Anthropomorphism in Software Agents:
Perceptions and Implications of Gendering Intelligent Personal Assistants

Emily Jane Harayda

A research paper submitted to the University of Dublin, in partial fulfillment of the requirements for the degree of Master of Science Interactive Digital Media.

May 2019

Declaration

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

I declare that the work described in this Research Paper is, except where other stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other University.

Signed: _____

Emily Jane Harayda

Date: May 2019

Permission to Lend and/ or Copy

I agree that Trinity College Dublin may lend or copy this research paper upon request.

Signed: _____

Emily Jane Harayda

Date: May 2019

Acknowledgements

I would like to thank Kerstin Ruhland for her guidance and support in the development of this research paper. I would also like to thank my friends and classmates who supported in the research without hesitation.

Abstract

This research paper explored how user experience and user satisfaction ratings may differ depending on the gender of the intelligent personal assistant (IPA). A study in the form of an interaction with an IPA, followed by a survey was conducted with participants over the age of eighteen to understand both user experience sentiments and user satisfaction. The study focused on the Google Assistant, due to its intentional attempt at avoiding a gendered name and personality, as well as its prevalence.

It was hypothesized that there would be a significant preference for female IPAs on the 6 scales of: 1) Attractiveness; 2) Perspicuity; 3) Efficiency; 4) Dependability; 5) Stimulation; and 6) Novelty, as well as on the overall satisfaction rating. It was also hypothesized that female participants who interacted with the female IPA would report higher levels across the 6 scales as well as on satisfaction, as compared to males, due to similarity biases.

Contrary to expectations, the results showed that IPA gender did not significantly impact user experience across any of the 6 UEQ scales, or overall user satisfaction. This result is in contrast to the literature examined. The lack of the impact of IPA gender on user experiences and satisfaction is a notable result, as it provides support for the development of gender-neutral technology. As people become increasingly reliant on devices such as IPAs for goal-oriented tasks and social support, it will be crucial to examine the ways in which technology is assigned anthropomorphic elements and to assess the negative social consequences of gendering.

Table of Contents

List of Abbreviations, Tables, Figures, and Illustrative Materials

Abbreviations

Term	Abbreviation
Intelligent Personal Assistant	IPA
User Experience Questionnaire	UEQ

List of Tables, Figures, and Illustrative Materials

1. Introduction

Intelligent personal assistants (IPAs) are vocal mobile software agents currently used by millions of people to perform human-like everyday tasks, on behalf of their users (Balakrishnan & Honavar, 2001; Skalski & Tamborini, 2007). Amazon's Alexa, Apple's Siri, or Google's Assistant, to name a few, perform tasks such as information searches, organizing schedules and providing entertainment (e.g. playing music). Although these tasks may appear straightforward and easily executed, IPAs have far deeper capabilities and are programmed to quickly learn and adapt to assist their users. Predictions for the future of IPAs indicate they will provide holistic enhancements to users' daily lives. The IPA will wake us up in the morning at the time of an alarm, with a pre-planned schedule for the day. The assistant will brief you on the weather for the day and read out your preferred news sources, emails, or messages to you while you get ready for work. While you work and attend meetings, the assistant may take notes for you and send debriefing emails, as well as making a dinner reservation for you and your family on Friday night. The assistant may provide companionship in lonely times and book holidays on your behalf when you show signs of exhaustion (Smith & Shum, 2018). IPAs have the potential to be an employee, a friend, or an extension of self and are therefore changing how we fundamentally interact with technology. Through providing emotional support and conversation, proactive assistance, and emotional comprehension, IPAs are computers that have so intimately embedded themselves into our social lives and minds (Turkle, 2007).

Predictions for the future of IPAs demonstrate how we have designed machines that are not solely goal-oriented intelligent devices, but social in their existence. Computers have shifted from being simply a platform to help us to achieve our objectives, into being communication devices and social entities within themselves (Guzman, 2017). A key enabler of this shift is the voice user interface. The voice user interface that enables communication between human and machine has freed us from the physical and allowed us to interact in a more natural way with technology.

In order to encourage the feeling of naturalness in communication, technology developers have programmed IPAs to follow human communication norms in a way that is a cultural process and not merely the exchange of information (Carey, 1989). Developers have taken this even further and given machines names, genders, and personalities.

While selecting name, gender, and personality characteristics for IPAs may seem an innocent way to enhance likeness between human and machine, it is not without consequences. Assigning a name and gender to a machine may alter expectations of the relationship between human and machine and opens a door for perpetuating gender stereotypes. IPAs are designed to mimic servility and

subservience and it is no coincidence that they are consistently gendered as female. Historically, the roles of secretaries and administrative assistants have been filled by women, conditioning society to expect a female voice for helpful tasks. Technology is now deepening the cultural bias around which gender is seen to work for, or respond to, the other. The comparison is stark between Siri and Alexa who function as personal assistants, obediently serving their users basic requests, contrasted with the high-powered tasks executed by male IPAs – IBM Watson or Salesforce's Einstein.

A place in technology which has been clearly defined as female – the intelligent personal assistant. IPAs exemplify the systemic sexism in the artificial intelligence that powers them. This leads to the central research question: What are user perceptions of gendered IPAs, and what are the implications of gendering?

Previous research has indicated that users are more comfortable interacting with a female IPA, potentially due to their preconceived or subconscious notions of a female's role of service in society. It is therefore crucial to examine how technology design plays a role in either altering or cementing these biases.

The purpose of this research paper will be to explore how user experience and user satisfaction may differ depending on the gender of the IPA. A study in the form of a survey was conducted with participants over the age of eighteen to understand both user experience sentiments and user satisfaction immediately following an interaction with either a male or female voice assistant. The study will focus on the Google Assistant, due to its intentional attempt at avoiding a gendered name and personality, as well as its prevalence. The Google Assistant accounts for 49.4% of the type of assistant that people have on their smartphones (Heitzman, 2017).

The research paper will commence with an in-depth literature review to examine the gendering of existing IPAs, user preferences for gender, and the findings from similar research studies. The literature review will be followed by the research methodology and the findings of the study conducted on participants. The paper will conclude by providing an analysis and discussion of the findings and discuss future topics for examination.

2. Literature Review

The following literature review will establish an understanding around IPAs; their technology, applications, market. It will then move onto an examination of their identities shaped by anthropomorphic traits, in particular gender.

Background to Intelligent Personal Assistants

Basic voice command capabilities have been present on computers for over a decade, but the technology for individual consumers commenced with Apple's integration of Siri into the iPhone in 2011. Since 2011, their usage, intelligence and investment into the technology has grown rapidly. IPAs are paving the way for more efficiency and ease in modern, everyday tasks through voice technology. The research firm Gartner predicted that 20% of our smartphone interactions will be with a virtual personal assistant by 2020, indicating a shift away from touch-based apps and towards voice-based interactions (Gartner, 2018).

IPAs are typically used to help users with tasks in everyday life. These could include personalized shopping suggestions, restaurant reservations, managing calendars, sending emails, taking notes, organizing travel, and providing directions while driving. The irony is that technology organizations have made their fortunes building capabilities that no human could do, and now they are investing hugely in developing machines to perform the most basic tasks that come naturally to humans.

The rise of IPAs clearly illustrates how our models of interaction are being simplified with technology, moving from punch cards to keyboards, to touch, and now to voice. Voice user interfaces can relieve us of tedious work by offering a more natural medium for people to interact with technology. The voice user interfaces essentially free us from physical technology such as screens and allows us to interact with machines in a more fluid, innate way (Ghorbel et al., 2004; Martin, 1976; Moller et al., 2006; Yerrapragada & Fisher, 1993). Conversation is the most natural interface as we converse every day and have done for thousands of years. Humans have transmitted knowledge and ideas orally – the voice is perhaps the most intuitive way for us to communicate.

The Technology Enabling Intelligent Personal Assistants

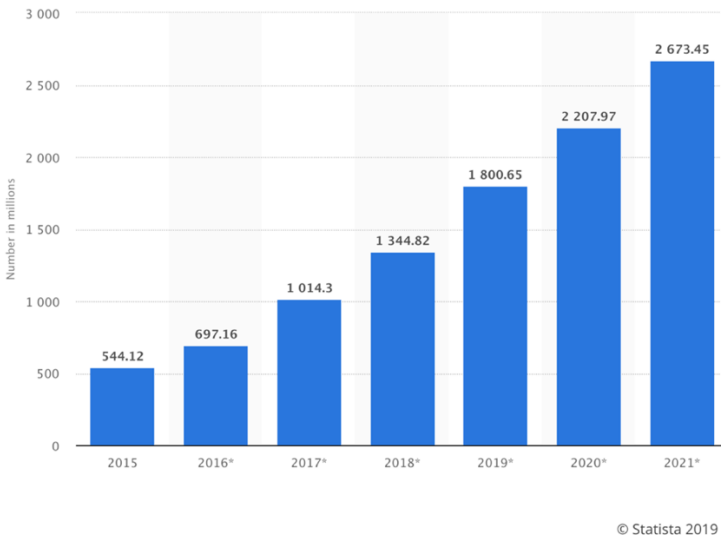
IPAs are mostly cloud-based, and mostly will not work accurately without an internet connection. IPAs work by combining specialized computer chips, microphones, and software to listen and respond to commands from the user. There are three steps involved in turning user commands into actions by the assistant. The assistant begins by converting the user's 'expression', or speech, to text through speech recognition technology. Speech recognition involves large amounts of data being fed

into machine learning models or statistical methods, which develop an understanding of the speech. Speech recognition technology can also be tailored to recognize different dialects or accents (Larson, 2010). Associated with a user's expression, is an intent. The speech data is used to establish this intent using natural language processing algorithms (Larson, 2010). Lastly, intent is turned to action when the assistant scans through available information; such as user input data, location data, and the search ability to access online sources in order to fulfil the user's intent.

The Intelligent Assistant Market

Since Apple's integration of Siri into the iPhone, the market for IPAs has grown rapidly along with increasing smartphone and smart speaker sales globally. These IPAs are not only available through our smartphones, but also can be found on a number of devices such as; televisions, watches, and cars (Apple, n.d.; Google, n.d.). As of mid-2017, there were an estimated one billion users worldwide and the number is only growing with time *(refer to Figure 1)* (Statistica, 2019).

Figure 1: Chart for Number of unique active virtual assistant users worldwide, 2015-2021 (in millions)



In terms of current active users of the most popular IPAs, Apple claims that over 500 million individuals used Siri in 2017, with 375 million users active on a regular basis (Apple, 2017). Usage statistics are reflected through the numbers, with consumer virtual personal assistant revenue in North America is expected to reach $296 billion USD by 2021 (Orbis Research, 2017).

Anthropomorphism in Intelligent Personal Assistants

As each technology company fights for the lead in the IPA space, they build specialized teams to differentiate their IPA not only by functionality but also by human-like traits, such as voice and

personality. The integration of such social, anthropomorphic traits is extremely significant to the design of IPAs.

Research has found that the integration of anthropomorphic traits in machines will influence how the user interacts with the machine. Social agency theory in the field of human-computer interaction states that enhancing the humanness of the machine will elicit a social response from the user, typically found in human to human interactions (Mayer, Sobko & Matutone, 2003; Moreno et al., 2001; Cassell & Tartaro, 2008; Kiesler & Sproull, 1997; MacDorman & Ishiguro, 2006). Enhancing the human qualities of an IPA can cause a user to infer that the non-human entity has human characteristics and therefore deserves human-like treatment (Purington, 2017).

Although IPAs do not have a 'self', they have been given a social form. As they function as the voice of a machine, they are both a social actor and a medium (Guzman, 2015; Guzman, 2017). Specifically, the role of voice in the computer interface assumes the role of a humanized agent, encouraging the user to treat it as a social actor and compare it to the value of a human (Nass et al., 1994; Norman, 1994; Reeves & Nass, 1996). Voice assists the user in assigning a unique identity to the machine (Laurel, 1997). Research has found that people recognize personality traits programmed into a technology and act toward that machine as if it were a human with similar characteristics (Lee & Nass, 2005; Nass et al., 1995). Users recognize aspects such as accents in IPAs and react subjectively by assigning personality traits on the basis of an ethnic identity established through an accent (Anisfeld et al., 1962; Birch & McPhail, 1997). Users also recognize gender in machine voices and respond to the machine as they would to a human of that gender (Nass & Brave, 2005). The process of assigning a gender to a machine anthropomorphizes the machine to the extent that a user may even believe an emotional relationship is possible.

In terms of user experience and satisfaction with IPAs, traditional measures may not be accurate. Due to the human-like traits assigned to IPAs, research argues that traditional measures of user experience place emphasis on the functional and goal-oriented aspects of interaction, rather than the social aspects. A modified approach to user experience measurement has taken the IPA as both a tool and a social actor. New measures may even focus on elements like trust, a key aspect of spoken interaction between humans (Nass & Moon, 2000). Recent studies on user interaction and satisfaction have shown that users report higher likeability, effective communication, and greater satisfaction levels with IPAs who they perceive to have enhanced human-like traits (Atkinson, Mayer, & Merrill, 2005; Mayer et al., 2003). On the other hand, this expectation of humanness may lead to unrealistic expectations of the assistant's capabilities. In evaluations of experience and satisfaction, social biases may play a role and work against the IPA. For example, the user may expect the

assistant to be able to answer questions and understand the surrounding context, and then may experience dissatisfaction when the assistant is unable to meet expectations (Luger, 2016).

Voice is a crucial input and output in the design of IPAs. The software behind IPAs is inherently difficult to understand as it is ephemeral (Chun, 2011). As the user cannot see how the IPA functions, they must rely on the IPA's voice to construct an understanding of how it works and how to interact with the software (Suchman, 2009; Guzman, 2015). There are specific ways that a user is instructed to speak with an assistant, which parallel human communication. Examples of this include how the user must activate the IPA through using its name – "Hi Siri", to awaken the IPA and make a request. Unlike with human-to-human voice communication, with an IPA there is not an expectation of mutual interest or attentiveness. IPAs do not require their users to put effort into being polite or entertaining during communication (Elliot & Hare, 2019). The norm of being agreeable during conversation is broken down when interacting with an assistant (Elliot & Hare, 2019). These communication norms may be broken down due to the user being in a position of power and complete dominance when communicating with an IPA.

Voice communication and power are intertwined, both in human-human communication and human-machine communication (Castells, 2013). When humans converse with each other, power, identity and social standing are established in relation to one another (Goffman, 1959). When humans converse with machines, power and social standing are similarly established. The IPA is a servile assistant, reinforcing a lower social status as compared to the user.

A key part of the interaction between user and IPA is the dialect of dominance and submission (Burgoon & Hale, 1984). If we now apply this to IPAs, IPAs are designed to be submissive through both their language and non-verbal aspects. To provide a non-verbal example, IPAs remain silent until the user engages them. The default method of communication is therefore a command from the user. Through the narrative of dominance and submission, IPAs can be seen to be sexualized in their service (Pierce, 2016). A user's fantasy may come to life through obtaining a gendered device, like Siri, to do anything the user wants as well as providing playful verbal responses. Upon receiving an inappropriate request from the user, such as a marriage proposition, Siri's responses to these inappropriate comments are to deflect the conversation. Woods (2018) sees these responses to be unsettling to women because they are familiar. Siri acknowledges the inappropriate comment from the user and attempts to defuse the situation by returning to ask the user what task they can do for them (Woods, 2018). The developers working to create Siri's responses to a user had to come up with these 'coping mechanisms' in order to deflect any inappropriate comments it receives (Woods, 2018).

Gendering Intelligent Personal Assistants

The technology field has a long history of not only assigning personalities to machines, but also genders. As gender is a social construct, unlike biological sex, it can be negotiated and even assigned to non-living entities (West & Zimmerman, 1987). When gender is assigned, it is a deliberate process about how best to relate to, assist, or persuade the user (Hester, 2016). A gender may essentially mask the true identity of the machine and assists with user communication and forming positive representations of the technology (Woods, 2018). This section of the literature review will examine how technologies have gender constructed within them and how technologies have become associated with the masculine or feminine.

We must first ask ourselves – why do we gender? On one hand, the process of gendering an IPA could be a conscious business decision, based on psychological reasons for preferring a certain gender's voice or characteristics. On the other hand, or in addition to, gendering may be an unconscious reiteration of society's existing stereotypes. Either way, it is by no coincidence that the majority of IPAs are gendered as female.

Establishing Gender through Voice

The most dominant form of gendering for an IPA is that of the voice. Female voices have and continue to dominate IPA interfaces. They are the default voice on Apple's Siri, Google Assistant, Amazon's Alexa, and others. Although there is an option to change the default voice on Google Assistant and Siri, it requires the user to actively seek out and change this setting. This would require the user to consciously distance from the subject of the interface in order to change the default voice (Phan 2017).

The default voice has been assigned to female which has an obvious appeal to stereotypes regarding women in industry. Feminine voices are typically associated to domestic and service roles as they are seen as more caring and emphatic (Sparks et al., 1997). A women's voice is therefore regarded as more suitable than a man's voice for a role that demands caring, service, or obedience. By this account, service is positioned as feminized labor, not because women make up the majority of the workforce but also because the image of the sphere is feminized (Hester, 2016). This is why people may never second guess why technological interfaces are clearly gendered as females.

Establishing Gender through Language

Gender can also be established through language and speech patterns. Research examining speech patterns has found that females typically use personal pronouns, such as 'I', at a greater frequency than males (Lenard, 2017). IPAs such as Siri and Alexa, both gendered as females, also use personal

pronouns at the same rate of human females (Hannon, 2016). For example, if Alexa mishears a request from the user, it will respond “I didn’t understand the question that I heard”. This response indicates that it is Alexa’s fault that it did not understand, rather than the users fault for asking a difficult or complex question.

Aside from solely male and female differences in the use of personal pronouns. Studies have also found that people who occupy a lower status in a relationship or within society, male or female, use a greater number of personal pronouns (Pennebaker, 2013). Pennebaker’s research also demonstrates how within any relationship between humans, regardless of gender, the person perceived to occupy a higher status uses fewer “I” words, and more “We-, You-, He-, She-, and They-“ words (Hannon, 2016).

The usage of language intimately connects gender and social status. The work that the IPAs execute on behalf of humans is seen as historically low status. The deliberate connection between female IPAs and the conventions of low-status administrative labour are obvious, and deeply rooted in social norms as well as gender stereotypes – representing what has been traditionally deemed ‘women’s work’ (Hester, 2016).

Feminization of Labor

Gender is more than masculine or feminine characteristics, it is an important marker of identity that helps us to relate to one another and is shaped by cultural norms and practices (Alesich & Rigby, 2017). Similarly, technology functions by drawing upon pre-existing assumptions, social practices, and cultural norms (Carey, 1990). The decision to assign a gender to technology links the two.

The selected gender for an IPA is usually correlated with the traditional gender of their work role or purpose. Traditionally, the gender of female has been strongly associated with that of the domestic and administrative spheres. Females have historically spent a greater amount of time in the home, an area in which work goes unnoticed compared to the realm of paid work outside the home (Adam, 1996). While this has evolved over time, with more participation of women in the workforce, the domestic sphere still relies upon stereotypes of femininity such as homemaking, care-taking, and other ‘pink collar’ labor (Woods, 2018; White, 2015).

Technology organizations use these existing stereotypes of femininity in the domestic and administrative spheres to make their devices more palatable for the public to adopt in their homes and in their offices (Woods, 2018). Technology companies promise an easier and more efficient lifestyle with their IPAs managing menial tasks and have exploited ideas about gender in their attempt to offer an effective service to users (Hester, 2016). While the IPA free a real woman from

her domestic or administrative tasks, it reinforces the existing power structures and the allocation of these tasks to women (Adam, 2005).

As interactions with IPAs are designed to mimic an employer to employee or master to servant style of relationship (Guzman, 2017), gendering the IPA as female reinforces that women are subordinate to males. The IPA therefore plays an important role in stabilizing or destabilizing particular conventions of gender and power (Oudshoorn, 2003). In this case, the IPA reinforces patriarchal gender stereotypes and male supremacy (Rothschild, 1983). For example, Alexa has been likened to both a wife and a companion who is confined to the living room or bedroom, interacting obediently but with a small bit of attitude (Brown, 2015). According to some users, Alexa can be a partner who provides companionship and eliminates the need for a human wife (Foner, 2015).

It cannot be ignored that the developers of intelligent systems are overwhelmingly male (Leavy, 2018). UNESCO (2017) found that on a global scale, women account for less than 29% of scientific researchers. This under-representation of women in scientific fields has corresponding effects of under-representation in their ideas. Rather, the technology behind IPAs is a reflection of the male developers own experiences, interests, and social conditioning – including bias. In order to avoid gender biased products influencing our society, a higher level of gender diversity in the workplace is required (Leavy, 2018).

Identities of Personal Assistants

There are significant differences among the IPAs, in terms of physical hardware they are found on but also in terms of their personality traits, proactiveness, and purposes. This section will mainly focus on the personality and gender of the most widely adopted IPAs.

Google Assistant: Google Assistant was launched in 2016 with an only female voice (Google, n.d.). In 2018, Google introduced a number of different voice options for the user – including male voices and different accents. In terms of default settings, the Google Assistant is gendered as female. The Google Assistant is marketed as a digital helper that seems human without pretending to be a human. It does therefore not have a human name or pretend to be a human in conversation (Google, n.d.). The personality given to the Google Assistant is conversational and positive.

Apple's Siri: Siri was first introduced on the iPhone 4S in 2011, after the technology was acquired by Apple (Apple, n.d.). Siri was launched with a female-only voice, and in 2013 Apple introduced male voices and regional accents to customize the assistant. A female voice actor provided the original vocal sounds for Siri and is now recognized by US media and audience as a female (Ravitz, 2013). Siri is marketed as a "humble personal assistant" that can help users get things done just by asking (Apple 2011). Apple had also released a guide book for interacting with Siri in 2012. Within the book,

it uses phrases such as “sweet-talk Siri into doing practically anything!” (Sadun & Sade, 2012). The choice of language implies there is both a sexual nature and power dynamic to the technology (Sadun & Sade, 2012).

Amazon Alexa: Amazon’s Alexa was launched in 2014 as a speaker paired with IPA software. In 2017, Alexa become widely available to Amazon customers and on a number of varied devices (Amazon, n.d.). Alexa was launched with a female only voice, but in 2018 Amazon released male voice options. Alexa’s name means the protector of defender of mankind in ancient Greek (Woods, 2018). In parallel to this, Alexa is marketed as a device that can manage the home and administrative spheres. In contrast to Siri and Google Assistant, Alexa lacks access to certain services such as email or messaging, however Alexa can integrate with other smart home devices and can make purchases on behalf of the user (Amazon, n.d.).

IBM Watson: IBM’s Watson was launched to the public in 2011. Its default voice setting is that of a male and it is marketed as an IPA of high capabilities that works along industry leaders (IBM, n.d.). Rather than the user talking directly to Watson, they speak to the device – e.g. talking to the hotel room to turn off the lights. Watson also enables users to build their own personal assistants with its technology (IBM, n.d.). IBM has differentiated Watson through marketing it as a ‘white collar service’ to businesses and does not target individual consumers. The target audience of Watson being companies and its gender of male provide a narrative that it is a male’s place in the workforce of advanced capabilities, in contrast with Alexa who is confined to the home.

Research Evaluating User Experience and Satisfaction with IPAs

The research question seeks to examine user experience and satisfaction with gendered IPAs. The evaluation of user satisfaction and user experiences are an essential part of any development process for technology, in particular ones that are so intimately intertwined with user’s daily lives as the IPA is. User satisfaction is widely adopted as a subjective measure of the quality of an experience interacting with an IPA. The ability to assess or measure this user experience and satisfaction provides crucial feedback and an understanding of the improvements needed in order to enhance the system. The evaluation of experiences and satisfaction of IPAs will become increasingly important as they are adopted into homes, cars, and workplaces at a greater rate. The IPA technology is coming ubiquitous, with the occasional user reporting that an emotional attachment has been developed between themselves and the IPA (Shead, 2017). The development of these emotional attachments may stem from the fact that IPAs no longer purely answer the user’s questions, they have a character and a personality. The IPA wants to know your name and be personal, in order to help you (Geller, 2012).

Previous research in the field has focused on measuring user satisfaction based upon interactions that are goal-oriented. For example, the user will be conducting an information search or requesting the IPA to complete a specific activity. In this case, researchers have found it difficult to establish what a 'correct' response from an IPA is due to the high personalization and contextualization (e.g. to a user's location) embedded within the response. These factors make measurement of user satisfaction a complex task. Due to this complexity, studies have evaluated IPA performance through examining factors such as whether or not the assistant cannot answer a user's query explicitly or has to redirect the user to a general mobile search engine result page (Kiseleva et al., 2016).

Other studies have focused on the participant's own characteristics that may impact experiences and satisfaction levels. A study conducted by Lapatovska et al. (2011) examined a range of variables that may affect user interactions with Alexa. These variables included user age, user proficiency with the IPA technology, task completion, and command understanding. It was found that age and user proficiency had no significant impact on satisfaction, while other variables had an impact on user experience and satisfaction levels (Lapatovska et al., 2011).

In comparison to existing research, this study has a strong focus on evaluating the overall experience of the user, rather than the ability of the IPA to accurately answer questions. This focus was made possible by the standardization of questions, the device, as well as the location that the questions were asked from. In order to evaluate user experiences and satisfaction with either a male or female IPA, participants are asked to interact with the assistant and then respond to the User Experience Questionnaire (UEQ). The UEQ is a widely used evaluation tool for examining the user experience in interactive products. The 6 scales of the questionnaire measure classical usability aspects such as; efficiency, perspicuity, dependability, as well as user experience aspects, such as originality and stimulation. The UEQ was originally created in 2005 using a data analytical approach to ensure practical relevance of the constructed scales, i.e. the scales were derived from data concerning a bigger pool of items (Schrepp, 2019). The UEQ comprises 26 questions which have the form of a semantic differential (Schrepp, 2019). This means that each item is represented by two terms with opposing meanings (Schrepp, 2019). The order of the terms is randomized per item and a 7-point scale is used to reduce the central tendency bias for items (Schrepp, 2019). Participants are then asked to respond to a 5-point Likert Satisfaction Scale, ranging from very satisfied to very dissatisfied. The focus of the research is a user's self-report of experience and satisfaction with the interaction – which involved treating the IPA as a search device but also as a social agent.

Past studies that have focused on social elements of voice user interfaces, or IPAs, have included research on accents, ethnicities, personalities, and, to some extent, gender. The research has shown that existing participant biases play a role in reported experiences and satisfaction levels.

Research by Niculescu et al. (2008) examined the impact of English regional accents on user acceptance of voice user interfaces. All participants in the study were Singaporean, with one group exposed to the Singaporean regional accent while the other was exposed to the foreign British accent. The researchers hypothesized that foreign non-native accents would be judged less favorably than local accents, as prior research has demonstrated that local accents appear to be more trustworthy and friendly (Giles, 1971; Brown et al., 1985; Luhman, 1990). This may be due to accent similarly being unconsciously interpreted as a sign of equal standing and values in society (Belch & Belch, 1993; Francis, 1991). In opposition to the researcher's hypotheses, participants indicated a preference for the British accent over their local Singaporean accent. Niculescu et al. (2008) discussed the attribution of these findings to cultural and psychological biases that exist. The widely recognized stereotypes of preferring a similar accent over a foreign one may be inconsistent in situations where these cultural and psychological biases are in play (Niculescu et al., 2008).

Research in the field has also focused on gender and its impact on user experiences. A study conducted by Nass & Brave (2005) at Stanford University demonstrated that gender and place of origin strongly effect user's experiences interacting with a voice interface. It was found that users were more trusting and reported greater levels of satisfaction with voice user interfaces which had the same gender and ethnicities of their own. This reinforces the strength of similarity bias where a person favors those who are more similar to themselves (Reeves & Nass, 1997; Nass & Brave, 2005).

Furthermore, Nass' (2011) research states that female voices tend to be more pleasing to an audience as compared to a male voice. This can be explained by the established phenomenon that the human brain is developed to like female voices, even from when a child is in the womb (Nass, 2011). However, Nass indicates that user preferences can be task-based. For example, participants may respond well to a male voice giving directions, due to the association of the male voice with social dominance and power, but for the majority of communication a female voice is preferred (Nass & Yen, 2010). It is therefore expected in the results of this study that there will be a preference for interacting with female IPAs, across both male and female participants. It is also expected that female participants will prefer interacting with females more than male participants prefer interacting with females, due to similarity biases which may come into play.

Voice itself is therefore a critical design issue of any IPA, as it strongly affects user perceptions (Nass & Brave, 2005). In conclusion, the exploration of designing IPAs that conform to stereotypes which have been ingrained in society may be unjustified and even detrimental to the goal of continually enhancing user experiences and satisfaction levels.

3. Methodology

Research Design

This study will explore the research question: What are user perceptions of gendered IPAs, and what are the implications of gendering?

Previous research has indicated that users are more comfortable interacting with a female voice assistant, potentially due to notions of a female's role of service in society. The purpose of this research paper will be to explore how user experience and user satisfaction may differ depending on the gender of the IPA.

It is hypothesized that there will be an overall preference for interacting with female IPAs across all participants. Furthermore, it is hypothesized that female participants will prefer interacting with females more than male participants prefer interacting with females, due to similarity biases.

Participants

A study in the form of an interaction followed by a survey was conducted with participants to understand both user experience sentiments and user satisfaction immediately following an interaction with either a male or female Google voice assistant. Participants were recruited through Trinity College Dublin Master's and undergraduate programs. Participants were over the age of eighteen and fluent English speakers. The participant breakdown was held even between male and female genders (refer to Figure 2).

Location

The study was conducted in Trinity College Dublin Westland Square computer labs.

Procedure

The experiment was conducted on one participant at a time. Participants were provided with a Participant Information Sheet and Consent Form which they were required to read and sign (refer to Appendix 1). It was made clear to participants that the study would take approximately twenty minutes and they would be able to withdraw at any time before, during, or immediately after the experiment. However, withdrawal was no longer possible after the questionnaire had been submitted due to no way for the experimenter to identify an individual participant in the dataset.

Participants were not told prior to the experiment that the focus was on how gendered voices impact user perceptions of IPAs. This gender-specific element was not revealed prior to the experiment in order to reduce any bias that may have impacted the results.

Participants sat in a quiet room with the researcher and a mobile device containing the Google Assistant. Participants were handed the first page of the Questionnaire, which contained background questions about the participant gender, age, current use of IPAs and frequency of use of IPAs (refer to Appendix 2).

Participants were then asked to interact with the IPA by asking it a series of 10 questions (refer to Figure 2). The 10 questions were designed to get the participant to think about the IPA as both a tool and a social agent (refer to Figure 2).

Figure 2: Table of Participant Question List for IPA Interaction

#	Question	Tool vs. Social Agent
1	What is the weather today?	Tool
2	What pizza restaurants are near me?	Tool
3	Is 'The Market Bar' still open?	Tool
4	How do you feel today?	Social Agent
5	What is the latest news on BBC news?	Tool
6	What is 10% of 75?	Tool
7	What food do you like to eat?	Social Agent
8	Can you play 'Another One Bites the Dust' by Queen?	Tool
9	What time is it?	Tool
10	What is your favorite color?	Social Agent

Half of the male participants interacted with a female voice assistant, with the other half interacting with a male voice assistant. Similarly, half the female participants interacted with a female voice assistant, and the other half interacted with a male voice assistant (refer to Figure 3). The Google assistant was set to the 'Red' voice setting for the female gendered assistant, and the 'Orange' voice setting for the male gendered assistant.

Figure 3: Table for Participant-IPA Interaction Breakdown

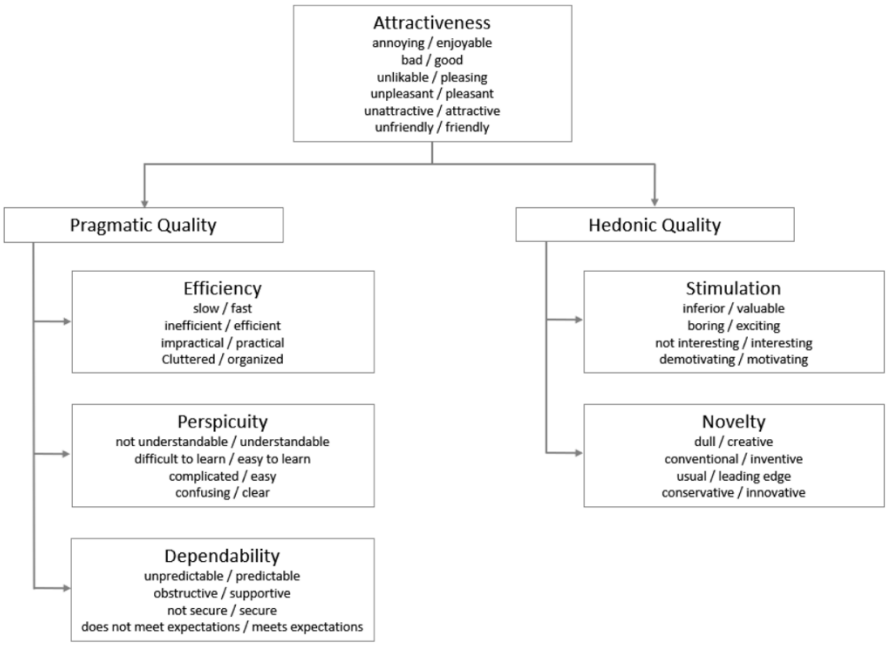
		IPA Gender:	
		Female	Male
Participant Gender	Female	8 participants	8 participants
	Male	8 participants	8 participants
Total		32 participants	

Data Collection Questionnaire

Following the interaction with the IPA, participants were asked to complete a paper UEQ (UEQ, n.d.) (refer to Appendix 2). The UEQ contains 6 scales overall with 26 items (refer to Figure 4 for the scale structure):

1. Attractiveness: Overall impression of the product in terms of likeability.
2. Perspicuity: Is it easy to get familiar with the product?
3. Efficiency: Can users solve their tasks without unnecessary effort?
4. Dependability: Does the user feel in control of the interaction?
5. Stimulation: Is it exciting and motivation to use the product?
6. Novelty: Is the product innovative and creative?

Figure 4: The scale structure of the UEQ and the scale items



Lastly, the participants were asked to rate their level of satisfaction with the interaction on a five (5) point scale ranging from very dissatisfied to very satisfied. The five-point scale was chosen as literature has indicated that five-point scales appear less confusing than seven-point scales and increase response quality along with reducing participant frustration level (Babakus & Mangold, 1992; Devlin et al., 1993; Buttle; 1996). It has also been suggested that a five-point scale is more appropriate for European surveys (Babakus & Mangold, 1992; Devlin et al., 1993; Buttle; 1996). This question on overall satisfaction is not part of the UEQ.

Participants were then given a debrief by the researcher to make them aware of the gender element to the experiment, as well as a reminder that once the paper questionnaire is submitted, they would be no longer able to withdraw from the experiment.

Data Management and Analysis

The raw data from the paper questionnaires (e.g. numbers ranging from 1 to 7) were inputted into an excel file for each of the 26 questions. As the order of positive and negative terms for each item were randomized in the questionnaire, data was first transformed to be used for statistical analysis. The raw data was transformed to range from positive 3 to -3. The positive 3 represents the most positive value and the negative 3 represents the most negative value.

Data analysis was conducted using SPSS. For each of the 6 scales, the mean was calculated across all responses. The means were then examined between groups of male and female participants, and between groups of male and female gendered IPAs.

A two-way MANOVA (Multiple Analysis of Variance) was conducted to examine the variation between groups. The two-way MANOVA allows for a test on each dependent variable (DV), to understand whether the scale result is changed by manipulating the independent variables (IV) which are the gender of participant and the gender of the voice IPA (refer to Figure 5 for the MANOVA structure). Results from the two-way MANOVA calculated a main effect and an interaction effect.

Figure 5: Two-Way MANOVA Structure

IV: Participant Gender	**IV: IPA Gender**	**DV: Scale 1**	**DV: Scale 2**	**DV: Scale 3**	**DV: Scale 4**	**DV: Scale 5**	**DV: Scale 6**
Female	Female	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Female	Male						
Male	Female						
Male	Male						

The two-way MANOVA was chosen as it allows for testing multiple dependent variables and can protect against Type 1 errors (i.e. rejecting a null hypothesis).

A separate two-way ANOVA was conducted for the user satisfaction scale rating, as it was not part of the UEQ. A two-way ANOVA is typically used to assess an interrelationship of two independent variables on a dependent variable.

Limitations

Sample size: the sample size of the experiment remains relatively small, reducing the overall power of the data collected and increasing the margin of error in results.

Sample profile: the profile of many of the participants may be similar (e.g. age, background, current academic status), which may not accurately represent wider population groups and the results that may have been obtained from a more diverse sample.

Data collection process: social desirability bias may be present in the questionnaire as users may under-report bad experiences and over-report good experiences with the IPA in order to please the researcher.

Data analysis: Although the MANOVA is a strong analysis model, the MANOVA is complex and may be difficult to understand which independent variables are affecting the dependent variables. In order for the MANOVA to produce accurate results, outliers must be carefully checked for in order to prevent Type 1 or Type 2 errors. Furthermore, one degree of freedom is lost with the addition of each new variable.

4. Results

Trinity College Dublin School of Computer Science & Statistics research ethical guidelines were abided by during the selection of participants and administration of the study. A research ethics application was approved by the Trinity College Dublin Research Ethics Committee in February 2019.

Two-Way MANOVA Assumptions for All Scales

The two-way MANOVA was utilized to test the interaction effect between the two independent variables of participant gender and IPA gender (each with two levels of male and female) and the 6 dependent variables. The 6 dependent variables were the 6 scales tested through the UEQ. 32 participants took part in the study, aged between 18 and 59. Participants were recruited through Trinity College Dublin undergraduate and graduate programs.

There were equal numbers of combinations of each of the levels of independent variables, with 16 males and 16 females. Half of the 16 males interacted with a male IPA and half interacted with a female IPA. Similarly, half of the 16 females interacted with a male IPA and half interacted with a female IPA. There was no missing data within the dataset.

Prior to conducting a MANOVA on all scales, certain assumptions were tested to ensure the MANOVA could be utilized appropriately on the dataset. These assumptions were:

1. Independence of Observations on Dependent Variables
2. Multivariate Normality: Tested using the Shapiro-Wilk test. Results from the Shapiro-Wilk test were found to be significant for the attractiveness scale, the perspicuity scale, and the efficiency scale at. This implies that the distribution is not normal for the three scales. However, results were insignificant for the dependability scale, the stimulation scale, and the novelty scale, assuming normality. A further test of Mahalanobis Distance was used by conducting a linear regression. The maximum value of 1.94 for Mahalanobis Distance was found to be less than the critical value, implying that the assumption of multivariate normality has been met.
3. Linearity: Linearity was examined between dependent variables at each level of the independent variables. Scatterplot matrices and bivariate correlations were used to test linearity. The Pearson Correlation values between dependent variables were all found to be less than 0.9, indicating that the variables are not multicollinear.
4. Homogeneity of Variance: Homogeneity of Variance was tested using Box's Test of Equality of Covariance Matrices. The p value was found to be insignificant (P = .001), therefore it is not

possible to reject the null hypotheses and it is assumed that covariance matrices of dependent variables are equal across groups.

The minor violations of MANOVA assumptions shaped the further analysis. For example, Pillai's Trace statistics were interpreted instead of Wilks' Lambda as there is evidence that Pillai's trace is more robust than the other statistics to violations of model assumption (Olson, 1974).

Summary of Results

The significance level of this statistical analysis is set at 0.05.

Figure 6: Summary Table of Results

Independent Variable	Dependent Variable	F-Test	Remarks
Participant Gender	Attractiveness	$F_{1,1} = 1.63$, $p > .05$	Not significant
	Perspicuity	$F_{1,1} = .14$, $p > .05$	Not significant
	Efficiency	$F_{1,1} = .99$, $p > .05$	Not significant
	Dependability	$F_{1,1} = .00$, $p > .05$	Not significant
	Stimulation	$F_{1,1} = 4.67$, $p < .05$	Estimated marginal mean for 'Female' (.94) was greater than for 'Male' (.28), indicating that belonging to the 'Female' participant group makes a higher rating on the Stimulation scale more likely.
	Novelty	$F_{1,1} = .90$, $p > .05$	Not significant
IPA Gender	Attractiveness	$F_{1,1} = 1.06$, $p > .05$	Not significant
	Perspicuity	$F_{1,1} = .55$, $p > .05$	Not significant
	Efficiency	$F_{1,1} = .15$, $p > .05$	Not significant
	Dependability	$F_{1,1} = .03$, $p > .05$	Not significant
	Stimulation	$F_{1,1} = 2.08$, $p > .05$	Not significant
	Novelty	$F_{1,1} = .04$, $p > .05$	Not significant
Participant Gender * IPA Gender	Attractiveness	$F_{1,1} = .51$, $p > .05$	Not significant
	Perspicuity	$F_{1,1} = 4.39$, $p < .05$	Estimated marginal means were similar across both 'Male' participants interacting with a male IPA (2.0) and interacting with a female IPA (2.3) and 'Female' participants interacting with a male IPA (2.4). However, 'Female' participants interacting with a female IPA had a lower mean value than other groups (1.7), indicating it is more likely for females interacting with female IPAs to report lower scores on the Perspicuity scale.
	Efficiency	$F_{1,1} = 1.17$, $p > .05$	Not significant
	Dependability	$F_{1,1} = .45$, $p > .05$	Not significant
	Stimulation	$F_{1,1} = .17$, $p > .05$	Not significant
	Novelty	$F_{1,1} = 1.21$, $p > .05$	Not significant

Figure 7: Chart Displaying Mean Values of the 6 Scales Across All Participant Groups

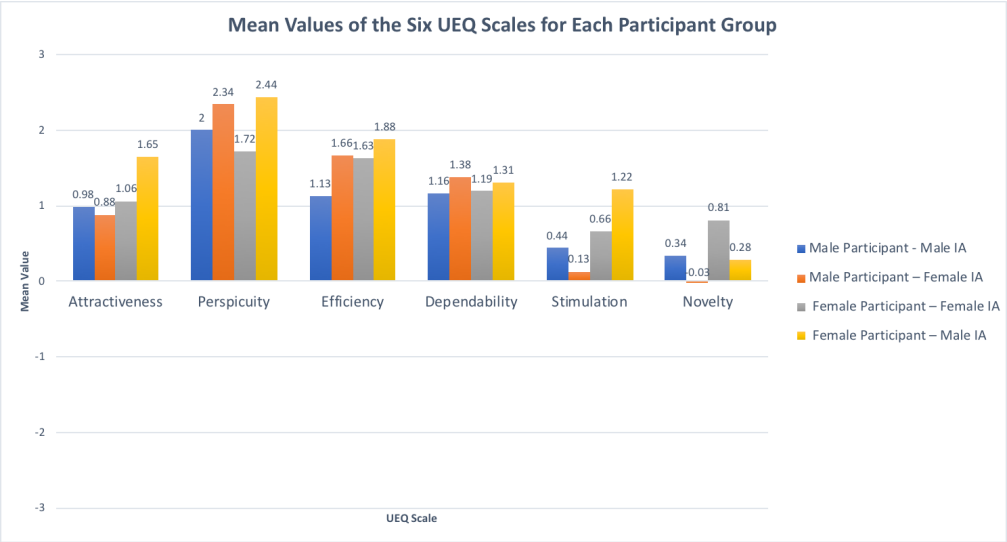


Figure 8: Chart Displaying Significant Value for the Stimulation Scale

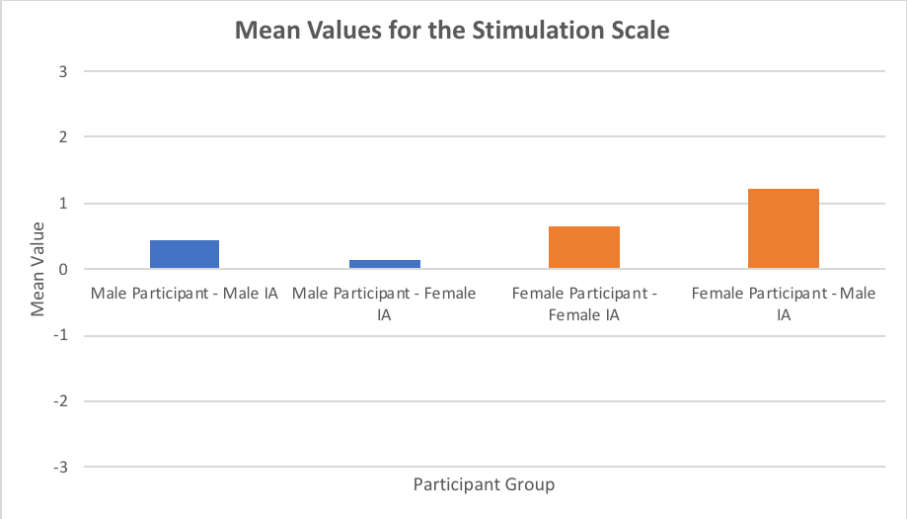


Figure 9: Chart Displaying Mean Values for the Perspicuity Scale

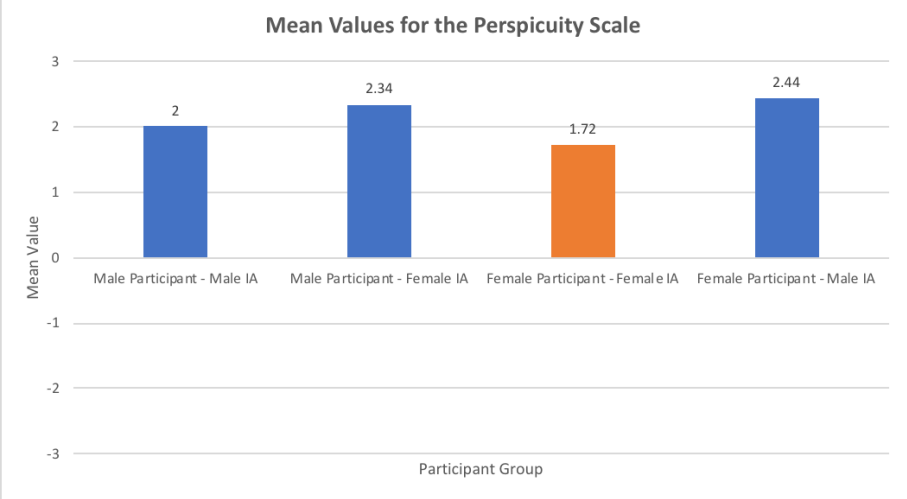


Figure 10: Table of Mean Values of the 6 Scales, Grouped by Participant and IPA Gender

	All Participant – All IPA	Male Participant – Male IPA	Male Participant – Female IPA	Female Participant – Female IPA	Female Participant – Male IPA
Attractiveness	1.14	0.98	0.88	1.06	1.65
Perspicuity	2.13	2.00	2.34	1.72	2.44
Efficiency	1.57	1.13	1.66	1.63	1.88
Dependability	1.26	1.16	1.38	1.19	1.31
Stimulation	0.61	0.44	0.13	0.66	1.22
Novelty	0.35	0.34	-0.03	0.81	0.28

Findings from Mean Values (refer to Figures 7 and 10):

- Attractiveness was reported higher amongst female participants interacting with male IPAs (1.65), as compared to the mean value across all participants (1.14).
- Stimulation was reported higher amongst females interacting with male IPAs (1.22), as compared to the mean value across all participants (0.61).

Two-Way MANOVA Results for All Scales

The null hypotheses for the statistical analysis are as follows:

- The means of the first independent variable, participant gender, are equal
- The means of the second independent variable, IPA gender, are equal
- There is no interaction effect between participant gender and IPA gender.

Multivariate Tests[a]

Pillai's Trace results were interpreted in the multivariate tests as there were violations of assumptions prior to conducting the two-way MANOVA. The multivariate tests were insignificant across all independent variables. It is therefore not possible to reject the null hypothesis and assume that the scales do not differ by participant gender, IPA gender, or the interaction between participant and IPA gender. Furthermore, the interaction effect of participant gender * IPA gender can be observed to be weaker than the effect of solely one independent variable on the dependent variables.

Levene's Test of Equality of Error Variances[a]

Levene's Test of Equality of Error Variances demonstrated that no main effects were found for the attractiveness scale, the perspicuity scale, the dependability scale, the stimulation scale, and the novelty scale ($P > .05$). The insignificant value means that the variability across the 6 scales is not significantly different. On the other hand, the efficiency scale was found to be significant ($P < .05$). This result indicates that participants found the efficiency of IPAs to greatly vary. The significant

value means that for the efficiency scale, the null hypothesis of equal variances is rejected and there is a difference between variances.

Tests of Between-Subjects Effects

Test of Between-Subjects Effects showed that participant gender was significant only for the stimulation scale ($P < .05$) (refer to Figure 8). This result indicates that participant gender may significantly predict participant's level of stimulation while interacting with an IPA. On the other hand, IPA gender was not significant across any of the scale ratings ($P > .05$). An interaction was found between participant gender and IPA gender, where the analysis demonstrates that the perspicuity scale was found to be significant ($P < .05$) (refer to Figure 9).

Post-hoc tests were not conducted as part of the two-way MANOVA as each independent variable only contained two levels, male and female. In order to determine what subgroup makes the dependent variable more likely, the 'Estimated Marginal Means' were examined to provide an indication of which independent variable makes a dependent variable more likely (refer to Figures 6 and 10).

In order to re-check the two-way MANOVA results, 6 two-way ANOVA's were conducted on the data. Results from the two-way ANOVAs were the same as results from the two-way MANOVA. Refer to Appendix 3 for detailed tables of results.

Two-Way ANOVA for Satisfaction Scale

A two-way ANOVA was conducted to examine the mean differences and interaction between participant gender and IPA gender on satisfaction ratings. The test of between-subjects effects demonstrates that there is no statistically significant effect of participant gender, IPA gender, or the interaction between participant gender and IPA gender on satisfaction. These results indicate that there is no significant difference between female and male participants in terms of their reported satisfaction after interacting with an IPA. The results also indicate that there is no significant difference between female and male gendered IPAs on reported satisfaction levels. Lastly, the interaction between participant gender and IPA gender did not impact the overall satisfaction levels reported by participants. Importantly, the results demonstrate that there is no preference for a particular gender of IPA. Refer to Appendix 4 for detailed tables of results.

5. Discussion and Conclusion

Discussion

This research paper explored how user experience and user satisfaction ratings may differ depending on the gender of the IPA. A study in the form of an interaction followed by a survey was conducted with participants over the age of eighteen to understand both user experience sentiments and user satisfaction immediately following an interaction with either a male or female voice assistant. The study focused on the Google Assistant, due to its intentional attempt at avoiding a gendered name and personality, as well as its prevalence.

It was hypothesized that there would be a significant preference for female IPAs on the 6 scales of: 1) Attractiveness; 2) Perspicuity; 3) Efficiency; 4) Dependability; 5) Stimulation; and 6) Novelty, as well as on the overall satisfaction rating. It was also hypothesized that female participants who interacted with the female IPA would report higher levels across the 6 scales as well as on satisfaction, as compared to males, due to similarity biases that previous research has identified may come into play.

Contrary to expectations, the results showed that IPA gender did not significantly impact user experience across any of the 6 UEQ scales, or overall user satisfaction. This result is in contrast to the literature examined, which stated a preference for the female voice across both male and female users. The lack of the impact of IPA gender on user experiences and satisfaction is a notable result, as it provides support for the development of gender-neutral technology. Gender is not only the voice that the IPA uses to communicate. As seen in the literature review, gender is a culmination of the audible voice, speech patterns, personality characteristics, appearance, and the name of the device itself.

Conclusion

In conclusion, if the trend of developing gendered technology is not halted, certain stereotypical beliefs and roles of genders will become further entrenched and cemented through technology. As people become increasingly reliant on devices such as IPAs for goal-oriented tasks and social support, it will be important to mitigate the negative social consequences of gendering. There is an acute need for the involvement of more women and diverse groups in the prototyping, design and development of software agents. Advancing women's careers in the area of computing is a right in itself, but also essential to ensure that historical power structures and roles of women are not continually reinforced through technology. Technology organizations must be increasingly challenged in their decisions to gender devices which touch people's lives so intimately.

Limitations and Further Research

This study is not without its limitations which must be discussed in relation to the findings. Firstly, the sample size of the experiment remains relatively small with a total number of 32 participants. The sample size reduces overall power and ability to generalize the findings to a larger population group. The sample profile was also quite similar in terms of key characteristics of participants. As participants were recruited through Trinity College Dublin academic programs, participants may be of similar ages, backgrounds, and educational level. The characteristics present in the sample may not accurately reflect a wider population group, and therefore results may have been different if the experiment had been conducted with a more diverse sample. In terms of the experiment itself, during the experiment participants had a limited and brief interaction with the IPA. The interaction was also closely dictated through a standardized set of questions, location, and device. These factors make for an unnatural setting as compared to how users may normally interact with their IPAs (e.g. in the car, wider range of queries, more personalized information).

Further research is required to explore how user experiences and satisfaction levels are impacted by gendered devices in specific social realms. For example, IPAs in the context of providing assistive support and companionship to the elderly or children should be evaluated in depth. Another topic worth exploring is studying the situations where there is a failed interaction between user and IPA, to explore whether or not a negative interaction is made worse or better with a gendered device. Lastly, further research could explore to what extent each element of the gendering process has an effect on user satisfaction or experiences and how to strive towards gender-neutral devices. This research could break down the naming of an IPA, the speech patterns, and voice, to understand how each element can be modified to appear gender-neutral to the user.

Bibliography

Adam, A. (1996). Constructions of Gender in the History of Artificial Intelligence, *IEEE Annals of the History of Computing*, 18(3), pp 47-53.

Adam, J. & Skubic, M. (2005). Introduction to the Special Issue on Human–Robot Interaction, *IEEE Trans. Systems, Man, And Cybernetics—Part A: Systems and Humans*, 35(4).

Alesich, S. & Rigby, M. (2017). Gendered Robots – Implications for Our Humanoid Future, *IEEE Technology and Society Magazine*, 36(2).

Anisfeld, M., Bogo, N., & Lambert, W. E. (1962). Evaluational Reactions to Accented English Speech, *Journal of Abnormal and Social Psychology*, 65(4), pp 223–231.

Apple (n.d.). *Siri does more than you ever. Even before you ask*. Retrieved from: https://www.apple.com/siri/.

Apple (2011). *Apple Launches iPhone 4S, iOS 5 & iCloud,* Retrieved from: https://www.apple.com/ie/newsroom/2011/10/04Apple-Launches-iPhone-4S-iOS-5-iCloud/.

Apple (2017). *HomePod arrives February 9, available to order this Friday,* Retrieved from: https://www.apple.com/ie/newsroom/2018/01/homepod-arrives-february-9-available-to-order-this-friday/.

Amazon (n.d.). Amazon Alexa, Retrieved from: https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011.

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice, *Contemporary Educational Psychology*, 30(1), pp 117–139.

Babakus, E. & Mangold, W.G. (1989). *Adapting the SERVQUAL scale to health care environment: an empirical assessment*, Enhancing Knowledge Development in Marketing, American Marketing Association, Chicago, IL, 195.

Balakrishnan, K., & Honavar, V. (2001). Evolutionary and neural synthesis of intelligent agents, *Advances in the evolutionary synthesis of intelligent agents,* pp 1–27.

Belch, G. E. & Belch, M. A. (1993). *Introduction to Advertising and Promotion: An Integrated Marketing Communications Perspective,* McGraw-Hill Inc., 2nd ed., Boston.

Birch, Dawn, & McPhail, J. (1997). *The impact of accented Speech in international advertisement,* Global Business Languages, Cultures and Cross Cultural Awareness in the Professions, 2.

Brown, B. L., Giles, H. & Thakerar, J. N. (1985). Speaker Evaluation as a Function of Speech Rate, Accent and Context, *Language and Communication*, 5 (3), pp 207–220.

Brown, J. (2015). *The amazon echo is more than a Bluetooth speaker – It's a bedtime buddy,* Retrieved from https://www.recode.net/2015/2/9/11558754/the-amazon-echo-is-more-than-a-bluetooth-speaker-its-a-bedtime-buddy.

Burgoon, J. K., & Hale, J. L. (1984). The fundamental topic of relational communication. *Communication Monographs*, 51, pp 193–214.

Buttle, F. (1996). SERVQUAL: review, critique, research agenda, *European Journal of Marketing,* 30(1), pp 8-32.

Carey, J. W. (1989). *Communication as culture: Essays on media and society*. New York, NY: Routledge.

Carey, J. W. (1990). Technology as a totem for culture: On Americans' use of high technology as a model for social order. *American Journalism*, 7(4), pp 242–251.

Cassell, J., & Tartaro, A. (2008). Intersubjectivity in human–agent interaction, *Interaction Studies*, 8(3), pp 391–410.

Castells, M. (2013). *Communication power*. Oxford, UK: Oxford University Press.

Chun, W. H. K. (2011). *Programmed Visions: Software and Memory,* Cambridge: MIT Press.

Devlin, S. J., Dong, H. K., & Brown, M. (1993). Selecting a scale measuring quality, *Marketing Research: A Magazine of Management*, 5(3), pp 12-17.

Elliot, A. & Hare, J. (2019). Will talking to AI voice assistants re-engineer our human conversations? *The Conversation*, Retrieved from: https://theconversation.com/will-talking-to-ai-voice-assistants-re-engineer-our-human-conversations-108922.

Foner, E. M. (2015). *Alexa, my love. Thy name is inflexible, but thou art otherwise a nearly perfect spouse*, Retrieved from https://www.amazon.com/gp/review/RJVDJIP1OE8?ref_=glimp_1rv_cl.

Francis, J. N. P. (1991).When in Rome? The Effects of Cultural Adaptation on Intercultural Business Negotiations*, Journal of International Business Studies*, pp 403–428.

Gartner (2018). *Hype Cycle for the Digital Workplace 2018*.

Geller, T. (2012). Talking to Machines, Technology, *Communications of the ACM*, 55(4).

Ghorbel, M., Segarra, M. T., Kerdreux, J., Keryell, R., Thepaut, A., & Mokhtari, M. (2004), Networking and communication in smart home for people with disabilities, *ICCHP 2004: Proceedings of the Ninth International Conference on Computers Helping People*, pp 937–944, Berlin: Springer-Verlag.

Giles, H. (1971). Patterns of Evaluation in Reactions to RP, South Welsh and Somerset Accented Speech, *British Journal of Social and Clinical Psychology*, 10, pp 280–281.

Goffman, E. (1959). *The presentation of self in everyday life*. New York, NY: Anchor Books.

Google (n.d.). *Google Assistant*, Retrieved from: https://assistant.google.com/.

Griggs, B. (2011). *Why computer voices are mostly female*, CNN.com, Retrieved from: https://edition.cnn.com/2011/10/21/tech/innovation/female-computer-voices/.

Guzman, A. L. (2015). *Imagining the voice in the machine* (Doctoral dissertation), University of Illinois at Chicago.

Guzman, A. L. (2017). Voices in and of the machine: Source orientation towards mobile virtual assistants, *Computers in Human Behavior*, 90, pp 343-350.

Hannon, C. (2016). Gender and Status in Voice User Interfaces, I*nteractions*, 23(3), pp 34-37.

Heitzman, A. (2017). How popular is voice search? *Higher Visibility,* Retrieved from: https://www.highervisibility.com/blog/how-popular-is-voice-search/.

Hester, H. (2016). Technology Becomes Her, *New Vistas*, 3(1), pp 46-50.

IBM (n.d.). *IBM Watson,* Retrieved from: https://www.ibm.com/watson.

Kiesler, S., & Sproull, L. (1997), Social human–computer interaction, *Human values and the design of technology*, pp 191–199, Stanford, CA: CLSI Publications.

Kiseleva, J., Crook, A. C., Williams, K., Zitouni, I., Awadallah, A. H., Anastasakos, T. (2016). Predicting user satisfaction with Intelligent Assistants, SIGIR 2016, *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval,* pp 45-54.

Lapatovska, I. , O'Brien, H. L., Rieh, S. Y. & Wildemuth, B. (2011). Capturing the complexity of information interactions: Measurement and evaluation issues. *Proceedings of the American Society for Information Science and Technology*, 48, pp 1-3.

Larson (2010) *Advances in Speech Recognition, Mobile Environments, Call Centers and Clinic*, Springer Science+Business Media, LLC.

Laurel, B. (1997), Interface agents: Metaphors with character, *Human values and the design of computer technology*, pp 207–219, Stanford, CA: CSLI Publications.

Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning, *ACM/IEEE 1st International Workshop on Gender Equality in Software Engineering,* University College Dublin.

Lee, K. M., & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, *7*(1), pp 31–45.

Lenard, D. B. (2017). Gender differences in the personal pronouns usage on the corpus of congressional speeches*, Journal of Research Design and Statistics in Linguistics and Communication Science,* 3(2), pp 161-188.

Luger, E. & Sellen, A. (2016). Like having a really bad pa: the gulf between user expectation and experience of conversational agents, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, pp 5286–5297.

Luhman, R. (1990). Appalachian English stereotypes: Language attitudes in Kentucky, *Language in Society*, 19, pp 331-348.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research, *Interaction Studies*, 7(3), pp 297–337.

Martin, T. B. (1976). Practical applications of voice input to machines, *Proceedings of the IEEE*, 64(4), pp 487–501.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice, *Journal of Educational Psychology*, 95(2), pp 419–425.

Möller, S., Krebber, J., & Smeele, P. (2006). Evaluating the speech output component of a smart-home system. *Speech Communication*, 48(1), pp 1–27.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2), pp 177–213.

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, Mass: MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers, *Journal of Social Issues*, 56(1), pp 81–103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the AMC/SIGCHI Conference on Human Factors in Computing Systems*, pp 72– 78, New York: ACM.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can Computer Personalities be Human Personalities? *International Journal of Human-Computer Studies*, *43*, pp 223– 239.

Nass, C., Moon, Y., Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers, *Journal of Applied Social Psychology*, 27(10), pp 864-876.

Nass, C., & Yen, C. (2010). *The Man Who Lied to his Laptop*, New York: The Penguin Group.

Niculescu, A., White, G., Lan, S. S., Waloejo, R. U., & Kawaguchi, Y. (2008). Impact of English Regional Accents on User Acceptance of Voice User Interfaces, *Institute for Infocomm Research*.

Norman, D. A. (1994), How might people interact with agents? *Communications of the ACM,* 37(7), pp 68–71.

Orbis Research (2017). *Global Virtual Personal Assistants (VPA) Market Segmentation, Size, Trends, Analysis, and Forecasts*, Orbis Research, Retrieved from: https://www.orbisresearch.com/contacts/request-sample/133667.

Oudshoorn, N. (2003). Clinical trials as a cultural niche in which to configure the gender identities of users: The case of male contraceptive development, *Users matter: The co-construction of users and technologies*, pp 209- 228, Cambridge, MA: The MIT Press.

Pennebaker, J. (2013). *The Secret Life of Pronouns: What Our Words Say About Us.* Bloomsbury Press, New York, 2013, 18.

Phan, T. (2017). The Materiality of the Digital and the Gendered Voice of Siri, *Transformations*, 1(29), pp 24-33.

Pierce, D. (2016). *Hey Siri, it's time to put up or shut up*, Retrieved from https://www. wired.com/2016/06/hey-siri-time-put-shut/.

Pierce, D. (2016). *I spent a week yelling at Siri in macOS Sierra,* Retrieved from https:// www.wired.com/2016/06/spent-week-yelling-siri-macos-sierra/.

Purington, A., Taft, J.G., Sannon, S., Bazarova, N.N., & Taylor, S.H. (2017). "Alexa is my new BFF". *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17,* pp 2853–2859. ACM Press, New York, New York, USA.

Ravitz, J. (2013). *'I'm the original voice of Siri',* CNN, Retrieved from http://www.cnn.com/2013/10/04/tech/mobile/bennett-siri-iphone-voice/index.html.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*, Cambridge: Cambridge University Press.

Rothschild, J. (1983). Foreword. In J. Rothschild (Ed). *Machina ex dea: Feminist perspectives on technology* (pp. vii-viii). New York, NY: Pergamon Press.

Sadun, E., & Sande, S. (2012). *Talking to Siri*, Indianapolis: Que Pub.

Schrepp, M. (2019). *User Experience Questionnaire Handbook*, UEQ Version 7.

Shead, S. (2017). *REPORT: 1 in 4 people have fantasised about Alexa, Siri, and other AI assistants*, Business Insider.

Skalski, P., & Tamborini, R. (2007). The Role of Social Presence in Interactive Agent-Based Persuasion, *Media Psychology*, *10*(3), pp 385–413.

Smith, B. & Shum, H. (2018). *The Future Computed: Artificial Intelligence and its role in society,* Microsoft.

Sparks, B. A., Bradley, G. L., & Callan, V. J. (1997). The impact of staff empowerment and communication style on customer evaluations: the special case of service failure, *Psychology & Marketing,* 14(5), pp 475-493.

Statistica (2019). *Number of unique active virtual digital assistants users worldwide, from 2015 to 2021 (in millions),* Retrieved from: https://www.statista.com/statistics/589071/worldwide-virtual-digital-assistants-active-unique/.

Suchman, L. A. (2009). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.), New York, NY: Cambridge University Press.

Turkle, S. (2007). Authenticity in the age of digital companions, *Interaction Studies*, *8*(3), pp 501–517.

UEQ (n.d.). User Experience Questionnaire, Retrieved from: https://www.ueq-online.org/.

UNESCO (2017). *Women in Science,* UIS UNESCO, Retrieved from: http://uis.unesco.org/en/topic/women-science.

West, C., & Zimmerman, D. H. (1987). Doing gender, *Gender & Society, 1*(2), pp 125–151.

White, M. (2015). *Producing women: The internet, traditional femininity, queerness, and creativity.* New York, NY: Routledge.

Woods, H. S. (2018). Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism*, Critical Studies in Media Communication,* 35(4), pp 334-349.

Yerrapragada, C., & Fisher, P. S. (1993). Voice controlled smart house*. ICCE 1993: International Conference on Consumer Electronics: Digest of Technical Papers*, pp 154–155, Piscataway, NJ.

Appendices

Appendix 1: Participant Consent Form

Trinity College Dublin
Informed Consent Form

Perceptions of Intelligent Personal Assistants

Lead Researcher:
Emily Harayda
Email: haraydae@tcd.ie
Location: School of Computer Science and Statistics, Trinity College Dublin

Background of research:
This study seeks to understand perceptions and user experience/ satisfaction with intelligent personal assistants. An intelligent personal assistant is a mobile software agent that can perform tasks or services on behalf of an individual.
Location of the study: The study will be conducted in the computer lab of the Trinity College Dublin Westland Square building, located at 8 Westland Square, Trinity College, Dublin 2.
Procedures of this study: In this experiment, you are asked to interact with an intelligent personal assistant through asking it a series of ten questions. The ten questions are:

1. What is the weather today?
2. What pizza restaurants are near me?
3. Is 'The Market Bar' still open?
4. How do you feel today?
5. What is the latest news on BBC news?
6. What is 10% of 75?
7. What food do you like to eat?
8. Can you play Another One Bites the Dust by Queen?
9. What time is it?
10. What is your favourite colour?

The ten questions should take approximately 5 minutes. After interacting with the intelligent personal assistant you will have to respond to a number of questions with a scale describing your impressions and your level of satisfaction with the interaction (as described in the Questionnaire sheet). You may take as long as you like to answer the questions. However, you will not be permitted to interact further with the intelligent assistant. Please do not hesitate to ask the experimenter if you have any questions before the experiment begins. The whole experiment will last approximately 20 minutes.

Publication: The data we collect will be analysed together with the data collected from other participants, and generalised results and conclusions will be drawn from these experiments may potentially be submitted for publication at conferences and/or scientific journals. Your name will not be used in any report or article.
Data Management: Data will be retained for a period of 5 years following the study. Data will be kept securely stored in two password protected folders on a hard drive. Physical copies of signed consent forms, participant information sheets, and questionnaires will be stored in a password protected safe.
Conflicts of interest: There are no conflicts of interest with regard to the research topic and with you.
Individual results may be aggregated anonymously and research reported on aggregate results.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that if the results of the research have been published, or my data has been fully anonymised so that it can no longer be attributed to me, then it will no longer be possible to withdraw.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- I have received a copy of this agreement.

By signing this document I consent to participate in this study, and consent to the data processing necessary to enable my participation and to achieve the research goals of this study. Secondary data analysis may also be utilised to provide a meta-analysis of perceptions of intelligent personal assistants.
PARTICIPANT'S NAME:
PARTICIPANT'S SIGNATURE:
Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.
RESEARCHERS CONTACT DETAILS: haraydae@tcd.ie
RESEARCHER'S SIGNATURE:
Date:

Appendix 2: Participant Information Sheet and Questionnaire

Trinity College Dublin
Participant Information Sheet

Perceptions of Intelligent Personal Assistants

Lead Researcher:
Emily Harayda
Email: haraydae@tcd.ie
Location: School of Computer Science and Statistics, Trinity College Dublin

Research background: You are invited to participate in this study on 'Perceptions of Intelligent Personal Assistants' which aims to identify how perceptions, user experience and user satisfaction levels may differ across participants.
Who is undertaking this experiment: This study is being carried out as part of the requirements for a M.Sc. in Interactive Digital Media at Trinity College Dublin.
Participant profile: We are seeking participants, over the age of 18, who are fluent in the English language.
Location of the study: The study will be conducted in the computer lab of the Trinity College Dublin Westland Square building, located at 8 Westland Square, Trinity College, Dublin 2.
Procedures: In this experiment, you are asked to interact with an intelligent personal assistant. An intelligent personal assistant is a mobile software agent that can perform tasks or services on behalf of an individual (e.g. Google Assistant). You will be asked to interact with the assistant by asking it a series of ten questions. The ten questions are:

1. What is the weather today?
2. What pizza restaurants are near me?
3. Is 'The Market Bar' still open?
4. How do you feel today?
5. What is the latest news on BBC news?
6. What is 10% of 75?
7. What food do you like to eat?
8. Can you play 'Another One Bites the Dust' by Queen?
9. What time is it?
10. What is your favourite colour?

The ten questions should take approximately 5 minutes.
After interacting with the intelligent personal assistant you will have to respond to a number of questions with a scale describing your impressions of the assistant and your level of satisfaction with the interaction (as described in the Questionnaire sheet). User satisfaction refers to your comfort level and acceptability of the assistant during the interaction. You may take as long as you like to answer the questions. However, you will not be permitted to interact further with the assistant.
Please do not hesitate to ask the experimenter if you have any questions before the experiment begins. The whole experiment will last approximately 20 minutes.
Before you begin the experiment, you will be asked to carefully read the informed consent and sign if you agree to continue with the experiment. Your participation in this study is entirely voluntary and you are able to withdraw without penalty. Your results will be kept strictly confidential, stored using a unique ID number and the experimenter will not be able to identify your data or link them with your personal details. As there is no way for us to identify an individual participant once they have submitted their data, it will no longer be possible to withdraw once you have submitted the questionnaire. The data we collect will be analysed together with the data collected from other participants, and generalised results and conclusions will be drawn from these experiments may potentially be submitted for publication at conferences and/or scientific journals. Your name will not be used in any report or article.

Data Management: Data will be retained for a period of 5 years following the study. Data will be kept securely stored in two password protected folders on a hard drive. Physical copies of signed consent forms, participant information sheets, and questionnaires will be stored in a password protected safe.
At the end of the experiment the researcher will explain more of the reasons behind the experiment and give you a chance to ask questions.
We do not anticipate any risks to participants in this study.
In an extremely unlikely event that illicit activity is reported during the study, these will be reported to appropriate authorities.

There are no conflicts of interest with regard to the research topic and with you.
If you have any further questions, please do not hesitate to ask the experimenter.

Participant Questionnaire

Please help us by filling in the below information. All answers to the questions are optional.

Sex: **Native English Speaker:** **Age:** .....

Male ☐ Female ☐ Yes ☐ No ☐

Do you currently use an intelligent assistant? If yes, which one?

None ☐
Google Assistant ☐
Siri ☐
Alexa ☐
Cortana ☐
Bixby ☐
Other ☐ _____ (please fill name of 'other' assistant)

How often do you interact with the assistant(s)?:

Never ☐
Less than once per year ☐
Yearly ☐
Monthly ☐
Weekly ☐
Hourly ☐

If you currently use an intelligent assistant, what activities do you mainly use the assistant for? (e.g. checking the weather, playing music, etc.)

[Participant Interacts with Assistant]
Following the interaction:

For the assessment of the assistant, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Sometimes you may not be completely sure about your agreement with an attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line. It is your personal opinion that counts. Please assess the assistant now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Source: User Experience Questionnaire (UEQ) (https://www.ueq-online.org/)

Overall, to what extent are you satisfied with your experience using the assistant?

Very Satisfied ☐ Satisfied ☐ Neither ☐ Dissatisfied ☐ Very Dissatisfied ☐

Debrief:

The experiment was intended to examine how gendered voices impact user perceptions of the intelligent personal assistant. The gender-specific element was not revealed prior to the experiment in order to reduce bias that may impact the results. Thank you. You have completed the experiment. If you have any further questions, please do not hesitate to ask the experimenter.

Appendix 3: Results – Two-Way MANOVA for UEQ

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Participant_Gender	Scale_1	1.458	1	1.458	1.631	.212	.055
	Scale_2	.070	1	.070	.137	.714	.005
	Scale_3	1.033	1	1.033	.994	.327	.034
	Scale_4	.002	1	.002	.004	.952	.000
	Scale_5	3.445	1	3.445	4.674	.039	.143
	Scale_6	1.221	1	1.221	.898	.351	.031
IA_Gender	Scale_1	.949	1	.949	1.061	.312	.037
	Scale_2	.281	1	.281	.547	.466	.019
	Scale_3	.158	1	.158	.152	.699	.005
	Scale_4	.018	1	.018	.033	.857	.001
	Scale_5	1.531	1	1.531	2.077	.161	.069
	Scale_6	.049	1	.049	.036	.851	.001
Participant_Gender * IA_Gender	Scale_1	.454	1	.454	.508	.482	.018
	Scale_2	2.258	1	2.258	4.393	.045	.136
	Scale_3	1.221	1	1.221	1.174	.288	.040
	Scale_4	.236	1	.236	.447	.509	.016
	Scale_5	.125	1	.125	.170	.684	.006
	Scale_6	1.643	1	1.643	1.208	.281	.041

Estimated Marginal Means

1. Participant_Gender

Dependent Variable	Participant_Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Scale_1	Male	.928	.236	.444	1.412
	Female	1.355	.236	.871	1.839
Scale_2	Male	2.172	.179	1.805	2.539
	Female	2.078	.179	1.711	2.445
Scale_3	Male	1.391	.255	.868	1.913
	Female	1.750	.255	1.228	2.272
Scale_4	Male	1.266	.182	.893	1.638
	Female	1.250	.182	.878	1.622
Scale_5	Male	.281	.215	-.158	.721
	Female	.938	.215	.498	1.377
Scale_6	Male	.156	.292	-.441	.753
	Female	.547	.292	-.050	1.144

2. IA_Gender

Dependent Variable	IA_Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Scale_1	Male	1.314	.236	.830	1.798
	Female	.969	.236	.485	1.454
Scale_2	Male	2.219	.179	1.852	2.586
	Female	2.031	.179	1.664	2.398
Scale_3	Male	1.500	.255	.978	2.022
	Female	1.641	.255	1.118	2.163
Scale_4	Male	1.234	.182	.862	1.607
	Female	1.281	.182	.909	1.654
Scale_5	Male	.828	.215	.388	1.268
	Female	.391	.215	-.049	.830
Scale_6	Male	.313	.292	-.285	.910
	Female	.391	.292	-.207	.988

3. Participant_Gender * IA_Gender

Dependent Variable	Participant_Gender	IA_Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Scale_1	Male	Male	.981	.334	.297	1.666

		Female	.875	.334	.190	1.560
	Female	Male	1.646	.334	.962	2.331
		Female	1.064	.334	.379	1.748
Scale_2	Male	Male	2.000	.253	1.481	2.519
		Female	2.344	.253	1.825	2.863
	Female	Male	2.438	.253	1.918	2.957
		Female	1.719	.253	1.200	2.238
Scale_3	Male	Male	1.125	.361	.386	1.864
		Female	1.656	.361	.918	2.395
	Female	Male	1.875	.361	1.136	2.614
		Female	1.625	.361	.886	2.364
Scale_4	Male	Male	1.156	.257	.630	1.683
		Female	1.375	.257	.848	1.902
	Female	Male	1.313	.257	.786	1.839
		Female	1.188	.257	.661	1.714
Scale_5	Male	Male	.438	.304	-.184	1.059
		Female	.125	.304	-.497	.747
	Female	Male	1.219	.304	.597	1.841
		Female	.656	.304	.034	1.278
Scale_6	Male	Male	.344	.412	-.501	1.188
		Female	-.031	.412	-.876	.813
	Female	Male	.281	.412	-.563	1.126
		Female	.813	.412	-.032	1.657

Appendix 4: Results – Two-Way ANOVA for Satisfaction Scale

Tests of Between-Subjects Effects

Dependent Variable: Satisfaction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2.844[a]	3	.948	1.464	.246	.136
Intercept	520.031	1	520.031	803.359	.000	.966
Participant_Gender	2.531	1	2.531	3.910	.058	.123
IA_Gender	.031	1	.031	.048	.828	.002
Participant_Gender * IA_Gender	.281	1	.281	.434	.515	.015
Error	18.125	28	.647			
Total	541.000	32				
Corrected Total	20.969	31				

a. R Squared = .136 (Adjusted R Squared = .043)