ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

Emotional Experience Design Strategy for In-Vehicle Intelligent Voice Assistant

Liu Xue[1,a,*], Liao Zicen[1,b]

[1]College of Architecture & ART, Hefei University of Technology, Hefei, China
[a]420913841qq.com, [b]liaozicen666@gmail.com
[*]Corresponding author

Abstract: *In the context of the rapid development of the intelligent automotive industry, the continuous exploration of new technologies and experiences in automotive intelligent cockpit products, and the increasing expectations of users for automotive intelligence, this paper analyzes the user experience level of in-vehicle intelligent voice assistants, and constructs a product experience model for in-vehicle voice assistants from the basic definition of the product to the design of fine-grained functional points. Based on the interactive characteristics of intelligent speech, combined with emotional design theory, this paper analyzes the emotional design of in vehicle intelligent speech assistants and proposes design strategies at different levels. Based on the technical link of the in-vehicle intelligent voice assistant, the design content of the in-vehicle voice assistant in the user experience design level was summarized; Based on the theory of emotional design, three levels of emotional design content for voice assistants were constructed. Based on the definition of emotional design levels for car voice assistants, targeted design strategies were proposed. Conclusion: Starting from the technology and interaction characteristics of in-vehicle voice interaction, the emotional design of in-vehicle voice interaction has been studied, which has reference significance for systematically improving the emotional experience of in-vehicle voice interaction products.*

Keywords: *voice interaction; Experience design elements; Emotional design; In vehicle voice assistant*

1. Introduction

In the context of automotive cockpits, human-computer interaction (HCI) has been evolving with the advancement of automation, computer technology, and artificial intelligence. The forms and content of interaction are continually changing.

Traditional automotive cockpits used physical hard buttons for interaction. The advantage of hard button interaction lies in its stronger user perception and clearer mapping between button position, operation, and function execution. However, as the number of functions increases, the complexity of the hardware interface also increases, raising the learning threshold for various interactions[1].

Graphical user interfaces (GUIs) are currently the mainstream form of interaction in intelligent automotive cockpits. Their advantages include high integration of interactions, convenient operation, and lower learning costs compared to hard buttons (as their operation logic is similar to that of smartphones and other smart mobile devices). GUIs can support more complex functions and interaction tasks, provide rich visual feedback. GUIs interaction in design focuses on aspects such as 2D interface and 3D rendering visual representation, information architecture, and interaction paths. However, their disadvantages include concentration of functions and information on the screen, increased complexity of interaction tasks with the increase in the number of functions and integration, and a heavier reliance on visual feedback, which increases cognitive load, posing safety risks in driving scenarios[2].

Natural interface interaction with multimodal and multi-device coordination is an important exploration direction for current intelligent automotive cockpits. This includes interaction forms such as voice interaction, gestures, virtual reality, and augmented reality, utilizing natural interaction methods closer to the real world with multi-channel information, allowing users to complete interactions in a more immersive, easier to learn, and easier to access manner. However, multimodal and multi-device interaction also brings new problems and challenges, including mistrust and resistance due to lack of user mental models, increased cognitive load due to redundant information transmission in multi-device channels, and information interaction management in complex scenarios with multiple users, devices,

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

and tasks, all of which impose more requirements on product designers.

Intelligent voice assistants are an important product form in the direction of intelligent natural interface interaction for intelligent cockpits. Voice interaction, as a natural interaction mode between humans, has advantages such as meeting users' subconscious natural interaction needs, lower cognitive load on the auditory channel during interaction, reducing interference with driving tasks compared to GUI interaction, shortening interaction paths, supporting multi-person interaction scenarios, providing more emotional feedback through language, expressing unstructured information, satisfying users' deep emotional needs, and establishing deep connections with products. However, the disadvantages of natural language interaction include the inability to guarantee 100% accuracy in speech and semantic recognition, errors in dialects, colloquial expressions, and continuous multi-turn dialogues, inability to ensure privacy as voice interactions in the cockpit are exposed to all passengers, limited structured information conveyed by voice interaction, low information density, and lower efficiency in executing complex interaction tasks. As an unrestricted interaction mode, its infinite expression space leads to a large amount of user dialogue that cannot be accurately identified and executed.

Therefore, for the unique product interaction form and features of intelligent voice assistants, it is necessary to propose interaction experience design strategies for intelligent voice assistants. Based on the interaction and technical features of intelligent voice interaction and the theory of emotional design analysis for in-vehicle voice assistants, the author proposes design strategies for emotional design at different levels of in-vehicle voice assistants.

Figure 1, from left to right, shows the general process of users using voice interaction, touch screen interface interaction, and hard button interaction to adjust the air conditioning temperature.

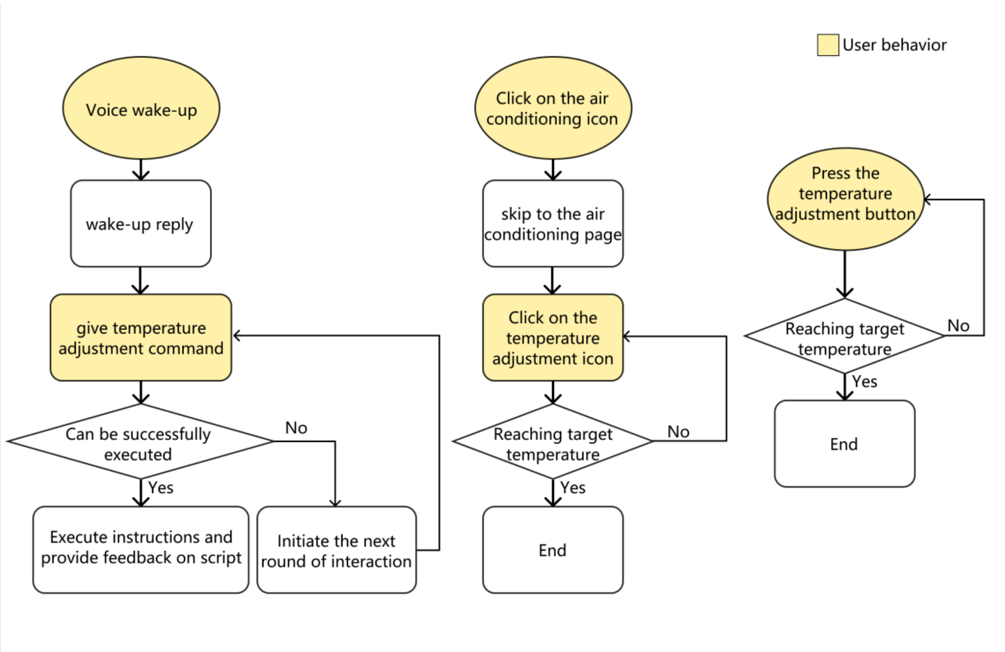

Figure 1: Typical User Interaction Processes under Different Interaction Channels - Taking Adjusting Car Air Conditioning as an Example.

2. In-Vehicle Voice Interaction: Technology Pipeline and Design Features

2.1. In-Vehicle Voice Interaction Technology Pipeline

The complete technology pipeline of automotive voice interaction can be roughly divided into three stages: voice input, voice processing, and voice output [1].

▪ Voice Input: This primarily includes audio reception and preprocessing, aiming to obtain clearer audio through noise reduction, voice enhancement, and similar techniques.

▪ Voice Processing: This stage involves processes such as speech-to-text recognition and semantic understanding, primarily utilizing AI technologies to comprehend and manage the spoken language.

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

▪ Voice Output: This stage involves dialogue management to output text and commands, followed by text generation and speech synthesis based on the generated text, providing users with feedback on the system's processing results.

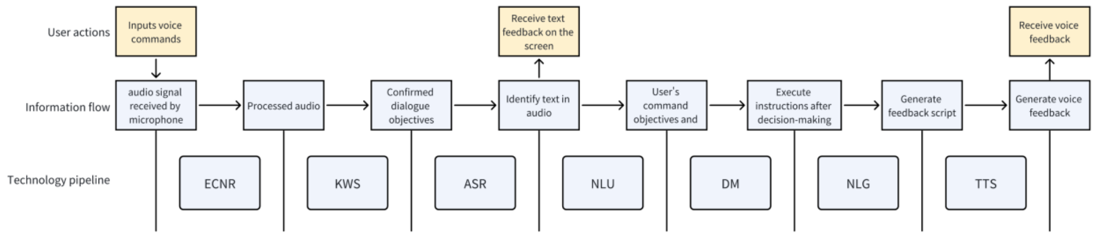

Figure 2: Vehicle intelligent voice assistant technology chain.

Figure 2 illustrates the interaction technology chain of in-car voice assistants, encompassing the entire process from the user's input of voice information, through the recognition and understanding of this information, to the generation and output of feedback. The following text will provide a detailed introduction to the key stages and definitions within this technology chain.

When a user speaks a voice command inside the vehicle, it is first received by the in-vehicle microphone array as an audio signal and then uploaded to the cloud for further processing. Before recognizing the speech, the audio needs to be processed to obtain clean audio, thus improving recognition accuracy. The role of the Echo Cancellation & Noise Reduction (ECNR) frontend speech processing module is to eliminate echo and reduce other noise. In a complete intelligent voice interaction process, the voice system is usually awakened through keyword spotting (KWS), transitioning from sleep to recognition state. The starting point of voice interaction is voice wake-up. After being awakened, the voice assistant performs real-time recognition of the received user speech, distinguishing valid voice commands and dialogue content, i.e., Automatic Speech Recognition (ASR) technology [3].

After recognizing the user's input voice commands as text, the intelligent voice interaction system needs to convert the text into understandable and quantifiable structured intents. The role of Natural Language Understanding (NLU) is to transform input text information into structured intents that the machine can understand. The structured information of NLU typically consists of three parts: domain, intent, and slots. The domain defines which category of services and functions the voice command needs to invoke. Different domains involve different technology pipelines and interaction channels. Users also have different experiential demands for voice commands in different domains. The intent specifies the specific functional points and objectives to which the voice command refers, while slots contain all the target parameters in the voice command.

Dialogue management (DM) plays a central role similar to the neural center in an intelligent voice interaction system, responsible for collecting information, analyzing it comprehensively, and outputting it. It manages the logic and status of the entire conversation and issues various control commands, serving as the starting point and hub of all business logic in the voice system. For example, when a user says "open the window," the intent recognition module outputs domain: vehicle control, intent: open the window, slots: open. The dialogue management module needs to collect and supplement information such as user information, vehicle speed, and window status. Based on the collected information and pre-set execution rules, it sends execution commands to the vehicle and generates feedback information in the form of words and sentences.

After completing information collection, the DM module outputs different types of feedback results. For some dialogue scenarios that depend on external information retrieval and service provision, the retrieval results output by DM need to be further organized into complete textual responses. Natural Language Generation (NLG) is responsible for generating the specific content that the voice system will reply with, converting non-verbal data into human-understandable language format. Thus, NLG can be seen as the reverse process of NLU. Text-to-Speech (TTS) is the final step in the speech recognition process. It parses the text generated by the NLG module, converts it into audio information, and outputs it to devices such as speakers.

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

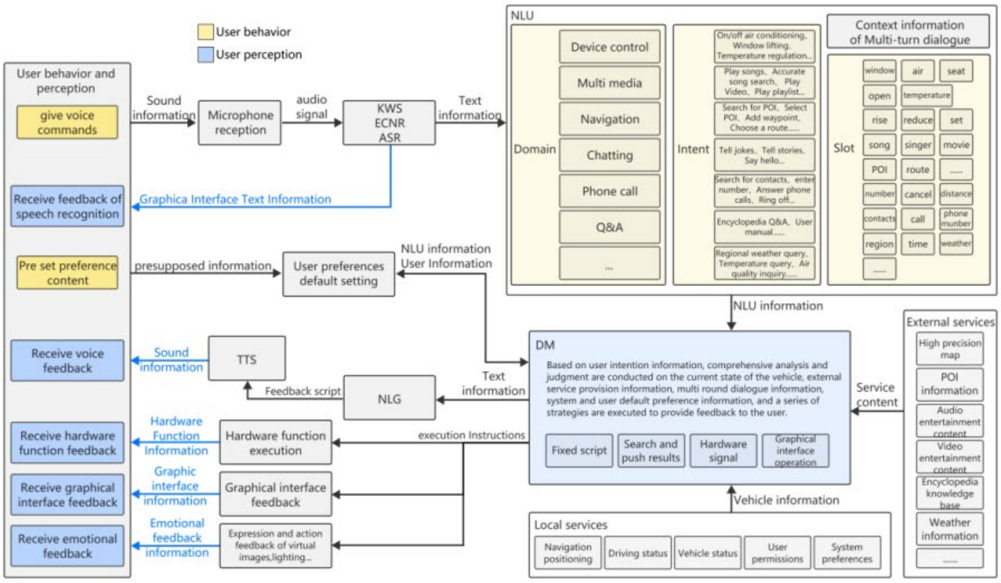

Figure 3: Information exchange process of in vehicle intelligent voice assistant.

Figure 3 demonstrates the information interaction relationships among the user, voice assistant, vehicle, and external services within the technology chain. It can be observed that NLU and DM are the core processes in the entire intelligent voice interaction, and they are also the main focus of intelligent voice interaction design.

2.2. In-Vehicle Voice Experience Design

2.2.1. Designing Role Profiles for In-Vehicle Voice Interaction

During conversations with various voice assistants, different characterizations of voice assistants will have different psychological effects on users. Users also apply their understanding of human society to their conversations with voice assistants. In design, the lack of human-like attributes such as intimacy and personalization in voice dialogue products can make users feel mechanical and awkward, leading to abandonment of use [4].

Establishing a persona and IP image for the voice assistant serves as a basic reference for designing dialogue functions and writing various dialogues. This helps designers of intelligent voice assistants decide on appropriate voice skills and the wording and grammar used in conversations in different scenarios. As the direct object perceived by users during conversation, users tend to project the role of the voice assistant as a real person and establish a user mindset for the voice assistant product and its brand over long-term dialogue interactions [5]. The interaction design of conversational intelligent assistants should meet people's expectations for roles such as assistants, companions, and pets, and defining the IP and persona of the voice assistant is the basis for meeting these requirements.

2.2.2. Functional Domain Division and Dialogue Norms

In the intelligent voice interaction pipeline, semantic understanding of user voice input includes domains, intents, and slots, where domains define the type of interaction goal for the user and the services needed to complete the interaction. Therefore, in the field of in-vehicle intelligent assistants, voice functions are mainly divided based on the types of services provided to users: phone, weather and date inquiry, navigation, entertainment (music, video, audiobooks), vehicle control, and chat, among others. Different domains provide different functions and information interaction content, and users have different expectations.

To ensure consistency in the user experience of the voice assistant and a unified perception of the persona, it is necessary to establish standards for dialogue writing based on role settings. This involves designing high-frequency basic dialogue scripts, including wake-up, greetings, farewells, etc., and standardizing the wording, information conveyed, text length, and text style for different functions in various domains. For example, for a professional, formal, and serious voice assistant, unnecessary modal particles should be minimized in dialogue design, and subjective expressions should be reduced in reply scripts. In contrast, for a voice assistant emphasizing emotional care and highlighting a friendly and cute

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

image, more informal and equal language should be used in addressing users, and emotional expressions should be added to pursue empathy with users in greeting and reply scripts.

2.2.3. Design of Functional Modules and Generic Dialogue Logic

In the process of designing in-vehicle voice assistants, it's essential to establish different functional modules within various domains, categorizing various atomic functions to provide users with complete and comprehensive functionality while ensuring convenience in technical implementation. For example, in the vehicle control domain, functional modules may include air conditioning adjustment, seat massage adjustment, seat position adjustment, window up/down adjustment, etc. In the media domain, this might include music search, video search, and audiobook playback.

In a complete voice interaction process, users interact with intelligent voice assistants in various complex dialogue scenarios and logics. To ensure natural and fluent dialogue interactions, it's necessary to design various generic dialogue logics at the framework level. Common dialogue flows typically include voice wake-up, semantic judgment, rejection of non-target dialogues and invalid voice commands, multi-turn dialogues, etc. These types of dialogue logics cover dialogue flows in all functions and scenarios, requiring designs that align with the most natural and fluent communication states for users, while also integrating various technical requirements and actively using new technologies to optimize dialogue performance. Comprehensive and unified dialogue logic design ensures that the voice assistant can use a consistent "thinking process" to converse with users in various interaction scenarios, helping users better understand and perceive the current status.

2.2.4. Intent Understanding and Voice Interaction Paths

At the structural level of intelligent voice assistant design, specific voice skills need to be designed. In the current framework of voice interaction technology, understanding the semantics of user utterances by intelligent assistants requires manual definition and annotation by designers, followed by AI model-based intent classification. Within a functional module, users often have various demand points and different language expressions and psychological expectations based on those demands. Therefore, designers need to deeply understand the user mental model, interface interaction model, and technical implementation model of task execution when decomposing user intents and designing voice interaction paths. The goal is to bridge the gap between user mental models and actual execution logic in the multi-device, multi-channel interaction process between voice and graphical interfaces. Designing voice interaction paths requires detailed analysis of various scenarios and states of users, vehicles, and environments based on current information and user goals, determining the types of broadcast scripts, dialogue rounds, and fallback scenarios for various ineffective feedback situations.

2.2.5. Auditory and Visual Feedback

As an intelligent interactive product in the intelligent car cockpit, the experience design of in-vehicle voice assistants includes not only output feedback at the voice level but also interaction feedback on graphical interfaces, emotional feedback from virtual representations of voice assistants, and multi-channel multi-device interaction feedback such as lights, sound effects, and vibrations[2].

The necessity of graphical interface design for in-vehicle voice assistants lies in supplementing the limitations of voice interaction itself. In scenarios where conveying complex and lengthy textual information, image information, and structured list information is necessary, voice alone may not effectively convey such information or may result in overly long broadcast times or overly complex abstract descriptions. Additionally, as an intelligent system, there are issues of explainability in the interaction process between users and voice assistants; users need to perceive how the voice assistant recognizes and understands the current conversation state to have more accurate psychological expectations for various voice assistant feedback. To address these issues, graphical interface feedback for in-vehicle voice assistants mainly focuses on two aspects of voice interaction:

- Displaying the real-time recognition status of the user's voice by the voice assistant when the user speaks a voice command, allowing users to perceive the current working status of the voice assistant and the recognition status of the voice.

- When executing specific commands, displaying pop-up windows with lists and checkboxes, providing users with the information needed to complete the interaction task, and offering the next step in the execution path.

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

2.3. Levels of In-Vehicle Voice Assistant Experience Design

In-vehicle intelligent voice assistants, due to their differences from traditional graphical interfaces in user interaction characteristics and technical implementation chains, have different requirements for product experience, as well as different contents and processes in experience design compared to graphical interface design methods. Therefore, based on the theory of user experience hierarchy, the author has divided and defined the elements of user experience for in-vehicle intelligent voice assistants and summarized the content of experience design for in-vehicle voice assistants in Figure 4[6].

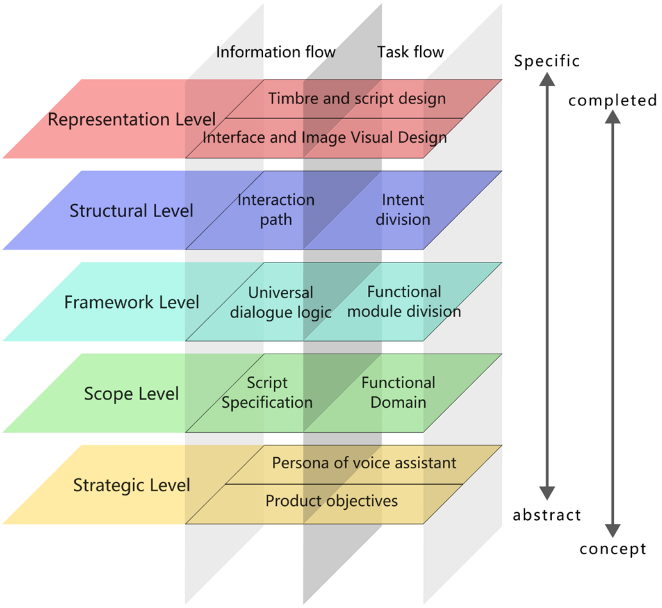

Figure 4: Hierarchical interaction design for vehicle voice assistant.

3. Emotional Design Model for In-vehicle Voice Assistants

3.1. Emotional Interaction Experience of In-vehicle Voice Assistants

In emotional design theory, emotional design is divided into three levels: instinctive, behavioral, and reflective[7]. These levels respectively focus on users' emotional experiences when they first encounter a product, during deep usage, and when establishing emotional connections with the product. In the usage scenarios of in-vehicle voice assistant products, such as sales demonstrations, long-term interactions with car owners, and interactions with other passengers, the emotional experience of in-vehicle voice assistants needs to be considered at all three levels.

The instinctive level design emphasizes short-term visual perceptions and emotional impacts. In the instinctive level design of in-vehicle voice assistants, it mainly depends on design at the presentation level and the definition of the persona of the voice assistant at the strategic level. Specifically, in the auditory channel, it focuses on the perception of the voice assistant's tone and the content of conversational scripts; in visual design, it mainly focuses on the design of the virtual image of the voice assistant.

The behavioral level emphasizes efficiency and enjoyment during product usage. In the behavioral level design of in-vehicle voice assistants, it mainly focuses on the experience of the entire dialogue process, requiring the provision of smooth, natural, and user-friendly voice interaction experiences to help users efficiently and accurately complete interaction tasks in various driving scenarios. Additionally, it can assist users in better understanding and familiarizing themselves with various functions inside the car through voice interaction.

The reflective level design emphasizes establishing long-term emotional bonds with users. To generate more empathy with users, the reflective level design of in-vehicle intelligent voice assistants focuses on users' long-term usage experiences, enhancing users' sense of identity with the voice assistant product and brand, and establishing deep emotional connections with users during long-term usage. In the users' minds, in-vehicle voice assistants have more complex identities, such as car steward, in-vehicle

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

companion, or in-vehicle pet. Reflective level design focuses on providing users with more emotional value under these identities.

3.2. Emotional Experience Design of In-vehicle Voice Assistants

in-vehicle voice assistants focus on different experience aspects under different emotional design levels, and these aspects correspond to different design contents at each level. Table 1 illustrates the content of emotional experience design for in-vehicle voice assistants.

Table 1: Emotional Experience Design for In-vehicle Voice Assistant.

Emotional Design Levels	Emotional Experience	Emotional Experience Design
The instinctive level	Visual image Synthesized Voice Selection Script Design	Representation Level: Image, voice, script Strategic Level: Persona definition
The behavioral level	Response Speed Recognition Accuracy Status Awareness Dialogue and Execution Logic	Representation Level: Script Design Structural Level: Interaction paths, intent recognition Framework Level: Dialogue logic, functional modules Scope Level: Functional domains
The reflective level	Emotional Companionship Social Value Identity Recognition Sense of Accomplishment	Representation Level: Image, script Framework Level: Functional modules Scope Level: Script standardization Strategic Level: Persona definition, product goals

4. Emotional Design Strategies for In-vehicle Voice Assistants

In the field of intelligent voice interaction design, Google has proposed design guidelines for conversational interactions[8], including: (1) Keep it concise; respect the user's time; (2) Build user trust; provide only the most natural way of communication and keep the conversation going; (3) Consider the context of the conversation; it should be contextually relevant and adaptable to the user's current needs and environment. (4) Sound pleasant without distracting the user's attention; (5) Engage novice users while continuing to attract expert users; (6) Take turns speaking; refrain from interrupting the user when it's their turn to speak; (7) Avoid guessing the user's intentions; provide factual information and let the user make decisions. These requirements include guidelines for conversational scripts, design requirements for interaction paths, and emotional requirements. Emotional interaction experience and information dissemination are important features that distinguish voice interaction from graphical interface interaction, and they are also one of the important goals of in-vehicle voice assistant product experience design. Therefore, based on the three levels of emotional design, this paper proposes interaction design strategies for in-vehicle voice assistants in the context of emotional design.

4.1. Instinctive Level Design Strategies

4.1.1. Synthetic Voice Design

For voice design, under the current commonly used Text-to-Speech (TTS) technology, designers can adjust parameters such as voice tone, speed, and intonation based on the role profile of the voice assistant. Apart from role positioning, requirements for synthetic voice include clear articulation, natural intonation, moderate speed, friendly tone, and establishing a sense of trust. Additionally, different preset voice options can be provided to users for selection, catering to various user types and usage scenarios.

4.1.2. Script Design

The information conveyed by voice assistant scripts typically includes: confirming the user's interaction task objective, confirming the current state of the interaction object, confirming the execution result of the interaction command, and providing explanations and guidance when commands cannot be directly executed. Script design can refer to Grice's maxims: (1) Quantity maxim: Only say truthful information; (2) Quality maxim: The information provided should satisfy the required amount for communication but not exceed it; (3) Relation maxim: Only say content relevant to the topic; (4) Manner

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

maxim: Speak clearly and avoid ambiguity.

4.1.3. Visual Image Design

Visual design of intelligent voice assistants needs to consider: (1) Interaction status display, conveying the current working status of the voice assistant to users (e.g., sleep, listening, analyzing), enhancing the user's understanding of the interaction status; (2) Brand tone, reflecting the brand and the tone and temperament defined and conveyed by the voice assistant's role profile; (3) Visual aesthetics, considering the aesthetic preferences of the target users and current popular visual design styles; (4) Emphasizing and balancing technological and emotional attributes, reflecting both the artificial intelligence technology and the caring and affinity in visual image design, requiring designers to consider both aspects simultaneously.

4.2. Behavioral Level Design Strategies

The design of the behavioral level experience can be considered from the perspective of the user journey of conversational interaction.

4.2.1. Voice Wake-up

Reduce the interaction threshold for user wake-up and provide various wake-up methods. This includes supporting replaceable wake-up words, allowing users to wake up the voice assistant with their preferred names; supporting skipping the wake-up process directly for frequently used voice commands (wake-up free); supporting one-shot interaction, where wake-up and task recognition execution are completed in one round of dialogue; and triggering voice interaction passively through preset dates or events (e.g., voice memos).

4.2.2. Speech Recognition

During the speech recognition phase, displaying the recognized text and the working status during recognition on the graphical interface can enhance the interpretability of the interaction process, allowing users to better understand the current dialogue status. In the intent recognition phase, improving the accuracy of user input recognition from various dimensions is required. In real natural language interaction, users' language expressions often do not follow fixed templates but involve a large amount of uncertain free language expression, making accurate recognition of various expressions a major challenge[9]. Generalizing recognition content includes: (1) Supporting a wider variety of sentence patterns; (2) Recognizing and providing feedback on single-entity words when users only say one target word, such as recognizing "air conditioning" directly as the intent to turn on the air conditioning based on the current air conditioning status; (3) Increasing fault tolerance, recognizing more accurate intents when users make inaccurate expressions; (4) Supporting colloquial and emotional expressions, such as mapping user feelings expressed as "too hot" or "too dark" to actionable intents.

4.2.3. Multi-turn Dialogue Design

Voice interaction skills can be divided into task-based dialogue, chat-based dialogue, and question-and-answer dialogue. In task-based and question-and-answer dialogues, users usually have clear task goals, and the interaction aims to exchange information or execute functions. In chat-based dialogue, user states are more similar to natural conversational states with no clear goals and require uncertainty in dialogue content. Therefore, in the process of dialogue task execution, users have different expectations for different types of dialogue turns.

Task-based dialogue skills require shortening the interaction process, completing the interaction task in one or two rounds of dialogue whenever possible. In contrast, chat-based dialogue skills require extending the interaction process to simulate active and engaging dialogue scenes similar to real human conversations. Strategies for shortening the interaction process in task-based dialogue include: (1) Reducing information confirmation steps; in task-based dialogues, commands issued by users often do not cover all the target parameters and requirements needed to complete the task. For example, in "adjusting the air conditioning temperature," the target air conditioning outlet position and temperature information are not specified. To complete the task within one round of dialogue, these types of information need to be supplemented, including obtaining user location information, sensory information, setting default values, obtaining system current status, obtaining external environmental information, and obtaining user preference information, etc.; (2) Merging or reducing the wake-up process for frequently used task-based dialogue skills; (3) Providing guidance paths for complex or unexecutable user requests to reduce selection steps and re-initiation of task processes. For chat-based tasks, designers often cannot

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

define clear interaction paths, thus relying more on expanding the voice assistant's knowledge base and accessing external databases or knowledge graphs. Additionally, based on the capabilities of large model technologies, establishing stronger long-term memory capabilities for multi-turn dialogues with better context understanding can allow the voice assistant to handle more complex and longer chat dialogue tasks[10].

4.2.4. Error Prevention Design

in-vehicle voice assistants need to consider the primary task of drivers during the interaction process, which is to safely drive. Therefore, it is necessary to establish permission control systems based on sound source positioning and voiceprint recognition, verify and confirm the activation and use of important modes and safety-related functions in driving scenarios, and provide danger warnings. For high-speed driving scenarios, targeted execution strategy designs for various functions are needed, such as automatically reducing window opening levels and limiting seat and rearview mirror adjustments during high-speed driving.

4.2.5. Proactive Interaction

Proactive interaction[11]: Based on memory of common functions and user adjustment goals, functions are intelligently pushed according to events, time, or location triggered by users, creating new task-based dialogue interaction processes through voice-initiated proactive interaction or text reminders on the graphical interface, which can further reduce the threshold for voice interaction initiation and improve the efficiency of voice interaction execution[12].

4.3. Reflection Level Design Strategies

4.3.1. Personification Design

Strengthening character shaping based on the existing voice assistant role profile to enhance personification and providing more personalized, customized, and memorable voice content as well as IP image design[13]. As a "more realistic entity," the voice assistant can have its own background story, preferences, strengths, and weaknesses. In script design for voice dialogue, more stylized and customized content can be formed, such as having its own catchphrases, commonly used sentence structures, etc., to create stronger memories and perceptions for users. When executing various voice tasks, informing users of the current supported voice commands based on the existing capabilities, allowing users to have a more comprehensive understanding of the voice assistant while fostering a sense of closeness. Inserting introductions to the voice assistant's self-character and background story in casual conversation skills to create a sense of closeness and trust while conveying the design philosophy of the product and brand to users.

4.3.2. Multi-channel Information Feedback

To simulate the rich and varied body and facial interactions in natural interactions between people during conversations, important methods for emotional information transmission for in-vehicle voice assistants include virtual image expression, motion design, and lighting design. For this type of emotional experience, customized designs based on the voice assistant's role profile are also required. Virtual image expressions and actions usually need to convey the interaction tasks and work status of the voice assistant, reflect more emotional and perceptual content, and resonate with the user's current experience[14]. As shown in Figure 5, NIO's voice assistant NOMI shows expressions and movements that accompany the music rhythm while users listen to music.

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

Source: https://hao.yiche.com/wenzhang/33708958

Figure 5: Expression and action feedback of NIO car.

4.3.3. Companion and Development Mechanism Design

Establish a companion and growth mechanism to let the voice assistant remember the user's names, preferences, etc. The memory content includes the names, voices, common positions, common function settings, common voice skills, and preferences for media and entertainment resources of multiple users in the car. After a certain period of learning and memory, the voice assistant can push functions and information content that are more in line with expectations and preferences to users. Making various abilities and user understanding enhancement processes perceptible to users, allowing users to perceive stronger feedback on growth and improvement during long-term use. For example, informing the user when performing a function that has not been pre-supported or memorized, "I don't know how to do that now, but I'm learning, and I'll be able to do it next month," or "I don't know now, can you tell me your preferences? I will remember them."

4.3.4. Social Ability Design

As a personified intelligent voice assistant, the interaction process with users can also be seen as a social process between humans and artificial intelligence. Therefore, meeting users' needs for identity recognition, value proof, etc., in social scenarios is also an important goal of the voice assistant's emotional experience. Based on the memory or pre-input of user information and social relationship chains, providing different greetings, greetings, special holiday or birthday wishes to different users in the car, and offering various special scene or special time-triggered Easter eggs, hidden functions; creating surprises and unexpected situations for users, satisfying users' self-identity, and group identity needs, can create memorable experience points during the long-term interaction between users and the voice assistant.

5. Conclusions

With the development of the smart automotive industry and artificial intelligence technology, in-vehicle intelligent voice assistants have become particularly important in the user experience of automotive cabins. As users delegate more complex tasks to voice assistants, they also have higher expectations for the experience. Therefore, the author, in conjunction with the smart in-vehicle voice technology chain, analyzed the complete process of in-vehicle voice assistant experience design from the perspective of user experience elements. They constructed an experience design model from the strategic layer to the presentation layer. Based on the analysis of in-vehicle voice assistant experience design, design strategies for emotional experiences, particularly prominent in voice experiences, were proposed. These strategies analyze the optimization design strategies for emotional experiences of in-vehicle voice assistants from the three levels of instinct, behavior, and reflection, providing a reference for the construction of the user experience model for in-vehicle voice assistant products and emotional experience design.

ISSN 2616-7433 Vol. 6, Issue 5: 78-88, DOI: 10.25236/FSST.2024.060511

References

[1] Wang X, Lei Q, Mei L, et al. (2021) A Review on Development Trends of Automotive Voice Interaction Technologies. Automotive Digest (Chinese), 2021(2): 9-15.
[2] YANG Ming-hao, TAO Jian-hua. (2018) Intelligent Method of Multi-model Information Fusion in Human-computer Interaction. Science in China: Information Science, 2018, 48(4): 433-448.
[3] LIU Quan, XU Wei, LI Shen-an. (2021) Development Trend of Automotive Intelligent Voice: Interactive Upgrade from Passive to Active. Auto Electric Parts, 2021(10):5-8.
[4] Chang, R. C. S., Lu, H. P., & Yang, P. (2018) Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. Computers in Human Behavior, 84, 194–210. https://doi.org/10.1016/j.chb.2018.02.025.
[5] SUN Yan-yan, LI Shi-yan, CHEN Xian-tao. (2019) Emotional Voice Interaction Design: Human Computer Interaction Research Map and Design Case of Baidu AI User Experience Department. Art & Design, 2019(11): 22-27.
[6] Garrett, J. J. (2022). The elements of user experience: User-centered Design for the Web and Beyond. 2nd ed. Upper Saddle River, NJ: Pearson Education.
[7] Donald A. Norman. (2004) Emotional Design, Basic Books, New York.
[8] Conversation design. https://developers.google.com/assistant/df-asdk/design.
[9] LIU J. LI Y L. (2019) Review of intent detection methods in humanmachine dialogue system. Computer Engineering and Applications, 2019, 55(12):1-7.
[10] Kraus J M, Nothdurft F, Hock P, et al. (2016) Human after all: Effects of mere presence and social interaction of a humanoid robot as a co-driver in automated driving//Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. 2016: 129-134.
[11] LI Jing-lu, SUN Xiao-hua, GUO Wei-wei. (2018) Proactive HMI Design Based on Smart Interaction. Journal of Graphics, 2018, 39(4): 668-674.
[12] Gai Sen, Lu Yi, Zhang Yinshuai, et al. (2023) Research on Proactive Interaction Design for Smart Home System. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(2): 230-237.
[13] Jizhou Huang, Haifeng Wang, Shiqiang Ding, and Shaolei Wang. (2022) DuIVA: An Intelligent Voice Assistant for Hands-free and Eyes-free Voice Interaction with the Baidu Maps App. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22). Association for Computing Machinery, New York, NY, USA, 3040–3050. https://doi.org/10.1145/3534678.3539030
[14] Zhang Fang, Xu Zihan. (2024) How to Perceive AI Communicators: Experimental Research on the Influence of Anthropomorphism on the Effects of Impression Formation of Bodiless Conversational Social Robots. Journalism and Mass Communication, 1-13[2024-04-22]. https://doi.org/10.15897/j.cnki.cn51-1046/g2.20240326.001.