Exploring Trust in Satellite Navigation Voices

VICTOR-CRISTIAN ZUGRAVU, University of Twente, The Netherlands

The tone of a voice and its familiarity can greatly influence the level of attention that drivers pay to the voice of a navigation system. To determine what voice suits best a satellite navigation system, an interface that could read different navigation prompts with a voice configured by a user was created. Previous research has indicated what type of voice people prefer in such a context, but none of them allowed the user to create their own ideal voice. The present study addresses this question through a co-design experiment. 17 European nationals were asked to create the voice they would trust most in a navigation environment using the interface. Their voice selection and their responses from a short semi-structured interview were analyzed. The findings confirm some of the claims made in previous studies, such that people trust voices with a lower pitch more, but the expected bias towards female voices was not met. It was revealed that only pitch remained relatively constant between participants, and that accent preference relied heavily on what country the participant was from. Results indicate that to achieve good user satisfaction, navigation system makers need to provide an ample selection of voices, especially when it comes to their rate of speech.

Additional Key Words and Phrases: GPS, Navigation, Voice, Pitch, Accent, Gender, Rate of speech, Nationality

1 INTRODUCTION

GPS systems were a game-changer when they were first introduced, and since then navigation systems have proven themselves to be a necessity for most modern drivers, for the better or worse [18]. Over time, navigation systems have evolved from simple pre-recorded phrases to high-quality text-to-speech synthesizers that can read new street names and be offered in a more varied fashion. Text-to-speech has been a topic of discussion in the field of Human-Computer Interaction for more than 20 years[15], and since it has become a mainstay in an increasing amount of GPS systems and not only[11]. In less than 20 years, in-car applications of text-to-speech synthesis has gone from being used in specialized GPS systems as seen in figure 1 to voicing personal assistants that replace the need to take your eyes off the road while driving, to improve your safety[17]. This is due to the fact that advancements in voice synthetizing technology, voice has increasingly become the medium through which we interact with the digital environment.

But how do we determine the role personal preferences play in what voice suits the context of satellite navigation devices? In his book titled "The Man Who Lied to His Laptop" published in 2010, Clifford Nass, a communications professor at Stanford University, wrote that BMW recalled their first GPS systems introduced in the 90s after German drivers refused directions from a female voice. That happened when the technology was still in its infancy, 30 years ago, when gender norms were not what they are today.

Fig. 1. Tom Tom 910, one of the first in-car Navigation Systems to feature text-to-speech synthesis technology [4]

Since then, the outlook on the situation has changed, and most navigation systems and in car entertainment systems use a calm female voice[9] that came standard in the system, as often people don't realize or don't care that they can alter the voice in their systems. This raises the question, what is the best course of action when designing a voice to be used to give navigation instructions? One can look at the industry standard and declare it good enough, but people tend to stop paying attention to things they have already gotten used to. In a world where ever more powerful technology creates the opportunity for technology to cater to each individual's needs, we need to take a step back and see what the people actually want.

Thus, to understand the thought process of satellite navigation users and to see what steps need to be taken towards creating a better navigation experience for everyone, the following research questions arise:

- What are the characteristics of the standard navigation system voice that people are used to? Do they contribute to a preconceived idea of what a satellite navigation voice should sound like?
- What voice characteristics are the most relevant in the selection for the voice of such a system?

2 RELATED WORK

2.1 Voice is better than a screen

Since its inception in the 70s, GPS has evolved from military application to civilian adoption beginning in the 90s, with studies into the effect of audio guidance taking place since the 80s[21]. These studies focused on how different delivery of navigation instructions affected drivers' focus and ability to navigate efficiently. The results unanimously concluded that verbal delivery of navigation instructions is better suited for navigation compared to visual aids in the form of either maps or digital screens, as it reduces visual

TScIT 41, July 5, 2024, Enschede, The Netherlands

distractions for the driver and helps them make fewer navigational errors and respond quicker to commands [12; 21].

2.2 Pleasant voices are important

Another type of research in the field has been focused on putting participants in simulated driving scenarios and letting them choose between 2 different voices, to check which one of them they would trust in situations where the voice guidance may contradict their environment. The voices were easily distinguishable from each other and selection was relatively straightforward. One study found that people prefer the more charismatic voice of Steve Jobs compared to Mark Zuckerberg [16], and that such a charismatic voice inspires more trustworthiness and quality. A similar study found that people trusted more the standard TomTom navigation Unit voice than the custom voice of Snoop Dogg as they found it more trustworthy because [14]. To distinguish what voice suits their navigation needs, people assign human-like traits to any voice they are given and it was shown that strong positive correlations exist between ratings for the trustworthiness, assertiveness, and clarity of a voice and the likelihood of that voice being selected for use every day [13]. From these studies, it became clear that the prosody of the voice is very important, and that in order to help participants focus on the characteristics of the voice such as pitch and speed, they would first need to interact with a voice model that has a pleasant prosody and is easy to follow.

2.3 What can we learn from voice assistants?

To help understand the choices that people make for GPS systems, we can analyze the overall sentiment toward other text-to-speech applications, such as voice assistants. Voice assistants are one of the most widely used applications of text-to-speech systems, with 25% of users using them daily[23]. We can leverage the studies made to analyze the voices of voice assistants because, similarly to the context of navigation, people depend on these systems to provide them with reliable information.

An intricate thing one observes when using the most common voice assistants such as Google Assistant, Siri, Cortana, or Alexa, is that they come by default with female-sounding voices. This is a clear indication of implicit stereotypes in voice assistants[22] and can be linked to female voices being perceived as more benevolent [6] and other stereotypical associations of traditionally feminine qualities, such as kindness and nurturance[2]. There is no conclusive proof of whether gender has a clear connection to trust[22], but there are arguments that virtual assistants perceived as being male are considered to be more competent than virtual assistants perceived as being female[5]. Additionally, lower pitch leads to a more persuasive voice, irrespective of gender [25].

2.4 What to expect

From the literature observed, previous studies on the matter have either been conducted in a real or simulated driving environment, where people were asked to make decisions under realistic, stressful situations. Those studies had the advantage that the feedback was directly applicable to the real world, but on the other side, the participants had no possibility to express their exact needs. This line of thought applies to most of the studies presented, as they were based on perception, rather than on co-design.

To take advantage of the participants not needing to focus on driving, the plan is to conduct a study where all the control over what the navigation voice sounds like is put in the hands of the user, such that they can configure a GPS voice that perfectly suits their needs. Looking at the previous subsection, we should observe a preference for low-pitch female voices, irrespective of the gender of the participant[22; 25]. Additional research has shown that people tend to trust accents they are familiar with[1]. That will be rigorously assessed, since people of multiple nationalities will take part in this study.

3 METHODOLOGY

In order to investigate people's preference for navigation guidance voices, we will conduct a co-design experiment in which participants are tasked to configure the voice they trust most in the context of navigation. For this task, an interface was constructed based on the Google Text-To-Speech API [7], which allows the participant to listen to various GPS prompts and then change the Speed, Pitch, Accent, and Gender of the voice. After the participants will use this interface, they will be interviewed to understand their choices and what motivates them. For analysis, the parameters chosen by each participant will be tested to check if they are in line with previous studies and if they render any new insights, and the sentiment towards navigation system voices will be subsequently subtracted .

3.1 The interface

The layout is simple, as seen in Figure 2. On the right, there are the 4 parameters the user can modulate: speed, pitch, accent and gender. On the left, there are navigation prompts the participant can click on to listen to the navigation instructions in the voice they created.

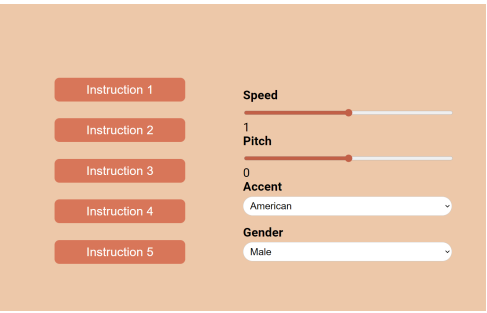

Fig. 2. The final design of the interface

The interface is designed to be as straightforward as possible, to help the participant concentrate on the task at hand. To aid to the participant's focus and make them more engaged, the color palette of the interface has red hues, as it is proven to make people more tense and focused [19].

Together with 2 colleagues who researched on the same track, we analyzed all the available text-to-speech APIs, and the most appropriate for this study was the Google Text-to-Speech API as it was used in previous research on trust of voices[22], and it has

voices with a high degree of configurability, with parameters such as Speed, Pitch, Accent, Gender(M/F), audio profiles and multiple voice types, as seen on Figure 3 .

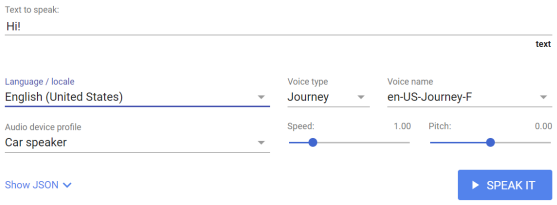

Fig. 3. All the configurable features of the Google text-to-speech API as presented on their website [7]

3.2 Preparing the voice parameters

3.2.1 Audio device profile. The Audio device profile parameter optimizes the playback of the generated voice for different types of hardware. The Car speaker profile was chosen as there was no designated profile for laptops, and this was the next best profile. Even though there is no specific research when it comes to audio profiles, the Car speaker profile seemed to output the most car-speaker-like audio out of all the profiles[3].

3.2.2 Accents. There are 4 English accents available in the API: Australian, British, American and Indian. All of them were selected for this project, as the aim was to offer as large of a selection as possible.

3.2.3 Pitch and speed. Google gives the option to modulate the speed of the voice between 0.25 and 4, but after testing out the system it was deduced that anything below 0.7 (equivalent to 135 wpm) and over 1.3 (260wpm) was unrealistic to be selected, as it would be too slow to be useful, or too fast to be understandable[24]. When it comes to pitch, the initial range of -20 to 20 was deemed too excessive toward the extremities because it distorted the voice too much, and it became difficult to adjust in small increments in the already designed layout. Thus, we settled on a range of -8 to 8 for the pitch. The speed is modifiable in increments of 0.05, the pitch in increments of 1, and their initial states are 1 and 0 respectively.

3.2.4 Voice type and name. Furthermore, considering that there will be 4 accents available for the participants, we had to select voice types that were available for all 4 accents. 3 voice types fit this requirement, Default, Neural2 and Wavenet. According to the Google website[3], WaveNet is their premium offering that is closest to bridging the gap between robot and human voices, Neural2 is based on technology used to create custom voices, which allows you to use a custom voice without needing to create one, and default mostly generates audio data by passing outputs through signal processing algorithms known as vocoders. Out of these 3 options, default sounds the most robotic and WaveNet the most human-like, but since WaveNet is already used in multiple Google applications[3], it was avoided so that participants would not choose a voice they easily recognize, to prevent bias. That is why the middle ground, Neural2 was chosen. Google does not specify genders for their voices, instead each voice name sounds like a particular gender. Below, in table1, the breakdown of voice names equivalent for each accent and gender present in the interface is shown, based on careful analysis of how each voice sounded.

Table 1. Voice Selection by Accent and Gender in Google Text-to-Speech API

Accent	Gender	Voice Name
American	Male	en-US-Neural2-D
	Female	en-US-Neural2-F
British	Male	en-GB-Neural2-B
	Female	en-GB-Neural2-C
Australian	Male	en-AU-Neural2-B
	Female	en-AU-Neural2-C
Indian	Male	en-IN-Neural2-B
	Female	en-IN-Neural2-A

To ensure that the voices selected were fit for the study, after the first 3 experiments concluded each participant was asked to rate the interface voice choices when compared with the other available voice names, and they all agreed that the voices chosen were the most representative for each accent.

3.2.5 Navigation Prompts. For the prompts used, a varied selection of scenarios was offered [Figure 4], to simulate multiple situations one might find themselves in while driving. The prompts are lengthier than usual navigation instructions, to give the participant enough time and context to assess each voice.

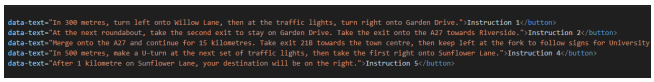

Fig. 4. The prompts for the interface

3.3 Participants

A total of 17 participants took part in this study. They were selected based on the requirement of having a good conversational level of English. To obtain a result that is more representative of the demographic that uses GPS systems, and to deter any biases in technology acceptance that may arise from age variance [20], the age range of participants is from 19 to 70. The majority of these participants (59%) were undergraduate students, and the other half of the participants had an average age of 49. The majority of the participants were male (82%). All participants were accustomed to Satellite Navigation, with the average use of this system being a few times per month. This was relevant, as not all participants had driving experience, although they made up a small share of the sample (12%). Each recorded interview took 6 minutes on average.

3.4 Measurements

3.4.1 Briefing questions. Some of the questions asked in the briefing, such as how often they drive, how often they use GPS navigation and what GPS navigation they use, are intended to help gauge if there are any differences in population preference. These preferences may arise based on the frequency that satellite navigation is used and which satellite navigation application is used, since they come with different navigation voices. The study also aims to measure the correlation between driving experience and perception of navigation system voices.

As there were no previous studies that considered these parameters, it is difficult to foresee if there will be a significant difference between participants based on these statistics. If there is a significant difference, it will be mentioned in the result section.

3.4.2 Quantitative results. Quantitative measurements are comprised of exact features of the voice each participant deemed most trustworthy, as well as the age and gender of the participant, and it is the first layer of understanding of the participants' choices.

3.4.3 Qualitative results. Qualitative results revolve around the 3 main questions that will be asked in the interview mentioned below in subsection 3.5, and the goal is to understand the overall sentiment toward navigation voices, what people want them to sound like, and whether already existing technology has predisposed them to expect and desire certain features, such as a feminine voice.

3.5 Procedure

Each participant will be briefed about the study when they agree to participate, and then they will sign the consent form. Following that, they will be asked to provide their age, gender, how often they drive, how often they use GPS navigation, and what GPS navigation they use. At this point, the audio recording will start.

Afterwards, the participants will be tasked to use the interface to create the navigation voice they would trust most. The participants will be assisted in using the interface by showing them how it works and explaining what output they should expect when manipulating the parameters. Additionally, the participants are encouraged to think aloud.

After the participants decide on a voice they like most, the characteristics of the voice are noted down in a database and we proceed with the interview.

The interview is semi-structured, and it focuses on 3 questions:

- Can you describe the voice you created, without necessarily using the parameters you manipulated?
- Which were the most important parameters in your selection?
- Do you have a preconceived idea of what a GPS voice should sound like? What are the characteristics of the standard navigation system voice that you are used to?

Each participant is encouraged to expand on their answers if possible, and to use natural language.

In the end, the participants are asked if they have any closing remarks, or suggestions for improvements. Finally, the recording stops.

4 RESULTS

4.1 Voice parameters

This section analyzes the voice parameters that participants chose for their preferred voice during the experiment. A strong preference for pitch can be observed, but not so much for speed.

4.1.1 Speed. The rate of speech range was between 0.85 and 1.2, which is equivalent to a range between 169 wpm and 230 wpm for the American accent, and approximately 10% more for the British accent. The lowest speed was equivalent to the speed of someone giving a lecture, the base (speed 1) was equivalent to a normal conversation [24], and the fastest speed was at the level of a sports commentator.

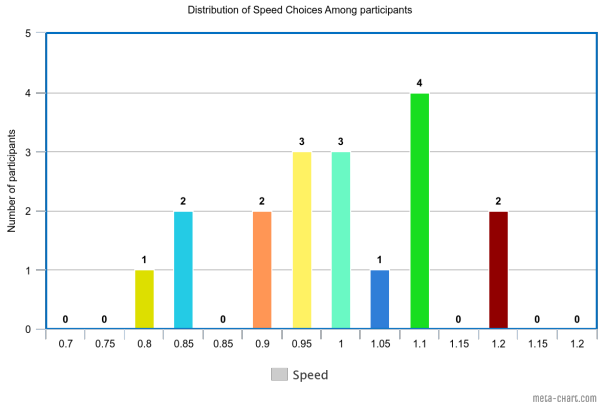

Fig. 5. The distribution of speed choices between participants

Looking at the chart above [5], we observe that 3 participants chose the base speed, 7 chose a slower rate of speech and 7 chose a faster rate of speech. The average speed was 1, although the most chosen speed was 1.1. From the following interviews, it could not be subtracted whether either choice had a stronger standing.

4.1.2 Accent. 11 Participants chose the American accent, 1 participant chose the Australian accent, and 4 participants chose the British accent. The data gains more significance when evaluating the participants' choices based on their nationality. Out of the 7 Dutch participants, 4 of them chose the British accent, the rest chose the American accent.

4.1.3 Gender. Out of the 17 participants, 47% chose a female voice, and 53% chose a male voice. There were too few female participants (18%) to be able to assess with certainty a difference in preference between the genders. The results were not expected, as previous studies indicate that drivers prefer female voices for guidance [9].

4.1.4 Pitch. There was a clear preference towards lower-pitch voices, as shown in Figure 6, with only one small exception between the participants, who generally considered lower-pitch voices as more trustworthy and calm.

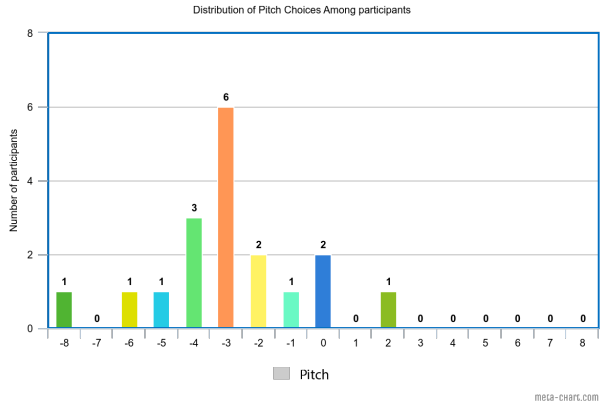

Fig. 6. Distribution of Pitch choices among participants

The consistency in the preference for lower pitch was in line with previous studies on the matter, where voices with lowered pitch were considered more assertive and trustworthy [10].

4.2 Interview data

In this section we will go though each of the main interview questions and see the data that was extracted from them.

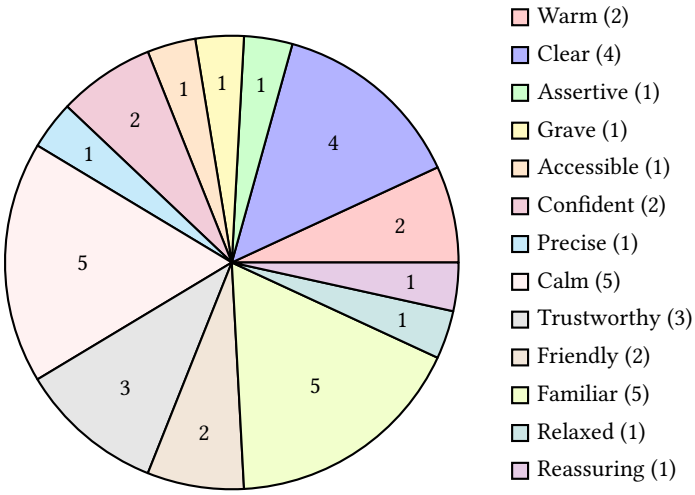

Fig. 7. How participants described the voices they created

Firstly, the participants were asked to describe the voice they created. In the pie chart above [7] all the adjectives mentioned by the participants, as well as the number of participants that used each adjective is displayed. Interesting findings regarding these responses will be presented in the Discussion section. Each participant used as many adjectives as they seemed fit, ranging from 1 to 3.

Next, when asked which characteristic of the voice was the most important in their choice, 2 participants could not decide between the parameters, both saying that for them it was important how calm the voice was. From the rest of the conversation, it could be subtracted that the factors that contributed to the calmness of the voice were the speed and pitch configuration. Since those choices were not conclusive, they were not included in the table showing the amount of participants who chose each parameter.

Characteristic	Number of Participants
Accent	6
Speed	5
Gender	3
Pitch	1

Table 2. Most Important Voice Characteristics for Participants

From the table above [2], accent emerges as the most important characteristic, followed closely by speed. Participants associated the accent mainly with the clarity, familiarity, friendliness, trustworthiness and accessibility of the voice, and speed was associated with clarity, precision and calmness. Some participants had strong feelings towards what gender the voice should be, either because they were used to one of them, or because they did not trust the opposite voice to what they chose. Finally, the pitch was also relevant, participants associating it with warmth, assertiveness, trustworthiness, precision, calmness and familiarity.

The last question asked, "Do you have a preconceived idea of what a SatNav should sound like? What are the characteristics of the standard navigation system voice that you are used to?" was met with a lot of noes (47%), partially because people do not pay much attention to how the voice sounds, or because they mute the GPS when using it. In fact, 35% of participants said they only use the visual aid when using a navigation system. As for the people who did have a voice they were used to, they said it was a female voice, usually of British accent, but that did not mean that they still preferred it.

4.3 Analysis

Looking at the numerical values, using the Shapiro-Wilk test, we can see that both speed and pitch are normally distributed (Speed: Statistic = 0.941, p-value = 0.331; Pitch: Statistic = 0.936, p-value = 0.282). To check whether there is a difference in voice preference between the group of under 40 years and the group of over 40 years, a one-sided ANOVA test was performed for speed and pitch. There were significant differences, but looking at the p-values (0.128544 for pitch and 0.569212 for speed) we can observe that they are far greater than 0.05, which means that especially because of the small sample size, these differences are far more likely to be due to variation than differences between the age groups.

This study should be replicated with a larger sample size to overcome the limitation of low observed power, since this limitation also affects the understanding of voice gender preference. Out of the 7

participants over the age of 40, 5 chose the female voice and 2 chose the male voice, whilst out of the 10 participants under the age of 40, 3 chose the female voice, whilst 7 chose the male voice. During the interviews most of the participants over the age of 40 attributed their choice to the fact that the female voice is the one that they are used to hearing in this context, which shows that driving experience and implicit navigation system usage experience do play a role in voice selection, especially as all participants over the age of 40 are daily drivers, and only 30% of the other participants drive daily.

Additionally, a correlation matrix [3] was constructed, to understand if there is a correlation between Age, Pitch and Speed. From the table, we can observe there is a moderate positive correlation between Age and Pitch, meaning that as age increases, the choice for pitch also increases. There is a negligible negative correlation between Age and Speed, and a slight negative correlation between Speed and Pitch, meaning that the faster a voice is, the more likely it is to have a lower pitch.

	Age	Pitch	Speed
Age	1.000000	0.407032	-0.098464
Pitch	0.407032	1.000000	-0.283281
Speed	-0.098464	-0.283281	1.000000

Table 3. Correlation Matrix

5 DISCUSSION

5.1 Limitations

For this study, due to the limited time frame and resources, most participants enrolled were from Romania (53%) and the Netherlands (35%). This was not intended, but only a result of the nationality of the researcher and the willingness of people to participate in the project on a short notice. In fact, the study had the larger scope to also analyze the role of ethnicity in the selection of the GPS voices, this is why the option to choose the Indian accent was available. The goal was to also interview international students from South-East Asia, to see whether they would choose what could be considered a more familiar accent, or if cultural influences played a role in them choosing another accent. If possible, a follow-up study will investigate into this matter.

5.2 Should a GPS voice be familiar?

An interesting finding from the interviews was that all the Eastern European participants in this study selected a voice with an American accent. Most said that it was due to it being the English accent that they most encounter in their day-to-day life, be it in media or other circumstances. This was expected [1], but then a few participants said that the American accent was the most unusual yet still understandable choice. Why would they choose it? Because an unusual voice made them pay more attention! To understand this sentiment better, we will look into the motivation of the best representative of this sentiment, the participant who chose the Australian voice.

First of all, they considered that the Australian accent is the most permissive when it comes to hiccups that may occur due to the voice being generated on the spot, and that it sounded friendly. Those are fair reasons, but then they said "For me, American, it's used everywhere, it's kind of very generic and it makes my ears kind of block it as a noise. So, if I would hear it when I drive I wouldn't really pay attention.". The reason that most chose the American accent, the familiarity, is the exact reason for which they didn't like it. A similar train of thought was presented by other participants, who didn't choose the British accent because, even though it was more familiar, it was too posh, arrogant, or heavy.

From these findings, we can observe a difference in mentality between the age groups, as younger people are more inclined to make a choice based on the stereotypes associated with their selection options [8; 26].

5.3 So what's the most important characteristic of a GPS voice?

It is not possible to crown any of the voice characteristics analyzed in this study as "the most important" as they all played an important role in the selection. From the interviews, it was obvious that first and foremost everyone wanted a clear voice. For a voice to be clear, it needs to have an easy-to-understand accent and accessible speed. These 2 parameters also turned out to be the 2 most relevant by participant rankings. Next, it is important to make the voice attractive. To do that, the participants lowered the pitch of the voice, which was an expected result in line with previous research in the field [25].

A surprising result was that most participants chose male voices, since previous studies have shown people to be accustomed to female voices in digital voice application contexts [22; 25], but it may be related to the fact that participants' gender was not evenly distributed, and that male voices seem more trustworthy to users [5]. In the planning stage of the study, deploying a gender ambiguous voice was considered, but after seeing that the Google API does not offer this option and that previous studies that implemented gender ambiguous voices using the Google API just offered lower pitch female voices or higher pitch male voices [22], the idea was dropped.

5.4 There is no voice that suits everyone

During the interviews, it was remarkable to see 2 parameters that remained consistent between all participants (pitch and accent) and 2 parameters that were impossible to predict before an interview (speed and gender). One of the participants, when talking about rate of speech, summed the general consensus pretty well: "You need information to flow fast when you drive, but at the same time, if they speak too fast, it doesn't really sound trustworthy." All participants expressed similar ideas, but everyone had a slightly different thought of what "speak too fast" means.

All in all, it is important to offer variety, such that everyone can find a voice that suits their needs and that they can be comfortable listening to, a voice that does not startle them, yet at the same time makes them pay attention.

6 CONCLUSION

In this study, it was discovered that speed, pitch, accent and gender all play important roles in how users perceive navigation instructions. There was no clear preferred configuration, but several points of interest were observed during the experiments. To obtain a clear voice, accent and rate of speech must be adjusted correlatively, as each accent has a slightly different rate of speech. People have different perceptions of what an acceptable speed is, varying between 169 wpm, the rate of speech of a teacher giving a lecture, and 230 wpm, the rate of speech of a sports commentator. Accent preference relies a lot on the users' origin, as participants born in The Netherlands had a clear preference for the British accent, whilst participants from Eastern Europe unanimously chose an American accent for their voice. These choices were made in accordance with the culture the participants were used to.

To make a voice more trustworthy, the overwhelming majority of the participants chose to lower the pitch, which was an expected result based on previous research [25]. The results of this study indicate that user age may be related to the preferred gender of the navigation system, due to the experience they have using them, and the standard voice that they come equipped with [22].

Some of the results in this study were inconclusive due to the limited amount of participants, such as the preference for different navigation voice genders between different age groups, or if a users' gender plays a role in their voice choice.

Further studies need to be conducted into how one can incorporate the findings of this study into already existing technology, and if it will make a real difference in user acceptance.

ACKNOWLEDGMENTS

I would like to thank my supervisors, Khiet Truong and Garcia Goo for their continued support, and I would also like to thank my colleagues Soban and Jump who are researching the same track, for the help received in building the interface.

REFERENCES

[1] BJW Beckers. The effects of text-to-speech gender and accent on the trustworthiness and listening experience of informational news content. 2021.
[2] Sandra Lipsitz Bem. Gender schema theory: A cognitive account of sex typing. *Psychological review*, 88(4):354, 1981.
[3] Google Cloud. Cloud text-to-speech: Convert text into natural-sounding speech. https://cloud.google.com/text-to-speech/docs/voice-types?_gl=1*1be3yse*_up*MQ..&gclid=CjwKCAjwm_SzBhAsEiwAXE2Cv41BZXW1UUvZXN0MtWf1c3ig6G-FYGbvFRjk7UDeapbBKU6n60V0nxoCkfcQAvD_BwE&gclsrc=aw.ds. Accessed: 2024-06-26.
[4] CNET. Tomtom go 910 gps receiver series review. https://www.cnet.com/roadshow/reviews/tomtom-go-910-gps-receiver-series-review/.
[5] Claus-Peter Ernst and Nils Herm-Stapelberg. Gender stereotyping's influence on the perceived competence of siri and co. 2020.
[6] Kylie L. Goodman and Christopher B. Mayhorn. It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics*, 106:103864, 2023.
[7] Google. Google cloud text-to-speech. https://cloud.google.com/text-to-speech/.
[8] Jonathan Haidt. Why the past 10 years of american life have been uniquely stupid. *The Atlantic*, 11:2022, 2022.
[9] Wang Ji, Ruijie Liu, and SeungHee Lee. Do drivers prefer female voice for guidance? an interaction design about information type and speaker gender for autonomous driving car. In Heidi Krömker, editor, *HCI in Mobility, Transport, and Automotive Systems*, pages 208–224, Cham, 2019. Springer International Publishing.
[10] Benedict C. Jones, David R. Feinberg, Lisa M. DeBruine, Anthony C. Little, and Jovana Vukovic. A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, 79(1):57–62, 2010.
[11] Ashlesha Vishnu Kadam. Text-to-speech in voice assistants: Challenges and mitigation strategies. *Journal of Engineering and Applied Sciences Technology. SRC/JEAST-272. DOI: doi. org/10.47363/JEAST/2023 (5)*, 183:2–4, 2023.
[12] Terry C. Lansdown. Visual demand and the introduction of advanced driver information systems into road vehicles. 1 1996.
[13] David R Large and Gary E Burnett. Drivers' preferences and emotional responses to satellite navigation voices. *International journal of vehicle noise and vibration*, 9(1-2):28–46, 2013.
[14] David R. Large and Gary E. Burnett. The effect of different navigation voices on trust and attention while using in-vehicle navigation systems. *Journal of Safety Research*, 49:69.e1–75, 2014. Proceedings of the International Conference on Road Safety (RSS2013).
[15] Lehlohonolo Mohasi and Daniel Mashao. Text-to-speech technology in human-computer interaction. In *5th conference on human computer interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI)*, pages 79–84, 2006.
[16] Oliver Niebuhr and Jan Michalsky. Computer-generated speaker charisma and its effects on human actions in a car-navigation system experiment-or how steve jobs' tone of voice can take you anywhere. In *Computational Science and Its Applications–ICCSA 2019: 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part II 19*, pages 375–390. Springer, 2019.
[17] R Ramnath, N Kinnear, S Chowdhury, and T Hyatt. Interacting with android auto and apple carplay when driving: The effect on driver performance. *IAM RoadSmart Published Project Report PPR948*, 2020.
[18] Jeff Robbins. Gps navigation…but what is it doing to us? In *2010 IEEE International Symposium on Technology and Society*, pages 309–318, 2010.
[19] Reza Shahidi, Rostam Golmohammadi, Mohammad Babamiri, Javad Faradmal, and Mohsen Aliabadi. Effect of warm/cool white lights on visual perception and mood in warm/cool color environments. *EXCLI J.*, 20:1379–1393, August 2021.
[20] Melanie K Smith and Monica Fabiola Coronel Padilla. Technology acceptance and readiness among different generations of urban touists. *Marketing & Menedzsment*, 57(1):37–48, 2023.
[21] Lynn A. Streeter, Diane Vitello, and Susan A. Wonsiewicz. How to tell people where to go: comparing navigational aids. *International Journal of Man-Machine Studies*, 22(5):549–562, 1985.
[22] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. Female by default? – exploring the effect of voice assistant gender and pitch on trait and trust attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery.
[23] Amrita S. Tulshan and Sudhir Namdeorao Dhage. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In Sabu M. Thampi, Oge Marques, Sri Krishnan, Kuan-Ching Li, Domenico Ciuonzo, and Maheshkumar H. Kolekar, editors, *Advances in Signal Processing and Intelligent Recognition Systems*, pages 190–201, Singapore, 2019. Springer Singapore.
[24] Li Wang. British english-speaking speed 2020. *Acad. J. Humanit. Soc. Sci*, 4:93–100, 2021.
[25] Tzu-Yang Wang, Ikkaku Kawaguchi, Hideaki Kuzuoka, and Mai Otsuki. Effect of manipulated amplitude and frequency of human voice on dominance and persuasiveness in audio conferences. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018.
[26] Halyna Yefymchuk. The conception of arrogance in british picture of the world. *Language, Culture, Politics. International Journal*, 1(1):177–186, 2016.

A USE OF AI TOOLS

During the preparation of this work the author used ChatGPT and Grammarly in order to speed up the process of getting accustomed to formatting a scientific paper in Latex (ChatGPT), correct simple grammatical mistakes (Grammarly), generate useful graphics for this paper, such as the pie chart and tables (ChatGPT), and to generate code to compute statistics (ChatGPT) . After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

B PARTICIPANT DATA

Pitch	Speed	Accent	Gender	Age	P Gender	Drives	Uses GPS	GPS App	Nationality
-3	1.05	A	F	47	F	daily	1/wk	Waze	RO
-3	0.9	A	M	45	M	daily	1/mo	Waze	RO
0	1	A	F	49	M	daily	3-4/yr	Waze	RO
-3	0.9	A	F	45	F	daily	1-2/mo	Waze	RO
0	1.1	A	F	45	M	daily	1/wk	Waze/ Maps	RO
-3	1.2	GB	M	21	M	2/mo	1/mo	Maps	NL
-8	1.1	A	M	19	F	never	3/wk	Maps	RO
-4	0.85	AU	M	21	M	1/yr	2-3/wk	Maps	CYP
-4	1.2	A	M	23	M	daily	1/mo	Maps	RO
-5	1.1	A	F	21	M	4/yr	2/wk	Maps	RO
2	0.85	A	F	20	M	1/mo	1/2mo	Apple Maps	RO
-1	1	GB	F	70	M	daily	daily	Maps	NL
-2	1.1	GB	M	24	M	daily	1/wk	Maps	NL
-2	1	A	M	43	M	daily	daily	Maps	NL
-3	0.95	A	M	20	M	1/mo	1/mo	Maps	NL
-4	0.95	GB	M	21	M	daily	daily	Maps/ Flitsmeister	NL
-3	0.95	A	F	20	M	never	5/mo	Apple Maps/ Maps	HU

Table 4. Quantitative data collected from the study. P Gender stands for Participant gender. The first 4 columns are the parameters chosen by each participant