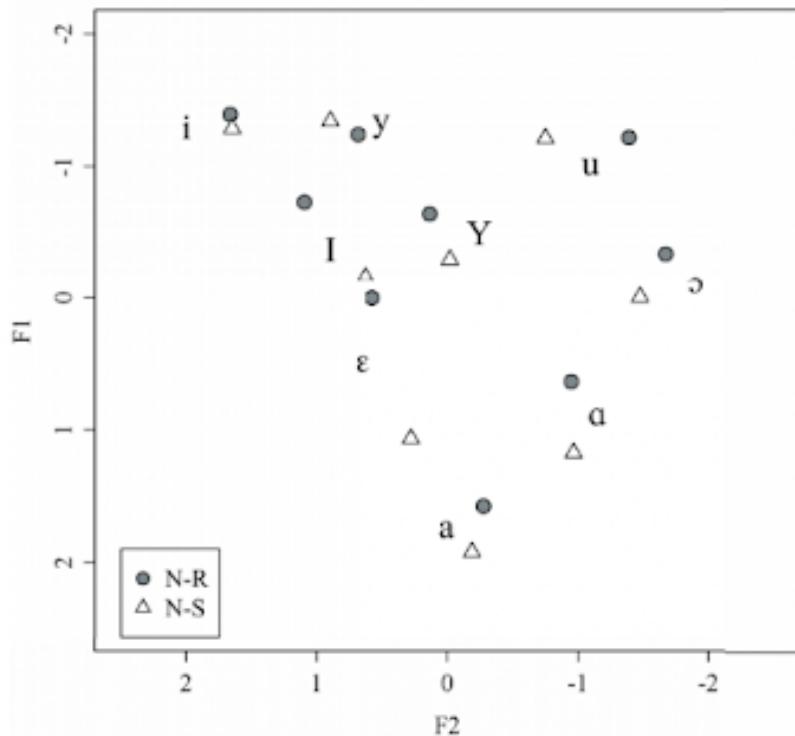
- Hoor je mij? →
- Hallo en welkom →
- Dit is een test →

At the bottom of each screenshot is a navigation bar with icons for back, forward, search, and other document-related functions.

Appendix E

Mean normalized F1 and F2 (at 50%) for the monophthongs in N-R and N-S, in spontaneous

speech.



Appendix F

Explanation of collaboration and division of labor with another researcher

This study was partly designed in collaboration with another researcher. A survey set-up, a prototype design/set-up/flow input, and a voice script were established together with another researcher. Furthermore, both researchers collaborated in contacting a speech therapist, making an appointment on campus to meet for recordings, reserving time to edit the audio recordings, and analyzing the audio in the speech analysis program Praat. Furthermore, both researchers collaborated on setting up a pretest. The tasks and time were equally divided between both researchers. This section goes into this in more detail

The prototype of the VA had to meet several requirements. First of all, a flow was developed based on codes and placed on a server. Four flows were developed, one of which was omitted from this study. The other researcher used this one. The flows are essentially the same; however, there is one flow with text input and three flows with speech input. One is filled with the synthesized speech of those three flows with speech input (it was omitted in this study). The other two types of speech were developed using a speech therapist.

A speech therapist was hired who voiced both an accented version and a standard Dutch version. Before this could have been accomplished, a flow was written out. The flow consisted of the spoken sentences, and response buttons clicked when respondents were exposed to a VA type. Then, in collaboration with the other researcher, applicants were recruited through LinkedIn and various Facebook groups. After a week, five individuals with a background in speech therapy had come forward. After a choice was made, an appointment was scheduled to record the sentences. The recordings for the speech VAs took place in the Dante building on the campus of Tilburg University. Finally, many flowers and chocolates thanked the speech therapist for her cooperation, and the recordings were edited and added to the already developed flows. After a week, the VAs were ready. The different types of speech

from the speech therapist were analyzed for differences in formats, and the pretest was deployed. A week later, the main study was plotted.