WHAT SPEAKS TO US

COMPARING ATTITUDES TO TEXT- AND VOICE-BASED PURCHASE REMINDERS

JAKOB NORDFELDT

Bachelor Thesis

Stockholm School of Economics

2019

What Speaks to Us: Comparing Attitudes to Text- and Voice-based Purchase Reminders

Abstract:

Despite the rapidly increasing usage of virtual personal assistants (VPAs), little research has been devoted to how consumers respond to commercial VPA messages. This study makes an initial contribution by analyzing responses to a purchase reminder. A 2x2 online survey-based experiment (n=195) exposed participants to different versions of a reminder to purchase batteries. Reminders differed in terms of stimuli mode (voice/text) and language complexity (high/low). The findings indicate that high-complexity reminders were viewed more favorably than low-complexity reminders. No significant difference was found between voice and text reminders. In addition, the attitude towards the reminder was shown to be positively related to purchase intentions, brand attitudes and usage intentions.

Keywords:

Voice technology, Virtual assistant, Marketing, Attitude, Reminder, AI, Linguistic devices, Smart assistant, Alexa, Google

Author:

Jakob Nordfeldt (23847)

Tutor:

Magnus Söderlund

Examiner:

Hanna Berg

Bachelor Thesis
Bachelor Program in Business & Economics
Stockholm School of Economics

List of definitions

Virtual Personal Assistant (VPA):	An algorithmic application designed to assist humans with various tasks through conversational interactions. Commercial examples include Amazon Alexa and Google Assistant.
Language complexity:	The level of sophistication of a message regardless of medium.
Stimuli modality:	The choice of communication medium for stimuli, e.g. voice or text.
Decoding:	The mental processing of a stimulus.
Acoustic cues:	The formants of sounds that distinguishes them from other sounds.
Natural language:	Language that has developed naturally through human interactions
Linguistic device:	Certain modes of expression used to add meaning to a message.

Table of Contents

"Piglet noticed that even though he had a very small heart,
it could hold a rather large amount of gratitude"
- A.A. Milner, "Winnie-the-Pooh"

Acknowledgements

Completing my thesis or degree would never have been possible without the support of people around me. In particular, I would like to extend my gratitude to:

Jonas Colliander for helping me narrow down my ideas and understanding the mindset needed for consumer research.

Susanne Magnusson at Acapela Group for helping me develop the voice-based stimuli

Roel Wijland for opening the world to voice technology in marketing, truly a Pandora's box for those who thirst knowledge

Kirsten Robertson, John Williams and **Janet Hoek** at the University of Otago for helping me develop by making an exception to allow me to pursue postgraduate courses

Internetstiftelsen and **GoTo10** for enabling me to share the insights with many more

Lina Reidarsdotter Källström for creative support and discussions.

Magnus Söderlund and the tutoring group for valuable discussions throughout the semester

Emelie Fröberg for the useful feedback regarding statistical procedures

David Vinnersten for helping me find my own voice well before embarking on studying voice technology.

CoolBlue in Amsterdam for helping me retrieve the thesis files from a damaged computer with only 2 days remaining to the deadline.

Friends and Family for dealing with my many and lengthy complaints.

Emma Perlelin, Jessica Ollus, Johanna Tell, Sigrid Hafskjold Holmgren, Axel Bryhn, Jakob Ringberg, Johan Faxner, Lina Reidarsdotter Källström, Lukas Källén, Alice D'Angelo, Alice O'Connor, Denusha Sharma, Molly Fröjdh, Nils Lager, Rejhan Dlir, Sebastian Stüben and **Tony Liang** for campaigning for ugly cats, helping me gather data and spreading some joy in the world.

1. Introduction

You could argue that it is the ultimate proof of human laziness. But you could just as well make a case for it being a testament to the ingenuity of our species. Either way, one thing stands clear. Humans have been dreaming of creating automated assistants for a very long time. Traces in the form of handicrafts and tales are scattered through ancient civilizations, from Mesopotamia to Greece (Mayor, 2018). One only has to consider contemporary works like Spike Jonze's *Her* (2013) to realize that the dream of creating assistants still permeates popular culture. What has changed, however, is the gap between the dream and the technological knowledge required to turn it into reality. With the rapid development of artificial intelligence and voice technology, science is on the brink of achieving the definitive cross-product of our ingenious laziness.

The development is powered by the success of voice assistants like Google Assistant, Amazon's Alexa or Apple's Siri. Voice technology is increasingly taking a place in the lives of everyday consumers. David Beckham is the face of the world's first "voice petition" to end malaria (Hobley & Gashe, 2019). Amazon employees are transcribing millions of recordings from Alexa users, including suspected sexual assault, in order to improve the language skills (Bell, 2019). For better or worse, voice technology is set to become an integral part of consumption in the 21st century. However, our understanding of its impacts remains limited. Therefore, this thesis aims to contribute to the knowledge of how changes in everyday consumption brought about by voice technology might affect consumers.

1.1. Background

To understand the changes, it is important to first grasp the preceding developments. This section provides an introduction to the how the role of a consumer has changed over the last years, and how this relates to the emergence of smart assistants.

1.1.1. The Evolving Nature of Consumer Interactions

Since the debut of one-way marketing communication theories and models, technology has helped shift the role of a consumer. Instead of passive recipients, consumers have increasingly become engaged co-creators of products, brands and experiences. Social media extended word-of-mouth "WOM" to what has been called "eWOM" (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004), giving more influence to consumers by enabling instant responses to marketing messages. Commercials now cause global backlashes and appraisals within a few minutes with the help of networks like Twitter (Handley, 2017). Some argue that consumers are using this increased power to demand that companies share their values and beliefs, regardless of connection to the core business (Baggs, 2019).

The increased power of consumers comes at a cost in the form of a reduced degree of privacy. Businesses extract extensive information from our online activities. This is perhaps best illustrated by Target's successful prediction of a young woman's pregnancy with the help of her search history (Hill, 2012). Furthermore, not only are we coming closer to businesses through our online presence, the relationship between consumers and brands is also becoming increasingly intimate physically. Through our ever-present smartphones, laptops and smartwatches, brands follow us to work and come home with us at the end. From coming with us on our day-to-day activities, these devices are now starting to take part in them actively. It is this transformation that has been powered by the development of Virtual Personal Assistants (VPAs).

1.1.2. The Emergence of Virtual Personal Assistants

Considering the ancient aspirations of creating "robot assistants" discussed earlier (Mayor, 2018), it should not come as a surprise that the first VPA was actually launched two decades before the personal computer (IBM Archives, 2003). Despite this early start, it is only recently that VPAs have appeared in everyday consumer contexts. Apple's purchase and subsequent launch of Siri, an offspring of SRI International, established VPAs as a standard features of smartphones.

Within 5 years, Samsung (2012), Microsoft (2014) and Google (2016) had all launched their respective smartphone VPAs to compete with Siri. Now, VPAs are moving out of our pockets and into our homes in the form of smart speakers. The number of VPAs in consumer homes has increased rapidly, particularly on the American and the Chinese markets (Cherian & Pounder, 2017; Clark, 2019) A global VPA population exceeding the global human population appears only a few years away (De Renesse, 2017).

VPAs perform an array of tasks including scheduling, information retrieval, communication, controlling other devices and making purchases on behalf of consumers (Jones, S., 2019; Merriman, 2019). Increasingly, VPAs integrate with existing smartphone and payment technology to enable new business models for retailers (Chen, C., Huang, Park, Tseng, & Yen, 2014). Google, for example, has recently developed the capability for their assistant to interact with businesses like hair salons on behalf of consumers, even programming it to utter strikingly human-like "mhm"s when asked to wait a moment (Solon, 2018). It is very likely that consumers will start using these capabilities more and more over the coming years, naturally increasing the importance and value of the VPA for all parties – similar to the advent of e-commerce in the early 2000s. However, there are some clear differences between consumption through traditional search engines like Baidu or Google and consumption through VPAs.

Figure 1illustrates the workings of VPA purchases, where the VPA algorithms scan the different online retailer alternatives to return a single recommendation – in this case a pack of AA batteries.

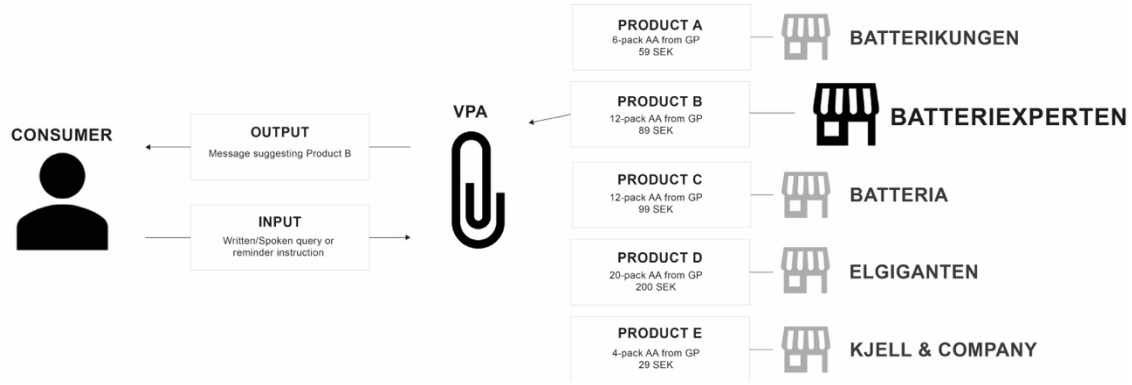


Figure 1. Visualization of the VPA Purchase process

Notably, the process is in stark contrast to conventional search engines by providing only one alternative instead of a list of recommended results. As such, it is to some extent a ceding of control from the consumers to the VPA providers, making it extremely important to have a high chance of being the one alternative that is recommended. This has led some marketing strategists to propose businesses need to transition from search engine optimization (SEO) to instead optimizing for the retrieval algorithms of VPAs (Rowe, 2017). Similarly, it has led legal scholars to raise the question of liability division between human and VPAs (Giancaspro, 2019), as well as attracted criticism towards VPA producers like Amazon who have been accused of promoting own products unfairly through Alexa (Creswell, 2018). But even though the market power implications may be extensive, the increasing commercial activity of VPAs remains an underexplored topic in marketing research.

1.2. Problem Area and Research Gap

Despite the increasing use of VPA technology by consumers, retailers and even governments (Clark, 2019), research is not giving much attention to the topic. A Scopus search for articles containing “voice” and “assistant” in the abstract within 45 leading journals in marketing, psychology and computing rendered only 12 results, half of which were scheduled/published in 2019. This makes it harder to understand the context in which the 21st century consumer acts, negatively impacting marketers as well as policy makers wishing to use voice technology. It seems that the transformation brought by social media that has instead become the central topic of marketing research. Searching the same 45 journals for abstracts containing “social media” generates 1065 results (see Appendix I).

Within research on VPAs, emphasis has been on user adoption and technical development, particularly for voice searches. A case study of Google's voice search showed the fast pace of technological development (Schalkwyk et al., 2010). A main feature in the case study, as well as in other research, is the goal of achieving "human-like" conversational abilities (Dubiel, 2018). This notion has also been extended to a social relationship perspective. Han and Yang (2018) used 304 survey samples to show that task attraction, social attraction and physical attraction are important factors for consumers' adoption of VPAs. Less focus has been on user satisfaction in VPA interactions. A limited study of 60 participants indicated that the importance of task completion and user effort vary between contexts (Kiseleva et al., 2016). Similarly, Lopatovska & Williams (2018) found that Alexa user satisfaction remained high even after failure to understand simple interactions like weather enquiries or commands to play music. Although valuable contributions to a new research area, these past studies have neglected to consider the importance of the communication medium of VPAs, which is either voice-based or text-based.

Past studies comparing text-based and voice-based VPA interactions appear to have been limited in terms of exploring different task domains. Among the few studies, most focus on information retrieval. By comparing spoken and written queries for a document retrieval VPA, Crestani & Du (2006) conclude that although the vocal interactions were longer, they did not necessarily contain any more useful information for the task of retrieving a document. However, the VPA used was considered basic at the time, and is an antiquity in comparison to today's highly complex algorithms. More recent studies, while admittedly also using rather simple VPA applications, have found similar results regarding the difference between text and voice input. An analysis of a commercial search engine's mobile app supported the notion of voice searches being closer to natural language than text searches (Guy, 2018). Studying a recommendation-providing VPA on a movie-website provided similar results (Kang et al., 2017). However, not only are the few comparative studies limited to information retrieval tasks – they also make little or no consideration of consumers' reactions to VPA outputs.

Existing research on consumer responses to VPA output has tended to focus only on voice output rather than comparative studies. Nass and Moon's (2000) seminal paper provided showed how machine-produced voice output is processed similarly to human voices. Subsequent contributions have found similar results in the context of smartphone VPAs (Jeong & Shin, 2015) and implicit evaluations of human and machine voices (Mitchell, Ho, Patel, & MacDorman, 2011). No major studies have investigated consumer attitudes to both VPA text and voice output. In addition to the lack of such comparative studies, a question that remains unanswered is how the attitude to VPA output is related to the linguistic content of the output. A recent review of the effects of linguistic devices on message decoding indicated that research on this topic is yet to be extended to a VPA context (Pogacar, Shrum, & Lowrey, 2018).

In addition, the review disregarded the use of humor, which is becoming a defining characteristic of conversational computing (Borenstein, 2019). Figure 2 visualizes the previously discussed research gap in the context of a consumer-VPA interaction.

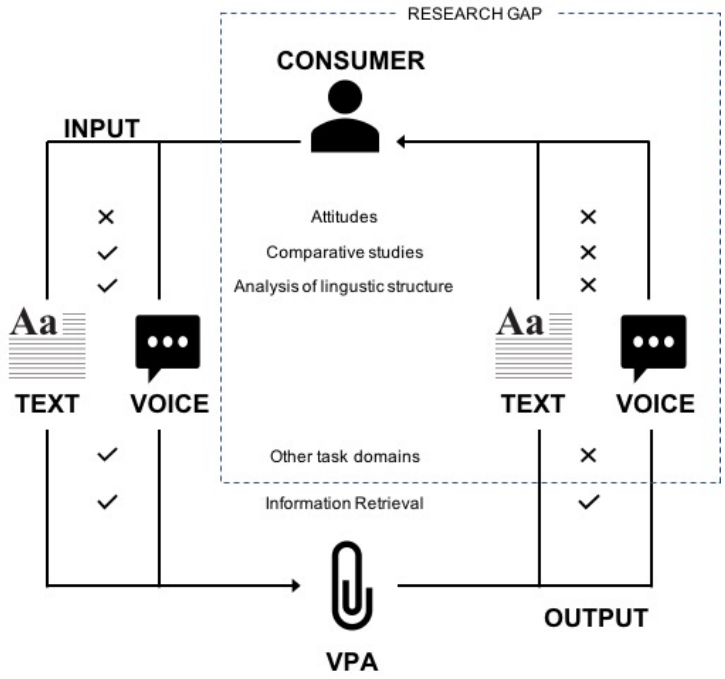


Figure 2. Visualization of research gap

1.3. Purpose and research question

The rapid increase of VPA technology in everyday consumption contexts (Clark, 2019) will lead to more and more consumers interacting with VPAs. Consumers will therefore receive more frequent messages from a VPA, often with commercial content. It is undeniably important for marketers to understand how consumers react to such messages. Therefore, this study aims to investigate what effect the communication medium and the message's linguistic content have on consumer attitudes to the message by answering the following questions:

How does the stimuli mode affect consumers' attitudes towards a VPA message?

How does the language complexity of a VPA message affect consumers' attitudes towards the message?

1.4. Delimitations

Several important delimitations were made in order to balance the ambition of the study with its scope. One such delimitation concerns the VPA output. There are several types of VPA messages; returns of information retrievals, weather forecasts and scheduling confirmations. However, this study focuses only on purchase reminders. In addition, this study only compares one type of text and one type of voice output. The effects of different visual presentations of text-based reminders as well as the differences between different voices for the same reminder will therefore not be analyzed. This is mainly due to practical reasons, as the questionnaire would become lengthy with several categories of manipulations. Similarly, this study investigates the effect of complex language in reminders only in a binary categorization of high-complexity or low-complexity, rather than using a continuous scale. Since the questionnaire was in Swedish, only responses of Swedish-speaking consumers were considered. Finally, for practical reasons the study only uses simulations of VPA messages rather than actual messages from an actual VPA like Siri or Alexa.

1.5. Expected contribution

Considering the vastness of the research gap discussed earlier, this study will contribute to the knowledge of how consumers react to different types of VPA messages. More specifically, it will be an initial contribution to understanding if consumers react differently to VPA messages depending on if they are voice-based or text-based. Furthermore, the study aims to contribute to the body of research on consumers' attitudes towards VPA messages overall, regardless of medium.

Another objective of the study is to extend the use of existing marketing research measurements such as attitudes and behavioral intentions to a VPA context. Given the robust body of marketing research using measurements of attitudes and intentions, it is expected that the previously defined relations between the measurements will also be valid in the context of VPAs.

Finally, the main contribution of this thesis is expected to come from bringing several different research areas together. By connecting existing research on linguistic devices, decoding of acoustic cues and consumer attitudes towards VPAs, this study aims to provide a foundation for a more integrated and multifaceted view of VPA technology. Such a perspective will prove useful as marketers strive to understand what the development of voice technology means for their daily activities.

2. Theoretical framework

This section aims to give a more detailed overview of how the study relates to previous research on the processing of human and machine voices, the effects of linguistic devices and previous studies comparing text and voice stimuli. Subsequently, the study's key measurements are defined and hypotheses are generated. Finally, background variables are discussed. The cited papers were mainly found through searches on Scopus and Google Scholar, relating to keywords such as "voice assistant", "smart assistant" and "attitudes".

2.1. Voice and sound in marketing

2.1.1. The processing of vocal stimuli

> "It's not what you say but *how* you say it"

The above proverb highlights a defining feature of speech; it consists of both linguistic and acoustic information. A more truthful version, however, would be that it is in fact *both* what you say and how you say it. Research has shown that both types of information are used to make personal attributions to a speaker (Apple, Streeter, & Krauss, 1979). This idea can also be found in many aspects of marketing theory. For instance, the Elaboration Likelihood Model (ELM) suggests stimuli are processed through a cognitive central path and a more affective peripheral path (O'Keefe, 2013; Petty & Cacioppo, 1984). In our proverb, the linguistic information (i.e. the "what") would thus be processed centrally while the acoustic information (i.e. the "how") is processed through the peripheral path. We will later return to why the proverb's argument is somewhat flawed. However, let us first dive deeper in the processing of acoustic information.

A major analysis of almost two thirds of the world's languages showed that certain sounds are used for similar vocabulary meanings across continents and language families (Blasi et al., 2016). In particular, it appeared that property words and words for body parts conformed to similar patterns. It thus stands clear that the acoustic properties of speech impact our decoding of messages – but how does this happen?

Let us first consider the rather simple associations between sounds and physical properties. A major study investigated how acoustic information is decoded to make attributions of size and shape to non-words (Knoeferle, Li, Maggioni, & Spence, 2017). Through two extensive experiments, it was shown that size attributions are more common for sounds with low tongue positions, greater jaw openings and the duration of a vowel. Shape attributions on the other hand were made primarily on the basis of lip rounding at the end of a vowel. The connections to oral movements would explain the similarity of sounds across unrelated languages found by Blasi et al. (2016).

A more complex type of acoustic decoding concerns the connection between emotion and sounds. In a comprehensive review of existing research, Scherer (1986) proposed a model based on the connections between emotional states and three major voice types; narrow-wide, lax-tense and full-thin. A subsequent study confirmed several aspects of the proposed model, with mean pitch rate, intensity and pauses being important predictors in decoding accuracy (Banse & Scherer, 1996). This implies that emotions characterized by a high level of intensity are easier to detect, which has been supported in later research (Juslin & Laukka, 2001).

Given the connection between acoustic properties and emotions as well as physical attributions, it should not come as a surprise that sound symbolism has been explored in marketing. In a pioneering study of sound symbolism, Klink (2000) found that a brand name's consonants and vowels communicate information about the size, strength and weight of the brand's product. This notion has subsequently been explored further in the context of brand names with numbers (Gunasti & Ross, 2010), repetition of sounds (Argo, Popa, & Smith, 2010), marketing strategy (Spence, 2012), gender attribution to brand names (Guèvremont & Grohmann, 2015), product attributes and pricing (Coulter & Coulter, 2010; Lowe & Haws, 2017) and comparisons of explicit and implicit willingness to pay (Pogacar, Kouril, Carpenter, & Kellaris, 2018).

However, since the decoding of a message is an individual process, it is inherently dependent on the context and motivation of the recipient. This implies that changing acoustic properties will have different effects in different settings. Based on the ELM, a limited experimental study of 221 participants found changing the intonation and intensity of an advertising message enhanced recipient attitudes only in low-involvement contexts (Gelinas-Chebat & Chebat, 1992). This suggests that as the ELM hypothesizes, the effects of acoustic peripheral cues are negatively related to the level of involvement. However, there is little agreement regarding what acoustic properties an advertising message should have. While Gelinas-Chebat and Chebat (1992) suggest low-intensity and low-intonation, others have proposed high syllable speeds and low pitch rates (Chattopadhyay, Dahl, Ritchie, & Shahin, 2003). Regardless of this inconsistency, it should be seen as clear that acoustic cues can be of great value for marketers, particularly in combination with visual stimuli (Chen, Y. & Spence, 2018)). As VPA interactions look set to increasingly incorporate several stimuli modes (Këpuska & Bohouta, 2018), it will be important for marketers to better understand the effects of acoustic cues. In addition, since the past studies have been limited to analyzing decoding of human speech, a question that remains unanswered is if the VPAs machine-synthesized voices are processed similarly.

2.1.2. Machine voices and society

Research on the decoding of machine-synthesized voices was initiated well before the commercial breakthrough of VPAs. A tendency to provide gender stereotypic responses was found to be strong even in the absence of gender-specific information (Nass, Moon, & Green, 1997). Similarly, robots more were viewed more favorably than when receiving male praise than female praise (Nass & Moon, 2000). Social norms thus appear to affect our decoding of human voices as well as machine voices. Later studies have supported this by looking at gender as well as acoustic cues such as intensity (Cheng, Tracy, Ho, & Henrich, 2016; Jeong & Shin, 2015; Lee, Nass, & Brave, 2000; Mitchell et al., 2011).

Given that the automatic decoding appears to be similar for human and machine voices, VPA technology should logically be a main interest for industries where the human nature of interaction is important but where resources are scarce. Indeed, the healthcare industry and research have explored several potential uses of voice technology (Cheng et al., 2016; Jeong & Shin, 2015; Lee et al., 2000; Mitchell et al., 2011). Some of the more notable include creating a better social life for elderly (Reis, Paulino, Paredes, & Barroso, 2017), increased mental health at workplaces (Kocielnik, Avrahami, Marlow, Lu, & Hsieh, 2018), home functionality for blind (Abdolrahmani, Kuber, & Branham, 2018) and detecting diseases such as Parkinsons (Lahmiri & Shmuel, 2019; Wu, Zhang, Lu, & Guo, 2019).

The development has been driven by new methods of computing affective properties encoded in voice (Eyben et al., 2016) as well as a more nuanced understanding of user satisfaction (Kiseleva et al., 2016). Notably, the aim for much of the development is to make the machine voices as human-like as possible, both in terms of acoustic output and message content. This is highlighted in a review of Google's work with voice searches (Schalkwyk et al., 2010), where availability and performance are established as the main criteria, with the latter having shown staggering improvements with increasing amounts of data. Research has subsequently echoed these ambitions as key thresholds for mass user adoption (Coskun-Setirek & Mardikyan, 2017; Dubiel, 2018; Han & Yang, 2018).

Acoustic properties can influence how advertising messages are perceived. The decoding of acoustic properties appears to be similar for human and machine voices. However, the acoustic information machine voices such as VPAs are far easier to control and manipulate than human voices. This highlights the marketing potential of paying attention to *how* something is said (Spence, 2012). Let us now turn our attention to the importance of *what* is said.

2.2. Linguistic structures in marketing

2.2.1. An integrated framework of language complexity and processing mode

The linguistic content of a message can be understood as being processed somewhere along a continuum of automatic and controlled processing (Moors, 2016), in line with the ELM. Based on this, Pogacar et al. (2018) proposed a framework for understanding linguistic devices' effects on consumers. By categorizing previous studies of linguistic devices in terms of processing mode and language complexity, the framework connects the processing to attitudes and persuasion.

Linguistic devices are described as being automatically processed when they meet one or more criteria (e.g. uncontrollable, not affected by cognitive capacity), implying that controlled processing is distinguished by its constraining effects on cognition and attention (Pogacar et al., 2018). Many of the acoustic cues discussed earlier, such as brand names using vowels to signify size or speed, would thus classify as automatically processed, while a brand name describing the products explicitly will be processed in a controlled manner. To exemplify, consider two fictive brand names for a monster truck producer; *Baum Trucks* and *HugeTruck*. Both names will signify a large-sized truck, but the former does so through an automatic processing mode while the latter requires more cognition through a controlled processing. Importantly, we will not necessarily be aware of the effects of the sound symbolism in the word *Baum*, while the intentions with the name *HugeTruck* are obvious to us. This is the defining contrast between the automatic and controlled ends of the processing continuum (Pogacar et al., 2018).

The complexity level of a linguistic device depends on its length. A linguistic device spanning a sentence is more complex than for example the fictive truck brand names. A good example of complex linguistic devices is the use of so-called "dispreferred markers", where including elements of negative messages can increase the trustworthiness of the message itself (Hamilton, Vohs, & McGill, 2014). It is thus possible for a complex linguistic device to contain less complex linguistic devices.

Various positive effects can be achieved by using linguistic devices. Pogacar et al.'s (2018) framework highlight enhanced persuasion, increased trustworthiness and a reduced effect of negative information as potential effects from different linguistic devices. Clearly, marketers also need to pay attention to *what* is being said. The underpinning logic of the proverb that initiated the discussion of this theoretical framework can therefore be questioned. Instead, research indicates that both what we say and how we say it are important in how the message is decoded and understood. What then, is the nature of this interaction?

2.3. Comparing stimuli modality

2.3.1. Processing and level of involvement

Although the number of studies comparing stimuli modes is limited, there are a few that add to our understanding of how messages are processed. For instance, Chaiken and Eagly's (1983) experimental study of American psychology students found the likeability of the communicator to be more important for audio-based messages than for printed messages. This underlines the interactive effects between automatically processed information (e.g. acoustic cues) and information that goes through controlled processing (e.g. certain linguistic devices). This can be seen as deviations from the ends of the "elaboration continuum" proposed by the ELM (Petty & Cacioppo, 1984). However, an important factor in where on the continuum stimuli fall is the involvement level of the recipient, which is not explicitly discussed by Chaiken and Eagly (1983).

The studies including the level of involvement tend to focus on the communicator rather than the recipient. An experiment letting participants provide feedback on an article showed that more critical and complex feedback was delivered mainly through speech while suggestions on trivial improvements were delivered by text (Chalfonte, Fish, & Kraut, 1991), suggesting a preference for embedding additional acoustic information through using speech rather than text in high-involvement settings. However, a contradictory result was found in an exploratory study of managers from eleven different organizations showed a preference for using speech for messages with little information (Trevino, Lengel, & Daft, 1987). This could be explained by differences in the level of involvement. Robert and Dennis (2005) argue that in high-involvement social settings, such as speaking face-to-face, the motivation to process a message exceeds the cognitive ability to process it. Consequently, the processing thus falls closer to the peripheral route on the "elaboration continuum" of the ELM, relying more on acoustic cues than on actual linguistic content. It would then be natural for communicators to prefer speech over text for delivering negative feedback, as was found by Chalfonte et al. (1991). Decoding speech will be more dependent on the acoustic properties of the message than the actual negative content, making it easier for the communicator to "compensate" by embedding positive emotion through acoustic cues. Such a counteracting effect would be hard to achieve with text. Similarly, it would explain why communicators prefer speech for low-information messages. Embedding acoustic cues makes it more likely that a low-information message is processed predominantly by the peripheral route, which requires less cognitive effort. How then, do the differences in decoding different stimuli modes relate to VPAs?

A clear similarity is that even when the recipient is a VPA communicators tend to embed additional information when speaking compared to when writing. An initial study analyzed interactions with a basic document retrieval VPA. Findings underlined spoken queries as longer and closer to "natural language" (Crestani & Du, 2006).

However, the limited sample size (n=12) decrease the robustness of the findings. But later research with more substantial samples have shown similar results, both in the context of a recommendation VPA on a movie website (Kang et al., 2017) as well as based on 500.000 queries from Yahoo's mobile search application (Guy, 2016). The substantial statistical evidence both serves as a support of Crestani and Du's (2006) conclusions and as highlighting an important difference between humans and VPAs as recipients of vocal stimuli.

For simple spoken messages, humans still far surpass VPAs in decoding speed and accuracy, stemming from the ability to use peripheral cues to process the information. However, for spoken messages with exhaustive information, these cues limit the processing capability for humans who will to some extent disregard *what* is being said in favor of *how* it is being said. No such limit exists for VPAs, who can simultaneously process a information-heavy message in its entirety. Considering the rapid improvements of VPA technology (Schalkwyk et al., 2010), a VPA surpassing human capabilities of decoding speech appears only a question of time. However, this study is concerned mainly with human responses to VPA output and will thus not consider the developments of VPA technology extensively. The theoretical contributions on the processing of vocal stimuli, the effects of linguistic devices and comparative studies of stimuli modality form the bases for comparing consumers' responses to a VPA message. The question thus becomes how to measure the responses.

2.3.2. Measurements

Increasingly saturated media landscapes has increased efforts to "ensure" the effectiveness of marketing messages. For instance, research has looked at attitudes towards advertisements of wristwatches (MacKenzie & Lutz, 1989). The study also confirmed a relationship between attitude towards the advertisement and brand attitude, a link that had been proposed earlier (MacKenzie, Lutz, & Belch, 1986).
Let us therefore define:

A_R: Attitude towards the reminder

A_B: Attitude towards the retailer brand in the reminder

In addition to the measurements of attitude, it is relevant to connect behavioral outcomes to the processing of the message. For practical reasons, actual behavior cannot be observed. However, behavioral intentions have repeatedly been connected both to attitudes and actual behavior (Barry & Howard, 1990; Lavidge & Steiner, 1961). Therefore, let us define the following measures for behavioral intentions:

I_P: Intentions to carry out the reminder's suggested purchase

I_U: Intentions to use the reminder type for non-purchase purposes

2.3.3. Development of hypotheses

A purchase reminder should be understood as containing little information, which based on earlier research suggests that purchase reminders are better suited for speech than text output (Robert & Dennis, 2005; Trevino et al., 1987). In addition, Chaiken and Eagly's (1983) findings suggest that using linguistic devices correctly may lead to larger positive effects for spoken messages than written messages. On the basis of these earlier findings, it is therefore believed that for a given purchase reminder, consumers will exhibit more favorable attitudes if it is communicated through speech than through text. It is hypothesized that:

H1a: A_R is higher for voice-based purchase reminders than for text-based purchase reminders

Just like any message, purchase reminders can have varying levels of language complexity without differing in the amount of objective information. It is thus possible to create a low-complexity condition and a high-complexity condition in line with what is discussed in Pogacar et al.'s (2018) framework. Their work demonstrates how increasing language complexity by using linguistic devices can increase persuasion and perceived source credibility; which are both likely to increase the attitude towards the reminder (Pogacar et al., 2018). It is therefore believed that for a purchase reminder containing a given amount of objective information, a high-complexity language will elicit more favorable attitudes than a low-complexity language:

H1b: A_R is higher for high-complexity reminders than for low-complexity reminders

A question that follows naturally is how the attitude to the purchase reminder may be related to other types of attitudes. Notably, findings have consistently shown a positive relationship between attitudes towards advertising messages and attitudes towards brands (MacKenzie et al., 1986; Samson & Voyer, 2012; Stayman & Brown, 1992). In this context, a purchase reminder can arguably be understood to be similar to an advertising message in that it is an unprompted message directed towards a consumer with the objective of triggering a purchase. With this approximation, the breadth of earlier research connecting advertising message attitudes to brand attitudes is an indicator that such a relationship is likely to exist for purchase reminders as well. Therefore, it is hypothesized that:

H2: A_R is positively related to A_B

Similar to brand attitudes, purchase intention is another frequently used measure in marketing science, stemming from hierarchy of effects model (Lavidge & Steiner, 1961). Subsequent developments of the model have provided further nuance in the relationship between attitudes, intentions and behavior (Barry & Howard, 1990), generally pointing to a positive relationship between attitudes and behavioral intention.

Applying this knowledge to the framework of purchase reminders, it is thus reasonable to believe that the attitude towards the brand will be positively related to the purchase intention. Similarly, it can be hypothesized that the attitude towards the reminder is likely positively related to the intentions to use that reminder type for non-purchase ends.

H3: A_B is positively related to I_P

H4: A_R is positively related to I_U

2.3.4. Background variables

Technological knowledge (T_K)

One of the most well-known connections between technology and attitudes is that proposed in the Technology Acceptance Model, which highlights perceived ease of use and perceived usefulness as key factors for explaining usage of new technology (Davis, 1989). This study will instead explore the level of knowledge of consumer technology (T_K) as a potential moderator or mediator of usage intentions.

Gender

The previous discussion of the decoding of machine voices showed social norms as being equally present for the genderless VPAs (Nass et al., 1997). It has also been shown that male voices are more influential when presenting recommendations, as well as that gendered voices trigger social identification processes leading to conforming with gender stereotypical responses (Lee et al., 2000). This may be explained by the decoding of status and intention through vocal signals, which humans engage in just as do most animals (Cheng et al., 2016). Considering that commercial VPAs like Siri and Alexa tend to be female, comparing the different gender groups' responses will be an additional area of interest.

Background noise (B_N)

Noise levels can impact both our creative ability and our purchase intentions positively as well as negatively (Mehta, Zhu, & Cheema, 2012). While visual stimuli can be ignored by looking away, it is harder to filter out conflicting vocal stimuli, reducing ability to process. Therefore, background noise level (B_N) will be explored as a potential mediator of attitudes and intentions.

2.3.5. Summary of hypotheses

The below box provides a visual overview of the hypotheses that are to be tested. Next, the methodology of the study will be discussed.

H1a:	Consumers will exhibit more favorable attitudes towards voice-based reminders than towards text-based reminders
H1b:	Consumers will exhibit more favorable attitudes towards reminders with high-complexity language than towards those with low-complexity language
H2:	Attitude towards the reminder is positively related to brand attitude
H3:	Brand attitude is positively related to purchase intentions
H4:	Attitude towards the reminder is positively related to usage intentions

3. Methodology

This section details the underlying reasons for the choice of scientific approach and thoroughly describes the experimental design. The insights of a preparatory study are discussed briefly as part of the main study's development.

3.1. Scientific approach

The basis for the scientific approach used is a positivist assumption of visual and auditory stimuli containing objective and measurable characteristics that have nearly universal meanings. This stems from the linguistic research on crossmodality and vowel sounds that was previously discussed (Blasi et al., 2016; Knoeferle et al., 2017). In coherence with a deductive approach, the previously derived hypotheses were tested using an experimental study. The decision to conduct an experimental study was taken with consideration to the suitability of a factorial (in this case 2x2) framework to test consumer reactions to different stimuli modes. A contributing factor was inspiration from similar studies employing this method (Park, Stoel, & Lennon, 2008). To perform statistical testing of the hypothesis, data was collected in line with Bryman and Bell's (2015) suggestions for quantitative methods. Figure 3 illustrates the main components of the research paradigm that this study adheres to.

Figure 3. Visualization of research paradigm

Alternative methods included doing a qualitative explorative study, or an ethnographic study observing user behavior with VPAs. However, an ethnographic study in the Swedish market's early stage would limit the generalizability of the findings. Making a qualitative explorative contribution to the field would definitely be merited considering the extent of the research gap. At last, the experimental and quantitative method's possibility to narrow down and measure a well-defined consumer aspect of VPA technology was ultimately held as better for comparing stimuli modes since it deviated less from the realism ontology that much of the theoretical framework is built on.

3.2. Experiment design

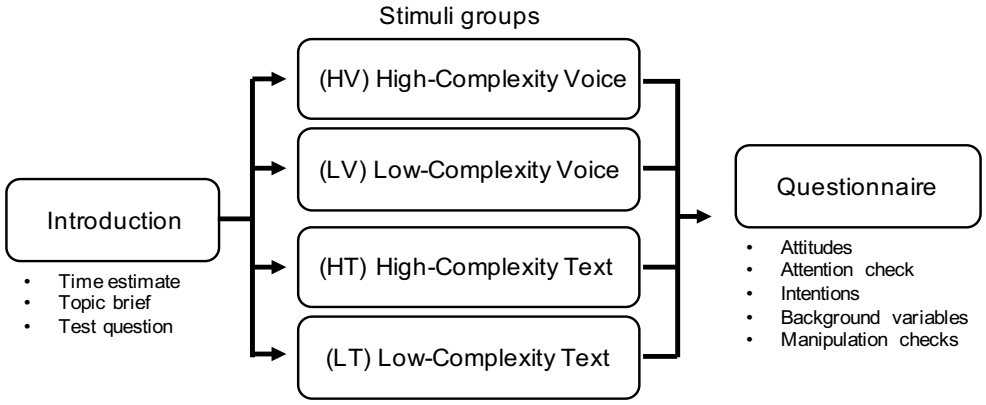


Figure 4. Overview of Experiment Design

The above figure highlights the three main components of the experiment. The experiment was constructed as a 2x2 framework where participants were randomly assigned to a stimuli group with a purchase reminder differentiated in terms of language complexity (high/low) and stimuli mode (voice/text). However, before being randomly assigned to a stimuli group, all respondents received the same introduction component, including a welcoming text detailing the topic and the estimated time of completion. In addition, the introduction component contained a test question asking participants to correctly identify a bell sound clip among three alternatives. This was done to minimize condition-dependent dropout for the voice stimuli groups (Zhou & Fishbach, 2016)).

Within the stimuli groups, respondents were instructed to imagine that they had themselves set the purchase reminder they received. The reminder was to buy batteries, with a suggested product (12-pack of AA-batteries by GP) from a suggested retailer (batteriexperten.com) at a determined price (89 kr. Including shipping). After exposure to the stimuli, all participants filled out the same questionnaire. The questionnaire mostly included Likert-scale questions designed to measure attitudes and intentions, but also included a "trap question" as an attention check as well as a section for background information and manipulation checks. Appendix III provides excerpts from the survey.

3.3. Stimuli development

3.3.1. Survey language and choice of product and brand

Even though the VPA technology is far more developed in English-speaking markets, the study was focused on Swedish-speaking consumers, and was thus entirely in Swedish. One reason for this was the suggested reluctance of Northern European consumers towards voice technology in public (Cherian & Pounder, 2017). In addition, the Swedish market was deemed interesting given its development, which can be seen through releases of new products (Wisterberg, 2019; Wittwång, 2019), trend reports highlighting the voice ecosystems (Kronborg Iversen, 2018) and researchers' prognostics (Juhlin, 2018) However, the decision was also taken with practical reasons in mind, as it was deemed easier to successfully carry out the language complexity manipulation in the author's native language.

The purchase reminder simulated a scenario of needing to remember to purchase batteries. Batteries were chosen both for the realistic nature of the scenario and for the high probability that most respondents would view this as a low-involvement purchase. Creating a low-involvement setting was desirable considering the previously discussed findings suggesting peripheral cues exert a stronger effect in such contexts (Chaiken & Eagly, 1983; Petty & Cacioppo, 1984). This was further enhance by choosing a retailer that was believed to be unknown for most consumers, which was later supported in a preparatory study.

3.3.2. Manipulation 1: Complexity level

The linguistic differences in the stimuli were developed based on Pogacar et al.'s (2018) framework on linguistic devices. To incorporate the linguistic devices into the stimuli, a low-complexity message was first developed; which is shown below:

> [Author's translation to English]
> Here's your reminder to buy batteries. My search generated the following alternative:
> 12-pack AA-batteries from GP for 89 kr including shipping from batteriexperten.com

Based on this low-complexity reminder, complex linguistic devices were subsequently incorporated to increase the complexity level based on Pogaacar et al.'s (2018) framework. This process is outlined in table 1.

Table 1. An overview of the incorporation of Pogacar et al.'s (2018) framework

Linguistic device	Complexity level	Processing mode	Text output	Suggested effects
Pronouns	Complex	Automatic	[…] annars får **du** slut på […] **Jag** hittade det […]	Induced self-referencing
Politeness	Complex	Automatic	**God förmiddag**! Kom ihåg […]	Lessens impact of negative information
Intensity	Complex	Automatic	God förmidag**! Kom ihåg** att […]	Increases perceived credibility of source if positive attitude
Analogy	Complex	Controlled	[…] slut på energi **lika snabbt som en student under tentaperioder.**	Enhances persuasion

In addition to the theoretically derived linguistic devices, inspiration was found in the trend of providing VPAs with capabilities of making jokes and contextual references (Borenstein, 2019; Must Share News, 2018; Titcomb, 2015), which has also been explicitly mentioned in patents (Surace et al., 2000). Recently, Google asked for Swedish customers to submit their best jokes and puns to be incorporated into their VPA (Bostrom, 2019). Consequently, the analogy in the reminder was developed as a humorous comparison to the stressful exam periods of university studies, generating the following message as the high-complexity condition.

> [Author's translation to English]
> Good morning! Remember to buy batteries today, or you'll run out of energy as quickly as a student during exam periods. I found this alternative for you: 12-pack AA-batteries from GP for 89 kr including shipping from batteriexperten.com

3.3.3. Manipulation 2: Stimuli modality

The voice-based reminder was developed by Acapela Group based on their standing on the Swedish-speaking market. Four different voice profiles were considered. Finally, a female profile named "Elin" was selected based on superior performance in the vocal output. Elin's voice was rather deep and stable for a female, but with a mean pitch high enough that it was easily identifiable as a female voice. For comparison, the recently

launched genderless VPA "Q" has a frequency range of 145-175 Hz (Genderless Voice, 2019). The voice profiles' frequency intervals are detailed further in table 2.

Table 2. Acoustic properties of Acapela's Swedish voices

Voice profile	Min pitch	Mean pitch	Max.
Elin (HV)	65 Hz	137 Hz	248 Hz
Elin (LV)	65 Hz	134 Hz	226 Hz
Elin (Overall)	64 Hz	144 Hz	366 Hz
Emma	66 Hz	185 Hz	457 Hz
Erik	65 Hz	110 Hz	237 Hz
Emil	65 Hz	117 Hz	269 Hz

So as to increase the degree of realism, the voice messages were complemented with a soft bell voice at the beginning and at the end of the message, similar to how e.g. Alexa functions. As the Qualtrics autoplay function appeared to limit OS compatibility, the audio clips required clicking the "Play" button. As the second-best alternative, the progress button was only revealed after the clip had been played, ensuring that clips were not ignored. The text-based reminders used a standard Qualtrics font. Corresponding to the bell sounds of the voice reminder, the text reminders included a header and a subheading (see Appendix III). Respondents were required to spend at least 10 seconds on the text stimuli page.

3.4. Insights from preparatory study

To test the survey design a preparatory study was conducted between March 27th and April 4th. The survey was shared through a private Facebook post and generated 56 valid responses. Appendix II contains a more detailed version of the responses' attitudes towards the stimuli. The findings of the preparatory study indicated that the manipulations had been interpreted correctly and that the attitudes different between stimuli groups. In addition, feedback from testers of the preparatory study underlined the relevance of including a measure relating to respondents' online concerns.

3.5. Main study

3.5.1. Measurements

Reminder Attitude (A_R) and Brand Attitude (A_B)

Immediately following exposure to the stimuli, respondents were asked to estimate their attitudes towards the reminder and the brand. This was measured through three-item Likert scale questions with the following three word pairs; bad-good, dislike-like and negative impression-positive impression. The antonyms were separated by 7 scale steps

ranging from e.g. 1=bad to 7 = good. The mean score of the three word pairs subsequently became the attitude measurement.

Purchase Intentions (I_P) and Usage Intentions (I_U)

In addition to the attitude measures, a 7-step Likert scale (*1 = completely disagree, 7 = completely agree*) was also the format used to approximate behavioral intentions. The purchase intentions were measured through presenting the respondent with two statements of intent:

> "I would make the purchase suggested in the reminder within one hour"
> "I would make the purchase suggested in the reminder the same day"

Since reminders are normally used to trigger action the same day, this was seen as a good approximation of purchase intentions in relation to the reminder.

Similarly, usage intentions were measured by letting respondents agree or disagree to statements of intended usage of the reminder type. The six different usages were cleaning, training, going to bed in time, remembering friends' birthdays, paying invoices and remembering scheduled meetings.

Technological knowledge (T_K)

Measuring subjective knowledge levels (e.g. technological know-how) comes with inherent uncertainty. Some respondents may overestimate their knowledge and others are likely to underestimate. To counter this, technological knowledge mas measured through separate questions estimating internal, external and objective evaluations of knowledge, as is suggested by Flynn and Goldsmith (1999). These are summarized in table 3.

Table 3. Overview of the different evaluations of T_K

Evalutation type	Question	Alternatives
Internal	"*I'm very knowledgeable when it comes to new technology*"	1 = completely disagree 7 = completely agree
External	"*Friends and family often turn to me with technological questions*"	1 = completely disagree 7 = completely agree
Objective	"*Click on the image* of a product developed by Google*"	A: Google Home B: Apple HomePod C: Amazon Echo

*The images used are attached in Appendix III

The last question was recoded so that the correct answer (A) resulted in a score of 7 and incorrect answers (B, C) were given 4 as a score. This enabled computing T_K as the average score of the three questions.

Online concerns (C_O)

Taking the feedback from testers of the preparatory study into account, the main study included a measure of respondents' online concerns. The measure consisted of three 7-step Likert scale questions placed late in the survey (*1 = Completely disagree, 7 = Completely agree)*;

> "I'm suspicious of large tech firms (e.g. Google, Apple and Amazon)"
> "I care a lot about my online integrity online"
> "Using smart assistants (like Siri or Alexa) increases the risk of being recorded"

3.5.2. Distribution

The distribution of the main study was more eventful than imagined, and is discussed more thoroughly in Appendix IV. The survey was reframed as a competition between three ugly cats dubbed "Fule Katt 2019" (Ugly Cat 2019). This included generating identities for the cats, creating a graphic profile for the competition, hosting a Facebook event, recruiting campaign teams for the different cats and printing campaign material. Inspiration was taken from New Zealand's "Bird of the year" competition, which annually attract attention, affection and donations to birds of varying beauty standards (Te Reo o Te Taiao, 2019). It was hoped that by providing an element of fun, respondents would be more inclined to share the survey with their friends.

The competition was launched on April 4th through a Facebook post and the creation of a public Facebook event. Within two days, the event had connected more than 700 Facebook members and the campaign teams had grown to engage 17 people. But the rapid sharing of the survey link resulted in a violation of Facebook's community rules, and the link was disabled on all Facebook-owned platforms. The Facebook event was shut down – proving a definite stop to the desired online spread of the survey. Fortunately, flyers with QR codes linking to the survey remained from the intense campaigning. The 460 flyers distributed between April 7th and April 15th only generated 36 valid responses. To combat time pressure and the risk of not yielding enough responses, 92 respondents from the online survey pool Prolific were paid £5.04/hour to complete the survey, leading to a total of 72 additional valid responses. Together with the 247 (87 valid) responses from the link, this last-resort strategy helped generate a sufficient respondent base for subsequent analysis.

3.5.3. Sampling of respondents

A convenience sample was used for this study, meaning that participants were mainly students or young professionals. All participants younger than 15 were omitted as kids have been shown to process certain auditory cues differently (Baxter & Lowrey, 2011; Baxter & Lowrey, 2014; Lopatovska & Williams, 2018). Since the study mainly aimed to investigate results of psychological processing valid for most adult consumers, the

convenience sample was not seen as a major drawback. Table 4 provides more information regarding the respondent sample:

Table 4. Overview of respondent sample

Variable	N	% of total sample
Gender		
Male	84	43.1%
Female	111	56.9%
Age		
15-20 years	22	11.3%
21-29 years	139	71.3%
30-40 years	20	10.3%
>40 years	16	8.2%
Occupation		
Student	130	66.7%
Employed	52	26.7%
Other	13	6.6%

Note: A more detailed version of this table is available in Appendix VI

3.5.4. Dropout analysis

Because of the test question in the survey introduction, a substantial dropout of respondents (28.5%) occurred between the introduction and the random assignment to a stimuli group. This meant that almost all participant dropouts occurred before a stimuli group was assigned. Only 3.5% of respondents dropped out after having been assigned a stimuli group.

3.6. Data analysis tools and tests

The surveys were distributed by the author using Qualtrics and later imported into SPSS for analysis. Dropout analysis and filtering was done in Microsoft Excel. As each stimuli group contained more than 30 unique respondents, a normal distribution was assumed (Bryman & Bell, 2015). A subsequent Levene's test for homogeneity of variance (see Appendix VIII) confirmed the suitability for a two-way analysis of variance (ANOVA) to be conducted for comparing stimuli group means (Bryman & Bell, 2015). Bivariate Pearson correlations were used as statistical tests for significant relationships between variables. Finally, analysis of possible mediation and/or moderation was conducted using model 1 of the PROCESS macro for SPSS (Hayes, 2013). A 5% significance level was used for all tests of significance in line with recommendations (Bryman & Bell, 2015). Using more stringent demands as proposed by some researchers (Dreber & Johanesson, 2018) was seen as desirable but infeasible.

3.7. Reliability and validity

3.7.1. Reliability

Filtering of respondents

The threat posed by low-quality responses to the reliability of measurements can be reduced with the help of various screening procedures (Jones, M., House, & Gao, 2015). A filtering process was thus established following the recommendations from Jones et al. (2015). The process included 6 stages; (1) speeding check, (2) discarding incomplete responses, (3) failed attention check, (4) straight lining, (5) manipulation check and (6) incompatible operating systems. Figure 5 provides an overview of how many responses were filtered out at what stage. A more detailed description of the filtering process is attached in Appendix V.

(1) The speeding check was developed through comparing the respondents' completion times to a benchmark time of 210 seconds. No upper limit on completion time was set considering the fact that many respondents are likely to open the link and start the survey but finish it later. In total, 41 respondents were discarded because of the speeding check.

(2) Responses were seen as incomplete if more than one question from the survey's main section were unanswered. No responses were discarded for unanswered background questions, such as age or gender. In total, 15 responses were discarded on the basis of incompletion.

(3) The survey included an attention check where participants were asked to select "completely disagree" to prove they were reading and answering the questions carefully. All eight responses failing to do so were subsequently discarded.

(4) One of the most common issues with Likert-scale questions is "straight lining", where a respondent rapidly selects the same alternative on the similar-looking scales (Jones et al., 2015). Out of the total 10 sections where straight lining could occur, participants straight lining through 3 or more sections were discarded. 10 responses (2.43%) qualified as straight liners.

(5) The manipulation checks placed late in the survey helped examine if a respondent could reasonably be believed to have understood the stimuli correctly. In total, the manipulation check stage of the filtering process led to 23 (5.60%) responses being discarded.

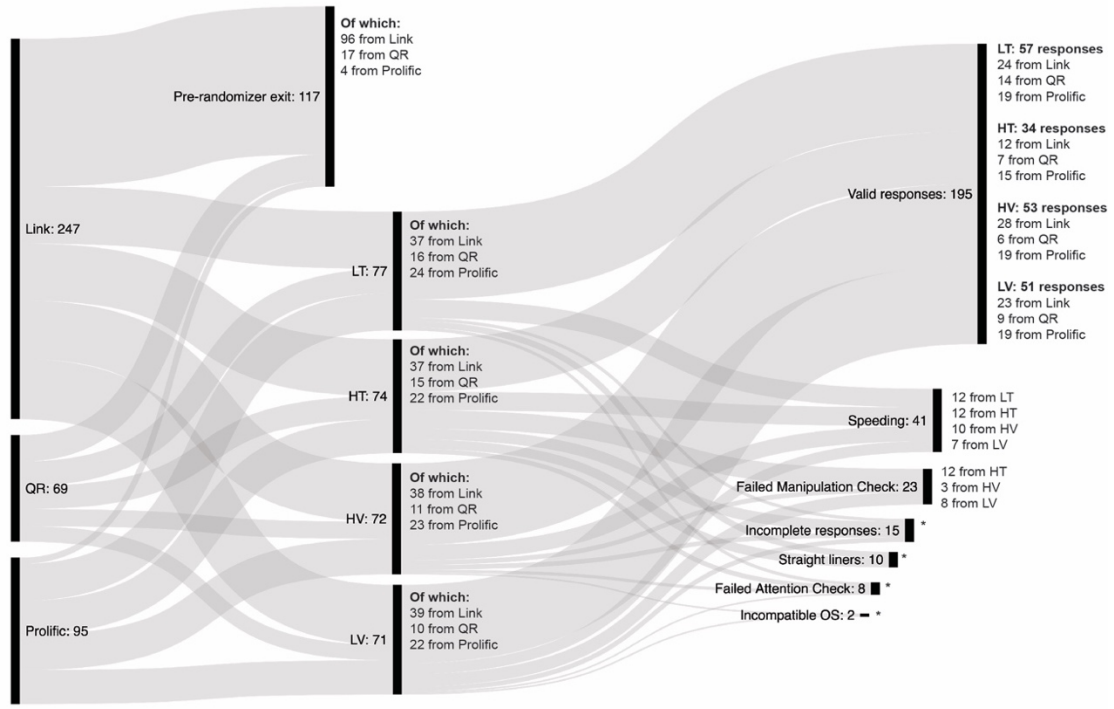


Figure 5: Sankey diagram of filtering process
(*Note*: A more detailed version of this diagram is available in Appendix V)

(6) Finally, it was discovered that some Android operating systems failed to play the audio files. Consequently, responses on the incompatible Android systems that were sorted into the voice condition were removed. 2 (0.49%) responses were removed because of this.

Ensuring random distribution

A question that remained is if the assignment of participants is to be considered random even in light of many dropouts (Zhou & Fishbach, 2016). Considering the large dropout (28.5%) before the assignment to stimuli groups, and the low dropout rate after assignment, it is unlikely that there have been any condition-specific dropouts. Even the relatively large filtering from stimuli group HT, as the responses removed were spread over multiple criteria and lacked common denominators such as a certain start date or device type.

Reliability of measurements

Most variables were measured using Likert scales. Using the parametric ANOVA to analyze Likert-scale items has been criticized on the basis of a faulty assumption of equidistant scale steps (Kaptein, Nass, & Markopoulos, 2010). However, since that critique mostly related to single item measurements, composite scores such as those used in this study remain suitable for parametric ANOVA. To test the reliability of the measures described previously, Cronbach's alpha was calculated. The results are shown in table 5.

Table 5: Internal consistency of measured variables

Variable	Number of items	α
A_R	3	0.938
A_B	3	0.903
I_U	6	0.785
I_P	2	0.787
C_O	3	0.618
T_K	3	0.703

As all but one $\alpha > 0.7$, it can be argued that the measurements of the dependent variables can be seen as reliable in line with academic norms (Bryman & Bell, 2015) The only variable not reaching the recommended level of internal consistency was online concerns (C_O), but this was not seen as a major issue given that it is a rather unexplored measure that is not central to the study.

Apart from internal consistency, the measurements' reliability can be examined in terms of test-retest reliability and inter-judge reliability. The former was seen as an unrealistic goal considering the scope of the study, and the latter was inapplicable for the experiment. The degree of internal consistency was thus held to be a sufficient indicator of the measurements' reliability.

3.7.2. Validity

This study does not aim to measure how the attitude is formed, which has been debated thoroughly as a both cognitive and affective process (Barry, 1987; Lavidge & Steiner, 1961). Instead, the study aims only to capture the end product of this evaluation; a likely combination of cognitive and affective processing. It can therefore be argued that the content validity of the attitude measurements is sufficient. The construct validity is sometimes described in terms in terms of nomological validity (Spiro & Weitz, 1990) and convergent validity (Cunningham, Preacher, & Banaji, 2001). Since exposing participants to a stimuli and immediately enquiring about their attitudes towards the stimuli has been a main feature of marketing research and theory for decades, varying both contextually and methodologically (MacKenzie & Lutz, 1989), measuring attitude towards a purchase reminder should be seen as having a high degree of nomological validity.

Behavioral intentions connect thought with action and have commonly been used to estimate actual behavior (Barry, 1987; Barry & Howard, 1990). As measuring actual behavior was deemed infeasible, this study only measures purchase and usage intentions. The Likert scale used to measure these intentions is thus considered to provide an acceptable degree of content validity. In terms of construct validity, the strong theoretical connection between attitudes and intentions should increase the overall construct validity of the measurements.

4. Results

In this section, the results of the experiment are presented and the hypotheses are tested chronologically. In addition, background variables are analyzed briefly.

4.1. Descriptive statistics

The experiment generated 195 valid responses after filtering low-quality responses. The average age of respondents was 26.7 years ($\sigma = 9.93$). Most respondents (91.8%) live in Sweden, and were either students (66.7%) or had full-time employment (26.7%). Most participants completed the survey on their mobile phone (55.9%), on a laptop (23.6%) or on a stationary computer (19%). Slightly more women (43.1%) than men (56.9%) completed the survey. Appendix VII provides an overview of the different stimuli groups. Below, tables 6 and 7 describes the descriptive statistics of the main measured variables for the total survey sample and the different groups.

Table 6. Overview of main measurements for the entire sample

Variable	μ	σ
Attitude towards the reminder (A_R)	4.38	1.56
Brand attitudes (A_B)	4.39	0.92
Purchase intentions (I_P)	3.74	1.62
Usage intentions (I_U)	3.86	1.36
Online concerns (C_O)	4.38	1.33
Technological knowledge (T_K)	5.02	1.26

Table 7. Descriptive statistics of attitudes and intentions between stimuli groups

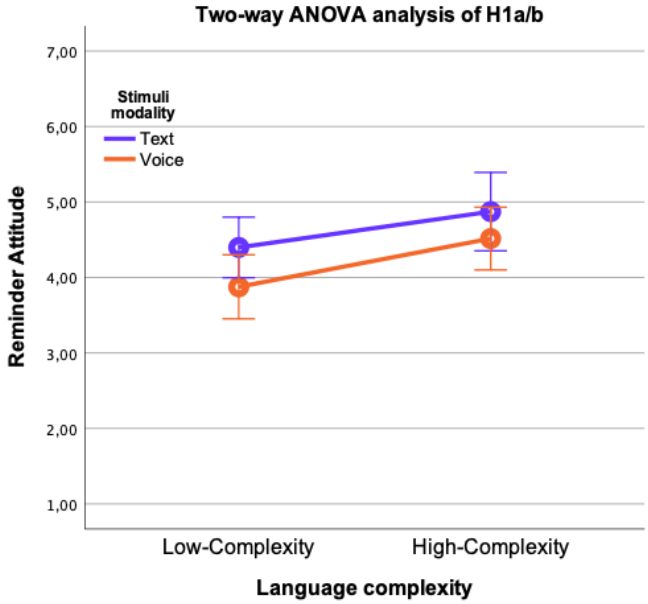
	LT n=57		LV n=51		HT n=34		HV n=53	
Attitudes and intentions	μ	σ	μ	σ	μ	σ	μ	σ
Attitude towards the reminder (A_R)	4.40	1.44	3.86	1.62	4.87	1.37	4.52	1.64
Brand attitude (A_B)	4.34	0.94	4.35	0.90	4.50	0.92	4.42	0.94
Purchase intentions (I_P)	3.61	1.53	3.67	1.76	4.06	1.67	3.76	1.56
Usage intentions (I_U)	3.93	1.19	4.08	1.27	4.05	1.19	3.44	1.64

Note: A more detailed version of this table is available in Appendix VII

4.2. Hypothesis testing

4.2.1. Differences in means: H1a/b

For H1a/b, a two-way analysis of variance (ANOVA) was calculated on respondents' attitudes towards the reminder. No significant main effect was found from the manipulation of stimuli modality, $F(1, 195) = 3.83$, $p = .052$. H1a was thus not supported. However, a significant main effect resulted from the language complexity manipulation, $F(1, 195) = 6.17$, $p < .05$. H1b was therefore supported. In addition, no significant interaction effect was reported, $F(1, 195) = .135$, $p = .714$. The results are shown visually below, with error bars at 95% confidence intervals.



4.2.2. Testing relationships: H2-H4

The hypotheses regarding relationships between the variables were tested using bivariate Pearson correlations with two-tailed tests for significance. The results are summarized in table 8.

Table 8. Overview of bivariate Pearson correlation coefficients

Variable	A_R	A_B	I_P	I_U
Attitude towards the reminder (A_R)	1.00			
Brand attitudes (A_B)	0.288***	1.00		
Purchase intentions (I_P)	0.333***	0.231**	1.00	
Usage intentions (I_U)	0.446***	0.177*	0.305***	1.00

Note: * $p < .05$ ** $p < .01$ *** $p < .001$

This shows that there is a significant positive relationship between A_R and A_B, $r(191) = .288$, $p < .001$. H2 was thus supported. Similarly, a positive correlation between A_B and I_P could be established, $r(191) = .231$, $p < .001$. H3 was thus supported. Finally, a positive relationship between A_R and I_U, $r(191) = .446$, $p < .001$, supporting H4.

In addition, a significant positive correlation, $r(191) = .333$ $p < .001$ was found between A_R and I_P. A similar correlation was found between I_U and I_P, $r(191) = .305$, $p < .001$.

4.3. Background variables and moderating effects

4.3.1. Gender

There were no significant differences between male and female A_R scores in the overall sample $F(193) = 3.690$, $p = .056$. The only stimuli group with a significant difference, $F(51) = 4.837$, $p < .05$, in A_R between male and female participants was the high-complexity voice condition. The findings are summarized in table 9.

Table 9. ANOVA of gender differences in A_R scores

Group	N	μ	σ	F	p
Total	*195*	*4.38*	*1.56*	*3.690*	*.056*
High-complexity voice	53	4.52	1.64	4.837	< .05
Low-complexity voice	51	3.86	1.62	2.241	.141
High complexity text	34	4.87	1.37	0.467	.499
Low-complexity text	57	4.40	1.44	0.472	.495

Note: A more detailed version of this table is available in Appendix VIII

In addition, a moderator analysis was run to see if gender exerts a moderating effect on the positive relationship between A_R and I_U. No such moderating effect was found either for the overall sample, $F(1,191)=1.218$, $p = .27$, for the voice condition $F(1,100)=0.374$, $p = .542$ or for the high-complexity voice stimuli group $F(1,49)=0.125$, $p = .726$.

4.3.2. Technological knowledge

It was believed that technological knowledge might be an important factor in determining the attitude towards the reminder as well as the usage intentions. However, analysis showed that there is no significant correlation between T_K and A_R, $r(195) = -0.05$, $p = .50$. The same was found to be true regarding I_U, $r(195)=0.12$, $p = .085$.

T_K was also tested as a potential moderator variable between A_R and I_U. This was not supported statistically, $F(1,191) = 2.74$, $p = .10$. However, it was shown that T_K moderated the effect between A_B and I_P, $F(1,191) = 5.59$, $p < .05$.

4.3.3. VPA usage

In general, participants were not frequent VPA users. Almost two thirds (62.1%) reported never using VPAs. Among those using VPAs, only 21.6% indicated daily usage of the assistants. Table 10 provides an overview of the usage. Notably, the most popular usage was information searches and weather forecasting. Only 4 (5.41%) respondents reported using VPAs to make purchases.

Table 10. Overview of VPA usage among participants

Usage	N	%
Freq. of usage	*195*	*100%*
Daily	16	8.3%
Weekly	25	12.8%
Monthly	33	16.9%
Never	121	62.1%
Usage form	*74*	*100%*
Information search	46	62.2%
Other	40	54.1%
Weather forecast	30	40.5%
Scheduling assistance	16	21.6%
Purchasing	4	5.41%

An ANOVA analysis showed that there was no significant difference in A_R between VPA users and non-users, $F(1, 193) = 0.457$, $p = .50$. Similarly, VPA Users were not found to be more inclined to use the reminder.

4.3.4. Online concerns

Analyzing correlations between online concerns and the other key variables, it was found that C_O only was correlated significantly to T_K, $r(195) = 0.144$, $p < .05$. C_O did not have a moderating effect on any other correlations. Through an ANOVA analysis it could furthermore be established that C_O did not significantly differ between the voice and text condition $F(1,193) = 0.447$, $p = .50$, nor between the high- and low-complexity conditions $F(1,193) = 2.783$, $p = .097$.

4.3.5. Background noise

The level of background noise was not found to significantly be related to any of the other measurements, or moderate any of the previously defined relationships.

4.4. Manipulation checks

To analyze whether the manipulations had been interpreted correctly, test of between-subjects effects were conducted for both manipulations. It could be shown that the manipulation of medium had the desired effect on perceived medium exposure $F(1, 191) = 6658.3$, $p<.01$, with no other source showing significant effects. Similarly, the manipulation of language complexity had the desired effect on perceiving the joke $F(1, 191) = 19054.1$, $p <.01$. This can be seen in table 11.

Table 11. Tests of between-subjects effects for the manipulations

Source of effect	F	p
Test for stimuli modality manipulation		
Intercept	1.857	.403
Manipulation of complexity	0.131	.779
Manipulation of medium	66538.3	< .01
Interaction	1.054	.306
Test for manipulation of language complexity		
Intercept	1.952	.395
Manipulation of medium	7.107	.228
Manipulation of complexity	19054.1	< .01
Interaction	0.190	.663

4.5. Summary of hypothesis testing

The below table provides an overview of what results the hypothesis testing generated.

Table 12. Summary of hypotheses and testing results

H1a	Consumers will exhibit more favorable attitudes towards voice-based reminders than towards text-based reminders	Not supported
H1b	Consumers will exhibit more favorable attitudes towards reminders with high-complexity language than towards those with low-complexity language	Supported
H2	Attitude towards the reminder is positively related to brand attitude	Supported
H3	Brand attitude is positively related to purchase intentions	Supported
H4	Attitude towards the reminder is positively related to usage intentions	Supported

5. Discussion

The purpose of this study was to compare consumers' attitudes and behavioral intentions for different types of VPA-like reminders. Contrary to what was predicted, there was no empirical support for differences in attitudes towards the reminder based on the stimuli mode. However, the hypothesis that reminders with high-complexity language would elicit more favorable reminder attitudes was supported. The findings also indicated a positive correlation between reminder attitude and the usage intentions as well as brand attitude, as was hypothesized. Similarly, the results confirmed a positive relationship between brand attitudes and purchase intentions. In addition, the results of the study showed a positive correlation between usage intention and purchase intentions.

5.1. Effects of stimuli modality and language complexity

The main outcome of the experiment was that there are no significant differences between attitudes towards text-based reminders and voice-based reminders. To some extent, this stands in contrast to the findings by Chaiken and Eagly (1983), who found that communicator likeability is more important than argument content for audio messages. It would therefore be expected that the language complexity would be less important for voice reminders than for text reminders. On the contrary, the difference in attitudes between high- and low-complexity language reminders was larger in the voice condition than in the text condition, although no significant interaction effect was found. While this study does not explicitly consider communicator likeability, the findings indicate that the content of an argument is not less important for voice-based messages than for text-based messages.

The positive effects on persuasion suggested by Pogacar et al.'s (2018) framework appeared to exert their hypothesized effects in the context of purchase reminders as well, since attitudes towards the reminder were higher for high-complexity reminders than for low-complexity reminders. The fact that this difference was more pronounced for voice reminders than for text reminders suggests that the effect of linguistic devices may be greater when spoken than when written. One possible explanation for this is the processing mode of such linguistic features. Following Pogacar et al.'s (2018) categorization of processing mode, the high-complexity condition included both controlled and automatic linguistic devices. The additional information embedded through acoustic cues may increase the effectiveness of the linguistic devices that are automatically processed. However, since no significant interaction effect was found between the stimuli groups, it is important to not draw unsupported conclusions regarding the differences between written and spoken linguistic devices.

In addition, even though the effect of the high-complexity language was larger for the voice condition than for the text condition, the attitude towards the HV reminder was still not higher than towards the HT reminder. Here, we should revisit Chaiken and Eagly's (1983) suggestion of communicator likeability as an important factor of persuasiveness in voice-based messages. It is entirely possible that the voice profile, Elin, was perceived as rather unlikeable, while the neutral text is unlikely to have elicited negative sentiments.

5.2. Intentions and commitments

The results confirmed the positive relationship between brand attitudes and purchase intentions that has repeatedly been reported in marketing research. This echoes earlier theoretical contributions connecting attitudes and behavioral intentions, such as the hierarchy of effects-model (Barry, 1987; Lavidge & Steiner, 1961). These well-known relations are arguably less interesting than analyzing the novel aspects of VPA technology. Given the current pace of VPA technology development, this study also set out to better understand consumers' intentions to use different types of reminders for different types of tasks. The results indicated that the attitudes towards the reminder is a predictor of usage intentions.

Within the usage intention items, there were some clear differences. Regardless of reminder type, respondents had higher intentions of using the reminder for remembering scheduled meetings, invoices to be paid and friends' birthdays. The intentions to use the reminder for remembering to clean, going to bed in time and remembering workouts were significantly lower. It is possible that this reflects human behavioral patterns to some extent. More specifically, it could show how we perceive responsibility and commitment differently depending on who will be impacted from failure to complete a task.

Meetings, invoices and friends' birthdays all reflect commitments towards other parties. Failure to meet expectations can directly affect relationships. Conversely, failing to remember to work out, getting enough sleep and cleaning can be understood to have a less direct effect on relationships, and more of a direct effect on our physical wellbeing. It is possible that we are reluctant to accept a reminder asking us to do things we expect ourselves to automatically remember, as it may serve as a reminder that we have procrastinated such tasks earlier. The reminder may actually remind us of our own failure more than serving as a helpful tool. External commitments thus appear more suitable for reminders than do internal commitments.

5.3. Understanding decoding differences

Decoding a message is highly individual, and it was therefore of interest to analyze differences in attitudes and intentions between different types of consumers. In the overall sample, there were no significant differences between male and female participants in attitudes or purchase intention. However, females had significantly higher intentions to use the reminder for external commitments. In addition, male participants showed a significantly higher degree of technological knowledge, as well as more pronounced online concerns.

Within the voice condition, females were more positive towards the reminder and found it more personal than did male respondents. No such differences were found within the text condition. This could possibly be explained by the findings by (Mitchell et al., 2011) showing that females prefer female voices, although the voice used was close to what is described as the "genderless" frequency range (Genderless Voice, 2019) (Genderless Voice, 2019). However, the results differ in that Mitchell et al. (2011) based their conclusions on an implicit association test (IAT) to measure direct attitudes towards the voice, while this experiment explicitly used Likert scales to measure direct attitudes towards the reminder.

Knowledge of consumer technology was not found to be an important variable in predicting attitudes or behavior. If technological knowledge is assumed to be positively related to perceived ease of use and perceived usefulness, this somewhat contradicts technology acceptance model (Davis, 1989). Similarly, it would be expected that usage intentions are higher for those with technological knowledge, which was not supported. Instead, technological knowledge only had a slight moderating effect between brand attitudes and purchase intentions. Online concerns varied only between male and female participants and had no moderating effects. Most participants did not use VPA technology at present, and for those who did, no difference was observed in terms of attitudes or behavioral intentions. This may be explained by the relatively low rates of VPA usage reported by the respondents. Given the early stage of the VPA market development in Sweden, it is possible that even technologically savvy consumers feel somewhat estranged to the newly released voice assistants. If this is the case, replicating this study in a later stage should generate results showing a stronger prediction value of technological knowledge in predicting reminder attitudes and behavioral intentions.

6. Conclusions and implications

This study investigated how consumer attitudes and intentions differ towards purchase reminders depending on the delivery medium and the language complexity. In conclusion, no significant differences were found from the different medium, while high-complexity language was shown to have a positive effect on attitudes towards the reminder. The attitude towards the reminder, in turn, was shown to be positively related to usage intentions, purchase intentions and brand attitudes.

To return to our proverb; in Elin's case it appears that *what* is said is more important than *how* it is said.

6.1. Theoretical contribution

This study extended the theoretical framework on comparing stimuli modality to a smart assistant context, adding to research that has focused mainly on comparing print messaged with audio or video messages (Chaiken & Eagly, 1983). Given the pace of technological development, existing concepts like attitudes and intentions will regularly be in need of analysis in new contexts, such as that of smart assistants. In addition, the study focuses more on the human decoding of a stimuli from smart assistants rather than on a whole interaction or on the input, adding to our understanding of the complex and evolving relationship between humans and smart assistants. However, the findings do not indicate any significant differences between attitudes towards text and voice reminders based on the medium alone.

However, the results contribute to the understanding of how linguistic structures and communication medium interact. In Pogacar et al.'s (2018) framework, instead of discussing the possibility of linguistic devices having stronger effects when delivered vocally, acoustic cues are described as a type of linguistic devices. Conversely, studies focusing on the impacts of acoustic cues make little regard to interactions between verbal and vocal content (Gelinas-Chebat & Chebat, 1992)). The findings of this study contribute to a better understanding of the interaction between delivery medium and language complexity by showing that the linguistic devices appear to be more effective when delivered vocally, even in the case of the rather neutral voice used in the experiment. This indicates a potential connection between Chaiken and Eagly's (1983) emphasis on communicator likeability, research on linguistic devices (Pogacar et al., 2018) and research focusing only on vocal properties of voice. In addition to providing more nuance to the relations between these research fields, this study adds to our understanding of linguistic devices through including an element of humor, which was omitted from Pogacar et al.'s (2018) framework. Here, the effectiveness of high-complexity language in purchase reminders should be understood as an indicator of the effectiveness of humor.

By showing that attitude towards the reminder is positively related to usage intentions for the reminder, the study adds to the existing body of research connecting attitude to behavioral intentions. Previous research has mainly revolved around connecting purchase intentions, brand attitudes and attitudes towards an advertisement (Barry, 1987). By putting these concepts into a VPA setting, this study aids in bringing existing theory into a contemporary context. In addition, the findings provide some insights into consumer preferences for VPA usage, which remains an area researched mainly by producers of smart assistants and consultancy firms. By showing a clear divide in preferences between reminders of external and internal commitment, this study indicates an opening to better understanding the 21st century consumer, who will likely interact with virtual entities on a daily basis. As usage becomes more widespread, it is possible that acceptance for reminders of internal commitment will also become more common.

6.2. Practical implications

For businesses working with conversational computing and the manufacturing of smart assistants, the findings in this study should reaffirm the high potential of a VPA with an almost-human level of linguistic capabilities. Participants were more positive towards reminders with high-complexity language, which translates into more favorable brand attitudes as well as stronger intentions to use both the reminder technology and to make the suggested purchase. Achieving a higher degree of language complexity is best done through gathering large amounts of data to train the algorithms, a process that has been ongoing for the last decade (Schalkwyk et al., 2010).

From a consumer perspective, the results are twofold. On the one hand, the findings suggest that we do not view voice-based purchase reminders differently than text-based purchase reminders. This supports the notion of a rational consumer, who evaluates arguments based on content and not peripheral cues. However, that very notion is also challenged by the fact that respondents were significantly more positive to purchase reminders using linguistic devices. This preference indirectly translates to purchase intentions, as the attitudes to the reminder were shown to be positively correlated to purchase intentions. Holding product information and price constant, consumers may be more likely to carry out the purchase if the reminder uses linguistic devices effectively. This is more in line with the idea of the irrational consumer using peripheral cues to make decisions. As this difference was larger for voice-based reminders than for text-based reminders, it begs the question how well-equipped consumers are to make rational purchase decisions in the face of an increasingly sophisticated voice technology.

This uncertainty combined with the high business potential of voice technology developments makes it important to consider societal implications of research in this domain. The results of this study indicate that even a technically solid voice assistant profile like Elin is not very potent in eliciting favorable attitudes that may translate into VPA usage. Instead, it will be important to pair a high-quality voice output with highly complex language skills to get users to adopt the technology. In the "race for voice" it is thus the largest of the large that are best suited to achieve success given their superior access to data – contrary to what is often argued (Richards, 2018).

6.3. Limitations and future research

VPA technology is inherently interactive, rocking back and forth between consumer input and VPA output. Since this study only considered responses to output, it does not provide an overall understanding of consumer-VPA interactions. Further limitations to the ecological validity include simulating a voice-based medium through an online survey that is inherently visual and text-based. In addition, the respondents were not frequent VPA users, and less than 6% had mad purchases through VPAs. By researching actual consumer behavior, such drawbacks can be overcome. Ideally, future research would form partnerships with VPA producers to gain access to the data recorded of actual behavior.

Another critique concerns the assumption of social presence. It is assumed that the level of social presence is higher for the voice condition than for the text condition, since those listening to the voice-based reminders cannot easily process additional vocal stimuli simultaneously. But all participants were engaged in identical online surveys except the stimuli, and all underwent the same voice-based test question after the survey introduction. This raises questions whether that assumption actually holds. However, that remains difficult to answer considering the insignificant difference between the voice and text condition. In addition, since some participants lived in countries other than Sweden, it is possible that they perceive the level of social presence differently and have different preexisting ideas influencing their attitudes towards VPA reminders. Future research could therefore attempt to see how attitudes to vocal and visual stimuli differ in settings with more clearly differentiated levels of social presence.

Finally, the study only uses a single voice profile, that is rather neutral. This limits the understanding of the interaction effect between acoustic cues and linguistic devices, which remains underexplored in the absence of additional research. Furthermore, only using a female voice makes it more difficult to draw conclusions on the relation between gender and automatic decoding. Future research would benefit from analyzing several different voice profiles in a commercial context.

Voice technology may fundamentally change consumer behavior, and not necessarily for the better. Understanding these effects needs to become a top priority in marketing.

References

Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). Siri talks at you: An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. Paper presented at the *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility,* 249-258. doi:10.1145/3234695.3236344 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056817246&doi=10.1145%2f3234695.3236344&partnerID=40&md5=84cfd011b535c4d89210b477393d8e8c

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715-727. doi:10.1037/0022-3514.37.5.715

Argo, J. J., Popa, M., & Smith, M. C. (2010). The sound of brands. *Journal of Marketing, 74*(4), 97-109. doi:10.1509/jmkg.74.4.97

Baggs, M. (2019). Gillette faces backlash and boycott over '#MeToo advert'. Retrieved from https://www.bbc.com/news/newsbeat-46874617

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614-636. doi:10.1037/0022-3514.70.3.614

Barry, T. (1987). The development of the hierarchy of effects: An historical perspective. *Current Issues and Research in Advertising, 10*(1-2), 251-295. doi:10.1080/01633392.1987.10504921

Barry, T., & Howard, D. (1990). A review and critique of the hierarchy of effects in advertising. *International Journal of Advertising, 9*(2), 121-135. doi:10.1080/02650487.1990.11107138

Baxter, S., & Lowrey, T. (2011). Phonetic symbolism and children's brand name preferences. *Journal of Consumer Marketing, 28*(7), 516-523. doi:10.1108/07363761111181509

Baxter, S., & Lowrey, T. (2014). Examining children's preference for phonetically manipulated brand names across two english accent groups. *International Journal of Research in Marketing, 31*(1), 122-124. doi:10.1016/j.ijresmar.2013.10.005

Bell, V. (2019). Alexa IS listening to your conversations. Retrieved from https://www.dailymail.co.uk/sciencetech/article-6910791/Alexa-listening-conversations.html

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America, 113*(39), 10818-10823. doi:10.1073/pnas.1605782113

Borenstein, S. (2019). A robot walks into a bar, doesn't get the joke: Struggling to teach humor to AI. Retrieved from https://www.latimes.com/business/technology/la-fi-tn-artificial-intelligence-humor-jokes-20190401-story.html

Bostrom, T. (2019). Google ber om svenskarnas bästa skämt – smarta högtalarna ska bli roligare. Retrieved from https://www.breakit.se/artikel/20154/google-ber-om-svenskarnas-basta-skamt-smarta-hogtalarna-ska-bli-roligare

Bryman, A., & Bell, E. (2015). *Business research methods* (Fourth edition ed.). Oxford: Oxford University Press.

Chaiken, S., & Eagly, A. H. (1983). Communication modality as a determinant of persuasion: The role of communicator salience. *Journal of Personality and Social Psychology, 45*(2), 241.

Chalfonte, B. L., Fish, R. S., & Kraut, R. E. (1991). Expressive richness: A comparison of speech and text as media for revision. Paper presented at the *Conference on Human Factors in Computing Systems - Proceedings,* 21-26. doi:10.1145/108844.108848 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84900563191&doi=10.1145%2f108844.108848&partnerID=40&md5=9dea0166e2ab91cc71b06b3dfed85f70

Chattopadhyay, A., Dahl, D. W., Ritchie, R. J. B., & Shahin, K. N. (2003). Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology, 13*(3), 198-204. doi:10.1207/S15327663JCP1303_02

Chen, C., Huang, T., Park, J. J., Tseng, H., & Yen, N. Y. (2014). A smart assistant toward product-awareness shopping. *Personal and Ubiquitous Computing, 18*(2), 339-349. doi:10.1007/s00779-013-0649-z

Chen, Y., & Spence, C. (2018). Audiovisual semantic interactions between linguistic and nonlinguistic stimuli: The time-courses and categorical specificity. *Journal of Experimental Psychology. Human Perception and Performance, 44*(10), 1488-1507. doi:10.1037/xhp0000545

Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General, 145*(5), 536.

Cherian, E., & Pounder, J. (2017). *Speak easy: The future answers to you.* (). London: J. Walter Thompson.

Clark, K. (2019). Here's mary meeker's 2019 internet trends report – TechCrunch. Retrieved from https://search.proquest.com/docview/2238168165

Coskun-Setirek, A., & Mardikyan, S. (2017). Understanding the adoption of voice activated personal assistants. *International Journal of E-Services and Mobile Applications, 9*(3), 1-21. doi:10.4018/IJESMA.2017070101

Coulter, K. S., & Coulter, R. A. (2010). Small sounds, big deals: Phonetic symbolism effects in pricing. *Journal of Consumer Research, 37*(2), 315-328. doi:10.1086/651241

Crestani, F., & Du, H. (2006). Written versus spoken queries: A qualitative and quantitative comparative analysis. *Journal of the American Society for Information Science and Technology, 57*(7), 881-890. doi:10.1002/asi.20350

Creswell, J. (2018). How amazon steers shoppers to its own products. Retrieved from https://search.proquest.com/docview/2058167011

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychol Sci, 12*(2), 163-170. doi:10.1111/1467-9280.00328

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly,* , 319-340.

De Renesse, R. (2017). Virtual digital assistants to overtake world population by 2021. Retrieved from https://ovum.informa.com/resources/product-content/virtual-digital-assistants-to-overtake-world-population-by-2021https://ovum.informa.com/resources/product-content/virtual-digital-assistants-to-overtake-world-population-by-2021

Dreber, A., & Johanesson, M. (2018). Vilka forskningsresultat kan vi lita på? *Ekonomisk Debatt, 46*(3), 17-28. Retrieved from http://www.econis.eu/PPNSET?PPN=1028139535

Dubiel, M. (2018). Towards human-like conversational search systems. Paper presented at the *Proceedings of the 2018 Conference on Human Information Interaction \& Retrieval,* New Brunswick, NJ, USA. 348-350. doi:10.1145/3176349.3176360 Retrieved from http://doi.acm.org/10.1145/3176349.3176360

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., . . . Truong, K. P. (2016). The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing, 7*(2), 190-202. doi:10.1109/TAFFC.2015.2457417

Flynn, L. R., & Goldsmith, R. E. (1999). A short, reliable measure of subjective knowledge. *Journal of Business Research, 46*(1), 57-66.

Gelinas-Chebat, C., & Chebat, J. (1992). Effects of two voice characteristics on the attitudes toward advertising messages. *Journal of Social Psychology, 132*(4), 447-459. doi:10.1080/00224545.1992.9924724

Genderless Voice. (2019). Meet Q. Retrieved from http://www.genderlessvoice.com/

Giancaspro, M. (2019). 'Alexa, call my lawyer!' are you legally liable if someone makes a purchase using your virtual assistant? Retrieved from http://theconversation.com/alexa-call-my-lawyer-are-you-legally-liable-if-someone-makes-a-purchase-using-your-virtual-assistant-109075

Guèvremont, A., & Grohmann, B. (2015). Consonants in brand names influence brand gender perceptions. *European Journal of Marketing, 49*, 101-122. doi:10.1108/EJM-02-2013-0106

Gunasti, K., & Ross, W. (2010). How and when alphanumeric brand names affect consumer preferences. *Journal of Marketing Research, 47*(6), 1177-1192. doi:10.1509/jmkr.47.6.1177

Guy, I. (2016). Searching by talking: Analysis of voice queries on mobile web search. Paper presented at the *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval,* 35-44. doi:10.1145/2911451.2911525 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84980319844&doi=10.1145%2f2911451.2911525&partnerID=40&md5=2ffa5246de361e3607868198d209eff7

Guy, I. (2018). The characteristics of voice search: Comparing spoken with typed-in mobile web search queries. *ACM Transactions on Information Systems, 36*(3) doi:10.1145/3182163

Hamilton, R., Vohs, K. D., & McGill, A. L. (2014). We'll be honest, this won't be the best article you'll ever read: The use of dispreferred markers in word-of-mouth communication. *Journal of Consumer Research, 41*(1), 197-212. doi:10.1086/675926

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management and Data Systems, 118*(3), 618-636. doi:10.1108/IMDS-05-2017-0214

Handley, L. (2017). Kendall jenner's pepsi ad pulled after twitter backlash. Retrieved from https://www.cnbc.com/2017/04/05/kendall-jenners-pepsi-ad-causes-twitter-storm-labeled-disrespectful.html

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York [u.a.]: Guilford Press. Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=025778167&sequence=000004&line_number=0002&func_code=DB_RECORDS&service_type=MEDIA

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). *Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet?* doi://doi.org/10.1002/dir.10073

Hill, K. (2012). How target figured out a teen girl was pregnant before her father did. Retrieved from https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6fd305066686

Hobley, R., & Gashe, V. (2019). David beckham launches the world's first voice petition to end malaria. Retrieved from https://www.malariamustdie.com/news/david-beckham-launches-worlds-first-voice-petition-end-malaria

IBM Archives. (2003). IBM archives: IBM shoebox. Retrieved from //www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html

Jeong, J., & Shin, D. (2015). It's not what it speaks, but it's how it speaks: A study into smartphone voice-user interfaces (VUI). Paper presented at the *Human-Computer Interaction: Interaction Technologies,* 284-291.

Jones, M., House, L., & Gao, Z. (2015). Respondent screening and revealed preference axioms: Testing quarantining methods for enhanced data quality in web panel surveys. *Public Opinion Quarterly, 79*(3), 687-709.

Jones, S. (2019). Amazon's latest smart speaker presents engagement, commerce opportunities for brands. Retrieved from https://www.luxurydaily.com/amazons-latest-smart-speaker-presents-engagement-commerce-opportunities-for-brands/

Jonze, S. (Producer), & Jonze, S. (Director). (2013). *Her.* [Video/DVD] Burbank, California, USA: Warner Bros. Pictures.

Juhlin, J. (2018). Digitala assistenter spås ta över – överallt. Retrieved from https://www.svt.se/nyheter/inrikes/roststyrningen-tar-over-eller

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*(4), 381-412. doi:10.1037/1528-3542.1.4.381

Kang, J., Condiff, K., Chang, S., Konstan, J. A., Terveen, L., & Harper, F. M. (2017). Understanding how people use natural language to ask for recommendations. Paper presented at the *RecSys 2017 - Proceedings of the 11th ACM Conference on Recommender Systems,* 229-237. doi:10.1145/3109859.3109873 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85030449678&doi=10.1145%2f3109859.3109873&partnerID=40&md5=2a0a11a13f5b61db4d7b2056af19fd2f

Kaptein, M. C., Nass, C., & Markopoulos, P. (2010). Powerful and consistent analysis of likert-type ratingscales. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2391-2394.

Kiseleva, J., Williams, K., Jiang, J., Awadallah, A. H., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding user satisfaction with intelligent assistants. Paper presented at the *CHIIR 2016 - Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval,* 121-130. doi:10.1145/2854946.2854961 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84974602953&doi=10.1145%2f2854946.2854961&partnerID=40&md5=723bb77b677284f04d5d7cbfc905a2fc

Klink, R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters, 11*(1), 5-20. doi:1008184423824

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports, 7*(1), 5562. doi:10.1038/s41598-017-05965-y

Kocielnik, R., Avrahami, D., Marlow, J., Lu, D., & Hsieh, G. (2018). Designing for workplace reflection: A chat and voice-based conversational agent. Paper presented at the *Proceedings of the 2018 Designing Interactive Systems Conference,* Hong Kong, China. 881-894. doi:10.1145/3196709.3196784 Retrieved from http://doi.acm.org/10.1145/3196709.3196784

Kronborg Iversen, H. (2018). Future of nordic retail. Retrieved from https://www.ey.com/se/sv/industries/consumer-products/ey-the-future-of-nordic-retail-2018

Lahmiri, S., & Shmuel, A. (2019). Detection of parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomedical Signal Processing and Control, 49*, 427-433. doi:10.1016/j.bspc.2018.08.029

Lavidge, R. J., & Steiner, G. A. (1961). A model for predictive measurements of advertising effectiveness. *Journal of Marketing, 25*(6), 59-62. doi:10.1177/002224296102500611

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender?: An experimental test of gender stereotype. Paper presented at the *CHI'00 Extended Abstracts on Human Factors in Computing Systems,* 289-290.

Lopatovska, I., & Williams, H. (2018). Personification of the amazon alexa: BFF or a mindless companion. Paper presented at the *Proceedings of the 2018 Conference on Human Information Interaction \& Retrieval,* New Brunswick, NJ, USA. 265-268. doi:10.1145/3176349.3176868 Retrieved from http://doi.acm.org/10.1145/3176349.3176868

Lowe, M. L., & Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *Journal of Marketing Research, 54*(2), 331-346.

MacKenzie, S. B., & Lutz, R. J. (1989). An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context. *Journal of Marketing, 53*(2), 48-65. doi:10.1177/002224298905300204

MacKenzie, S. B., Lutz, R. J., & Belch, G. E. (1986). The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. *Journal of Marketing Research, 23*(2), 130-143. doi:10.1177/002224378602300205

Mayor, A. (2018). *Gods and robots : The ancient quest for artificial life*. Princeton: Princeton University Press. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=5507640

Mehta, R., Zhu, R., & Cheema, A. (2012). Is noise always bad? exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research, 39*(4), 784-799. doi:10.1086/665048

Merriman, C. (2019). Alexa in-skill purchasing rolls out to the UK and beyond. Retrieved from https://www.theinquirer.net/inquirer/news/3075092/alexa-in-skill-purchasing-uk

Mitchell, W. J., Ho, C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? explicit–implicit evaluations of synthesized speech support a

new HCI model of impression management. *Computers in Human Behavior, 27*(1), 402-412.
Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67* doi:10.1146/annurev-psych-122414-033550
Must Share News. (2018, 07/13/). 19-year-old tries to flirt with 'ask jamie' AI chatbots on government websites. *Must Share News* Retrieved from https://mustsharenews.com/ask-jamie/
Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.
Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*(10), 864-876. doi:10.1111/j.1559-1816.1997.tb00275.x
O'Keefe, D. (2013). The elaboration likelihood model. In J. Dillard, & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 137-149). United States: SAGE Publications, Inc. Retrieved from https://search.credoreference.com/content/entry/sagepersu/the_elaboration_likelihood_model/0
Park, J., Stoel, L., & Lennon, S. J. (2008). Cognitive, affective and conative responses to visual simulation: The effects of rotation in online product presentation. *Journal of Consumer Behaviour: An International Research Review, 7*(1), 72-87.
Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research, 11*, 668. Retrieved from https://search.proquest.com/docview/1293569629
Pogacar, R., Kouril, M., Carpenter, T. P., & Kellaris, J. J. (2018). Implicit and explicit preferences for brand name sounds. *Marketing Letters, 29*(2), 241-259. doi:10.1007/s11002-018-9456-7
Pogacar, R., Shrum, L. J., & Lowrey, T. M. (2018). The effects of linguistic devices on consumer information processing and persuasion: A language Complexity × Processing mode framework. *Journal of Consumer Psychology, 28*(4), 689-711. doi:10.1002/jcpy.1052
Reis, A., Paulino, D., Paredes, H., & Barroso, J. (2017). Using intelligent personal assistants to strengthen the elderlies' social bonds. Paper presented at the *Universal Access in Human–Computer Interaction. Human and Technological Environments,* 593-602.
Richards, R. (2018). The race for voice tech is on: A fertile ground for startups. Retrieved from https://hax.co/2018/09/21/the-race-for-voice-tech-is-on-startups-paving-the-way/
Robert, L. P., & Dennis, A. R. (2005). Paradox of richness: A cognitive model of media choice. *IEEE Transactions on Professional Communication, 48*(1), 10-21.
Rowe, E. (2017). Conversational computing delivers the ultimate frictionless experience. Retrieved from https://www.publicissapient.com

Samson, A., & Voyer, B. G. (2012). Two minds, three ways: Dual system and dual process models in consumer psychology. *AMS Review, 2*(2), 48-71. doi:10.1007/s13162-012-0030-9

Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., . . . Strope, B. (2010). Your word is my command: Google search by voice: A case study. *Advances in speech recognition: Mobile environments, call centers and clinics* (pp. 61-90) doi:10.1007/978-1-4419-5951-5_4 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84885728886&doi=10.1007%2f978-1-4419-5951-5_4&partnerID=40&md5=4ba26f7a8c04130e5dab0fef3d4d5b09

Scherer, K. R. (1986). Vocal affect expression. A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165. doi:10.1037/0033-2909.99.2.143

Solon, O. (2018, -05-08T20:13:57.000Z). Google's robot assistant now makes eerily lifelike phone calls for you. *The Guardian* Retrieved from https://www.theguardian.com/technology/2018/may/08/google-duplex-assistant-phone-calls-robot-human

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology, 22*(1), 37-54. doi:10.1016/j.jcps.2011.09.004

Spiro, R. L., & Weitz, B. A. (1990). Adaptive selling: Conceptualization, measurement, and nomological validity. *Journal of Marketing Research, 27*(1), 61-69. doi:10.1177/002224379002700106

Stayman, D. M., & Brown, S. P. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research, 19*(1), 34-51. doi:10.1086/209284

Surace, K. J., White, G. M., Reeves, B. B., Nass, C., Campbell, M., Albert, R., & Giangola, J. (2000). In Eloqui Voice Systems, LLC Oracle America Inc(Eds.), *Voice user interface with personality*. United States:

Te Reo o Te Taiao. (2019). Bird of the year. Retrieved from https://birdoftheyear.org.nz/

Titcomb, J. (2015, -07-01). The best of siri: 11 funny responses from the iPhone's virtual assistant. *The Telegraph* Retrieved from https://www.telegraph.co.uk/technology/apple/11709991/The-best-of-Siri-11-funny-responses-from-the-iPhones-virtual-assistant.html

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness, and media choice in organizations: A symbolic interactionist perspective. *Communication Research, 14*(5), 553-574.

V. Këpuska, & G. Bohouta. (2018). Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). Paper presented at the *Annual Computing and Communication Workshop and Conference (CCWC),* 99-103. doi:10.1109/CCWC.2018.8301638

Wisterberg, E. (2019). Google home till sverige – är det starten på en ny revolution? Retrieved from https://www.breakit.se/artikel/15670/google-home-till-sverige-ar-det-starten-pa-en-ny-revolution

Wittwång, A. (2019). Google nest hub lanseras i sverige – röststyrd assistent med bildskärm. Retrieved from https://www.sweclockers.com/nyhet/27605-google-nest-hub-lanseras-i-sverige-roststyrd-assistent-med-bildskarm

Wu, K., Zhang, D., Lu, G., & Guo, Z. (2019). Joint learning for voice based disease detection. *Pattern Recognition, 87*, 130-139. doi:10.1016/j.patcog.2018.09.013

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493.

7. Appendix

7.1. Appendix I: Detailed Scopus Search

Table 13. Overview of Scopus search in terms of years published

Year	“voice” + “assistant”	“social media”
2019	6	123
2018	2	199
2017	2	192
2016	0	218
2015	1	118
2014	0	108
2013	0	64
2012	0	27
Before 2012	1	16

Table 14. Detailed results of Scopus search results

Publication	"voice + assistant"	"social media"
Computers In Human Behavior	3	441
IEEE Access	1	85
Journal of Business Research	0	68
Journal of the Association for Information Science and Technology	0	49
Journal of Interactive Marketing	0	43
European Journal of Marketing	0	35
Psychology and Marketing	1	30
MIS Quarterly Management Information Systems	0 0	23
Scientific Reports	0	22
Journal of Organizational and End User Computing	0	21
Personal and Ubiquitous Computing	1	21
Industrial Management and Data Systems	1	20
Journal of Consumer Marketing	0	20
Journal of Political Marketing	0	20
Journal of Advertising Research	0	19
ACM Transactions On Information Systems	0	17
International Journal of Human Computer Interaction	2	17
International Journal of Research in Marketing	0	13
International Journal of Retail and Distribution Management	1	12
Journal of Marketing	0	12
Medical Reference Services Quarterly	1	10
Journal of Marketing Research	0	8
Journal of the Acacemy of Marketing Science	0	8
Journal of Consumer Research	0	7
Management Science	0	7
Journal of Social Psychology	0	6
Journal of Personality and Social Psychology	0	5
Information Systems and E-business Management	0	4
Journal of Experimental Social Psychology	0	4
Journal of Computing and Information Science in Engineering	0	3
Journal of Language and Social Psychology	0	3
Pattern Recognition	0	3
Annual Review of Psychology	0	2
Emotion	0	2
International Journal of E Services and Mobile Applications	1	2
Journal of Retailing	0	2
Journal of Consumer Psychology	0	1

7.2. Appendix II: Preparatory Study

To test the survey design a preparatory study was conducted between March 27th and April 4th. The survey was shared through a private Facebook post and generated 56 valid responses. The findings of the preparatory study indicated that the manipulations had been interpreted correctly and that the attitudes different between stimuli groups. In addition, feedback from testers of the preparatory study underlined the relevance of including a measure relating to respondents' online concerns. It was argued that negative views of major tech companies and their collection of personal data could lead to biased survey answers.

Table 15. Descriptive statistics of participants in preparatory study

	HT n=11	LV n=17	HT n=13	HV n=16
Mean age	22.7	23.1	23.3	23.4
A_R	5.27	3.04	4.82	4.04
A_B	4.33	4.16	4.00	4.40

Note: 68 % female, 31 % male

7.3. Appendix III: Main Study

Below, excerpts of the main study is shown; some from what it looked like for mobile phone respondents and others for those responding using a computer.

7.3.1. Introduction to survey

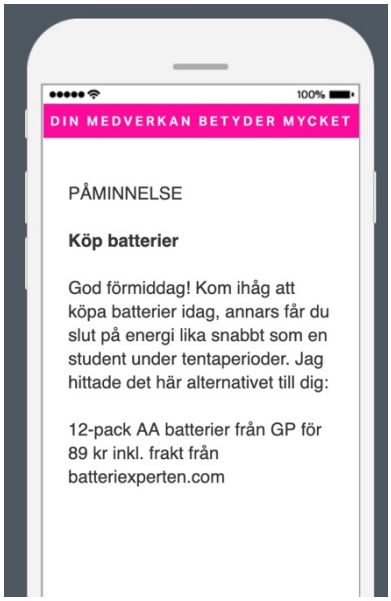
Page 1: Survey landing page

Page 2: Introduction to test question

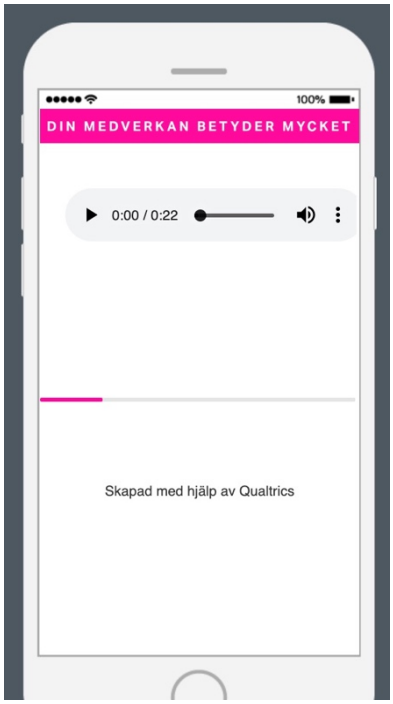
Page 3: Test question

7.3.2. Stimuli

Page 4: Intro to stimuli



Page 5: HT Stimulus



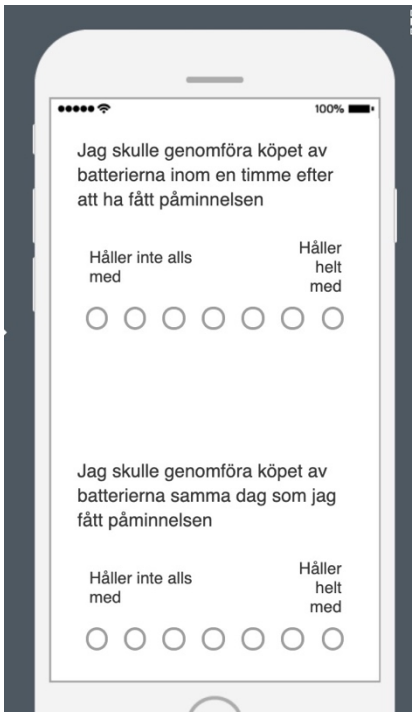
Page 6: HV Stimulus

7.3.3. Questionnaire

Page 7: Reminder attitude

Page 8: Brand attitude

Page 9: Intro to intention questions

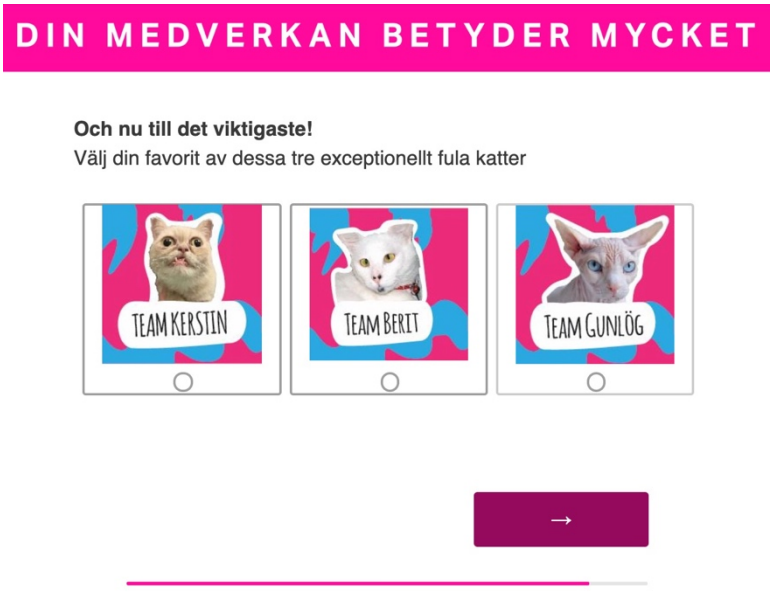


Page 10: Purchase intentions

Page 11: Usage intentions

Page 12: Choosing smart speaker

Page 13: Manipulation check



Page 14:Vote for cat

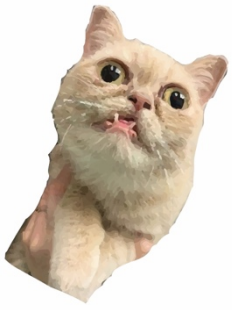
Page 15: Survey end message

7.4. Appendix IV: Distribution of Main Study

The distribution of the main study became much more eventful than what was originally planned. To reduce the possibility of speeding through the many similar Likert scale questions, the preparatory study had included a question asking respondents to choose the ugliest out of three ghastly cats. This yielded an unexpected amount of positive feedback, as participants expressed a strong desire to know more about the cats and the reason for their appearance.

With this in mind, the cats became the focal point of the distribution. The survey was reframed as a competition between the three cats dubbed "Fule Katt 2019" (Ugleh Cat 2019". This included generating identities for the cats, creating a graphic profile for the competition, hosting a Facebook event, recruiting campaign teams for the different cats and printing campaign material. Inspiration was taken from New Zealand's "Bird of the year" competition, which annually attract attention, affection and donations to birds of varying beauty standards (Te Reo o Te Taiao, 2019). Another source of inspiration was the "fun theory", suggesting that entertainment value can lead to participation. It was hoped that by providing an element of fun, respondents would be more inclined to share the survey with their friends. At the end of the survey, respondents were able to cast their vote and also choose to join the campaign teams of their preferred cat.

So as to avoid any copyright issues with the images used, they were redrawn digitally into detailed vector graphic using Adobe Illustrator CC 2017 (see examples below). The graphic profile was developed to appear energetic and humorous. The printed campaign material included large QR codes directing to the survey.

The competition was launched on April 4th through a Facebook post and the creation of a public Facebook event. Within two days, the event had connected more than 700 Facebook members and the campaign teams had grown to engage 17 people. But the rapid sharing of the survey link resulted in a violation of Facebook's community rules, although no specification was provided regarding what rule was violated. However, the link was disabled on all Facebook-owned platforms, meaning that the link could not be

shared nearly as easily. The Facebook event was shut down – proving a definite stop to the desired online spread of the survey.

Fortunately, many flyers with QR codes linking to the survey remained from the intense campaigning. These were ineffectively distributed at various locations around universities in Stockholm. The 460 flyers distributed between April 7th and April 15th only generated 36 valid responses. To combat time pressure and the risk of not yielding enough responses, 92 respondents from the online survey pool Prolific were paid £5.04/hour to complete the survey, leading to a total of 72 additional valid responses. Together with the 247 (87 valid) responses from the link, this last-resort strategy helped generate a sufficient respondent base for subsequent analysis.

Below, examples of the flyers used for the campaigning are seen.

7.5. Appendix V: The Filtering Process

The speeding check was developed through comparing the respondents' completion times to a benchmark time. The benchmark time was set by recording 8 author response times, 2 per stimuli condition, and adding a safety margin. The author responses were completed as fast as possible to still be able to read instructions and questions properly, and returned an average response time of 179 seconds. It was assumed that no respondent could complete the survey in a meaningful way faster than this. However, to further increase the reliability, a safety margin of 31 seconds was added, effectively extending the assumption threshold to 210 seconds. No upper limit on completion time was set considering the fact that many respondents are likely to open the link and start the survey but finish it later. In total, 41 respondents (9.98%) were discarded because of the speeding check.

Responses were seen as incomplete if more than one question from the survey's main section were unanswered. No responses were discarded for unanswered background questions, such as age or gender. In total, 15 responses (3.65%) were discarded on the basis of incompletion.

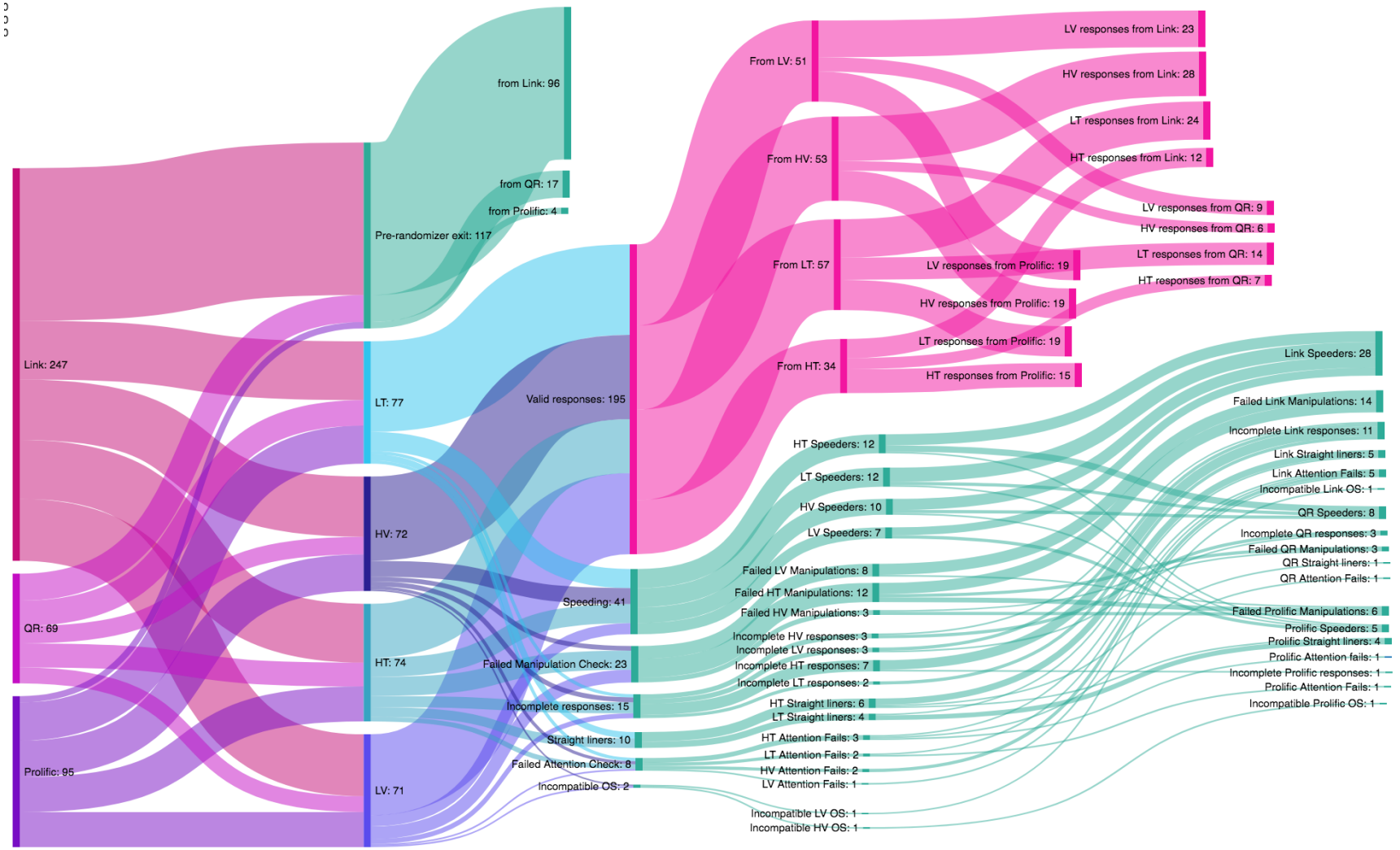
The survey included an attention check in the form of a "trap question" where participants were asked to select "completely disagree" to prove they were reading and answering the questions carefully. All responses failing to do so were subsequently discarded, amounting to an additional 8 (1.95%) omissions.

One of the most common issues with Likert-scale questions is "straight lining", where a respondent rapidly selects the same alternative on the similar-looking scales (Jones et al., 2015). Out of the total 10 sections where straight lining could occur, participants straight lining through 3 or more sections were discarded. 10 responses (2.43%) qualified as straight liners.

The manipulation check placed late in the survey helped examine if a respondent could reasonably be believed to have understood the stimuli correctly. Given the clear division of voice and type stimuli conditions, the 4 (0.97%) cases where the reminder was said to be equally voice-based and text-based were discarded. In addition, participants exposed to voice stimuli answering closer to "completely disagree" than to "completely agree" on the voice manipulation check were also discarded. For the same question, responses from the text condition that were closer to "completely agree" were discarded. In total, 3 responses (0.73%) were removed on the basis of the voice condition manipulation check. A corresponding approach for the text condition manipulation check yielded 2 (0.49%) omissions. Finally, participants were seen as having failed the manipulation check for the language complexity level if their responses were closer to the "wrong" alternative. This led to 14 (3.41%) of responses being discarded. In total, the manipulation check stage of the filtering process led to 23 (5.60%) responses being discarded.

Finally, it was discovered that some Android operating systems failed to play the audio files. Consequently, responses on the incompatible Android systems that were sorted into the voice condition were removed. 2 (0.49%) responses were removed because of this.

On the following page, a more detailed version of the Sankey diagram is presented.


Link: 247
QR: 69
Prolific: 95
Pre-randomizer exit: 117
from Link: 96
from QR: 17
from Prolific: 4
LT: 77
HV: 72
HT: 74
LV: 71
Valid responses: 195
Speeding: 41
Failed Manipulation Check: 23
Incomplete responses: 15
Straight liners: 10
Failed Attention Check: 8
Incompatible OS: 2
From LV: 51
From HV: 53
From LT: 57
From HT: 34
LV responses from Link: 23
HV responses from Link: 28
LT responses from Link: 24
HT responses from Link: 12
LV responses from QR: 9
HV responses from QR: 6
LT responses from QR: 14
HT responses from QR: 7
LV responses from Prolific: 19
HV responses from Prolific: 19
LT responses from Prolific: 19
HT responses from Prolific: 15
HT Speeders: 12
LT Speeders: 12
HV Speeders: 10
LV Speeders: 7
Failed LV Manipulations: 8
Failed HT Manipulations: 12
Failed HV Manipulations: 3
Incomplete HV responses: 3
Incomplete LV responses: 3
Incomplete HT responses: 7
Incomplete LT responses: 2
HT Straight liners: 6
LT Straight liners: 4
HT Attention Fails: 3
LT Attention Fails: 2
HV Attention Fails: 2
LV Attention Fails: 1
Incompatible LV OS: 1
Incompatible HV OS: 1
Link Speeders: 28
Failed Link Manipulations: 14
Incomplete Link responses: 11
Link Straight liners: 5
Link Attention Fails: 5
Incompatible Link OS: 1
QR Speeders: 8
Incomplete QR responses: 3
Failed QR Manipulations: 3
QR Straight liners: 1
QR Attention Fails: 1
Failed Prolific Manipulations: 6
Prolific Speeders: 5
Prolific Straight liners: 4
Prolific Attention fails: 1
Incomplete Prolific responses: 1
Prolific Attention Fails: 1
Incompatible Prolific OS: 1

7.6. Appendix VI: Background information of respondent sample

Table 4. Overview of respondent sample

Variable	N	% of total sample
Gender		
Male	84	43.1%
Female	111	56.9%
Age		
15-20 years	22	11.3%
21-29 years	139	71.3%
30-40 years	20	10.3%
>40 years	16	8.2%
Occupation		
Student	130	66.7%
Employed	52	26.7%
Other	13	6.6%
Completed years of higher study		
0 years	42	21.5%
1 year	28	14.4%
2 years	40	20.5%
3 years	46	23.6%
4 years	23	11.8%
5 years	9	4.6%
6 years	5	2.6%
7 years	1	0.5%
>7 years	1	0.5%
Living in Sweden	179	91.8%
Living abroad*	16	8.2%
Response context		
At home	128	65.6%
At work/study place	54	27.7%
In public transport	5	2.6%
Outdoors	4	2.1%
Other**	3	1.5%
Device		
Mobile phone	109	55.9%
Laptop	46	23.6%
Stationary computer	37	19.0%
Tablet	3	1.5%

*Japan, Canada, US, France, UK, South Africa

** Café, in a car (not driving), at a friends house

7.7. Appendix VII: More detailed descriptive statistics

Table 7. Descriptive statistics of attitudes and intentions between stimuli groups

	LT n=57		LV n=51		HT n=34		HV n=53	
Attitudes and intentions	μ	σ	μ	σ	μ	σ	μ	σ
Attitude towards the reminder (A_R)	4.40	1.44	3.86	1.62	4.87	1.37	4.52	1.64
Bad – Good	4.58	1.46	4.25	1.66	4.82	1.57	4.60	1.71
Dislike – Like	4.26	1.52	3.47	1.95	4.65	1.63	4.28	1.81
Negative – Positive	4.35	1.52	3.90	1.57	5.15	1.37	4.66	1.62
Brand attitude (A_B)	4.34	0.94	4.35	0.90	4.50	0.92	4.42	0.94
Negative – Positive	4.35	0.97	4.47	1.10	4.50	1.16	4.43	1.07
Dislike – Like	4.32	0.95	4.20	0.96	4.41	0.89	4.36	1.13
Bad – Good	4.35	1.04	4.39	0.87	4.59	0.89	4.47	0.99
Purchase intentions (I_P)	3.61	1.53	3.67	1.76	4.06	1.67	3.76	1.56
I would make the purchase the same hour…	3.04	1.57	3.10	1.86	3.24	1.92	2.96	1.64
I would make the purchase the same day…	4.18	1.72	4.24	1.98	4.88	1.81	4.55	1.87
Usage intentions (I_U)	3.93	1.19	4.08	1.27	4.05	1.19	3.44	1.64
Remember workout	2.77	1.65	3.57	1.92	3.47	1.99	3.00	2.18
Go to bed on time	3.00	1.95	3.49	1.94	3.41	2.13	3.09	2.12
Friends' birthdays	5.05	1.65	4.76	2.07	4.59	2.20	4.08	2.31
Paying invoices	5.11	1.99	4.76	1.72	4.79	2.19	3.87	2.31
Cleaning	2.54	1.45	3.22	1.76	2.88	1.87	2.58	1.82
Scheduled meetings	5.11	1.79	4.76	1.85	5.15	1.73	4.02	2.13

7.8. Appendix VIII: Additional Statistical Calculations

Table 16. Levene's test for homogeneity of variances (based on mean)

	Levene statistic	p-value
Reminder attitude ($A_{R)}$)	.691	.559
Brand attitude (A_B)	.060	.981
Purchase intentions (I_P)	.661	.577
Usage intentions	4.48	< .01

Table 17. ANOVA calculations of gender differences in A_R scores

Gender	N	μ	σ	F	p
Total	195	4.38	1.56	3.690	.056
Male	84	4.13	1.47		
Female	111	4.56	1.61		
High-complexity voice	53	4.52	1.64	4.837	<.05
Male	20	3.90	1.56		
Female	33	4.89	1.60		
Low-complexity voice	51	3.86	1.62	2.241	.141
Male	25	3.53	1.43		
Female	26	4.21	1.75		
High complexity text	34	4.87	1.37	0.467	.499
Male	21	5.00	1.27		
Female	13	4.67	1.56		
Low-complexity text	57	4.40	1.44	0.472	.495
Male	18	4.20	1.24		
Female	39	4.49	1.53		