Research on Personality Traits of In-Vehicle Intelligent Voice Assistants to Enhance Driving Experience

Xinyi Zhou and Yangshuo Zheng[✉]

Wuhan University of Technology, Wuhan 430000, China
zhengyangshuo@163.com

Abstract. Despite the rapidly expanding research on in-car intelligent voice assistants, relatively little attention has been paid to the study of their personality traits and how to improve the user's emotional experience by cutting through from this perspective. In this study, first, an exploratory interview-based study is conducted to identify key traits that elicit driver emotions and five personality traits that can be used in in-car intelligent voice assistants: emotional enrichment, emotional expressiveness, persona congruence, empathic expressiveness, and user impressions. Second, an online survey-based study is conducted to investigate how the above five performance dimensions that focus on the degree of personification of intelligent assistants can be incorporated into smart car voice interaction systems in different scenarios. Finally, the ideal use cases for each personality trait are investigated, as well as how in-car intelligent voice assistants can apply the corresponding trait dimensions according to the current real-time driving situation. This study explores how the five personality trait dimensions can enhance driving users' emotional experience and emotional connection with intelligent voice assistants by adding a pleasant attitude, a sense of pleasure, and can accurately grasp users' emotional state and provide proactive services conducive to emotional appeasement and safe driving in navigation.

Keywords: In-car intelligent voice assistant · Personality traits · Driving emotional experience

1 Car Intelligent Voice Assistant Concept

Voice is the most commonly used interaction method in people's daily life, and with the development of intelligent AI technology, it is also gradually applied in car products. Voice assistants help to free users' hands and eyes during the driving process, thus enhancing driving safety and enhancing the emotional experience of driving users. People increasingly rely on maps to navigate, check road conditions and find nearby points of interest in their daily travels, while users use their hands to control or view the screen while driving, which poses a great safety risk. The emergence of in-car intelligent voice assistant greatly liberates the user's hands and eyes during the driving process and improves driving safety. It not only supports the whole process of voice wake up, but also can quickly and accurately understand user commands and propose effective solutions.

1.1 Voice Assistant Definition

A voice assistant is a human-machine dialogue program embedded in a hardware device or APP software to assist users to use the functions on the host device or program by voice. A complete human-machine dialogue includes the front-end processing of sound signals, the conversion of sound into text for machine processing, and the conversion of textual language into sound waves using speech synthesis technology after the machine generates the language, thus forming a complete human-machine voice interaction.

1.2 Development Trend of In-Car Intelligent Voice Assistant

Smart cars are vehicles that can use ICT (information and communication technology) and artificial intelligence to interact with users autonomously. With innovations in the areas of electric vehicles, mobile connectivity and autonomous driving, smart cars have captured the attention of many interaction designers. The importance of improving the emotional experience of smart cars is reflected in its ability to improve driving pleasure, efficiency and safety.

With the development of artificial intelligence and voice recognition technology, intelligent voice is becoming the main vehicle interaction tool, in the car HMI, the car voice assistant provides auxiliary functions for the driver. As the third living space, i.e. another living space after home and work, intelligent voice assistant is the core content, which helps users to control the whole intelligent cabin and provide services to users through intelligent voice assistant. The development of AI technology and hardware upgrade has given intelligent devices more and more perception channels, and also continuously enhanced the ability to output information through various media. Based on research results in the fields of intelligent remote technology speech, natural language and machine translation, in-car voice assistant technology has been greatly improved in fluency, quality, fidelity and naturalness. Beyond the perception and output layer, the improvement of emotional computing capability has enabled the machine to make a qualitative leap in the cognitive layer.

With the development of technology and hardware upgrades, smart devices are given more and more perception channels, and also continue to enhance the ability to output information through various media. Beyond the perception and output layer, the improvement of emotional computing capability enables machines to make a qualitative leap in the cognitive layer. They understand users far better than before, and are far more articulate than before. Based on design understanding and practice, it is believed that the voice assistant experience is showing three trends in terms of interaction channels and interaction objects: the expression of information services incorporating multi-channel experience, dialogue close to natural human instincts, and the ability to interact with emotions.

Expression of Information Services Converge Multi-Channel Experience. In addition to voice channels, computer technology has expanded interaction channels such as face and spaced gestures, while traditional interaction methods such as touch and knob have their own advantages in terms of operation accuracy, information output efficiency and technical cost. Multi-channel fusion can bring into play the advantages and scenario

applicability of different channels to express information services more naturally and maximize efficiency.

The navigation scenario is the most commonly used driving scenario. In the navigation scenario, based on the map speech of touch screen mobile devices, the collaborative output of visual information can effectively compensate for the defects such as invisible and not easily memorized by voice, and improve the user's understanding of the voice interaction experience. The first time the voice interaction capability was introduced, the interaction form of map voice followed the industry's common closed dialogue flow in order to reduce cognitive costs. This form also has limitations with the expansion of voice supported map functions: (1) The form is independent and closed and does not integrate with the scene, which will interfere with navigation and affect driving safety. (2) The original information expression of the map cannot be used, and the results need to be presented separately in the conversation stream. This not only affects the expansion of complex requirements such as road calculation, but also increases the maintenance cost of design and development.

Dialogue is Close to Natural Human Instincts. Continuous dialogue and can be interrupted at any time, in line with the instinctive awareness of daily communication, but most of the current products voice interaction is still not natural enough: to initiate a dialogue, you need to wake up first and then issue commands in a quiet environment, and the main focus is on "One question and one answer". With the emergence of full-duplex wake-up-free voice technology, the user's instructions can be predicted and judged by contextual information, eliminating the intermediate wake-up process and achieving more natural and smooth multi-talk. Before achieving natural dialogue, effective cognitive education is a necessary way to lower the threshold for users to use speech. due to the "invisibility" of voice information, there is the defect of weak discoverability of skills, which makes users often ignore the use. At the same time, the basic research of voice map found that the primary reason for not using voice is that users are not used to the operation method.

In the user's mental model, the concept of voice as a tool determines that users will use it only when they have the intention, which inevitably affects the frequency and regularity of practice required for learning and habit formation. Therefore, Map Voice Skill Center proposes the concept of Xiaodu growth, completing daily tasks and using rewards to help Xiaodu grow, so that users can learn skills and develop habits quickly. The tasks are mainly organized in personalized, hot and level dimensions, for example, in personalization, priority is given to displaying user error-prone instruction tasks and solving the problem of instruction expression through repeated learning.

Ability to Interact Emotionally. Language is a symbol of human intelligence, and users will have an "Empathy" effect on voice products. emotion computing enables products to process relevant data such as human facial expressions, body movements and various psychological parameters such as heartbeat, pulse, brain waves, etc. Through machine learning algorithms such as emotion analysis, and finally calculate the emotional state of a person by combining external environmental information. Then give three-dimensional emotional feedback from the hardware level, GUI and VUI level to realize emotional interaction.

Car intelligent voice assistant active service currently includes three main scenarios, such as safety advice, road conditions and destination services. For example, when the user drives for a long time at high speed, it provides the nearest rest service area; when the road is crowded, it suggests a suitable route; and when the destination is near, it recommends a convenient parking spot. We provide valuable proactive services at key touchpoints during the driving process to improve driving safety and establish a trustworthy emotional connection.

With more sensors to obtain human body-related data in the future, it can accurately grasp the user's emotional state and provide active services in navigation that are conducive to emotional calming and safe driving.

2   Design Research

2.1   Exploratory Study

Objectives and Methods. An exploratory study was conducted to identify personality traits of voice assistants applicable to smart car interactions, with the aim of identifying key traits that elicit driver emotions and personality traits that can be used in in-car intelligent voice assistants.

Sample. The study was conducted on 10 drivers, 7 males and 3 females, aged 23 to 55 years old, who had experience with in-vehicle voice-activated products. all had experience with frequent use of in-vehicle intelligent voice interaction products within three months to one year.

Procedures. The study involved an interview process consisting of two parts. first, contextual interviews were conducted in the participants' cars to allow users to generate design ideas. the mobile app voice search for place names and voice navigation functions were tested. the steps for testing the voice search place name function included: first step: voice input the place name; second step: select the destination based on voice prompts; and third step: start navigation. the voice navigation function test includes voice announcement: route location, full road length, estimated time required, real-time road conditions, monitoring reminders, service station reminders, etc. next, a structured interview was conducted and the scenario should be needs-based. participants were asked to think in advance about what they imagined in-car intelligent voice interaction would be like. seven driving scenarios were included: (1) entering the vehicle (2) waiting for a traffic signal (3) receiving guided navigation (4) refueling the car (5) dozing off (6) in-car entertainment (7) parking. questions in the contextual interviews included, "how would you communicate with the in-car intelligent voice assistant in each situation?" and "how would you like the in-car intelligent voice assistant to respond anthropomorphically?" etc.

Results. For in-car voice interactive navigation, the needs of participants are as follows: (1) basic needs: in-car intelligent voice assistant can meet the navigation function (2) desired needs: in-car intelligent voice assistant can understand the driver's needs and give timely feedback (3) excited needs: in-car intelligent voice assistant can have personality traits and communicate with the driver like a friend or family member without hindrance.

The structured interviews for in-car intelligent voice interaction resulted in the following conclusions: key traits that elicit driver emotions and five personality traits that can be used for in-car intelligent voice assistants: emotional enrichment, emotional expressiveness, persona consistency, empathic expressiveness, and user impressions.

With the advent of natural language interaction, the first human can use the tool according to his or her customary needs, and it is important to create an appropriate "personification" of the intelligent voice product during conversational interaction. These results provide initial insights into how to use in-car intelligent voice assistants for automotive interactions, which may be designed to bring more emotional experiences to users.

2.2   Online Survey

Objectives and Methods. Second, an online survey-based study was conducted to investigate how the above five performance dimensions that focus on the degree of personality of the intelligent assistant can be incorporated into smart car voice interaction systems in different scenarios. This resulted in the desire to elucidate the most desirable use of each personality trait in smart car voice assistants and how each personality trait can be applied to different instances. An online survey was therefore conducted to investigate whether different in-car intelligent voice assistants are appropriate for application scenarios in specific driving situations, and if so, why these scenarios are preferable.

Survey Design. The survey was designed to evaluate five application scenarios using two independent variables (different driving situations, different in-car intelligent voice assistant personality traits). By using fictional scenarios to familiarize users with the new situations in the survey, selective representations of real user scenarios can help unravel the complexities and conflicts that exist in the real world.

By constructing a short narrative covering all aspects of car use, including task-oriented situations (e.g., refueling, listening to music), specific contexts (e.g., first meeting, return), or covering a period of use (e.g., adaptation), it is easier to imagine being in these scenarios. Questions include "Does the assistant need to empathize with the driver's negative emotions when the vehicle is in an undesirable driving situation?", "How should the assistant respond when a human presents a need that cannot be met by the current intelligent assistant?" and "How do you want the assistant to successfully identify the user's emotional state?" etc.

Five representative driving situations were selected based on previous studies: (1) adapting to the car (2) listening to music (3) refueling (4) parking the car (5) returning to the car. In selecting these situations, both realistic and common situations (e.g., parking) were considered, rather than rarely encountered situations (e.g., dropping the car).

By creating different usage scenarios, these scenarios were specifically explained and expressed in a familiar representation so that participants could easily understand and immerse themselves in them. Testers used scales to indicate their preferences for personality traits of in-car intelligent voice assistants in different application scenarios.

Respondents were also asked to assess the usefulness of each scenario. These scores were used to understand why scenario-specific personality traits were preferred.

Results. The test results showed that the most popular personality trait was empathic expressiveness, while the next most preferred was emotional expressiveness and emotional enrichment. This was closely followed by Personality Consistency. The user impression personality traits of the in-car intelligent voice assistant were the least important.

3 How In-Car intelligent Voice Assistant Personality Traits Enhance the Driving Experience

3.1 Emotional Enrichment

Joy and sorrow examine the emotional richness of intelligent assistants, most of the voice assistants nowadays are a tool-based product, on top of which personality is attached. Take Gaode voice navigation as an example, the experience of using this product is undoubtedly done pleasantly, and the interaction with them is full of fun: "There is a sharp turn ahead, do you understand? It's a sharp turn, probably the sharpest sharp turn in the Eastern Hemisphere". By using the voices of celebrities and internet celebrities or using jingles that people are familiar with and orchestrated for the current driving scenario, the driving fun is enhanced and the driver's emotional experience is enriched.

But in the actual business, it is difficult for us to make a similar design when users are interacting with the assistant. The reason is as follows: users have already managed their expectations for the product when they choose the voice package of Gaode Map. And if it is the first time to use the intelligent assistant users do not have an expectation management. Some times using some jingle or slang, users who don't understand it will have an inexplicable feeling, and sometimes even because of regional and cultural differences may produce offense to certain kinds of groups. Therefore it is safest to try to choose a design that faces the general public.

Designed for the masses, meaning professional, only positive emotions can generally exist, with little negative feedback, and often only the joy of a task being completed. "If the Internet is stuck, does the in-car intelligent voice assistant assistant need to feel the anger in sync with the user?" "If the user makes an unreasonable request, what level of flirtation needs to be used?" These are all questions that will exist in the design. Generally speaking, professional assistants do not have negative emotions and do not flirt freely. Once you choose to design for the masses, often the design is more limited and the safe thing to do is to keep only positive emotions. Moderate is safe, but it seems to have less human light. And Gao De voice navigation processing solution is when the user is familiar with, let the user make a choice, their own management of their own expectations, perhaps a solution.

3.2 Emotional Enrichment

Emotional expressiveness here refers to the expressiveness and infectiousness of the voice assistant in expressing emotions. Assuming that the direction of emotion is joy,

anger, sorrow, sadness, fear and shock, how to express and what intensity is appropriate? The computer expressions that can be listed are: text, expressions, voice, sound effects, images, light effects, and even the robot's body movements. The more of these ways are overlaid, the richer their expressiveness. It is important to be able to demonstrate the emotional expressiveness that you can show in order to impress the user on key occasions.

3.3   Personality Consistency

Once the persona of the in-vehicle intelligent voice assistant is defined, its behavior habits, timbre, speech speed, and language content expression should be guaranteed to conform to the consistency. Different people have independent personalities as individuals with very distinct personas. The performance when facing problems, values, language expressions, logical sequence, and the position of interest they are in must be based on the previous persona performance. Although each person is a complex individual, but by and large also fluctuate within a certain range of values, so that it is a reasonable character setting, will not give rise to a sense of dismay. The logic of the intelligent voice assistant, timbre and speech rate towards is based on the same voice model setting, is easier to ensure consistency, and the difficulty is in the presentation of the language content level. "I'm sorry! I can do something else for you" "Hey, I can't do this yet, but I will try to learn to serve my master soon" This is the words expressed by different personalities, if you frequently switch between different language styles, it is easy to have The persona is inconsistent and fails in shaping the personality of the assistant's character.

3.4   Empathic Expressiveness

Emotional intelligence and empathy are high-level competencies - responding to the user's descriptions and corresponding content. "Empathy is a psychological phenomenon in which people actively project their true feelings onto the things they see. "Empathy" is a kind of feeling, standing in the perspective of others to think about the problem, is one of the necessary communication skills. The act of empathy requires feeling and observing first, and then responding. How to successfully identify the user's emotional state? From the machine level, various components are responsible for collecting and various technologies are responsible for analyzing. For example, visual recognition analysis, audio track analysis, text understanding, and even brainwave signal acquisition technology can do to analyze emotions and the corresponding degree. When the user is happy, sad or anxious, how should the voice assistant do empathic feedback?

Users rarely have mood swings, and when they have violent mood swings (ecstasy, anger, sadness), if the assistant can show some empathy and resonate with them with the same frequency of emotions, it can enhance the user's emotional experience. And empathic performance, invariably considers the four dimensions of the previously mentioned capabilities.

3.5 User Impressions

The product is able to manage the psychological expectations of the user and successfully shape an image. In other words, what kind of brand impression the assistant has in the user's mind. In the past, branding required a lot of effort from various departments such as product, operation, marketing, business, brand, and channel to expose and maintain. Nowadays, conversational interaction has more opportunities and personalization is easier. At present, the most shipped voice products on the market is the smart speaker, the user interaction with these smart speakers compared to traditional hardware products has changed fundamentally, because the anthropomorphic interaction form of voice dialogue, it is easier to attach personality, and then to convey the brand impression. In the actual use of the process, occasionally because of some playful words of the speaker to make themselves laugh out, this use of the process to generate positive emotions, the formation of pleasant memories, and then promote the user's willingness to use, to enhance tolerance and trust.

But in many smart speaker products, the performance of personality traits shaped by much the same, most speakers are still functional, business, service state, once experienced it is difficult to leave a deep impression on the user. And the design of the product, if the user experience a period of time, but also can not leave any impression, it is a great failure.

4 Research Conclusion

The purpose of this paper is to improve the in-vehicle voice interaction experience, shape a voice image personality that is more emotional and more in line with user expectations, highlight the humanization and intelligence of the voice assistant, improve the intelligence of the in-vehicle assistant, and assist in optimizing the user driving experience.

After describing what the definition of an in-car intelligent voice assistant is, I hope to think further about, "Emotions need to be rich, how does the assistant handle & apply negative emotions?" "Strong emotional expression is needed, how to grasp the proportion of the assistant's emotions?" "How to do the assistant's persona selection and how to ensure consistency in feedback?" "The empathic performance of the assistant, how to identify as well as feedback?" "How does the product impress the user and shape the brand?". All five personality traits of the in-vehicle intelligent voice assistant are interrelated and yet exist independently. It is easy to make requests, but difficult to have a methodology to apply to specific examples. For example, when the alarm clock wakes up the user, the content could be a cyclic alarm or a repeater voice announcement, or a variety of flirtatious ways to stimulate the user to get up.

Benefiting from the further evolution of voice capabilities, information and services continue to flow around the user rather than the medium. People's demand for natural, emotional and personalized is more prominent than in any previous era, and the voice experience will be more real-time and versatile. However, with the advent of the smart driving era, in-vehicle voice products are developing rapidly, but research on issues related to in-vehicle voice interaction design is still relatively small. In addition to problems such as poor environmental network, improper user operation and environmental

noise, existing voice assistants have serious homogenization, rigid language and lack of personality in auditory experience, which cannot bring good emotional experience for users. At present, the majority of products voice interaction is still not natural enough: waking up in a relatively single way, you need to wake up the voice assistant in a quiet environment by voice call and then issue commands. The voice assistant lacks the ability of multi-round interaction, and it is difficult to maintain the state of dialogue with the user after a wakeup, and the dialogue effect is not natural enough. Understanding the user's intention is basically only through the voice, but not through the multimodal information acquisition to identify the user's intention. Although the voice assistant has evolved from "one question and one answer" to "continuous question and answer" and "emotional dialogue", the "emotional dialogue" relies on The "emotional dialogue" relies on keyword recognition technology, and once it is separated from certain specific keywords, it will return to the state of indifferent machine. In-car voice assistant is difficult to understand human emotion and situation, give caring and compassionate response, can not provide immediate emotional support and long-term emotional companionship. Due to the "invisible" nature of voice information, there is a weakness of skill discoverability, resulting in users often ignoring the use. Users often have high expectations of voice interaction, and people will unconsciously compare voice assistants with real people when communicating with language, trying to understand voice assistants with the thinking habits of the human brain, which will inevitably lead to many times users feel that the results of human-computer dialogue do not meet expectations.

This paper focuses on the exploratory research of analyzing the personality traits of in-car voice assistants to enhance user driving experience, and the subsequent step-by-step advancement is needed to improve the interaction fluency and intention understanding degree of in-car voice assistants intelligence.

References

1. Braun, M., et al.: Improving driver emotions with affective strategies. Multimodal Technol. Interact. **3**(1), 21 (2019). https://doi.org/10.3390/mti3010021
2. Gruber, D., Aune, A., Koutstaal, W.: Can semi-anthropomorphism influence trust and compliance? In: Proceedings of the Conference on Technology, Mind, and Society, pp. 1–6. ACM, New York (2018). https://doi.org/10.1145/3183654.3183700