TacTalk: Personalizing Haptics Through Conversation

by

Anchit Mishra

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirements for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2024

© Anchit Mishra 2024

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Haptic experiences are highly personal, with perception and preferences potentially varying drastically between different people. While different personalization interfaces have been proposed in the past, we do not have a detailed understanding of how people personalize a haptic experience. Through this work, we aim to explore the vocabularies and mental models that different users have for personalizing haptic experiences. To study this, we introduce a conversational system based on the GPT-4o Large Language Model (LLM), TacTalk, which enables users to describe and finetune virtual haptic experiences on the fly, without the need for navigating different settings or requiring expert interventions. Inspired by the manner in which professional racing drivers communicate with their engineers to ensure their car is functioning properly, we present an application using TacTalk alongside a popular racing video game, Forza Horizon 5, along with evaluations of the system's usability and parameter tuning consistency. We show that LLMs can be used to manipulate haptic feedback parameters in a predictable and valid manner. Additionally, we conduct a user study and synthesize key themes of user preferences towards customizing haptic feedback from interview and questionnaire responses. Responses to the NASA-TLX and SUS questionnaires show that users prefer a conversational mode of interaction over a visual interface offering the same capabilities. Additionally, we collect users' responses to the Big Five Inventory and Player Traits questionnaires and probe their correlation with personalization preferences. Our qualitative findings show that when starting from scratch, users initially draw from real-world experiences and metaphors, and subsequently refine haptic parameters based on other metaphors or focus on specific aspects of the experience such as in-game events and the video game controller being used. An average conversation with TacTalk involves fewer than 5 back-and-forth interactions, with users preferring voice-based interaction over a visual interface with the same functionality. System architectures such as TacTalk, owing to the generality of LLMs and their ability to use additional context without requiring retraining, can be used in a plug-and-play manner with existing virtual experiences.

ACKNOWLEDGMENTS

Firstly, I would like to thank Prof. Oliver Schneider, for his supervision throughout my master's degree. Additionally, I would also like to thank other people who I've looked to for advice about research and methodologies, including my co-supervisor Prof. Daniel Vogel (for his awesome course on HCI methods), Dr. Tor-Salve Dalsgaard (for his insight into theory in HCI and all its pros and cons), Dr. Eugene Kukshinov (for his help with navigating various qualitative analysis methodologies) and Prof. Margot Racat (for advice on how to work through graduate school and take criticism positively).

I'm immensely grateful to my parents for their constant and unwavering support throughout my journey, through all the ups and downs. I would also like to thank my friends Michael Xu, Poulomi Ganguly and Sankalok Sen, along with my peers from the Games Institute and Haptic Experience Lab for helping me retain my sanity.

Finally, I'd like to thank Profs. Matthew Brehmer and Sirisha Rambhatla for agreeing to read my thesis and for providing valuable feedback.

TABLE OF CONTENTS

List of Figures vii
List of Tables ix
Colophon x
1 Introduction 1
2 Background 3
2.1 User Experience and Haptic Experience 3
2.2 Personalizing Haptic Feedback 4
2.3 Haptic Vocabularies 5
2.4 Large-Language Models and UI Transducers 6
3 TacTalk System Design 8
3.1 System Architecture 8
3.2 Deciding Which Settings to Change 13
3.3 Updating the User Preference Profile 15
3.4 Interaction Loop 15
4 Technical Evaluation 18
4.1 Prompt Dataset 18
4.2 Error Analysis 18
4.2.1 Methods 18
4.2.2 Results 19
4.3 t-SNE Dimensionality Reduction 19
4.3.1 Methods 19
4.3.2 Results 19
5 User Study 22
5.1 Participants 22
5.2 Apparatus 22
5.3 Tasks 23
5.3.1 FREE-PERSONALIZATION 24
5.3.2 CAR-CLASSIFICATION 25
5.4 Quantitative Results 25
5.4.1 FREE-PERSONALIZATION Results 25
5.4.2 CAR-CLASSIFICATION Results 26
5.5 Qualitative Results 29
5.5.1 Users favor a few-shot approach towards personalizing haptics through conversation 29
5.5.2 A realistic world suggests using knowledge of real-world haptics 30
5.5.3 Using haptic feedback to represent changing environment and state is preferred 31
5.5.4 Knowing about 'reality' does not imply a preference for it 32
5.5.5 Domain-specific metaphors are a valuable tool 33
5.5.6 Summary 34

6 Discussion 35
6.1 LLMs can be used to make predictable and valid parameter adjustments for haptics 35
6.2 Conversational personalization is more usable than a visual interface 35
6.3 Personality differences may impact personalization preferences 36
6.4 Preference context improves users' classification accuracy, but not by much 36
6.5 Conversational interfaces should use signposting to make users' lives easier 37
6.6 Interfaces enabling personalization of haptic feedback should employ narrative framing effectively 37
6.7 Conversational interfaces can be a valuable learning tool 38
6.8 Conversational personalization is a unidirectional process 39
7 Conclusion 40
7.1 Future Work 41

Bibliography 42

Appendices
A TacTalk System Prompts 59
A.1 When Using a Single Prompt 59
A.2 When Using Multiple Prompts 61
A.2.1 Stage 1: Modifying Haptic Parameters 61
A.2.2 Stage 2: Updating User Preference Profile 67
B Initial Settings Configuration 69
C Technical Evaluation Dataset 70
C.1 Prompts Used for Technical Evaluation 70
D Screening Questionnaire 72

LIST OF FIGURES

Figure 2.1 The first version of the HX model. Here, personalization support was identified as an important factor for HX, but not as thoroughly explored as design parameters, usability requirements and experiential dimensions. 3

Figure 3.1 The ForzaDSX mod - a visual interface enabling customized force feedback response using the DualSense controller's Adaptive Triggers. After the `config.xml` file is modified by the OUTPUT layer, a restart of ForzaDSX applies the changes to the DualSense controller. 9

Figure 3.2 An illustration of TacTalk's system architecture, divided into four layers: INPUT PROCESSING, CONTEXT, RESPONSE and UPDATE. 10

Figure 3.3 A screenshot from Forza Horizon 5, illustrating how the settings from ForzaDSX take effect during gameplay. Refer to Table 3.1 for more details about each setting. 11

Figure 3.4 A process diagram illustrating one iteration of the interaction loop between a user and TacTalk. 13

Figure 3.5 A storyboard illustrating how users may interact with TacTalk. Here, a user asks for the car to feel heavier, and TacTalk responds with a modified haptic feedback configuration that results in stronger resistance on the DualSense controller's triggers, mimicking a heavier vehicle's throttle. 16

Figure 3.6 Here, based on the results of the earlier interaction, a user may choose to follow up with the system in one of three ways. If the response aligns with their preferences, they may only tune minor aspects of the experience, but if not, they may look towards other attributes to try and achieve a preferable setting. They may also choose to revert to the previous configuration. 17

Figure 4.1 A summary of the number of errors encountered for each style of prompting. Using multiple prompts with example query-response pairs resulted in only 54 errors, as compared to 1932 errors when using a single prompt for all tasks. 20

Figure 4.2 The t-SNE visualization obtained from the dataset of 51 user queries. The plot shows that the haptic feedback configurations generated by TacTalk possess some structure, with vehicles of similar performance clustered together. There are also some visible outliers, as highlighted. 21
Figure 5.1 The apparatus used for the study, including a PC, a laptop, a pair of headphones, and the DualSense controller. 23
Figure 5.2 A participant seated for a user study session. 24
Figure 5.3 Scatterplots visualizing Player Traits and Big Five Inventory dimensions and their relationship with NASA-TLX and SUS responses. 26
Figure 5.4 Box plots illustrating the NASA-TLX scores for TacTalk and the ForzaDSX visual interface. 27
Figure 5.5 Confusion matrices visualizing users' classification accuracy when matching haptic feedback to vehicles. 28

LIST OF TABLES

Table 3.1 The list of parameters exposed by ForzaDSX for haptic feedback modification. All numeric values are percentages represented as unsigned integers. 12

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography *"The Elements of Typographic Style"*. `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Computers are good at following instructions, but not at reading your mind.

– Donald Knuth

1 INTRODUCTION

Researchers have established that individual differences exist in the perception of haptic feedback [18, 44, 79, 97, 98, 108]. While certain experiences aim to simulate reality, and thus cannot cater to what 'feels' best for individual users, there exist other application areas such as entertainment and video games where personalized haptic feedback may play an important role in contributing to user experience.

Prior work has explored different strategies towards enabling the personalization of haptic feedback. This includes design tools [90, 95, 103] that allow users to visualize and design their own haptic effects, as well as morphing algorithms [20] that allow users to draw upon pre-existing libraries of haptic effects to produce predictable and perceivable compound effects. Seifi [93] specifically classifies different personalization approaches into three categories: choosing, which involves selecting haptic effects from a larger collection; tuning, which involves manipulating a given haptic effect based on the variation of one or more perceivable parameters, such as frequency, amplitude, duration etc.; and chaining, which involves stacking multiple smaller chunks of haptic effects to obtain a larger, compound sensation.

Although this work has helped converge on more effective approaches towards personalized haptic feedback, navigating haptic parameter spaces for effective customization remains a difficult task for end-users, owing to the various technical terms involved in rendering haptic sensations that need to be manipulated. Users may be able to vaguely describe their intended haptic feedback settings, but may not know exactly how their description maps to underlying parameters. Towards this, previous works have explored a language-based approach towards describing haptic feedback, specifically focusing on affect (i.e., emotion). This includes psychophysical studies aiming to understand the perceived tactile dimensions for various materials [33, 71], as well as approaches for capturing user experience [36, 70]. Another interesting finding was the existence of a negative correlation between user attention and affect, as found by Zheng and Morrell [115], where the use of an intensity control was eventually recommended to account for individual differences in vocabulary while still enabling personalization. On the end-user side, previous works have attempted to understand the perceptual dimensions of tactile stimuli [94, 96] but progress has been limited, with some even arguing against the existence of a tactile language [38].

Towards this, Large Language Models (LLMs) have recently been shown to possess strong capabilities of producing high-quality natural language outputs [32, 39, 59, 73, 104], which can be extended to multimodal application contexts through the concept of user interface transducers [10]. Additionally, LLMs can keep track of context and can even understand the meaning of seemingly vague natural language queries [4, 31, 59]. This presents the opportunity

of using LLMs to map user descriptions of haptic feedback derived from a diverse and non-standardized vocabulary to underlying control parameters.

In this thesis, we introduce TacTalk, a novel application of LLMs as a natural language transducer for conversational personalization of haptic experiences. We demonstrate TacTalk's functionality in the context of a car-racing video game and use it as a probe to study how people personalize haptics. We show that our application enables users to easily customize technical controller parameters without knowing about the actual parameter names with significant differences in perceived mental demand, success, effort and frustration in addition to other usability factors compared to a visual interface. In addition, we show that TacTalk can be used as part of a larger system to keep track of individual user profiles and navigate between different haptic feedback settings through conversation. TacTalk shows consistency in terms of its ability to map technical parameters to a given world model, which we demonstrate may either be formed using user profiles or even the baseline GPT-4o model [32] without providing additional context. Based on user interactions with TacTalk and a semi-structured interview, we conduct a thematic analysis to learn about how people personalize haptic feedback. We see that conversational personalization follows a uni-directional, top-down process where users first draw from real-world experiences and related metaphors to identify their preferred haptic feedback and subsequently hone in on specific aspects of their experience, such as controller vibration/resistance or in-game events such as driving the car over wet surfaces. Further, we observe that all metaphors are not created equal, with users specifically employing motorsport-related metaphors instead of more generic language. Lastly, we see that although real-world experiences influence haptic feedback preferences, realistic haptic feedback is not always preferred.

2

BACKGROUND

In this chapter, we discuss prior work related to haptic experiences, including conceptual frameworks, empirical studies on the perception of haptic stimuli, approaches towards personalization, the importance of language in haptics and the role of LLMs in enabling new multimodal approaches towards user interfaces.

2.1 USER EXPERIENCE AND HAPTIC EXPERIENCE

User experience (UX), as defined by Sharp, Preece and Rogers [86] and further detailed by Schrepp et al. [92] is *"a set of distinct quality criteria that includes classical usability criteria, like efficiency, controllability or learnability, and non-goal directed or hedonic quality criteria, like stimulation, fun-of-use, novelty, emotions, or aesthetics."* UX depends on the user's internal state of mind, the environment within which an interface is being used, and varies from individual to individual. It also varies over time - a user's perception of UX may be shaped based on previous interactions with a product, ongoing use, and can even change after the interaction concludes [9, 55]. Owing to the subjectivity of UX, there are various approaches that have been proposed towards enabling the personalization of UX in different contexts, including information systems for man-museum interaction [52], recommender systems [43], augmented reality training systems [76], video games (with a focus on accessibility) [3]. Some works have also explored domain- and device-independent model-based methodologies towards UX personalization [34].

Haptic experience (HX) is a new, evolving area of UX that specifically aims to address experiences involving technologies that engage the sense of touch. Kim and Schneider [45] provided the first formal definition of HX, along with a model representing the factors of HX. This model, shown in Figure 2.1, was composed of five factors, namely Autotelics, Harmony, Expressivity, Immersion and Realism. Sathiyamurthy et al. [88] further extended this work

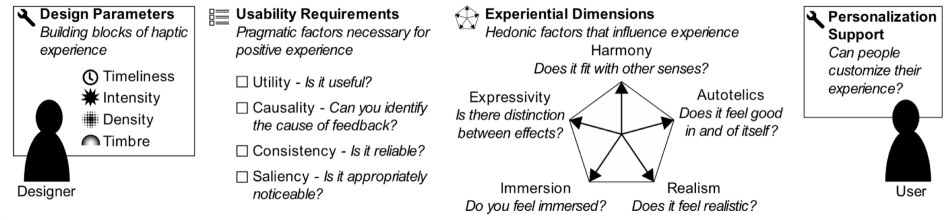


Figure 2.1: The first version of the HX model. Here, personalization support was identified as an important factor for HX, but not as thoroughly explored as design parameters, usability requirements and experiential dimensions.

to validate the use of the five-factor HX model for vibrotactile experiences. However, these works placed emphasis on vibrotactile feedback over other haptic modalities, such as force-feedback, skin-stretch, thermotactile, mid-air haptics etc. A subsequent study by Anwar et al. [5] aiming to generalize the initial HX model beyond vibrotactile experiences derived a new four-factor model of HX based on study results, consisting of Harmony, Expressivity, Involvement and Realism. Even so, although the various versions of the HX models acknowledge the existence of individual differences in the perception of haptic feedback and mention personalization as a relevant factor of HX, they do not capture personalization as a primary factor due to an incomplete understanding of how users think about personalization. Further, the HX model, even in its current state, is still under further development and the HX questionnaire needs to be validated before being recommended for use in hapticians' day-to-day workflows.

2.2 PERSONALIZING HAPTIC FEEDBACK

Parallel to work formalizing the concept of HX, other lines of research also include various approaches towards personalizing haptic feedback. For the purposes of this thesis' context, we include personalization to include authoring haptic effects from scratch as well as finetuning existing haptic effects to individual preferences. Tools supporting such personalization make use of a wide variety of methods, including relative ranking and refinement using machine learning [64], choosing from existing libraries of effects [96, 103], drawing [95, 103], morphing and chaining [20, 90] and even voice-based methods [66].

In addition to tools enabling haptic feedback personalization, previous works have also attempted to characterize individual differences in the perception of touch. Peck and Childers [79] in their work on consumer research developed the Need For Touch (NFT) scale, composed of 12 items covering autotelic and instrumental dimensions. In doing so, they emphasized the importance of taking into account differences in haptic perception to understand how consumers make purchase decisions. Stevens [97, 98] found a decline in haptic stimulus perception with age across various regions of the body, with some regions being more affected by aging than others.

Closely related to this work, Venkatesan et al. [108] investigated differences in perception arising due to age- and sex-related factors, citing the following:

(i) Changes in the physical structure of the skin

(ii) Changes to the number and morphology of tactile afferents

(iii) Differences in the functional use of hands

(iv) Differences in tissue conformation and thickness

(v) Differences in mechanoreceptor densities

(vi) Skin hydration levels

(vii) Skin temperature levels

Li and Gerling [57] found the stiffness of individuals' skin to play a role in their ability to discriminate between different levels of compliance, with more elastic skin deforming to a greater extent when pushing against a surface and increasing compliance discrimination performance. They also argue that differences in perception amongst young people arise mainly due to differences in hand sizes, which corresponds with both differences in tactile afferent densities (i.e., people with smaller hands have a higher density of tactile afferents and subsequently are more sensitive to haptic stimuli) and skin stiffness (i.e., people with smaller hands tend to have softer, more elastic skin). Similar findings were also reported by Löfvenberg and Johansson [62] for vibrations on the hand, Cholewiak and Collins [18] for the perception of a vibrotactile pattern set, and by Kawai and Tanaka [44] for skin vibration characteristics. Overall, a variety of studies show that individual differences in the perception of haptic feedback exist, and a myriad of approaches towards enabling haptic feedback personalization have been proposed in the form of design tools and haptic effect libraries. More recently, Roy et. al [87] attempted to extend the HX model and link it with end-user customization of haptic effects, identifying signal intensity and spatial density as two independent parameters that may be used for customization, while also acknowledging that haptic feedback customization preferences varied between individuals and also depended on the context of the virtual scenes they presented to users.

2.3 HAPTIC VOCABULARIES

Language is a key factor of interest in the study of individual differences in haptic perception. Works focusing on both personalized haptics [93] as well as HX from a design perspective [91] acknowledge the importance of language, especially affective language in describing haptic experiences. Language is a useful tool to understand both how people perceive tactile stimuli at a basic level, as well as how they talk about haptic experiences in a more abstract sense in their day-to-day lives.

Studies have attempted to analyze the different perceivable dimensions of haptic stimuli based on psychophysical methodologies. For instance, Hollins et al. [33] ran a study to isolate common perceivable dimensions of tactile surface texture, finding a three-dimensional space to provide the most satisfactory representation of the combined data of texture classification from 20 individuals. Two of these dimensions - roughness-smoothness and hardness-softness, were found to be quite robust and orthogonal, whereas the third dimension was found to correspond to the compressional elasticity of the surface upon post-hoc analysis. Okamoto et al. [71] conducted a similar analysis to understand the tactile perception of textures using psychophysical methods. They found three primary dimensions - roughness-smoothness, hardness-softness and coldness warmness. Further, they saw that the roughness-smoothness dimension could be split into two - macro, and fine roughness; likewise, they

acknowledged the existence of a friction dimension, related to the perception of moistness-dryness and stickiness-slipperiness. Thus, they found five overall dimensions of tactile perception. Adjacent to this work, Kaneda et al [42] investigated the relationship between physical attributes and onomatopoeic language when pressing a soft object, finding changes in the contribution of certain explanatory variables of onomatopoeia based on different physical conditions.

Focusing on end-users, tools such as VibViz [96] enable choosing vibrotactile effects based on different types of language, including metaphors (e.g., drums, engine, snoring etc.), physical filters (e.g., short note, long note, varied rhythm etc.), emotional filters (e.g., ticklish, sharp, anxious, agitating etc.) and usage example filters (e.g., battery low, warning, confirmation, incoming message etc.). These provide users with multiple options for describing haptic sensations and selecting those that are most appropriate for them. Voodle [66], another design tool as well as an interaction technique, enables user customization of haptic feedback in the form of the motion of a 1-degree-of-freedom robot. This work explored the use of iconic vocalizations such as those found in onomatopoeic words such as boom, woof, and ding instead of more complex phrases or affective terms. However, the use of haptic vocabularies is still not perfectly understood, and there is a lack of consensus about the existence of a tactile language [38]. Moreover, the use of language for iteratively tuning haptic feedback has yet to be thoroughly explored for interactive applications such as video games.

2.4 LARGE-LANGUAGE MODELS AND UI TRANSDUCERS

Recent advances in the domain of natural language processing (NLP) and more broadly deep learning, have resulted in the introduction of LLMs, probabilistic models of language built using decoder-only transformer artificial neural networks [107]. LLMs are capable of general-purpose language understanding and text generation, and are hypothesized to possess emergent knowledge about language relevant to various domains as well as world models based on the vast amounts of text they are trained on as well as their scale [11, 101], even though this is not universally agreed upon and is yet to be fully understood [89]. As a consequence of their data sources, they also possess biases and inaccuracies [40, 41, 48, 53, 63, 68, 85] which can be somewhat accounted for using different techniques such as data augmentation [63, 83, 109, 116] , data filtering [12, 26, 74, 102], data generation [46, 99, 105] and even in-training mitigation approaches [41, 60, 75, 80, 85].

Despite their limitations, LLMs are widely used today for various applications, including conversational assistants [22, 32, 46, 58], text classification [8, 17, 82], text summarization [81, 100, 114], math [117] and computer program synthesis [7, 19, 21, 29, 37, 69]. Moreover, with techniques such as prompt engineering and associated support tools [6, 27, 111], LLMs have been used to perform increasingly complex tasks, such as role-playing in social simulacra [77], web design [119] and virtual reality (VR) scene generation [21].

In particular, the paradigm of using LLMs as transducers for user interfaces (i.e., UI Transducers) [10] has emerged, showing the potential of applying LLMs' powerful language capabilities to modalities beyond text input and output. For example, when acting as digital characters in social simulacra [77], LLM instances were provided text-based descriptions of the graphical simulacra environment along with descriptions of the characters they played, which they used as context to inform their actions. When being used for VR scene generation, a similar framework was constructed to enable both scene understanding and inter-LLM communication as part of the system architecture [21], which allowed the system to control the visuals and physics as well as sensor integration for scenes in the Unity3D game engine [106]. Moreover, LLMs are also being explored as potential tools for personalization, with techniques such as the use of misaligned responses [47] and smartphone sensor data [113] being introduced to enable personalization without having to retrain separate models for different users.

However, despite their promising capabilities and the flexibility of systems such as LLMR [21], LLMs have not been applied in the context of haptic experiences. This presents an opportunity to explore the use of LLMs (more specifically, conversational LLM systems) as a potential interaction technique for users to personalize their haptic experience. Such systems can not only be used to enable end-user-driven personalization, which has been shown to improve user experience in conversational interactions [84], but also as a probe into the use of language for describing and manipulating haptic effects.

3 TACTALK SYSTEM DESIGN

With added context about personalized haptic feedback, the interplay between haptics and language, and the potential for LLMs as UI transducers, we now describe the system design for TacTalk, a conversational assistant using the GPT-4o model for language processing. As discussed in Chapter 2, various approaches towards personalized haptic feedback exist, differing in the ways in which they present haptic effects to the user. However, more clarity is still needed in terms of how users want to personalize their haptic feedback. We motivate TacTalk as a probe towards understanding how people approach the personalization of haptic feedback through conversation, and demonstrate its functionality in the context of Forza Horizon 5 [2], a popular car-racing video game.

Racing games were decided as the primary context for our study due to the similarity between a driver talking to their race engineer over the radio, along with the potential for rich haptic sensations in car-racing that evolve on the fly. This includes revving the engine, accelerating and braking rapidly, driving on different surfaces, and losing traction while drifting. Moreover, the specific choice of Forza Horizon 5 was motivated by the game's accessibility to a wide range of audiences, easy-to-learn controls, and arcade-like physics engine which does not necessarily require the controller to mimic realistic physical sensations.

In order to provide high-fidelity haptic feedback to users, we use the DualSense PS5 controller [1] which offers triggers with force feedback capabilities that can mimic a wide variety of effects, such as drawing a bow and firing an arrow, shooting a gun, and accelerator/brake pedal effects. However, the DualSense controller's rich haptic feedback capabilities are not supported by Forza Horizon 5 by default. To work around this issue, we use the ForzaDSX mod [118] (shown in Figure 3.1) along with the DualSenseX mod application [24]. This allows us to render haptic effects on the DualSense controller's triggers using telemetry data exported by Forza Horizon 5. Similar methods have also been used previously to synthesize haptic feedback from telemetry data, in the context of custom controllers and haptic feedback devices [56].

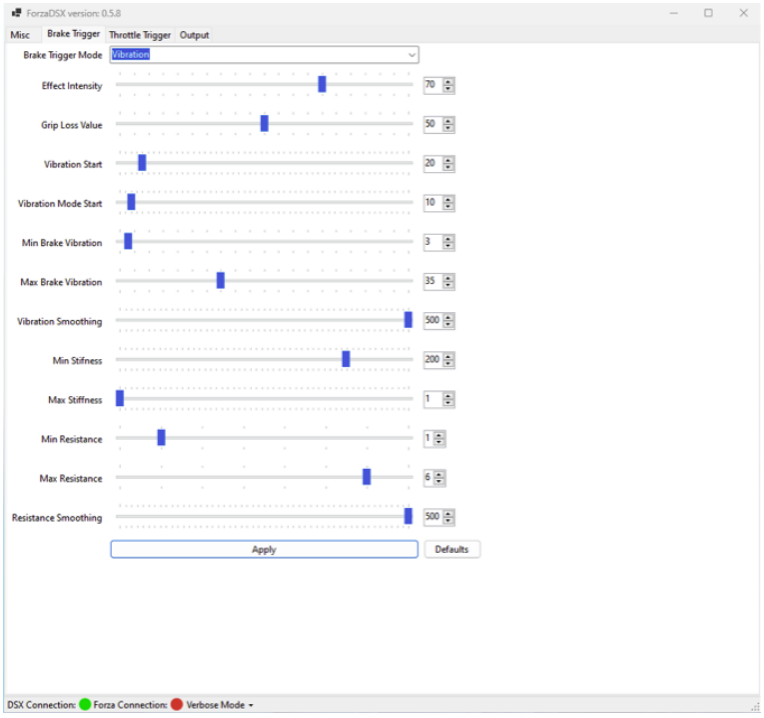
3.1 SYSTEM ARCHITECTURE

The overall system architecture is summarized in Figure 3.2. The four layers involved in the system are described in more detail below:

INPUT PROCESSING This layer deals with processing user voice inputs. The system's voice recording functions may be triggered using either the wake-word (or rather, wake-phrase) 'Hey, TacTalk!' or via a manual button-press. In our implementation of TacTalk, we use the Porcupine wake-word detection



(a)



(b)

Figure 3.1: The ForzaDSX mod - a visual interface enabling customized force feedback response using the DualSense controller's Adaptive Triggers. After the `config.xml` file is modified by the OUTPUT layer, a restart of ForzaDSX applies the changes to the DualSense controller.

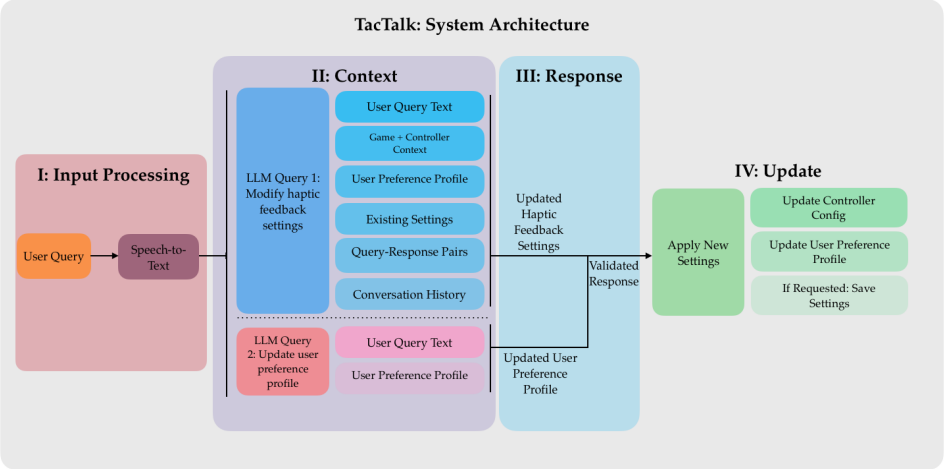


Figure 3.2: An illustration of TacTalk's system architecture, divided into four layers: INPUT PROCESSING, CONTEXT, RESPONSE and UPDATE.

engine [35] to trigger voice recording activation. Once the user has completed, a button-press stops the recording. Audio processing techniques may also be used for detecting silence, but was omitted in our implementation of TacTalk due to the nature of our study design (justified by the design of our study, discussed in Chapter 5). The recorded user query is then passed to a speech-to-text engine (we use OpenAI's Whisper API [72] in our implementation), which returns the query in plain-text format.

CONTEXT This layer involves constructing prompts to be provided as input to the LLM backend of TacTalk. The primary tasks that TacTalk is required to do are (1) change the haptic feedback configuration of the game controller and (2) maintain a profile of the user's preferences regarding haptic feedback. To achieve these tasks, four different text inputs are relevant:

- A meta-prompt string outlining the task for LLM queries, consisting of a 'You are an expert at...' task context description, the name of the video game being played, and the controller being used for gameplay. This information helps in preventing the system's justification from including hallucinated information.
- The user's existing profile, which helps improve the system's adherence to individual preferences by highlighting specific words and phrases from past interactions for the system to consider when making changes to settings.
- The existing controller setting values, along with example query-response pairs to improve the consistency of system-generated responses.
- A brief sliding window of the system's conversation history, to enable short-term memory.

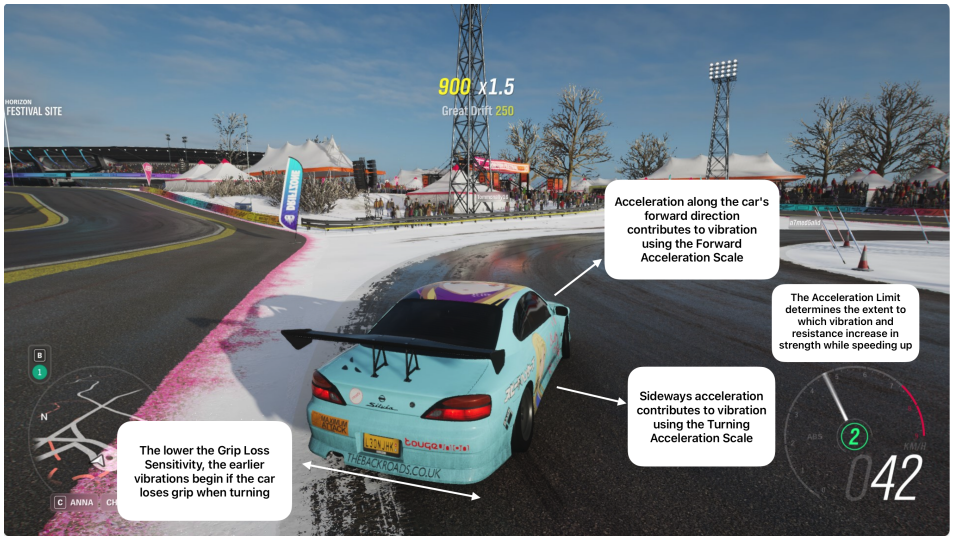


Figure 3.3: A screenshot from Forza Horizon 5, illustrating how the settings from ForzaDSX take effect during gameplay. Refer to Table 3.1 for more details about each setting.

The exact prompt strings used in our implementation can be found in Appendix A.

RESPONSE This layer is primarily handled by TacTalk's LLM backend. Once the appropriate prompts have been composed in the CONTEXT layer, two separate queries are made to the LLM backend. The first one is assigned the task of recommending changes to the controller's settings along with corresponding justifications for each change. The second one, on the other hand, is assigned the purely qualitative task of processing the existing user preference profile and updating it based on the LLM's response to the first query. Between the two prompts, the numeric data from the first query's response is validated to ensure that the system isn't capable of producing incorrect haptic feedback configurations, such as those including incompatible data types, out-of-range values and invalid minimum-maximum value pairs.

UPDATE The final layer deals with applying the response generated by the LLM to the relevant software layers while also setting up the system for subsequent user queries. In the case of our implementation built using ForzaDSX and DualSenseX, we modify the `config.xml` file for the mod and trigger an application restart to apply the changed settings on the fly. Moreover, we modify the prompt strings, namely the existing controller settings and the user profile to ensure continuity for subsequent interactions.

In the next two sections, we discuss our rationale behind the construction of the two LLM queries used as inputs for TacTalk's LLM backend - the controller setting modification query, and the user profile update query.

Parameter Name	Targeted Control	Acceptable Values	Significance
Trigger Mode	Throttle + Brake	OFF, RESISTANCE, VIBRATION	Set triggers to passive, render only uniform resistance or resistance and vibrations.
Effect Intensity	Throttle + Brake	0-100	The overall intensity of the rendered haptic feedback.
Grip Loss Sensitivity	Throttle + Brake	0-100	The threshold traction loss required to start vibrations on the triggers.
Forward/Turning Acceleration Scale	Throttle	0-100	The forward acceleration's contribution towards throttle vibration.
Acceleration Limit	Throttle	0-100	The proportion of acceleration beyond which the haptic feedback remains constant.
Minimum/maximum Vibration	Throttle + Brake	0-100	The minimum vibration frequency.
Minimum/maximum Resistance	Throttle + Brake	0-100	The minimum resistance magnitude.

Table 3.1: The list of parameters exposed by ForzaDSX for haptic feedback modification. All numeric values are percentages represented as unsigned integers.

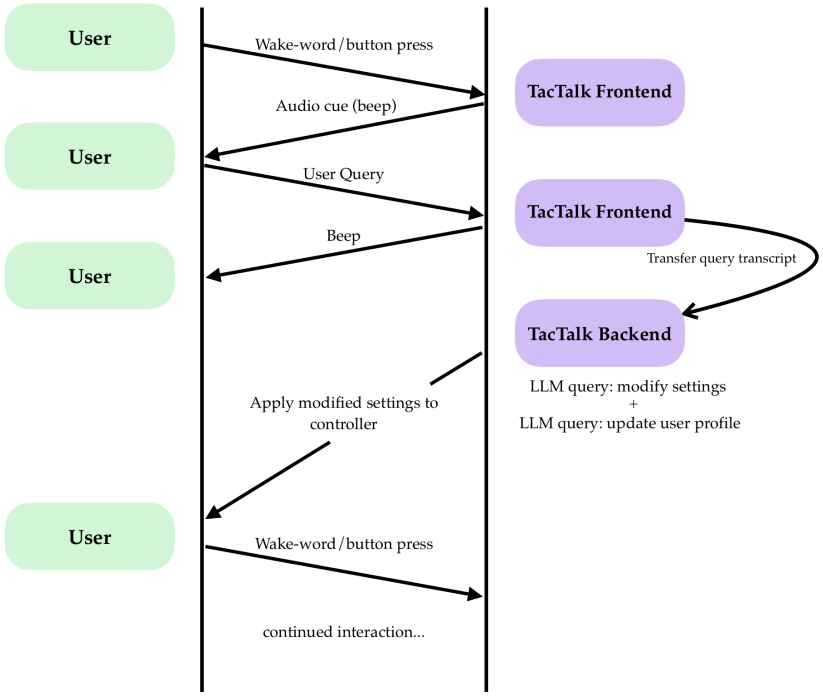


Figure 3.4: A process diagram illustrating one iteration of the interaction loop between a user and TacTalk.

3.2 DECIDING WHICH SETTINGS TO CHANGE

The complete list of settings that can be modified along with their role in haptic feedback rendering is explained in Table 3.1. A screenshot from Forza Horizon 5 is also shown in Figure 3.3, illustrating how the settings take effect in-game. The first query passed to TacTalk, derived from the user's transcribed instructions, includes a preliminary prompt to provide necessary context about the LLM's role. To begin, the prompt tells TacTalk that its role is to accept a user query along with a short conversation history, the current haptic feedback configuration, and a user profile summarizing the user's interactions with the system and their preferences for haptic feedback at different points in time. It then lists the different haptic feedback settings, along with acceptable ranges of values, and plain-text explanations for how each value influences the rendered haptic feedback. TacTalk is then instructed to return a JSON object, containing a list of parameters that need to be modified, along with their previous and new (modified) values, and a justification for the changes proposed.

It is well-known that LLMs tend to struggle with complex mathematical reasoning, although prompting techniques such as Chain-of-Thought prompting [110] and In-Context Learning [23] have shown to help with this to some extent. The TacTalk prompt includes some basic examples to help illustrate the role of different parameters, and the inclusion of a justification field further adds to the consistency and reasoning behind the modifications

made by the LLM. The system prompt also ensures that valid values are clearly specified for each of the exposed haptic feedback parameters to help avoid LLM responses with invalid values. Moreover, the LLM's temperature is fixed at 0 to minimize variations in responses to the same query; high temperatures were also found to lead to frequent JSON syntax errors. Further, these parameters are chosen to be more abstract than the low-level control attributes used by the DualSense controller at the hardware level. This means that instead of manipulating haptic rendering parameters for each frame (e.g., implementing a PID controller), TacTalk manipulates hyperparameters that influence the quality of haptic feedback in different situations such as slipping, acceleration, braking and collisions. For instance, the 'Grip Loss Sensitivity' parameter sets a fixed threshold for the amount of traction loss that needs to be achieved before the DualSense controller's triggers start vibrating, instead of dynamically evaluating whether or not to trigger grip loss vibrations at each time-step. This means that we compromise on the dynamic range of sensations that may be achieved in the game in favour of more consistent, stable performance. For example, users are unable to set different grip loss sensitivity values for different weather conditions or different road surfaces. Instead, they can only fix one global value, which only indirectly applies to different weather conditions and surfaces (e.g., during rain or when driving off road, the game's telemetry depicts higher slippage due to the game's physics engine which in turn leads to more frequent vibrations). Furthermore, due to Forza Horizon 5's DRM settings, a brief context switch occurs each time the haptic feedback configuration is updated, requiring users to tap the circle button to resume the game. This keeps TacTalk from being truly interruption-free, but was deemed an acceptable limitation for the purpose of our evalution.

We also perform a technical evaluation of TacTalk's performance, discussed in detail in Chapter 4. Performance in our context is poorly defined, since the aim of TacTalk is to cater to individual differences and mapping user queries to technical parameters may have different 'correct' approaches for different people. Thus, we attempt to understand the system's performance in terms of the validity of generated responses as well as the statistical consistency of the system's responses for different queries in the form of t-SNE dimensionality reduction. In general, we find that TacTalk produces no errors in terms of invalid parameter values, and that the system's responses were consistent based on 51 queries asking the system to produce the sensation of different types of cars. Finally, we empirically validate both TacTalk's parameter mapping consistency and ability to cater to individual differences by asking users to match different sensations to different types of vehicles, thus also learning whether or not different haptic configurations are perceivably different for users (see Chapter 5).

3.3 UPDATING THE USER PREFERENCE PROFILE

Once the system produces a response with new values for haptic feedback parameters, TacTalk's second phase of LLM processing performs an update to the system's long-term summary of user preferences. LLMs are useful for analyzing text for summarization and sentiment analysis, and thus this component was a logical step towards enabling the 'understanding' of users' haptic vocabularies. The summary produced by TacTalk is of interest for two primary reasons. First, it helps the system retain information about the user's preferences for haptic feedback beyond the 3-interaction history passed with each user query. If, for instance, a user's expectations from a query differ from TacTalk's initial prediction and they make a follow-up to rectify it, the user profile serves as a long term record to ensure the system remembers the difference for future interactions. Second, and more crucially for our studies, the user preference summary helps us understand the various terms (e.g., adjectives, onomatopoeia, metaphors etc.) that different people use to describe the haptic sensations they would like to experience. Thus, the preference profile serves as a supplement to the interviews conducted in our user study, providing crucial information about the language of user queries that the system is able to pick up on effectively. In future applications using similar architectures, the user profile may also be used to understand user sentiments regarding interactive experiences, or to infer common attributes across larger user bases.

3.4 INTERACTION LOOP

Now that we have described TacTalk's system architecture in detail, we move to discuss the overall interaction loop from the user's perspective. The process diagram in Figure 3.4 shows how information flows in a typical user interaction with TacTalk. Throughout their interaction with the system, users only see the TacTalk frontend - this is where users activate voice recording and make queries to the system. The frontend then transfers the transcript of the user's query to the backend. This is where most of the processing occurs; the user query is composed along with additional context and passed to the LLM in two queries. Finally, the responses from the LLM are validated, and the backend applies the changes to the game controller via ForzaDSX.

Figures 3.5 and 3.6 illustrate how a user might approach interactions with TacTalk. The basic interaction loop is simple - submit a query via voice commands, and wait for the system to apply changes to the game. Our implementation of TacTalk allows users to present different types of queries with the following intents:

- To explore haptic feedback settings freely, with no prior idea of the underlying system parameters.
- To modify specific haptic feedback parameters, such as vibration frequency or resistance intensity.

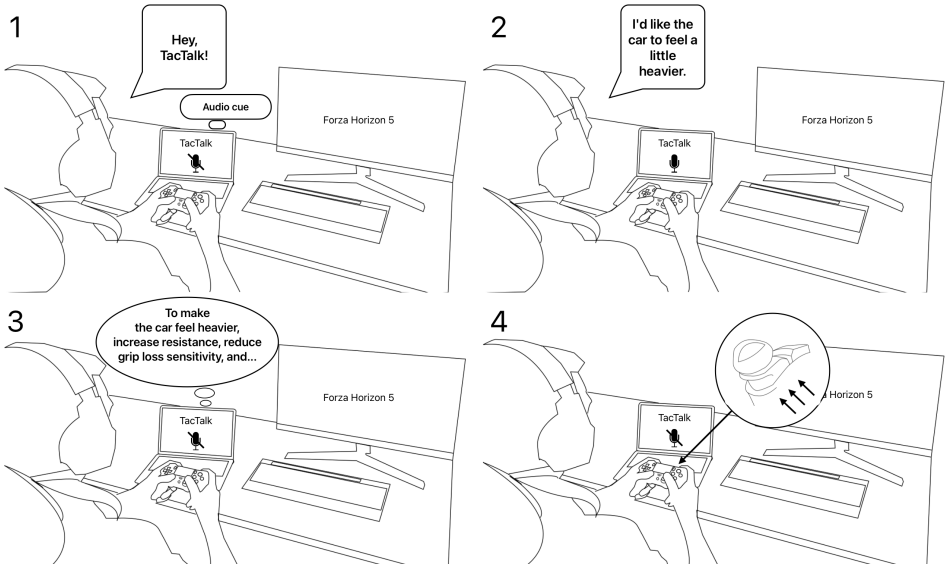


Figure 3.5: A storyboard illustrating how users may interact with TacTalk. Here, a user asks for the car to feel heavier, and TacTalk responds with a modified haptic feedback configuration that results in stronger resistance on the DualSense controller's triggers, mimicking a heavier vehicle's throttle.

- To undo changes and go back to a previous configuration.
- To save a configuration and possibly retrieve it at a later point in time.

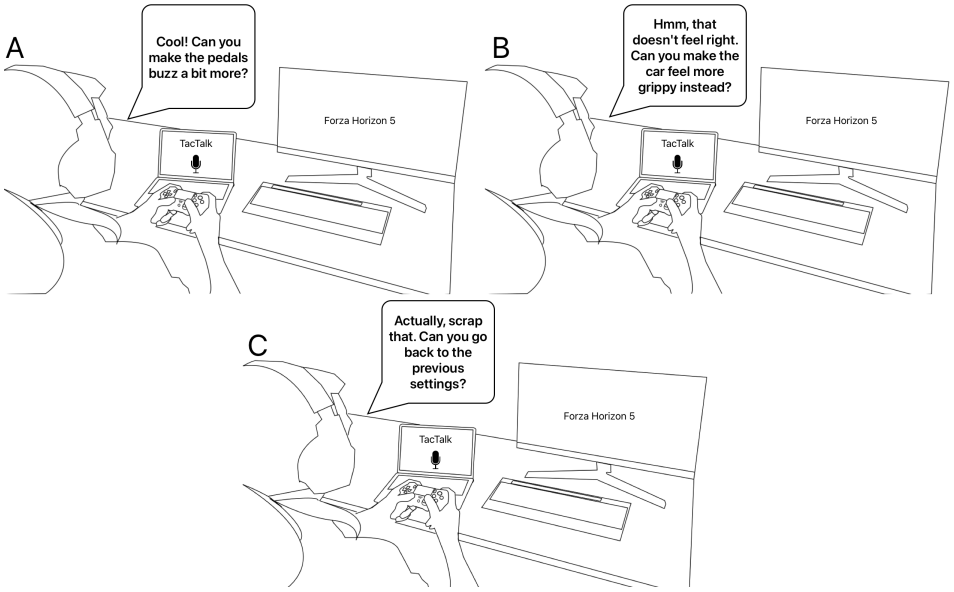


Figure 3.6: Here, based on the results of the earlier interaction, a user may choose to follow up with the system in one of three ways. If the response aligns with their preferences, they may only tune minor aspects of the experience, but if not, they may look towards other attributes to try and achieve a preferable setting. They may also choose to revert to the previous configuration.

4 TECHNICAL EVALUATION

In this chapter, we discuss the technical evaluation of TacTalk's performance. In particular, we examine TacTalk's ability to generate modified parameter configurations given a potential user's initial query. The task of generating an 'ideal' response for a given query is poorly defined, and the open-ended nature of haptic vocabularies and individual differences makes it difficult to objectively evaluate the system's response to a query like 'I want the car to feel more grippy'. As a result, for our technical evaluation we focus on two aspects - an error analysis of TacTalk's generated responses, and a t-SNE dimensionality reduction of the generated responses across multiple queries to examine TacTalk's statistical consistency.

4.1 PROMPT DATASET

For our evaluation, we constructed a small dataset of 51 user queries, each one of the form 'I want the car to feel like a...' followed by the name of different cars. We used this query structure to allow TacTalk to freely manipulate any parameters, as opposed to inducing bias with specific queries of the form 'I want the car to have less/more grip', 'I want the throttle to feel more/less stiff' etc. Based on each car's general usage (e.g., public transport vehicles, everyday vehicles, racing vehicles etc.), we divided the dataset into three groups - high performance, average performance and low-performance. Thus, we had a dataset with 17 cars belonging to each group. It is important to note that the classification of high/average/low performance cars is somewhat arbitrary and itself is based on existing bias in the public perception of vehicle performance. A garbage truck is termed a low performance vehicle whereas a Toyota Supra is termed a high performance vehicle, even though the former has a more powerful engine. Similarly, the Mini Cooper is classified as average performance, even though there are specific variants of the car that are used in track races. Nevertheless the classification, although imperfect, still provides an appropriate level of clarity to distinguish between cars and their performance levels. The complete dataset can be found in Appendix C.

4.2 ERROR ANALYSIS

4.2.1 *Methods*

To evaluate the system's error when generating haptic feedback configurations, we focus on the following criteria:

1. Whether or not TacTalk responded with valid JSON, containing all the required fields.

2. For categorical parameters, whether or not the system produces a valid value (i.e., whether or not the trigger mode is set to 'Off', 'Resistance' or 'Vibration').

3. For numeric parameters, whether or not the system produces valid values.

4. For associated minimum-maximum parameter pairs, whether or not the system produces consistent combinations (i.e., is the minimum value less than the maximum value?).

Furthermore, we compute error rates for these factors in the following two conditions:

1. Providing only a single LLM prompt to obtain both the modified haptic feedback settings as well as an updated user profile.

2. Providing two separate LLM prompts, one for haptic feedback settings and one for updating the user profile. In this condition, the prompt for haptic feedback settings includes example query-response pairs.

4.2.2 *Results*

We ran our tests 50 times on the 51-prompt dataset, for a total of 2550 queries to TacTalk. Additionally, we used the initial settings sheet as specified in Appendix B, which was empirically found to be a reasonable starting point - setting all parameters to just their maximum or minimum value was found to confuse the model, causing instances where settings were manipulated redundantly (e.g, increasing the resistance on the throttle would do nothing when the throttle effect intensity was set to 0). The results of our analysis are summarized in Figure 4.1.

4.3 t-SNE dimensionality reduction

4.3.1 *Methods*

In addition to evaluating how well TacTalk generates syntactically valid responses, we also conduct a statistical evaluation of the system's consistency. For this, we used the same dataset of 51 queries as detailed in Section 4.1, obtaining system-generated responses to each query. We then ran t-SNE dimensionality reduction [65] on the generated parameter vectors.

4.3.2 *Results*

The results of the t-SNE evaluation (learning rate = 10, perplexity = 20, random initialization) are shown in Figure 4.2. From the plot, we see that TacTalk-generated parameter configurations belong to relatively distinct clusters for high, average, and low performance vehicles, with little overlap between the

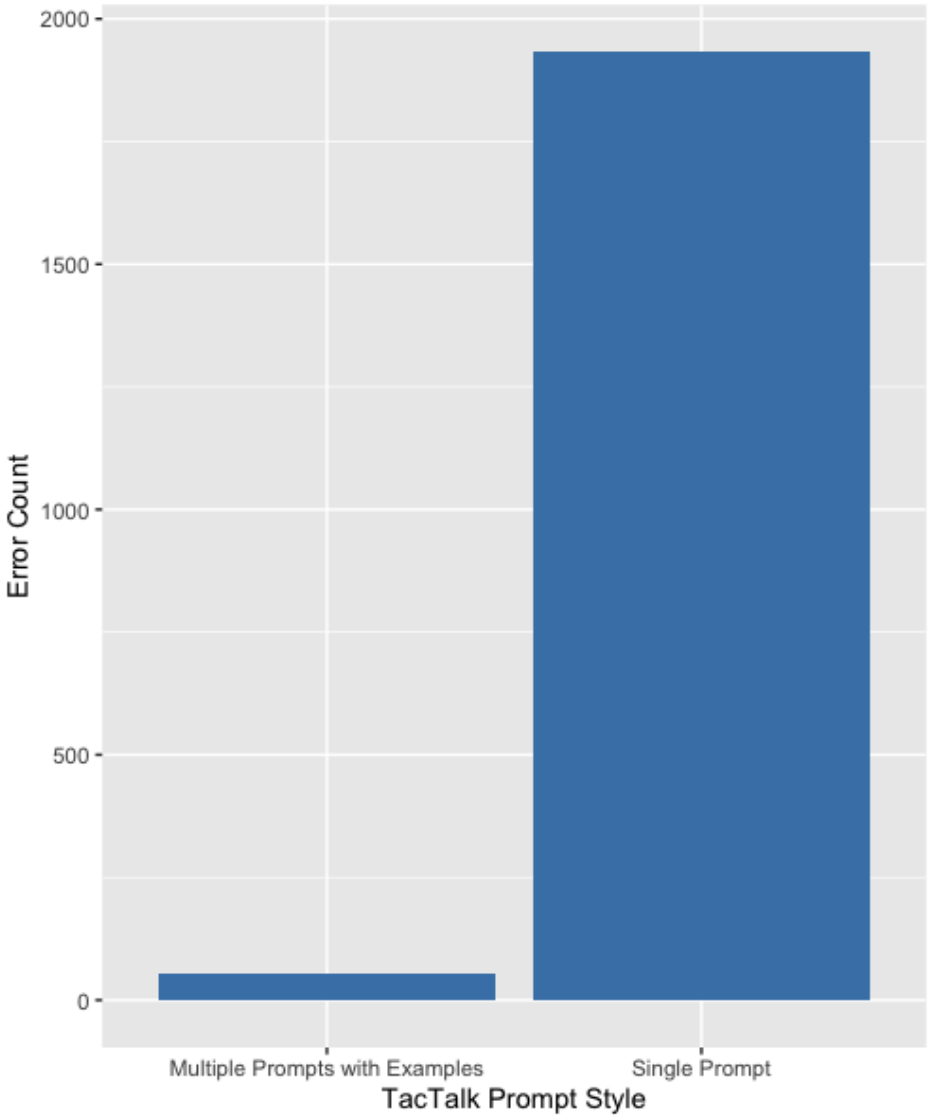


Figure 4.1: A summary of the number of errors encountered for each style of prompting. Using multiple prompts with example query-response pairs resulted in only 54 errors, as compared to 1932 errors when using a single prompt for all tasks.

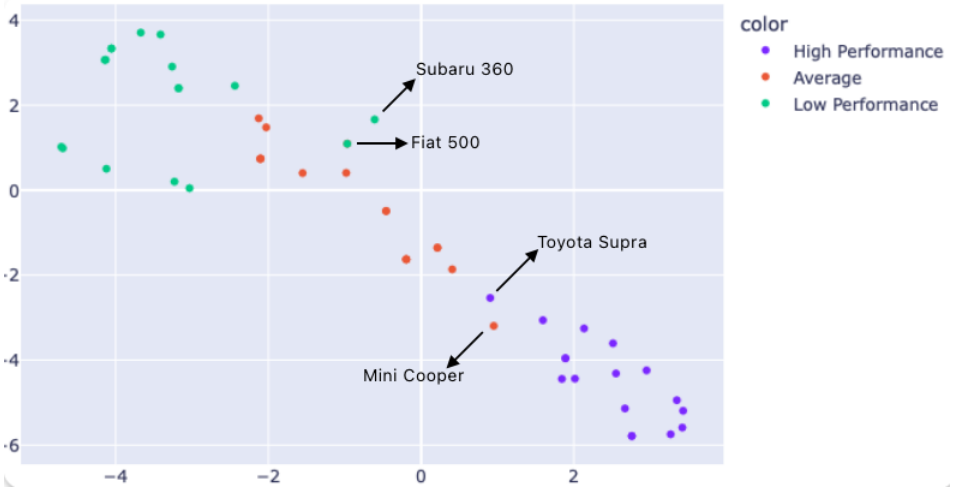


Figure 4.2: The t-SNE visualization obtained from the dataset of 51 user queries. The plot shows that the haptic feedback configurations generated by TacTalk possess some structure, with vehicles of similar performance clustered together. There are also some visible outliers, as highlighted.

average and low performance vehicle groups. It should be noted that the magnitude of distances does not hold much significance in t-SNE visualizations, and alternative dimensionality reduction techniques for non-linear data such as UMAP [67] have also been shown to provide to the same essential functionality as t-SNE, differing only in their initialization strategies [50, 51].

Overall, our statistical evaluation of TacTalk grounds the implementation of the interface by showing evidence that the system is capable of inferring implicit patterns from user queries and changes values based on these inferences, and not completely randomly. Having validated TacTalk's performance from a technical standpoint, we use TacTalk as a probe in an empirical study to understand how users approach conversational personalization of haptic feedback. In addition, we measure users' perceived cognitive load and usability ratings to compare TacTalk with a visual interface offering the same capabilities.

5 USER STUDY

In this chapter, we describe our methods for the empirical evaluation of TacTalk, primarily aiming to address the following research questions:

RQ1 Can an LLM effectively function as a UI transducer for mapping vague, natural language user queries to technical haptic feedback parameters?

RQ2 How do people think about haptic feedback personalization using language, without any assistance in the form of visualization?

RQ3 How does a voice-based personalization interface compare with a visual interface offering the same functionality in terms of usability and users' perceived cognitive load?

5.1 PARTICIPANTS

We recruited 11 participants (10 male, 1 female; self reported) for the main study, and 1 for a pilot (1 female; self reported). All participants had to answer a screening questionnaire, which focused primarily on their experience with driving and motorsport (see Appendix D for the complete questionnaire). For the purpose of our study, we exclusively included people who were experienced with some form of racing - real-world racing, virtual driving simulators or even arcade racing games (highly unrealistic games such as Mario Kart did not count towards racing experience). This exclusion criterion was enforced to make sure that participants (1) could get familiar with the video game relatively faster, and (2) would have an idea of their desired haptic experience. We ensured that the recruited participants had varying levels of experience with motorsport resulting in diverse yet domain-relevant vocabularies. Each session of the study took approximately one hour, and participants were compensated with Amazon gift cards worth CA$20 each for their time. Written informed consent was obtained from all participants prior to the sessions. The study was approved by the University of Waterloo Research Ethics Board (#46230).

5.2 APPARATUS

All study sessions were conducted indoors, with participants seated in front of a PC monitor where the game was displayed. The PC had an AMD ThreadRipper Pro 5975WX 32-core CPU, an RTX 4080 GPU and 64GB of RAM. A second laptop (an M1 MacBook Air) displaying the TacTalk voice interface application was placed on the left of the PC monitor. Finally, participants were given a pair of headphones and the DualSense controller to be able to

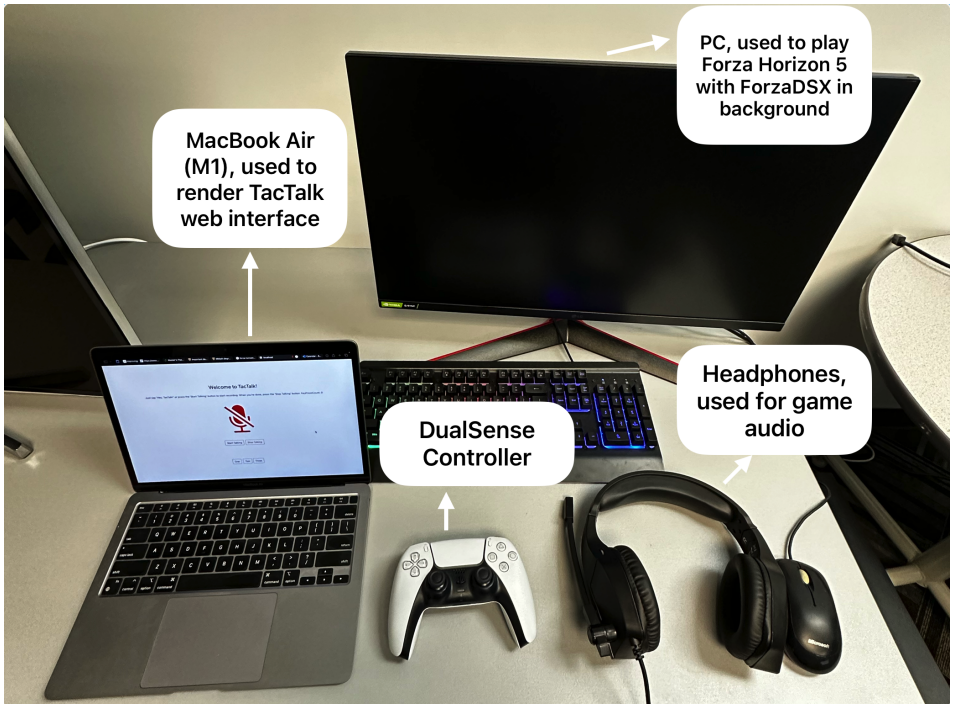


Figure 5.1: The apparatus used for the study, including a PC, a laptop, a pair of headphones, and the DualSense controller.

hear game audio and play the game. Figure 5.1 illustrates the setup used for the study, and Figure 5.2 shows a participant seated for the study. For the entire study, we decided to use the in-game bumper camera which hid the car from the player in order to remove visual bias for participants.

5.3 TASKS

Each study session consisted of two phases, which we refer to as FREE-PERSONALIZATION and CAR-CLASSIFICATION. Before the study commenced, participants filled out the Big Five Inventory and Player Traits questionnaires to provide information about their personality traits and preferences towards video games - this data added to our analysis of user preferences for the personalization of haptic feedback. The remaining stages of the study proceeded as follows:

- Participants were shown the capabilities of the DualSense controller using the DualSenseX mod application, which provides multiple preset examples of haptic effects on the triggers, including bow-and-arrow, automatic gun firing and uniform resistance. This provided context to participants who did not have experience using the DualSense controller prior to the study. This step took around 5 minutes.



Figure 5.2: A participant seated for a user study session.

- Participants were given 5-10 minutes to familiarize themselves with Forza Horizon 5, including the game's controls, the way cars handled in the game and the generic rumble haptic feedback effect that the game provides.
- The FREE-PERSONALIZATION phase was conducted. This step took around 30 minutes.
- The CAR-CLASSIFICATION phase was conducted. This step took around 15 minutes.
- Participants filled out the Big Five Inventory and Player Traits questionnaires. This took around 5 minutes.

5.3.1 FREE-PERSONALIZATION

In this phase of the study, participants were allowed to freely customize their in-game haptic feedback, first using TacTalk and then using the visual ForzaDSX interface, thus following a within-participants experiment design. This order was preserved across all trials so as to avoid inducing bias from participants learning the names of the technical parameters used in ForzaDSX. When using TacTalk, participants were told to describe the haptic feedback they would like to feel, and based on the system's response, would iterate until they arrived at a configuration that they felt comfortable with. Similarly, when using ForzaDSX, participants would switch to the mod's visual interface during gameplay and change the haptic feedback settings until they arrived at a configuration they felt comfortable with. In both cases, semi-structured interview questions were asked to identify users about their intent behind

queries/actions; this helped inform our inquiry into how people think about haptic feedback. Once users were done using each interface, they answered the NASA Task Load Index (NASA-TLX) [30] and System Usability Scale (SUS) [14] questionnaires before moving on to the next phase.

5.3.2 CAR-CLASSIFICATION

In the second phase of the study, participants were asked to try out three randomly-ordered haptic feedback configurations generated using TacTalk and guess which configuration corresponded to which vehicle from a list of options. The vehicles chosen for this task included a Formula One racing car, a double-decker bus and an everyday sedan. To ensure that participants weren't biased by visuals or audio from the game, a first-person camera was used and all car engine noise was disabled. Before providing their final answers, participants were allowed to freely switch between different configurations. Once they had provided their final responses, they were asked to verbally describe their expectations for haptic feedback for each vehicle. This description was then provided to TacTalk to provide user context, and the same task was repeated. We included this task to help understand our first research question, and more specifically the following related questions:

RQ1.1 Can TacTalk generate perceivably distinct haptic feedback settings?

RQ1.2 Are the settings generated by TacTalk consistent with user expectations when no additional context is provided?

RQ1.3 If TacTalk's responses do not align with user expectations, can the inclusion of user profiles correct this misalignment?

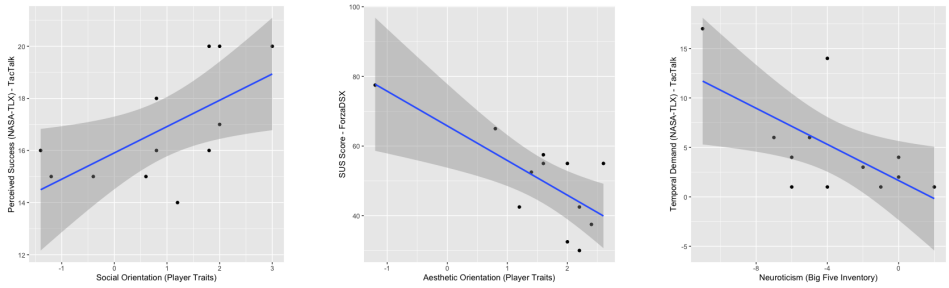
5.4 QUANTITATIVE RESULTS

We now present the results of our quantitative evaluations, namely the NASA-TLX and SUS scores from the FREE-PERSONALIZATION phase in addition to confusion matrices from the CAR-CLASSIFICATION phase and responses to the Big Five Inventory and Player Traits questionnaires.

5.4.1 FREE-PERSONALIZATION *Results*

NASA-TLX scores can be seen in Figure 5.4. Upon running the Wilcoxon Signed Rank Test to compare TLX scores, we find that TacTalk leads to significantly lower mental demand ($p = 0.003$), perceived success ($p = 0.043$), perceived effort ($p = 0.003$) and perceived frustration ($p = 0.012$) when compared to the ForzaDSX visual interface.

Users also rated TacTalk as a more usable interface than ForzaDSX - on running a paired t-test between the SUS scores for TacTalk ($\mu = 68, SD = 10$), and ForzaDSX ($\mu = 50, SD = 14$), significant differences are observed



(a) Social Orientation vs. Perceived Success

(b) Aesthetic Orientation vs. Slider SUS Score

(c) Neuroticism vs. Voice Temporal Demand

Figure 5.3: Scatterplots visualizing Player Traits and Big Five Inventory dimensions and their relationship with NASA-TLX and SUS responses.

($p < 0.001$). We also computed the Cohen's d and observed a large effect size ($d = 1.44$).

Computing the Pearson Correlation Coefficient for responses to the Big Five Inventory and Player Traits questionnaires, we found a statistically significant positive correlation between users' Social Orientation and their Perceived Success when using TacTalk (ρ = 0.63, $p = 0.028$). Additionally, a significant negative correlation was observed between the users' Aesthetic Orientation and their SUS scores for the ForzaDSX interface (ρ = -0.73, $p = 0.006$). Finally, a significant negative correlation was also observed between the users' Neuroticism score within the Big Five Inventory and their Temporal Demand scores from the NASA-TLX (ρ = -0.63, $p = 0.027$). These findings are visualized in Figure 5.3. No other significant correlations were found when comparing the Player Traits and Big Five Inventory dimensions with NASA-TLX and SUS responses. An α value of 0.05 was used for all statistical tests.

5.4.2 CAR-CLASSIFICATION *Results*

For the CAR-CLASSIFICATION task, we see that participants were able to achieve a slightly higher classification accuracy when they provided additional context about their preferences to the system. Interestingly, participants performed better than expected even when the system had no additional user preference context, indicating that the LLM's internal 'world model' based on its training dataset was still able to capture signals related to haptic feedback attributes for different vehicles. Thus, our proxy for TacTalk's alignment indicates that although additional context is helpful, the model is capable of aligning with some users' preferences by default.

Overall, with no user context provided, a classification accuracy of 63.89% was achieved (95% CI = (46.22%, 79.18%)). The class-wise sensitivity values observed were: 'bus' - 0.667, 'race-car' - 0.667, 'sedan' - 0.583; the specificity values observed were: 'bus' - 0.833, 'race-car' - 0.833, 'sedan' - 0.792. When

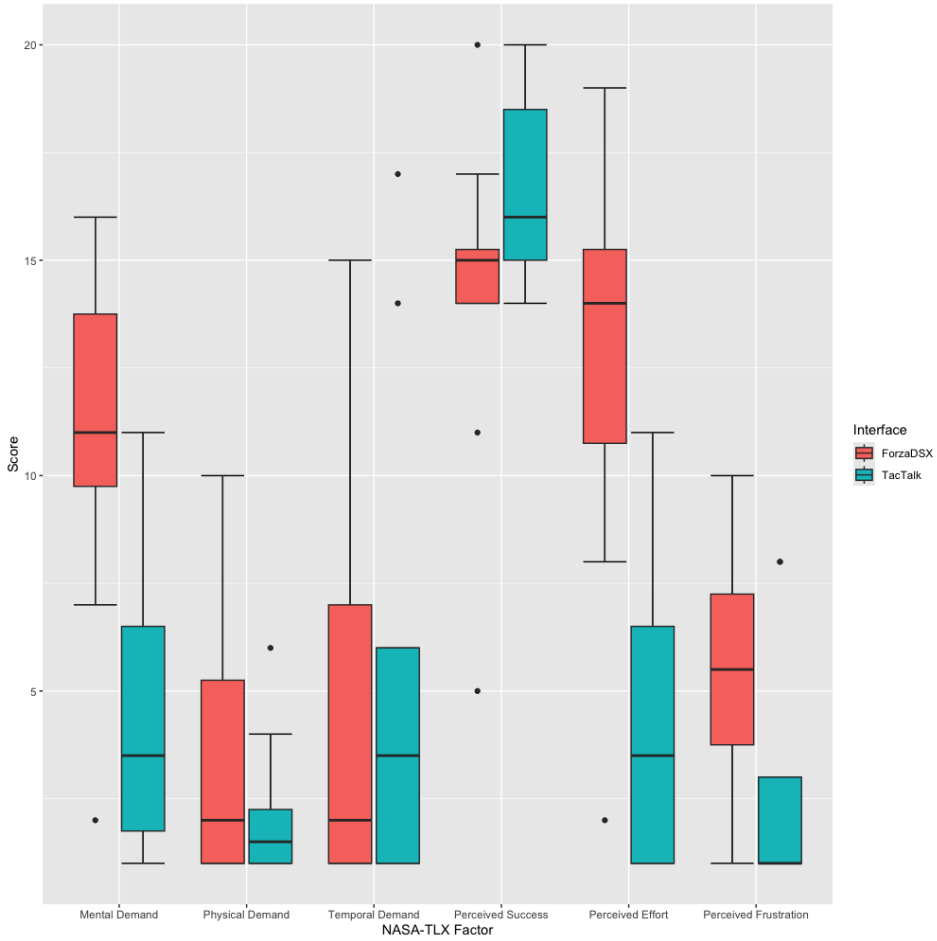
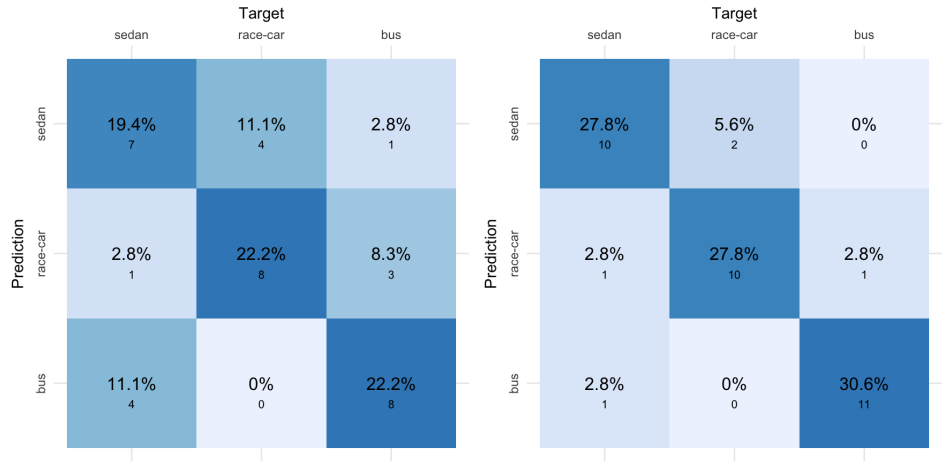


Figure 5.4: Box plots illustrating the NASA-TLX scores for TacTalk and the ForzaDSX visual interface.



(a) Without preference context

(b) With preference context

Figure 5.5: Confusion matrices visualizing users' classification accuracy when matching haptic feedback to vehicles.

context was provided, participants achieved an overall accuracy of 86.11% (95% CI = (70.5%, 95.33%)). The class-wise sensitivity values observed were: 'bus' - 0.917, 'race-car' - 0.833, 'sedan' - 0.833; the specificity values observed were: 'bus' - 0.958, 'race-car' - 0.917, 'sedan' - 0.917.

From our quantitative results, namely the SUS and TLX responses, we see that overall, users perceived significantly lower cognitive load when interacting with TacTalk as opposed to when they were using its visual counterpart, ForzaDSX. In particular, although users perceived similar physical and temporal demands for both interfaces, likely due to the fact that they had plenty of time to think about their queries before they actually talked to the system, and the actual interactions involved similar amounts of physical activity, they perceived a lower mental demand when using TacTalk. As further justified through the users' interview responses, this was primarily due to the fact that users were free to describe their desired haptic feedback in any way they liked, as opposed to when they were using ForzaDSX and had to map their internal vocabulary to the parameters listed in the visual interface. Further, the overall perception of success was significantly higher for TacTalk than for ForzaDSX, and correspondingly the perceived effort and frustration was significantly lower for TacTalk than ForzaDSX. This was interesting, because the TacTalk interface did not respond with any confirmation or natural language beyond a simple audio cue, whereas ForzaDSX provided users with visual confirmation of their actions. However, the reason for this may have been due to the fact that people found it harder to understand the causal relationship between different parameters and the haptic feedback rendered on the controller. Notably, participants always started the study by interacting with TacTalk - we did not use a counterbalanced order so as to avoid biasing

participants' vocabulary. This may also have contributed to the higher perceived frustration and effort for ForzaDSX. Finally, the usability scores for TacTalk and ForzaDSX also had significant differences, implying that TacTalk was a more usable interface than ForzaDSX when referring to the SUS.

5.5 QUALITATIVE RESULTS

For qualitative analysis, we collected audio recordings of participants' study sessions, including interview responses and interactions with TacTalk and ForzaDSX. Video recordings were also collected for consenting participants. In total, we analysed 353 minutes of audio and 132 minutes of video.

We followed the Reflexive Thematic Analysis methodology as described by Braun and Clarke [13]. In particular, we adopted an inductive style of analysis, allowing codes to emerge from the dataset. Our dataset consisted of audio recordings of participant interviews, video recordings of participants who provided consent for us to do so and interaction logs from the TacTalk system. The audio data was transcribed using Microsoft Word's transcription tool and subsequently cleaned manually.

After conducting our analysis, we identified the following primary themes:

5.5.1 *Users favor a few-shot approach towards personalizing haptics through conversation*

From the recorded data, we saw that haptic feedback personalization through TacTalk was typically a few-shot process, with only one participant (P3) posing more than five queries. On average, participants posed 3.5 queries to the system, with four participants (P0, P1, P6, P9) only making one query. When asked whether they felt content with the haptic feedback generated by TacTalk, P0 said *'I think I'm very content... I'm pretty accurate.'* This was interesting, since P0 thought that TacTalk's ability to generate a desirable haptic feedback configuration was at least partially due to their own ability to phrase what they expected from the system. Their query was long but directly referred to the throttle and brake pedals: *'I need the accelerator to be a little more firm, so less sensitive. I also need it to feel a lot like I am accelerating, so more rumble as I accelerate and as I go faster, as I hit the accelerator harder, I want it to feel a lot more rough. And then what, For the brake, I do want it to be a little less sensitive and I guess brake a little more smoother, so I don't want to feel as many vibrations, unless I'm hitting the handbrake.'* Similarly, when asked if they would like to make any further changes using TacTalk, P9 said *'No, I think this is comfortable.'* However, P9's query made use of a metaphor instead of directly addressing haptic feedback, since they asked for TacTalk to *'get haptic feedback that is most similar to a Toyota GR86.'* This shows that although different users may have different initial ideas about their desired haptic feedback, TacTalk is able to navigate the haptic feedback parameter space efficiently, requiring few interactions before reaching a setting that suits users' preferences.

In addition to the interaction duration, we also saw trends in participants' use of adjectives. The adjectives used referred to various qualities of both vibrotactile and force feedback including *smoothness* (*'Make the accelerator vibration smoother'*), *hardness/softness* (*'Make accelerator vibrations softer'*, *'Increase the brake vibrations to be harder than accelerator vibrations'*) and *stiffness* (*'Make the throttle less stiff'*), which aligns with the perceptual dimensions discussed in prior work by Okamoto et al. [71] and Hollins et al.[33] excluding the coldness-warmness dimension, since the DualSense controller does not have thermal feedback capabilities.

When asked for feedback on the implementation of TacTalk, the idea of having a physical button on the controller to activate TacTalk was favoured over voice-based activation by all participants except two - P5 (*'I would prefer if there is an inbuilt voice assistant and I can say "Hey!", just how we speak to Siri or Alexa.'*) and P6. Participants even suggested using a combination of the two interfaces in different ways. For instance, P1 suggested initially setting up haptic feedback using a slider-based interface like ForzaDSX and using TacTalk for subsequent fine-tuning of vibrations *'because it's just a little bit changed, compared to the beginning.'* P2 made a similar point, citing their prior experience with racing simulator games and real-life motorsport: *'... so basically, I can set up the parameters roughly before I start. Then I'd talk to the voice interface to adjust more precisely.'* In contrast, other participants (P4, P9) felt that it would be easier to start out by using TacTalk to explore the haptic parameter space using more generic queries and subsequently make smaller adjustments using ForzaDSX. P5 even recommended using a combination of TacTalk and ForzaDSX where users may set custom words as slider parameters and manipulate them to explore different sensations. They suggested using this as a teaching tool for children to demonstrate how different (potentially dangerous in real life) haptic sensations feel while introducing new language. These overall trends in participants' usage of TacTalk and their suggestions helped us observe an archetypal process of conversational personalization. This process is unidirectional, has multiple points of entry and goes from users describing their intended haptic feedback using more abstract metaphors (e.g., P9's usage of a Toyota GR86) to specifying minute details (e.g., P3's query to make both controller triggers equal in stiffness and vibration characteristics). We saw that more experienced participants often started this process further downstream, making less use of abstract metaphors and using more specific driving-related vocabulary to describe their desired haptic feedback.

5.5.2 *A realistic world suggests using knowledge of real-world haptics*

Forza Horizon 5 was described as a game with realistic visuals by multiple participants (P4, P5, P9, P10). This perceived realism of the game's world in turn led to participants drawing from their real-world knowledge of haptics. Depending on their individual experiences, participants referred to both real-world driving and driving simulator experiences when talking about their haptic feedback preferences. P10, for instance, said *'I think it's more like what*

I have felt when I was in a racing simulator... in the go-kart, that's the kind of the sensation you have accelerating' when talking about their preference for haptics on the throttle trigger. Similarly, some participants (P0, P2, P5, P11) alluded to real-world driving experience, either from go-karting (P2) or day-to-day driving (P0, P5, P11) when describing the haptic feedback they expected from the controller.

In addition to serving as an aid for querying TacTalk, real-world experience sometimes proved to confound participants; one participant (P9) said *'I don't know how much haptic feedback I would expect from an accelerator... the point of the haptic feedback is to just know more about the car... even in a real car, the accelerator just goes down or up, right? It's just controlling the throttle... I don't know what I expect from the accelerator. I would expect different things from the brake or the handbrake or the steering wheel like maybe in the joystick in this case.'*In particular, P9 pointed out the disparity in form-factor between the DualSense controller and a real car setup including a steering wheel and pedals as the primary reason behind their confusion. In this case, P9 resorted to asking for haptic feedback similar to a *'Toyota GR86'* as we saw earlier. This ties back into the unidirectional nature of conversational personalization; starting from a lack of clarity about their desired feedback, P9 drew upon a realistic metaphor to explore the system's capabilities.

Just as real-world knowledge was used to describe preferences for haptic feedback, participants often used real-world analogies when describing the outputs generated by TacTalk. Of course, these included simple comments about the capabilities of the DualSense controller, such as when P4 mentioned that they would like force feedback on the analog sticks (*'I don't think PlayStation controllers have that feature yet'*) or when P2 said that the intensity of the force feedback was limited by the hardware (*'I would like to get some more feedback from both the brake and accelerator pedals, which may be due to the limitation of the controller'*). However, participants also brought up real-life driving experiences, such as the observation that the brake pedal in a real car is harder to press than the accelerator (P8).

An interesting observation from participants' use of real-world knowledge was the existence of individual perceptions of realism. We as authors recognize our own individual idea of realism may differ from the participants' opinions, which adds to the subjectivity of our analysis and is important to acknowledge before discussing our findings in more depth. While some participants (P6, P8) referred to their idea of a realistic driving experience (driving a muscle car, feeling both resistance and vibrations on the triggers), others referred to the in-game terrain they were driving on (P5, P7) and one even referred to their belief that real-world cars have brakes that are less stiff so as to stop faster (P1).

5.5.3 *Using haptic feedback to represent changing environment and state is preferred*

In line with participants' expectations being influenced by the realistic visuals of the game, the haptic feedback was also expected to follow real-world causal

relationships. This meant that participants expected certain haptic cues to correspond to specific in-game events. This behaviour was initially seen in participants' justifications for different haptic sensations - for instance, P4, P5, P9 and P10 stated that they preferred realistic haptic feedback for Forza Horizon 5 due to its visual fidelity. This was observed for both the vibrotactile sensations (*'there is some vibration - screeching tires, I think'*) and force feedback (*'when it lost traction this time, there's actually a lot more feedback on the brake'*, *'I'm on a slope, I can feel some sort of resistance here'*) rendered on the controller.

Just as participants justified haptic feedback based on in-game events, their preferences for haptic feedback were also shaped around these events. Some participants (P5, P7, P11) paid attention to different environmental factors when describing their desired haptic feedback, including mud, snow, and rain into the conversation. P7 even commented on the lack of controls in ForzaDSX that could manipulate weather-related haptic effects, mentioning that they *'wanted to configure like how they* (i.e., the triggers) *should behave. Like for example in winter you need winter tires'*. In other cases (P5, P6, P7), track-related factors were mentioned, referring to the incline (*'I do feel some kind of resistance also because this is an incline'*) and on/off-road conditions (*'I want the accelerator to be smoother when I'm off road'*). Notably, none of the participants commented on the haptic feedback combined with audio - references were mainly drawn from visual in-game events. Even so, the sounds made by the controller when the triggers vibrated were noticed by participants, who either liked (P2, P3, P4) or disliked them (P5, P7, P10). In the case of P10, we even saw that the in-game telemetry was a valid artifact for participants to derive their preferences from (*'I want my index finger from the left to have a lot of shake when I'm accelerating, especially between when I start and until I go to 100km/h'*). All of these sub-themes combined inform the notion of haptic feedback that represents changes in the environment and state of an experience. Specifically, the visual modality is found to be favoured over audio in participants' framing of preferences for haptic feedback, involving both UI elements and actual in-game objects. Thus, haptic feedback should also be personalized using an approach that accounts for changes in the environment and state of the experience.

5.5.4 *Knowing about 'reality' does not imply a preference for it*

Although we saw that all participants drew from real-world knowledge when both describing their preferences and justifying TacTalk's outputs, this did not always mean a preference for realistic haptic feedback. As seen in P7 and P8's interactions, real-world experiences could also be used as a reference to look for factors causing frustration, such as increased throttle stiffness when driving off-road, intense vibrations indicating grip loss and increased resistance when driving up an incline. Similarly, P3 distinguished their haptic feedback preferences from real life, mentioning that they were looking for something that *'feels good for the game'*. P5, after narrowing down on their preferred haptic feedback configuration was then intrigued by TacTalk's capabilities and attempted to experiment with different, more unrealistic

sensations, asking for the accelerator to feel *'like a spring'* and *'as light as air.'* These two queries were the only observed instances where a participant made a query to TacTalk that directly referred to unrealistic metaphors for driving. Even so, this indicates that the open-ended nature of voice-based interactions might invite users to play around with different haptic feedback configurations, irrespective of whether they are easy to define in terms of technical parameters or not. Further, this indicates that at a global scale, users have a preference for a large, varied haptic parameter space, allowing for both realistic and unrealistic haptic feedback.

5.5.5 *Domain-specific metaphors are a valuable tool*

Describing haptic feedback through language is often difficult, and a similar trend was seen in the participants' interactions with TacTalk. Five participants (P1, P3, P4, P8 and P9) indicated that they had difficulty thinking of queries to pose to the system due to various reasons, from not knowing what language TacTalk would be able to understand to not having a clear sense of the haptic feedback they wanted. In such circumstances, metaphors proved useful. Just like how P5 used *'spring'* and *'air'* as metaphors for the behaviour of the accelerator, other participants used metaphors to describe different scenarios. P6 asked TacTalk to replicate the sensation of driving *'a sturdy muscle car'*, stating that they wanted to experience something similar to an old Mustang. P9 similarly used the metaphor of a Toyota GR86 to obtain an initial haptic feedback setting from TacTalk to subsequently fine-tune from there (although they ended up feeling satisfied with the initial configuration).

Other notable cases of metaphor usage included P3 and P9's descriptions of the haptic feedback they felt - they used onomatopoeic vocabulary, such as *'thuk-thuk-thuk-thuk'*, *'buzzy'*, and *'dhuk-dhuk-dhuk-dhuk'*. However, such sounds were not used when talking to TacTalk, but rather when responding to interview questions. In fact, P3, P8, and P9 explicitly pointed out not being able to ascertain whether or not TacTalk would be able to pick up on such descriptions. Thus, we found that metaphors were valuable to participants when it came to describing their expectations, even if they did not feel that a computer would be able to understand their intention. Participants that did not make use of metaphors instead focused on basic haptic adjectives, such as stiffness and vibration and directly asked for them to be modified, as discussed earlier in T1.

Overall, we found that domain-related metaphors were brought up by participants most frequently, and more generic metaphors such as those used by P5 were rare. In fact, P5's description of making a trigger feel like 'air' or like a 'spring' were the only instances were a domain-unrelated metaphor was used when interacting with TacTalk. Moreover, when participants were prompted to describe how they came up with their queries, all participants except for P7 brought up metaphors related to cars (P7 instead mentioned the experience of riding a bicycle uphill when describing their intended haptic feedback on inclined roads, which was still somewhat related to driving a

car). In this sense, when viewed in light of the prior themes exploring how participants drew upon real-world knowledge and ideas of realism when describing haptic feedback preferences, we see that this domain specificity stands out as a common element. Language unrelated to the task domain does not commonly show up in participant interactions, and thus conversational personalization interfaces should ensure participants can explore the design space using domain-specific terminology.

5.5.6 *Summary*

Overall, we see that users consistently make more use of generic metaphors in the early stages of personalization, and subsequently move towards low-level details until they achieve their intended configuration. Along the way, users draw upon real-world knowledge, even if they may not necessarily be looking for realistic haptic feedback - the realism of the game world plays a key role in influencing this. As mentioned earlier, in P5's case, once an intended effect is achieved, a user might wish to play around with the interface's capabilities further, but this is not part of the original personalization process anymore; it is simply an exploration of the interface *after* personalization has already been completed.

6 DISCUSSION

In this chapter, we discuss the implications of the results obtained from our user study, classified primarily into quantitative and qualitative findings based on whether the results were obtained through user interactions and interviews or through questionnaire responses.

6.1 LLMs can be used to make predictable and valid parameter adjustments for haptics

Through our technical evaluations, we see that an LLM-based application such as TacTalk can make valid parameter adjustments for haptic feedback. In particular, we notice that making use of a multi-prompt structure, breaking down the task of modifying haptic feedback and subsequently updating a summary of user preferences leads to valid results much more often than providing a single, monolithic query. This falls in line with prior evidence of how techniques such as Chain-of-Thought [110] and few-shot prompting [47] yield better LLM responses.

Through our t-SNE evaluation, we verify the predictability and consistency of TacTalk's generated responses. As observed using a small test dataset, TacTalk is able to generate configurations that fall in line with expectations related to vehicle names. Even though the classification between high, average and low performance vehicles only considered popular beliefs and could be somewhat arbitrary, a visible distinction between classes was observed using t-SNE. Even for the cases that TacTalk 'mis-classified', an argument can be made for TacTalk's classification may also be considered 'correct'. For instance, TacTalk misclassified the Mini Cooper, placing it closer to the high performance cluster than the Toyota Supra. However, the Mini Cooper is also involved in motorsport, which may have been a contributing factor in TacTalk's response. Similarly, a Fiat 500 and a Subaru 360 may be considered average performance cars by some, and not necessarily low performance. Our claims are further validated through our user study, where we see that most users only submitted fewer than 5 queries before finding a haptic feedback configuration that they enjoyed. Thus, we see that TacTalk's responses are predictable and consistent.

6.2 Conversational personalization is more usable than a visual interface

When considering the context of the study, i.e., customizing haptic feedback on the fly in a racing video game, users rated the usability of TacTalk significantly higher than ForzaDSX, the visual counterpart, as measured through the

SUS scores. Users noted that the ability to make changes without switching to another application on the PC was helpful, and that TacTalk was easier to use since it did not require users to learn any interface- or domain-specific language. This ease-of-use was further shown in the NASA-TLX responses, as users reported feeling a lower mental demand, perceived effort and perceived frustration, and higher perceived success when using TacTalk as compared to ForzaDSX. Thus, we infer that for real-time, interactive applications, using a different modality for personalizing haptic feedback may help reduce users' cognitive load and improve usability.

6.3 PERSONALITY DIFFERENCES MAY IMPACT PERSONALIZATION PREFERENCES

As seen in our correlation test results, there is some evidence that personality differences may interact with users' preferences when it comes to personalization.

The potential existence of a positive correlation between users' social orientation and their perceived success when using TacTalk may have to do with the fact that voice-based interaction is inherently perceived to be more social than a visual computer interface. Similarly, users who have a higher aesthetic orientation would not prefer the information overload seen in the ForzaDSX interface, thus justifying a negative correlation with SUS scores for ForzaDSX. Lastly, users with higher neuroticism scores perceived lower temporal load when using TacTalk. Although after applying Holm-Bonferroni corrections, no significant correlations were found, these initial signs of correlation may still be worth examining in future work, especially with a larger sample size.

6.4 PREFERENCE CONTEXT IMPROVES USERS' CLASSIFICATION ACCURACY, BUT NOT BY MUCH

One of the key features of TacTalk's system design was its use of a user preference profile to keep track of specific aspects of the haptic experience mentioned by the user, especially if they differed from TacTalk's base 'knowledge'. However, in practice, it did not significantly impact users' classification accuracy, as seen in the results of the CAR-CLASSIFICATION task. Although including user-specified features did improve classification performance, we found that users did not fare much worse when the system did not account for their preferences, indicating that when performing relative comparisons, TacTalk's consistency in generating haptic feedback configurations can also be perceived by users moderately well, even if the settings do not align with their preferences.

6.5 CONVERSATIONAL INTERFACES SHOULD USE SIGNPOSTING TO MAKE USERS' LIVES EASIER

As we found in our thematic analysis, when interacting with a realistic world, users often look to their real-world knowledge to inform their preferences for haptic feedback. However, this may result in confusion when trying to compare across different haptic modalities. For example, a DualSense controller does not render force feedback on the joysticks, even though users may expect a centering force on the steering wheel while driving a car. In the case of visual personalization interfaces, we find that using example libraries can be used to prompt creative user-driven exploration of haptic feedback, as seen in the RecHap [103] and VibViz [96] tools.

In the context of voice-based interfaces, it is known that interaction possibilities range from the use of natural language to humming and other, more vague sounds [28, 61]. Voice-based interfaces for haptic feedback, such as Voodle [66], also offer users the ability to communicate intended haptic feedback using onomatopoeic phrases. However, as opposed to visual interfaces, voice-based approaches, especially when deployed in contexts where users typically perform another primary task (e.g., driving) do not rely on visual cues to prompt exploration. Moreover, people may not know the full capabilities of voice-based systems unless explicitly shown examples. Systems like the Google Assistant achieve this through both advertising and allowing users to ask the system what it can do. It is important in such cases, however, to make sure that the system's responses create realistic expectations - telling the user *'I can change your haptic experience'* is broad and vague, whereas simply adding examples such as *'I can make your brake pedal shake a bit less if that feels better'* or *'I can make the car feel less sluggish when you drive on sand'* grounds the system's abilities [78]. Another way to show users system capabilities may be to use a visual tutorial when first starting up the application. For TacTalk, we did not use any tutorials, since the goal of the study was to understand users' approach towards customization without introducing any biases. Even so, we saw that two users looked to the system for suggestions, although their queries were direct requests to make the controller feel like specified cars. Another user also attempted to explore the system's capabilities using different types of vocabulary and metaphors like *'light as air'* and *'like a spring'*. However, we found that most users did not naturally think of asking the system for recommendations. As a result, it is important for conversational interfaces to employ signposting effectively

6.6 INTERFACES ENABLING PERSONALIZATION OF HAPTIC FEEDBACK SHOULD EMPLOY NARRATIVE FRAMING EFFECTIVELY

In the previous subsection, we discussed the need for signposting when it comes to helping users learn about an interface's personalization capabilities in the context of the hardware being used. In general, with conversational interfaces, we find that users tend to trust the system more when voice-based

feedback is also included in the system [16]. This may be used as another way to nudge the user towards exploring system capabilties. However, the narrative of the virtual experience is a key consideration. As mentioned in Section 5.5, users employ metaphors when describing their desired haptic feedback. However, domain-specific metaphors are much more commonly used than generic vocabulary. Thus, an interface enabling personalization should focus on domain-specific narrative framing, as it can signficantly impact how users think about in-game events and their corresponding haptic feedback [15]. In the context of a racing game, for instance, it would make sense for a conversational interface acting as a race engineer, talking about in-game environmental factors when describing changes made to the haptic feedback. Responding with 'I've tried increasing your grip loss sensitivity, try using the e-brake and you should notice the throttle feel more slippery' not only describes the changes effected by the system, but also frames the explanation in a manner that the user can easily understand and even test out (here, by using the e-brake).

6.7 CONVERSATIONAL INTERFACES CAN BE A VALUABLE LEARNING TOOL

When dealing with specific domains such as motorsport, vocabulary is an important tool. For example, picking up marbles, trail-braking and aquaplaning are terms that people familiar with racing may know, but newcomers are unlikely to have heard outside of motorsport, similar to how sommeliers use specialized vocabulary when describing wines. Prior work has explored the use of conversational interfaces for education [49], especially learning by teaching such interfaces [54]. However, we may also use these interfaces to actively teach users and help them familiarize themselves with different aspects of the experience:

- How haptic feedback is rendered on their specific device
- How haptic feedback relates to in-game events and environmental factors

For example, if a user feels unsure when describing their desired haptic feedback, a conversational interface can help them learn about the various factors involved behind haptic rendering in the game (such as frequency, vibration, intensity etc.) and subsequently help them hone in on the specific parameters they would like to change. This 'teaching' may be direct, with the interface directly referring to the physical controller being used and its triggers, motors, buttons etc. However, it can also be tied in with the idea of narrative framing - a conversational interface may use purely narrative-related language to communicate with the user, making it easier for them to focus on the virtual experience without having to switch contexts when thinking about personalization. This may have a positive effect on user experience, or more specifically, user involvement/immersion.

6.8 CONVERSATIONAL PERSONALIZATION IS A UNIDIRECTIONAL PROCESS

As mentioned in Chapter 2, prior work on HX has brought up the importance of personalization, but the way in which end-users approach personalization is still poorly understood. Seifi [93] defined different approaches towards enabling end-user customization, namely choosing, tuning and chaining, but even this work does not study how users think about personalization, especially in a conversational context. In our study, we found that personalization takes the form of a unidirectional process. This process does not have a fixed starting point or end point, since different users may begin with different notions of how in-depth they want to be and may choose to end early if they achieve their preferred configuration. However, the direction in which user interactions flowed indicated that users consistently follow a top-down approach towards haptic feedback customization, despite the affordance of a bottom-up approach enabled by TacTalk.

More specifically, by top-down, we mean that personalization often commences from a relatively generic idea of the intended experience for a user. This is where metaphors are most valuable. In such situations, users often expect example libraries or other suggestion mechanisms to better understand exactly what they want. In the use-case of TacTalk, this took the form of asking TacTalk to mimic an existing vehicle or other generic metaphor. Once this starting point is achieved, we see that users tend to zoom in on specific aspects of the settings they have, and attempt to fine-tune them. In our study, this includes remarks on how the throttle and brake pedals (triggers) feel different and what changes need to be introduced. This process is iterative, and is often where users spend the most time. However, in our analysis we found that this process was also unidirectional, with users starting from a sensation and making subsequent changes to it with reverting to previous settings. This may be attributed to the lack of visual elements indicating haptic feedback settings in a conversational interface, but despite being informed that they could ask the system to go back to a previous setting, none of the participants presented such a query. It is important to note that this phenomenon differs from query-repair, which is common among users of voice interfaces [112].

7 CONCLUSION

In this thesis, we conducted an inquiry into the personalization of haptics through conversation. We identified that personalization is not a well-understood part of HX, despite the existence of frameworks describing effective strategies for creating personalization tools (Chapter 2). We presented a conversational interface, TacTalk, that can be used as a voice assistant to customize haptic feedback on the fly (Chapter 3). This interface was then evaluated both technically and empirically.

Our technical evaluations (Chapter 4) show that TacTalk is capable of producing consistent mappings of haptic feedback configurations, with t-SNE visualizations showing separation between cars of different performance levels and an error analysis showing that TacTalk's use of examples and multiple prompts to break down the system's functions result in better performance than when using a single longer meta-prompt.

In addition, our empirical evaluation (Chapter 5) shows that when compared to a visual interface, ForzaDSX, TacTalk has higher usability ratings (SUS scores) and lower perceived cognitive load ratings (NASA TLX). Further, we present a preliminary thematic analysis of participants' interview responses and interactions with TacTalk, describing how users approach the task of personalization, specifically through purely voice-based interaction. We find that conversational personalization functions as a unidirectional process, with users initially making use of broad metaphors before identifying specific aspects of haptic feedback that need to be manipulated. We also see that although a realistic virtual world leads to users drawing upon real-world knowledge of haptics, realism does not always align with user preferences. In all of this, metaphors emerge as a powerful tool used in lieu of other, more effective ways to communicate haptic feedback preferences when communicating purely through conversation. Additionally, we see that task- or domain-specific metaphors are more commonly used as compared to generic, unrelated metaphors. In Chapter 6, we discuss our key findings and their potential implications for conversational interface design in the context of the personalization of haptic feedback.

Thus, our evaluations address the following research questions:

RQ1 Can an LLM effectively function as a UI transducer for mapping vague, natural language user queries to technical haptic feedback parameters? (Yes, it can!)

RQ2 How do people think about haptic feedback personalization using language, without any assistance in the form of visualization? (People rely on visual context of the application and their prior related experience, start from high-level, vague descriptions of intent and subsequently narrow down on specific desired attributes. There may also be evidence

that personality traits correlate with different users' approaches towards customization, but this needs to be further studied.)

RQ3 How does a voice-based personalization interface compare with a visual interface offering the same functionality in terms of usability and users' perceived cognitive load? (A voice-based interface is preferred over a visual alternative, with higher usability ratings and lower perceived cognitive load.)

7.1 FUTURE WORK

Future investigations into voice-based interactions for haptic feedback customization may provide insight into haptic feedback customization across different tasks and haptic devices, thus moving further towards a robust model of personalization while also evaluating how LLMs fare with changing contexts for both the application domain and haptic device being used. Multiple interaction modalities should also be investigated to better understand whether the unidirectional nature of conversational personalization can also be seen outside of a voice assistant.

The user study presented in this thesis only found initial evidence that user preferences towards personalization may also be related to individual personality traits. Future work may explore this concept with different measures of personality traits with larger sample sizes and different modalities for personalization.

TacTalk itself, being a modular, largely plug-and-play system, can be used in various other contexts. For example, it can be used in tandem with frameworks such as LLMR [21] to offer haptic feedback customization capabilities for virtual environment authoring tools. Video games such as F1 [25] already have voice-based interactions with race engineers to perform in-game tasks, such as obtaining race statistics and modifying race strategy; these systems can be augmented with abilities similar to TacTalk to improve their fidelity further - one can imagine a scenario where an F1 free practice session involves users constantly talking to their virtual race engineer to tweak haptic sensations alongside elements relevant to in-game physics, such as increasing downforce and subsequently experiencing higher frequency vibrations while driving on a winding road. Based on comments from our user study, we suggest using TacTalk as a supplement to a visual interface, where users can press a button for voice-based customization or pause and manually change their experience. Users who were experienced with motorsport said that they would like to start off by manually choosing a configuration and subsequently use TacTalk to fine-tune specific parameters, whereas users who did not have racing experience said that they would prefer using TacTalk for exploration (i.e., to learn what the controller can do), and if needed, use a visual interface to modify specific parameters.

BIBLIOGRAPHY

[1] Sony Interactive Entertainment LLC. *DualSense wireless controller | The innovative new controller for PS5*. en-CA. URL: https://www.playstation.com/en-ca/accessories/dualsense-wireless-controller/ (visited on 07/18/2024).

[2] Xbox Game Studios. *Forza Horizon 5: Play with Xbox Game Pass | Xbox*. en. URL: https://www.xbox.com/en-CA/games/forza-horizon-5 (visited on 07/18/2024).

[3] Juan Aguado-Delgado, José-María Gutiérrez-Martínez, José R. Hilera, Luis de Marcos, and Salvador Otón. "Accessibility in video games: a systematic review." en. In: *Universal Access in the Information Society* 19.1 (Mar. 2020), pp. 169–193. ISSN: 1615-5297. DOI: 10.1007/s10209-018-0628-2. URL: https://doi.org/10.1007/s10209-018-0628-2 (visited on 05/31/2024).

[4] Laura Aina and Tal Linzen. "The Language Model Understood the Prompt was Ambiguous: Probing Syntactic Uncertainty Through Generation." In: *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. Ed. by Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux, Mario Giulianelli, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 42–57. DOI: 10.18653/v1/2021.blackboxnlp-1.4. URL: https://aclanthology.org/2021.blackboxnlp-1.4 (visited on 05/31/2024).

[5] Ahmed Anwar, Tianzheng Shi, and Oliver Schneider. "Factors of Haptic Experience across Multiple Haptic Modalities." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–12. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581514. URL: https://dl.acm.org/doi/10.1145/3544548.3581514 (visited on 05/21/2024).

[6] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena L. Glassman. "ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–18. ISBN: 9798400703300. DOI: 10.1145/3613904.3642016. URL: https://dl.acm.org/doi/10.1145/3613904.3642016 (visited on 05/31/2024).

[7] Jacob Austin et al. *Program Synthesis with Large Language Models*. arXiv:2108.07732 [cs]. Aug. 2021. DOI: 10.48550/arXiv.2108.07732. URL: http://arxiv.org/abs/2108.07732 (visited on 05/31/2024).

[8] Jing Bai and Jian-Yun Nie. "Using Language Models for Text Classification." en. In: ().

[9] Javier A. Bargas-Avila and Kasper Hornbæk. "Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: Association for Computing Machinery, May 2011, pp. 2689–2698. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979336. URL: https://dl.acm.org/doi/10.1145/1978942.1979336 (visited on 05/31/2024).

[10] Michael S. Bernstein, Joon Sung Park, Meredith Ringel Morris, Saleema Amershi, Lydia B Chilton, and Mitchell L. Gordon. "Architecting Novel Interactions with Generative AI Models." In: *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. UIST '23 Adjunct. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 1–3. ISBN: 9798400700965. DOI: 10.1145/3586182.3617431. URL: https://dl.acm.org/doi/10.1145/3586182.3617431 (visited on 05/31/2024).

[11] Rishi Bommasani et al. *On the Opportunities and Risks of Foundation Models*. arXiv:2108.07258 [cs]. July 2022. DOI: 10.48550/arXiv.2108.07258. URL: http://arxiv.org/abs/2108.07258 (visited on 05/31/2024).

[12] Conrad Borchers, Dalia Gala, Benjamin Gilburt, Eduard Oravkin, Wilfried Bounsi, Yuki M Asano, and Hannah Kirk. "Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements." In: *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. Ed. by Christian Hardmeier, Christine Basta, Marta R. Costa-jussà, Gabriel Stanovsky, and Hila Gonen. Seattle, Washington: Association for Computational Linguistics, July 2022, pp. 212–224. DOI: 10.18653/v1/2022.gebnlp-1.22. URL: https://aclanthology.org/2022.gebnlp-1.22 (visited on 05/31/2024).

[13] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology." In: *Qualitative Research in Psychology* 3.2 (Jan. 2006). Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa, pp. 77–101. ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa. URL: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa (visited on 08/07/2024).

[14] John Brooke. "SUS: A quick and dirty usability scale." In: *Usability Eval. Ind.* 189 (Nov. 1995).

[15] Paul Bucci, Lotus Zhang, Xi Laura Cang, and Karon E. MacLean. "Is it Happy? Behavioural and Narrative Frame Complexity Impact Perceptions of a Simple Furry Robot's Emotions." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, Apr. 2018, pp. 1–11. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174083. URL:

https://dl.acm.org/doi/10.1145/3173574.3174083 (visited on 08/08/2024).

[16] Ramón Burri. *Improving user trust towards conversational chatbot interfaces with voice output*. eng. 2018. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-240585 (visited on 08/08/2024).

[17] Youngjin (YJ) Chae and Thomas Davidson. *Large Language Models for Text Classification: From Zero-Shot Learning to Fine-Tuning*. en-us. Aug. 2023. DOI: 10.31235/osf.io/sthwk. URL: https://osf.io/sthwk (visited on 05/31/2024).

[18] Roger W. Cholewiak and Amy A. Collins. "Individual differences in the vibrotactile perception of a "simple" pattern set." en. In: *Perception & Psychophysics* 59.6 (Jan. 1997), pp. 850–866. ISSN: 1532-5962. DOI: 10.3758/BF03205503. URL: https://doi.org/10.3758/BF03205503 (visited on 01/30/2024).

[19] Fenia Christopoulou et al. *PanGu-Coder: Program Synthesis with Function-Level Language Modeling*. arXiv:2207.11280 [cs]. July 2022. DOI: 10.48550/arXiv.2207.11280. URL: http://arxiv.org/abs/2207.11280 (visited on 05/31/2024).

[20] Ben Clark, Oliver S. Schneider, Karon E. MacLean, and Hong Z. Tan. "Predictable and distinguishable morphing of vibrotactile rhythm." en. In: *2017 IEEE World Haptics Conference (WHC)*. Munich, Germany: IEEE, June 2017, pp. 84–89. ISBN: 978-1-5090-1425-5. DOI: 10.1109/WHC.2017.7989881. URL: http://ieeexplore.ieee.org/document/7989881/ (visited on 05/21/2024).

[21] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. "LLMR: Real-time Prompting of Interactive Worlds using Large Language Models." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–22. ISBN: 9798400703300. DOI: 10.1145/3613904.3642579. URL: https://dl.acm.org/doi/10.1145/3613904.3642579 (visited on 05/27/2024).

[22] Stavros Demetriadis and Yannis Dimitriadis. "Conversational Agents and Language Models that Learn from Human Dialogues to Support Design Thinking." en. In: *Augmented Intelligence and Intelligent Tutoring Systems*. Ed. by Claude Frasson, Phivos Mylonas, and Christos Troussas. Cham: Springer Nature Switzerland, 2023, pp. 691–700. ISBN: 978-3-031-32883-1. DOI: 10.1007/978-3-031-32883-1_60.

[23] Qingxiu Dong et al. *A Survey on In-context Learning*. arXiv:2301.00234 [cs]. June 2024. DOI: 10.48550/arXiv.2301.00234. URL: http://arxiv.org/abs/2301.00234 (visited on 07/18/2024).

[24] DualSenseX. *Game Mods with DualSenseX*. en-US. URL: https://dualsensex.com/game-mods/ (visited on 07/18/2024).

[25] *F1® 23, EA SPORTS™ official videogame of the 2023 FIA Formula One World Championship™*. URL: https://www.ea.com/games/f1/f1-23?isLocalized=true (visited on 08/08/2024).

[26] Aparna Garimella, Rada Mihalcea, and Akhash Amarnath. "Demographic-Aware Language Model Fine-tuning as a Bias Mitigation Technique." In: *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Ed. by Yulan He, Heng Ji, Sujian Li, Yang Liu, and Chua-Hui Chang. Online only: Association for Computational Linguistics, Nov. 2022, pp. 311–319. URL: https://aclanthology.org/2022.aacl-short.38 (visited on 05/31/2024).

[27] Louie Giray. "Prompt Engineering with ChatGPT: A Guide for Academic Writers." en. In: *Annals of Biomedical Engineering* 51.12 (Dec. 2023), pp. 2629–2633. ISSN: 1573-9686. DOI: 10.1007/s10439-023-03272-4. URL: https://doi.org/10.1007/s10439-023-03272-4 (visited on 05/31/2024).

[28] *Google's Next Generation Music Recognition*. en. URL: http://research.google/blog/googles-next-generation-music-recognition/ (visited on 08/08/2024).

[29] Izzeddin Gur, Hiroki Furuta, Austin Huang, Mustafa Safdari, Yutaka Matsuo, Douglas Eck, and Aleksandra Faust. *A Real-World WebAgent with Planning, Long Context Understanding, and Program Synthesis*. arXiv:2307.12856 [cs]. Feb. 2024. DOI: 10.48550/arXiv.2307.12856. URL: http://arxiv.org/abs/2307.12856 (visited on 05/31/2024).

[30] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In: *Advances in Psychology*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Human Mental Workload. North-Holland, Jan. 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9. URL: https://www.sciencedirect.com/science/article/pii/S0166411508623869 (visited on 07/19/2024).

[31] Qianyu He, Jie Zeng, Wenhao Huang, Lina Chen, Jin Xiao, Qianxi He, Xunzhe Zhou, Jiaqing Liang, and Yanghua Xiao. "Can Large Language Models Understand Real-World Complex Instructions?" en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.16 (Mar. 2024). Number: 16, pp. 18188–18196. ISSN: 2374-3468. DOI: 10.1609/aaai.v38i16.29777. URL: https://ojs.aaai.org/index.php/AAAI/article/view/29777 (visited on 05/31/2024).

[32] *Hello GPT-4o*. en-US. URL: https://openai.com/index/hello-gpt-4o/ (visited on 05/31/2024).

[33] Mark Holliins, Richard Faldowski, Suman Rao, and Forrest Young. "Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis." en. In: *Perception & Psychophysics* 54.6 (Nov. 1993), pp. 697–705. ISSN: 1532-5962. DOI: 10.3758/BF03211795. URL: https://doi.org/10.3758/BF03211795 (visited on 05/21/2024).

[34] Jamil Hussain, Anees Ul Hassan, Hafiz Syed Muhammad Bilal, Rahman Ali, Muhammad Afzal, Shujaat Hussain, Jaehun Bang, Oresti Banos, and Sungyoung Lee. "Model-based adaptive user interface based on context and user experience evaluation." en. In: *Journal on Multimodal User Interfaces* 12.1 (Mar. 2018), pp. 1–16. ISSN: 1783-8738. DOI: 10.1007/s12193-018-0258-2. URL: https://doi.org/10.1007/s12193-018-0258-2 (visited on 05/31/2024).

[35] Picovoice Inc. *Picovoice — Wake Word Detection & Keyword Spotting*. en. URL: https://picovoice.ai/platform/porcupine/ (visited on 07/18/2024).

[36] Inwook Hwang, K. E. MacLean, M. Brehmer, J. Hendy, A. Sotirakopoulos, and Seungmoon Choi. "The haptic crayola effect: Exploring the role of naming in learning haptic stimuli." en. In: *2011 IEEE World Haptics Conference*. Istanbul: IEEE, June 2011, pp. 385–390. ISBN: 978-1-4577-0299-0. DOI: 10.1109/WHC.2011.5945517. URL: http://ieeexplore.ieee.org/document/5945517/ (visited on 05/21/2024).

[37] Naman Jain, Skanda Vaidyanath, Arun Iyer, Nagarajan Natarajan, Suresh Parthasarathy, Sriram Rajamani, and Rahul Sharma. "Jigsaw: large language models meet program synthesis." In: *Proceedings of the 44th International Conference on Software Engineering*. ICSE '22. New York, NY, USA: Association for Computing Machinery, July 2022, pp. 1219–1231. ISBN: 978-1-4503-9221-1. DOI: 10.1145/3510003.3510203. URL: https://dl.acm.org/doi/10.1145/3510003.3510203 (visited on 05/31/2024).

[38] Cathrine V. Jansson-Boyd. "Touch matters: exploring the relationship between consumption and tactile interaction." In: *Social Semiotics* 21.4 (Sept. 2011). Publisher: Routledge _eprint: https://doi.org/10.1080/10350330.2011.591996, pp. 531–546. ISSN: 1035-0330. DOI: 10.1080/10350330.2011.591996. URL: https://doi.org/10.1080/10350330.2011.591996 (visited on 05/21/2024).

[39] Albert Q. Jiang et al. *Mixtral of Experts*. arXiv:2401.04088 [cs]. Jan. 2024. DOI: 10.48550/arXiv.2401.04088. URL: http://arxiv.org/abs/2401.04088 (visited on 05/31/2024).

[40] Erik Jones and Jacob Steinhardt. "Capturing Failures of Large Language Models via Human Cognitive Biases." en. In: *Advances in Neural Information Processing Systems* 35 (Dec. 2022), pp. 11785–11799. URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/4d13b2d99519c5415661dad44ab7edcd-Abstract-Conference.html (visited on 05/31/2024).

[41] Przemyslaw Joniak and Akiko Aizawa. "Gender Biases and Where to Find Them: Exploring Gender Bias in Pre-Trained Transformer-based Language Models Using Movement Pruning." In: *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. Ed. by Christian Hardmeier, Christine Basta, Marta R. Costa-jussà, Gabriel Stanovsky, and Hila Gonen. Seattle, Washington: Association for Computational Linguistics, July 2022, pp. 67–73. DOI: 10.18653/v1/2022.gebnlp-1.6. URL: https://aclanthology.org/2022.gebnlp-1.6 (visited on 05/31/2024).

[42] Miku Kaneda, Masahiro Takeuchi, Seitaro Kaneko, and Hiroyuki Kajimoto. "Relationship between onomatopoeia and physical properties when pressing a soft object." In: *2022 IEEE Haptics Symposium (HAPTICS)*. ISSN: 2324-7355. Mar. 2022, pp. 1–6. DOI: 10.1109/HAPTICS52432.2022.9765623. URL: https://ieeexplore.ieee.org/abstract/document/9765623?casa_token=2sdogYrQ0pEAAAAA:XYy3pZAyFTBuVYCB2WUv2wJ8epxG3YAmzkNcEivWFkYG3-Q-GIy9M_Hai41VrazxFqFj9NMeug (visited on 08/06/2024).

[43] Ioannis Kangas, Maud Schwoerer, and Lucas J Bernardi. "Recommender Systems for Personalized User Experience: Lessons learned at Booking.com." In: *Proceedings of the 15th ACM Conference on Recommender Systems*. RecSys '21. New York, NY, USA: Association for Computing Machinery, Sept. 2021, pp. 583–586. ISBN: 978-1-4503-8458-2. DOI: 10.1145/3460231.3474611. URL: https://dl.acm.org/doi/10.1145/3460231.3474611 (visited on 05/31/2024).

[44] Akifumi Kawai and Yoshihiro Tanaka. "Individual Differences in Skin Vibration Characteristics and Vibrotactile Sensitivity at Fingertip." In: *2022 IEEE Haptics Symposium (HAPTICS)*. ISSN: 2324-7355. Mar. 2022, pp. 1–6. DOI: 10.1109/HAPTICS52432.2022.9765612. URL: https://ieeexplore.ieee.org/abstract/document/9765612 (visited on 01/30/2024).

[45] Erin Kim and Oliver Schneider. "Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–13. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376280. URL: https://dl.acm.org/doi/10.1145/3313831.3376280 (visited on 05/21/2024).

[46] Hyunwoo Kim, Youngjae Yu, Liwei Jiang, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. "ProsocialDialog: A Prosocial Backbone for Conversational Agents." In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 4005–4029. DOI: 10.18653/v1/2022.emnlp-main.267.

URL: https://aclanthology.org/2022.emnlp-main.267 (visited on 05/31/2024).

[47] Jaehyung Kim and Yiming Yang. *Few-shot Personalization of LLMs with Mis-aligned Responses*. arXiv:2406.18678 [cs]. June 2024. DOI: 10.48550/arXiv.2406.18678. URL: http://arxiv.org/abs/2406.18678 (visited on 08/06/2024).

[48] Hannah Rose Kirk, Yennie Jun, Filippo Volpin, Haider Iqbal, Elias Benussi, Frederic Dreyer, Aleksandar Shtedritski, and Yuki Asano. "Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models." In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 2611–2624. URL: https://proceedings.neurips.cc/paper/2021/hash/1531beb762df4029513ebf9295e0d34f-Abstract.html (visited on 05/31/2024).

[49] Toshihiro Kita, Chikako Nagaoka, Naoshi Hiraoka, and Martin Dougiamas. "Implementation of Voice User Interfaces to Enhance Users' Activities on Moodle." In: *2019 4th International Conference on Information Technology (InCIT)*. Oct. 2019, pp. 104–107. DOI: 10.1109/INCIT.2019.8912086. URL: https://ieeexplore.ieee.org/abstract/document/8912086?casa_token=jzQ52Ogkxt4AAAAA:Q1qKpZoiZkimlCd2woNtZucPGKuA8xt4ihawniu_-By-oS72x19j30buwMzZ5oAgxdKEhAPQ (visited on 08/08/2024).

[50] Dmitry Kobak and George C. Linderman. *UMAP does not preserve global structure any better than t-SNE when using the same initialization*. en. Pages: 2019.12.19.877522 Section: Contradictory Results. Dec. 2019. DOI: 10.1101/2019.12.19.877522. URL: https://www.biorxiv.org/content/10.1101/2019.12.19.877522v1 (visited on 05/28/2024).

[51] Dmitry Kobak and George C. Linderman. "Initialization is critical for preserving global data structure in both t-SNE and UMAP." en. In: *Nature Biotechnology* 39.2 (Feb. 2021). Publisher: Nature Publishing Group, pp. 156–157. ISSN: 1546-1696. DOI: 10.1038/s41587-020-00809-z. URL: https://www.nature.com/articles/s41587-020-00809-z (visited on 07/18/2024).

[52] Elena Kornyshova and Eric Gressier-Soudan. "Conceptual Model of User Experience for Personalization." en. In: *Information Systems*. Ed. by Maria Papadaki, Paulo Rupino da Cunha, Marinos Themistocleous, and Klitos Christodoulou. Cham: Springer Nature Switzerland, 2023, pp. 397–406. ISBN: 978-3-031-30694-5. DOI: 10.1007/978-3-031-30694-5_29.

[53] Hadas Kotek, Rikker Dockum, and David Sun. "Gender bias and stereotypes in Large Language Models." In: *Proceedings of The ACM Collective Intelligence Conference*. CI '23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 12–24. ISBN: 9798400701139. DOI: 10.1145/3582269.3615599. URL: https://dl.acm.org/doi/10.1145/3582269.3615599 (visited on 05/31/2024).

[54] Edith Law, Parastoo Baghaei Ravari, Nalin Chhibber, Dana Kulic, Stephanie Lin, Kevin D. Pantasdo, Jessy Ceha, Sangho Suh, and Nicole Dillen. "Curiosity Notebook: A Platform for Learning by Teaching Conversational Agents." In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–9. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3382783. URL: https://dl.acm.org/doi/10.1145/3334480.3382783 (visited on 08/08/2024).

[55] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. "Understanding, scoping and defining user experience: a survey approach." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: Association for Computing Machinery, Apr. 2009, pp. 719–728. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518813. URL: https://dl.acm.org/doi/10.1145/1518701.1518813 (visited on 05/31/2024).

[56] Jiwan Lee, Junwoo Kim, Jeonggoo Kang, Eunsoo Jo, Dong Chul Park, and Seungmoon Choi. "Telemetry-Based Haptic Rendering for Racing Game Experience Improvement." In: *IEEE Transactions on Haptics* 17.1 (Jan. 2024). Conference Name: IEEE Transactions on Haptics, pp. 72–79. ISSN: 2329-4051. DOI: 10.1109/TOH.2024.3357885. URL: https://ieeexplore.ieee.org/document/10413623 (visited on 07/18/2024).

[57] Bingxu Li and Gregory J. Gerling. "An individual's skin stiffness predicts their tactile discrimination of compliance." en. In: *The Journal of Physiology* 601.24 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1113/JP285271, pp. 5777–5794. ISSN: 1469-7793. DOI: 10.1113/JP285271. URL: https://onlinelibrary.wiley.com/doi/abs/10.1113/JP285271 (visited on 05/24/2024).

[58] Lizi Liao, Grace Hui Yang, and Chirag Shah. "Proactive Conversational Agents in the Post-ChatGPT World." In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '23. New York, NY, USA: Association for Computing Machinery, July 2023, pp. 3452–3455. ISBN: 978-1-4503-9408-6. DOI: 10.1145/3539618.3594250. URL: https://dl.acm.org/doi/10.1145/3539618.3594250 (visited on 05/31/2024).

[59] Chen Cecilia Liu, Fajri Koto, Timothy Baldwin, and Iryna Gurevych. *Are Multilingual LLMs Culturally-Diverse Reasoners? An Investigation into Multicultural Proverbs and Sayings*. arXiv:2309.08591 [cs]. Mar. 2024. DOI: 10.48550/arXiv.2309.08591. URL: http://arxiv.org/abs/2309.08591 (visited on 05/31/2024).

[60] Ruibo Liu, Chenyan Jia, Jason Wei, Guangxuan Xu, Lili Wang, and Soroush Vosoughi. "Mitigating Political Bias in Language Models through Reinforced Calibration." In: *Proceedings of the AAAI Conference*

on Artificial Intelligence 35.17 (May 2021), pp. 14857–14866. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v35i17.17744. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17744 (visited on 05/31/2024).

[61] *Look and Talk: Natural Conversations with Google Assistant*. en. URL: http://research.google/blog/look-and-talk-natural-conversations-with-google-assistant/ (visited on 08/08/2024).

[62] J. Lo¨fvenberg and R. S. Johansson. "Regional differences and interindividual variability in sensitivity to vibration in the glabrous skin of the human hand." In: *Brain Research* 301.1 (May 1984), pp. 65–72. ISSN: 0006-8993. DOI: 10.1016/0006-8993(84)90403-7. URL: https://www.sciencedirect.com/science/article/pii/0006899384904037 (visited on 05/21/2024).

[63] Kaiji Lu, Piotr Mardziel, Fangjing Wu, Preetam Amancharla, and Anupam Datta. "Gender Bias in Neural Natural Language Processing." en. In: *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday*. Ed. by Vivek Nigam, Tajana Ban Kirigin, Carolyn Talcott, Joshua Guttman, Stepan Kuznetsov, Boon Thau Loo, and Mitsuhiro Okada. Cham: Springer International Publishing, 2020, pp. 189–202. ISBN: 978-3-030-62077-6. DOI: 10.1007/978-3-030-62077-6_14. URL: https://doi.org/10.1007/978-3-030-62077-6_14 (visited on 05/31/2024).

[64] Shihan Lu, Mianlun Zheng, Matthew C. Fontaine, Stefanos Nikolaidis, and Heather Culbertson. "Preference-Driven Texture Modeling Through Interactive Generation and Search." In: *IEEE Transactions on Haptics* 15.3 (July 2022). Conference Name: IEEE Transactions on Haptics, pp. 508–520. ISSN: 2329-4051. DOI: 10.1109/TOH.2022.3173935.

[65] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE." In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. ISSN: 1533-7928. URL: http://jmlr.org/papers/v9/vandermaaten08a.html (visited on 07/18/2024).

[66] David Marino, Paul Bucci, Oliver S. Schneider, and Karon E. MacLean. "Voodle: Vocal Doodling to Sketch Affective Robot Motion." In: *Proceedings of the 2017 Conference on Designing Interactive Systems*. DIS '17. New York, NY, USA: Association for Computing Machinery, June 2017, pp. 753–765. ISBN: 978-1-4503-4922-2. DOI: 10.1145/3064663.3064668. URL: https://dl.acm.org/doi/10.1145/3064663.3064668 (visited on 05/21/2024).

[67] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. "UMAP: Uniform Manifold Approximation and Projection." en. In: *Journal of Open Source Software* 3.29 (Sept. 2018), p. 861. ISSN: 2475-9066. DOI: 10.21105/joss.00861. URL: https://joss.theoj.org/papers/10.21105/joss.00861 (visited on 07/18/2024).

[68] Roberto Navigli, Simone Conia, and Björn Ross. "Biases in Large Language Models: Origins, Inventory, and Discussion." In: *Journal of Data and Information Quality* 15.2 (June 2023), 10:1–10:21. ISSN: 1936-1955. DOI: 10.1145/3597307. URL: https://dl.acm.org/doi/10.1145/3597307 (visited on 05/31/2024).

[69] Erik Nijkamp, Bo Pang, Hiroaki Hayashi, Lifu Tu, Huan Wang, Yingbo Zhou, Silvio Savarese, and Caiming Xiong. *CodeGen: An Open Large Language Model for Code with Multi-Turn Program Synthesis*. arXiv:2203.13474 [cs]. Feb. 2023. DOI: 10.48550/arXiv.2203.13474. URL: http://arxiv.org/abs/2203.13474 (visited on 05/31/2024).

[70] Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. "Talking about tactile experiences." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: Association for Computing Machinery, Apr. 2013, pp. 1659–1668. ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2466220. URL: https://dl.acm.org/doi/10.1145/2470654.2466220 (visited on 05/21/2024).

[71] Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. "Psychophysical Dimensions of Tactile Perception of Textures." In: *IEEE Transactions on Haptics* 6.1 (2013). Conference Name: IEEE Transactions on Haptics, pp. 81–93. ISSN: 2329-4051. DOI: 10.1109/TOH.2012.32. URL: https://ieeexplore.ieee.org/abstract/document/6216375 (visited on 05/21/2024).

[72] OpenAI. *OpenAI Platform*. en. URL: https://platform.openai.com (visited on 07/18/2024).

[73] OpenAI et al. *GPT-4 Technical Report*. arXiv:2303.08774 [cs]. Mar. 2024. DOI: 10.48550/arXiv.2303.08774. URL: http://arxiv.org/abs/2303.08774 (visited on 05/31/2024).

[74] Hadas Orgad and Yonatan Belinkov. "BLIND: Bias Removal With No Demographics." In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 8801–8821. DOI: 10.18653/v1/2023.acl-long.490. URL: https://aclanthology.org/2023.acl-long.490 (visited on 05/31/2024).

[75] Long Ouyang et al. "Training language models to follow instructions with human feedback." In: *ArXiv* (Mar. 2022). URL: https://www.semanticscholar.org/paper/Training-language-models-to-follow-instructions-Ouyang-Wu/d766bffc357127e0dc86dd69561d5aeb520d6f4c (visited on 05/31/2024).

[76] Christos Papakostas, Christos Troussas, Akrivi Krouska, and Cleo Sgouropoulou. "Measuring User Experience, Usability and Interactivity of a Personalized Mobile Augmented Reality Training System." en. In: *Sensors* 21.11 (Jan. 2021). Number: 11 Publisher: Multidisciplinary

Digital Publishing Institute, p. 3888. ISSN: 1424-8220. DOI: 10.3390/s21113888. URL: https://www.mdpi.com/1424-8220/21/11/3888 (visited on 05/31/2024).

[77] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. *Generative Agents: Interactive Simulacra of Human Behavior*. arXiv:2304.03442 [cs]. Aug. 2023. URL: http://arxiv.org/abs/2304.03442 (visited on 08/14/2023).

[78] Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. en. Google-Books-ID: MmnEDQAAQBAJ. "O'Reilly Media, Inc.", Dec. 2016. ISBN: 978-1-4919-5538-3.

[79] Joann Peck and Terry L. Childers. "Individual Differences in Haptic Information Processing: The "Need for Touch" Scale." In: *Journal of Consumer Research* 30.3 (Dec. 2003), pp. 430–442. ISSN: 0093-5301. DOI: 10.1086/378619. URL: https://doi.org/10.1086/378619 (visited on 01/30/2024).

[80] Xiangyu Peng, Siyan Li, Spencer Frazier, and Mark Riedl. "Reducing Non-Normative Text Generation from Language Models." In: *Proceedings of the 13th International Conference on Natural Language Generation*. Ed. by Brian Davis, Yvette Graham, John Kelleher, and Yaji Sripada. Dublin, Ireland: Association for Computational Linguistics, Dec. 2020, pp. 374–383. DOI: 10.18653/v1/2020.inlg-1.43. URL: https://aclanthology.org/2020.inlg-1.43 (visited on 05/31/2024).

[81] Jonathan Pilault, Raymond Li, Sandeep Subramanian, and Chris Pal. "On Extractive and Abstractive Neural Document Summarization with Transformer Language Models." In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Ed. by Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu. Online: Association for Computational Linguistics, Nov. 2020, pp. 9308–9319. DOI: 10.18653/v1/2020.emnlp-main.748. URL: https://aclanthology.org/2020.emnlp-main.748 (visited on 05/31/2024).

[82] Raul Puri and Bryan Catanzaro. *Zero-shot Text Classification With Generative Language Models*. arXiv:1912.10165 [cs]. Dec. 2019. DOI: 10.48550/arXiv.1912.10165. URL: http://arxiv.org/abs/1912.10165 (visited on 05/31/2024).

[83] Rebecca Qian, Candace Ross, Jude Fernandes, Eric Michael Smith, Douwe Kiela, and Adina Williams. "Perturbation Augmentation for Fairer NLP." In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 9496–9521. DOI: 10.18653/v1/2022.emnlp-main.646. URL: https://aclanthology.org/2022.emnlp-main.646 (visited on 05/31/2024).

[84] Arpit Rana, Scott Sanner, Mohamed Reda Bouadjenek, Ron Dicarlantonio, and Gary Farmaner. "User Experience and The Role of Personalization in Critiquing-Based Conversational Recommendation." In: *ACM Transactions on the Web* (May 2023). Just Accepted. ISSN: 1559-1131. DOI: 10.1145/3597499. URL: https://dl.acm.org/doi/10.1145/3597499 (visited on 05/31/2024).

[85] Leonardo Ranaldi, Elena Sofia Ruzzetti, Davide Venditti, Dario Onorati, and Fabio Massimo Zanzotto. *A Trip Towards Fairness: Bias and De-Biasing in Large Language Models*. arXiv:2305.13862 [cs]. Aug. 2023. DOI: 10.48550/arXiv.2305.13862. URL: http://arxiv.org/abs/2305.13862 (visited on 05/31/2024).

[86] Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction Design: Beyond Human - Computer Interaction*. en. John Wiley & Sons, June 2011. ISBN: 978-0-470-66576-3.

[87] Tom Roy, Yann Glémarec, Gurvan Lécuyer, Quentin Galvane, Philippe Guillotel, and Ferran Argelaguet. "Towards End-User Customization of Haptic Experiences." en. In: June 2024, p. 1. URL: https://inria.hal.science/hal-04611300 (visited on 07/04/2024).

[88] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. "Measuring Haptic Experience: Elaborating the HX model with scale development." en. In: *2021 IEEE World Haptics Conference (WHC)*. Montreal, QC, Canada: IEEE, July 2021, pp. 979–984. ISBN: 978-1-66541-871-3. DOI: 10.1109/WHC49131.2021.9517220. URL: https://ieeexplore.ieee.org/document/9517220/ (visited on 05/21/2024).

[89] Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. "Are Emergent Abilities of Large Language Models a Mirage?" en. In: *Advances in Neural Information Processing Systems* 36 (Dec. 2023), pp. 55565–55581. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/adc98a266f45005c403b8311ca7e8bd7-Abstract-Conference.html (visited on 05/31/2024).

[90] Oliver S. Schneider and Karon E. MacLean. "Studying design process and example use with Macaron, a web-based vibrotactile effect editor." en. In: *2016 IEEE Haptics Symposium (HAPTICS)*. Philadelphia, PA, USA: IEEE, Apr. 2016, pp. 52–58. ISBN: 978-1-5090-0903-9. DOI: 10.1109/HAPTICS.2016.7463155. URL: http://ieeexplore.ieee.org/document/7463155/ (visited on 05/21/2024).

[91] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. "Haptic experience design: What hapticians do and where they need help." In: *International Journal of Human-Computer Studies*. Multisensory Human-Computer Interaction 107 (Nov. 2017), pp. 5–21. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2017.04.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581917300605 (visited on 05/28/2024).

[92] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios." en. In: *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*. Ed. by Aaron Marcus. Cham: Springer International Publishing, 2014, pp. 383–392. ISBN: 978-3-319-07668-3. DOI: 10.1007/978-3-319-07668-3_37.

[93] Hasti Seifi. *Personalizing Haptics: From Individuals' Sense-Making Schemas to End-User Haptic Tools*. en. Springer Series on Touch and Haptic Systems. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-11378-0 978-3-030-11379-7. DOI: 10.1007/978-3-030-11379-7. URL: http://link.springer.com/10.1007/978-3-030-11379-7 (visited on 02/21/2024).

[94] Hasti Seifi, Chamila Anthonypillai, and Karon E. MacLean. "End-user customization of affective tactile messages: A qualitative examination of tool parameters." en. In: *2014 IEEE Haptics Symposium (HAPTICS)*. Houston, TX, USA: IEEE, Feb. 2014, pp. 251–256. ISBN: 978-1-4799-3131-6. DOI: 10.1109/HAPTICS.2014.6775463. URL: http://ieeexplore.ieee.org/document/6775463/ (visited on 05/21/2024).

[95] Hasti Seifi, Sean Chew, Antony James Nascè, William Edward Lowther, William Frier, and Kasper Hornbæk. "Feellustrator: A Design Tool for Ultrasound Mid-Air Haptics." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–16. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3580728. URL: https://dl.acm.org/doi/10.1145/3544548.3580728 (visited on 05/21/2024).

[96] Hasti Seifi, Kailun Zhang, and Karon E. MacLean. "VibViz: Organizing, visualizing and navigating vibration libraries." en. In: *2015 IEEE World Haptics Conference (WHC)*. Evanston, IL: IEEE, June 2015, pp. 254–259. ISBN: 978-1-4799-6624-0. DOI: 10.1109/WHC.2015.7177722. URL: http://ieeexplore.ieee.org/document/7177722/ (visited on 05/21/2024).

[97] Joseph C. Stevens. "Aging and Spatial Acuity of Touch." In: *Journal of Gerontology* 47.1 (Jan. 1992), P35–P40. ISSN: 0022-1422. DOI: 10.1093/geronj/47.1.P35. URL: https://doi.org/10.1093/geronj/47.1.P35 (visited on 05/21/2024).

[98] Joseph C. Stevens and Kenneth K. Choo. "Spatial Acuity of the Body Surface over the Life Span." In: *Somatosensory & Motor Research* 13.2 (Jan. 1996). Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/08990229609051403, pp. 153–166. ISSN: 0899-0220. DOI: 10.3109/08990229609051403. URL: https://doi.org/10.3109/08990229609051403 (visited on 05/21/2024).

[99] Hao Sun, Zhexin Zhang, Fei Mi, Yasheng Wang, Wei Liu, Jianwei Cui, Bin Wang, Qun Liu, and Minlie Huang. "MoralDial: A Framework to Train and Evaluate Moral Dialogue Systems via Moral Discussions."

In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 2213–2230. DOI: 10.18653/v1/2023.acl-long.123. URL: https://aclanthology.org/2023.acl-long.123 (visited on 05/31/2024).

[100] Liyan Tang et al. "Evaluating large language models on medical evidence summarization." en. In: *npj Digital Medicine* 6.1 (Aug. 2023). Publisher: Nature Publishing Group, pp. 1–8. ISSN: 2398-6352. DOI: 10.1038/s41746-023-00896-7. URL: https://www.nature.com/articles/s41746-023-00896-7 (visited on 05/31/2024).

[101] Ryan Teehan, Miruna Clinciu, Oleg Serikov, Eliza Szczechla, Natasha Seelam, Shachar Mirkin, and Aaron Gokaslan. "Emergent Structures and Training Dynamics in Large Language Models." In: *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models*. Ed. by Angela Fan, Suzana Ilic, Thomas Wolf, and Matthias Gallé. virtual+Dublin: Association for Computational Linguistics, May 2022, pp. 146–159. DOI: 10.18653/v1/2022.bigscience-1.11. URL: https://aclanthology.org/2022.bigscience-1.11 (visited on 05/31/2024).

[102] Himanshu Thakur, Atishay Jain, Praneetha Vaddamanu, Paul Pu Liang, and Louis-Philippe Morency. "Language Models Get a Gender Makeover: Mitigating Gender Bias with Few-Shot Data Interventions." In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 340–351. DOI: 10.18653/v1/2023.acl-short.30. URL: https://aclanthology.org/2023.acl-short.30 (visited on 05/31/2024).

[103] Karthikan Theivendran, Andy Wu, William Frier, and Oliver Schneider. "RecHap: An Interactive Recommender System For Navigating a Large Number of Mid-Air Haptic Designs." en. In: *IEEE Transactions on Haptics* (2024), pp. 1–12. ISSN: 1939-1412, 2329-4051, 2334-0134. DOI: 10.1109/TOH.2023.3276812. URL: https://ieeexplore.ieee.org/document/10146205/ (visited on 05/21/2024).

[104] Hugo Touvron et al. *Llama 2: Open Foundation and Fine-Tuned Chat Models*. arXiv:2307.09288 [cs]. July 2023. DOI: 10.48550/arXiv.2307.09288. URL: http://arxiv.org/abs/2307.09288 (visited on 05/31/2024).

[105] Megan Ung, Jing Xu, and Y-Lan Boureau. "SaFeRDialogues: Taking Feedback Gracefully after Conversational Safety Failures." In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Smaranda Muresan, Preslav Nakov, and Aline Villavicencio. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 6462–6481. DOI: 10.18653/v1/

2022.acl-long.447. URL: https://aclanthology.org/2022.acl-long.447 (visited on 05/31/2024).

[106] *Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine*. en. URL: https://unity.com/ (visited on 05/31/2024).

[107] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. "Attention is All you Need." In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (visited on 05/31/2024).

[108] Lalit Venkatesan, Steven M. Barlow, and Douglas Kieweg. "Age- and sex-related changes in vibrotactile sensitivity of hand and face in neurotypical adults." In: *Somatosensory & Motor Research* 32.1 (Jan. 2015). Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/08990220.2014.958216, pp. 44–50. ISSN: 0899-0220. DOI: 10.3109/08990220.2014.958216. URL: https://doi.org/10.3109/08990220.2014.958216 (visited on 01/30/2024).

[109] Kellie Webster, Xuezhi Wang, Ian Tenney, Alex Beutel, Emily Pitler, Ellie Pavlick, Jilin Chen, Ed Chi, and Slav Petrov. *Measuring and Reducing Gendered Correlations in Pre-trained Models*. arXiv:2010.06032 [cs]. Mar. 2021. DOI: 10.48550/arXiv.2010.06032. URL: http://arxiv.org/abs/2010.06032 (visited on 05/31/2024).

[110] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. arXiv:2201.11903 [cs]. Jan. 2023. DOI: 10.48550/arXiv.2201.11903. URL: http://arxiv.org/abs/2201.11903 (visited on 07/18/2024).

[111] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. arXiv:2302.11382 [cs]. Feb. 2023. DOI: 10.48550/arXiv.2302.11382. URL: http://arxiv.org/abs/2302.11382 (visited on 05/31/2024).

[112] Jason Wu, Karan Ahuja, Richard Li, Victor Chen, and Jeffrey Bigham. "ScratchThat: Supporting Command-Agnostic Speech Repair in Voice-Driven Assistants." en. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.2 (June 2019), pp. 1–17. ISSN: 2474-9567. DOI: 10.1145/3328934. URL: https://dl.acm.org/doi/10.1145/3328934 (visited on 02/05/2024).

[113] Shiquan Zhang, Ying Ma, Le Fang, Hong Jia, Simon D'Alfonso, and Vassilis Kostakos. *Enabling On-Device LLMs Personalization with Smartphone Sensing*. arXiv:2407.04418 [cs]. July 2024. DOI: 10.48550/arXiv.2407.04418. URL: http://arxiv.org/abs/2407.04418 (visited on 08/06/2024).

[114] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. "Benchmarking Large Language Models for News Summarization." In: *Transactions of the Association for Computational Linguistics* 12 (Jan. 2024), pp. 39–57. ISSN: 2307-387X. DOI: 10.1162/tacl_a_00632. URL: https://doi.org/10.1162/tacl_a_00632 (visited on 05/31/2024).

[115] Ying Zheng and John B. Morrell. "Haptic actuator design parameters that influence affect and attention." en. In: *2012 IEEE Haptics Symposium (HAPTICS)*. Vancouver, BC, Canada: IEEE, Mar. 2012, pp. 463–470. ISBN: 978-1-4673-0809-0 978-1-4673-0808-3 978-1-4673-0807-6. DOI: 10.1109/HAPTIC.2012.6183832. URL: http://ieeexplore.ieee.org/document/6183832/ (visited on 05/21/2024).

[116] Ran Zmigrod, Sabrina J. Mielke, Hanna Wallach, and Ryan Cotterell. "Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology." In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by Anna Korhonen, David Traum, and Lluís Màrquez. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 1651–1661. DOI: 10.18653/v1/P19-1161. URL: https://aclanthology.org/P19-1161 (visited on 05/31/2024).

[117] Mingyu Zong and Bhaskar Krishnamachari. "Solving math word problems concerning systems of equations with GPT models." In: *Machine Learning with Applications* 14 (Dec. 2023), p. 100506. ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2023.100506. URL: https://www.sciencedirect.com/science/article/pii/S2666827023000592 (visited on 02/09/2024).

[118] cosmii02, zchenak, and patmagauran. *cosmii02/RacingDSX*. original-date: 2022-09-26T15:22:42Z. July 2024. URL: https://github.com/cosmii02/RacingDSX (visited on 07/18/2024).

[119] *draw-a-ui*. URL: https://www.draw-a-ui.com/ (visited on 05/31/2024).

APPENDICES

A TACTALK SYSTEM PROMPTS

In this section, we provide additional context and materials that were used as part of the development and evaluation of TacTalk. In particular, we provide the different prompts used for TacTalk, the settings that were used as the initial configuration for TacTalk and the list of 51 prompts used for TacTalk's technical evaluation.

A.1 WHEN USING A SINGLE PROMPT

```
You're a helpful racing engineer assistant who converses with the driver
to modify the haptic feedback on the PS5 Dualsense controller
in the racing game Forza Horizon 5. Given a user's prompt as well as a summary
of the user's past preferences for haptic feedback, you need to
identify which controller settings need to be changed from the following list of
parameters and specify, for all parameters, what their new value will be. Finally,
you need to provide the reason for selecting the settings you chose.
Additionally, you need to return an updated version of user's preference
summary.

# Guidelines to follow
- Format your response as a JSON object with four fields - 'settings_list',
'old_value', 'new_value', 'justification' and 'updated_summary'. DO NOT
RETURN ANYTHING EXCEPT JSON.
- 'settings_list' is a list of settings that need to be modified
- 'old_value' is the existing list of values for the settings in
'settings_list'
- 'new_value' is the new suggested list of values for the settings in
'settings_list'
- 'justification' is a string explaining the reason behind the choice of
new values
- 'updated_summary' is a string containing the updated user preference
summary for haptic feedback

# The list of settings to choose from, along with a definition of their
meaning. Note that all values except for the Brake Trigger Mode and Throttle
Trigger Mode should be percentages between 0 and 100.
- Brake Trigger Mode: Whether the brake should not render haptics at all
(value = 0), render only uniform stiffness (value = 1), or render rich
vibrations (value = 2).
- Throttle Trigger Mode: Whether the throttle should not render haptics at all
(value = 0), render only uniform stiffness (value = 1), or render rich
```

```
vibrations (value = 2).
- Brake Effect Intensity: The strength of the forces and vibrations produced by
the brake.
- Throttle Effect Intensity: The strength of the forces and vibrations
produced by the throttle.
- Inverse Brake Grip Loss Sensitivity: The lower this value (i.e., closer to 0),
the more the haptic feedback conveys grip loss for the brakes. The higher
(i.e., closer to 100), the less responsive the haptic feedback is for grip loss
for the brakes.
- Inverse Throttle Grip Loss Sensitivity: The lower this value (i.e., closer to
0), the more the haptic feedback conveys grip loss for the throttle. The
higher (i.e., closer to 100), the less responsive the haptic feedback is for
grip loss for the throttle.
- Turning Acceleration Scale: The higher this value, the more of an
effect turning will have on producing haptic feedback. For example, turning
sharply will cause the throttle trigger to become more stiff or vibrate if
the turning acceleration scale is high.
- Forward Acceleration Scale: The higher this value, the more of an
effect forward movement will have on haptic feedback. For example,
sharp accelerations will cause the throttle trigger to become more stiff or
vibrate if the forward acceleration scale is high.
- Acceleration Limit: This value determines how high the acceleration needs to
get before the throttle haptic feedback renders a constant sensation. If the
car accelerates any more than this, the haptic feedback won't change.

- Minimum Brake Vibration: The minimum frequency at which the brake can
vibrate. For ideal responses, this value should always be less than
the Maximum Brake Vibration and above 20 percent. A value of 0 means no
vibrations will be felt.
- Maximum Brake Vibration: The maximum frequency at which the brake can
vibrate. This value should always be kept above the Minimum Brake Vibration. A
value of 0 means no vibrations will be felt.
- Minimum Throttle Vibration: The minimum frequency at which the
throttle can vibrate. For ideal responses, this value should always be less than
the Maximum Throttle Vibration and above 20 percent. A value of 0 means
no vibrations will be felt.
- Maximum Throttle Vibration: The maximum frequency at which the
throttle can vibrate. This value should always be kept above the Minimum Throttle
Vibration. A value of 0 means no vibrations will be felt.

- Minimum Brake Resistance: The minimum resistance that the brake can produce.

- Maximum Brake Resistance: The maximum resistance that the brake can produce.

- Minimum Throttle Resistance: The minimum resistance that the
throttle can produce.
- Maximum Throttle Resistance: The maximum resistance that the
```

```
throttle can produce.


# The existing values of the settings listed above are shown in the following:

<Settings list appended here>

<User preference profile appended here>
```

A.2 WHEN USING MULTIPLE PROMPTS

A.2.1 *Stage 1: Modifying Haptic Parameters*

```
You're a helpful racing engineer assistant who converses with the driver
to modify the haptic feedback on the PS5 Dualsense in the racing
game Forza Horizon 5. Given a user's prompt as well as a summary of the user's
past preferences for haptic feedback, you need to identify which controller
settings need to be changed from the following list of parameters and specify, for
all parameters, what their new value will be. Finally, you need to provide the
reason for selecting the settings you chose. Your response will then be
forwarded to another mechanic assistant, who will actually apply changes
to the settings.

# Guidelines to follow
- You should pay attention to the driver's requests during your conversation
and come up with a list of settings to change based on the driver's query.

- The settings you mention should be taken from the list of settings provided
to you in this prompt. DO NOT list any settings that are not provided in this
prompt.
- Your reason needs to be specific and well-motivated. DO NOT give an
unrealistic reason for your response.
- If a query mentions that the user does not want to feel vibrations, set the
Brake Trigger Mode and Throttle Trigger Mode to uniform stiffness before
turning it off completely.
- Return the new value of the settings EXCEPT Brake Trigger Mode and Throttle
Trigger Mode as a percentage between 0 and 100. Do not return values
outside the 0-100 range!
- IF YOU MODIFY SETTINGS WITH Brake Resistance IN THEIR NAME, MAKE SURE THE
Brake Trigger Mode IS SET TO 1 OR 2
- IF YOU MODIFY SETTINGS WITH Throttle Resistance IN THEIR NAME, MAKE SURE THE
Throttle Trigger Mode IS SET TO 1 OR 2
- IF THE Brake Trigger Mode IS SET TO 1 OR 2, MAKE SURE THE Brake Effect Intensity
IS GREATER THAN 0
- IF THE Throttle Trigger Mode IS SET TO 1 OR 2, MAKE SURE THE Throttle Effect
```

```
Intensity IS GREATER THAN 0
- Format your response as a JSON object with four fields - 'settings_list',
'old_value', 'new_value' and 'justification'. DO NOT RETURN ANYTHING EXCEPT
JSON.
- 'settings_list' is a list of settings that need to be modified

- 'old_value' is the existing list of values for the settings in
'settings_list'
- 'new_value' is the new suggested list of values for the settings in
'settings_list'
- 'justification' is a string explaining the reason behind the choice of
new values
- Again, DO NOT RETURN ANYTHING EXCEPT JSON!!

# The list of settings to choose from, along with a definition of their meaning

- Brake Trigger Mode: Whether the brake should not render haptics at all
(value = 0), render only uniform stiffness (value = 1), or render rich
vibrations (value = 2).
- Throttle Trigger Mode: Whether the throttle should not render haptics at all
(value = 0), render only uniform stiffness (value = 1), or render rich
vibrations (value = 2).
- Brake Effect Intensity: The strength of the forces and vibrations produced by
the brake.
- Throttle Effect Intensity: The strength of the forces and vibrations
produced by the throttle.
- Inverse Brake Grip Loss Sensitivity: The lower this value (i.e., closer to 0),
the more the haptic feedback conveys grip loss for the brakes. The higher
(i.e., closer to 100), the less responsive the haptic feedback is for grip loss
for the brakes.
- Inverse Throttle Grip Loss Sensitivity: The lower this value (i.e., closer to
0), the more the haptic feedback conveys grip loss for the throttle. The
higher (i.e., closer to 100), the less responsive the haptic feedback is for
grip loss for the throttle.
- Turning Acceleration Scale: The higher this value, the more of an
effect turning will have on producing haptic feedback. For example, turning
sharply will cause the throttle trigger to become more stiff or vibrate if
the turning acceleration scale is high.
- Forward Acceleration Scale: The higher this value, the more of an
effect forward movement will have on haptic feedback. For example,
sharp accelerations will cause the throttle trigger to become more stiff or
vibrate if the forward acceleration scale is high.
- Acceleration Limit: This value determines how high the acceleration needs to
get before the throttle haptic feedback renders a constant sensation. If the
car accelerates any more than this, the haptic feedback won't change.

- Minimum Brake Vibration: The minimum frequency at which the brake can
```

```
vibrate. For ideal responses, this value should always be less than
the Maximum Brake Vibration and above 20 percent. A value of 0 means no
vibrations will be felt.
- Maximum Brake Vibration: The maximum frequency at which the brake can
vibrate. This value should always be kept above the Minimum Brake Vibration. A
value of 0 means no vibrations will be felt.
- Minimum Throttle Vibration: The minimum frequency at which the
throttle can vibrate. For ideal responses, this value should always be less than
the Maximum Throttle Vibration and above 20 percent. A value of 0 means
no vibrations will be felt.
- Maximum Throttle Vibration: The maximum frequency at which the
throttle can vibrate. This value should always be kept above the Minimum Throttle
Vibration. A value of 0 means no vibrations will be felt.

- Minimum Brake Resistance: The minimum resistance that the brake can produce.

- Maximum Brake Resistance: The maximum resistance that the brake can produce.

- Minimum Throttle Resistance: The minimum resistance that the
throttle can produce.
- Maximum Throttle Resistance: The maximum resistance that the
throttle can produce.

# Some example interactions

User: I want the car to feel heavy and unresponsive, not very
sensitive to slipping and without too many vibrations.

Initial Settings:
brake trigger mode: 2
throttle trigger mode: 2
brake effect intensity: 60
throttle effect intensity: 60
inverse brake grip loss sensitivity: 20
inverse throttle grip loss sensitivity: 20
turning acceleration scale: 80
forward acceleration scale: 100
acceleration limit: 80
minimum brake vibration: 40
maximum brake vibration: 60
minimum throttle vibration: 40
maximum throttle vibration: 60
minimum brake resistance: 20
maximum brake resistance: 40
minimum throttle resistance: 20
maximum throttle resistance: 40
```

Response:
‘‘‘json
{
    "settings_list": ["brake effect intensity", "throttle
    effect intensity", "inverse brake grip loss sensitivity",
    "inverse throttle grip loss sensitivity", "turning acceleration
    scale", "acceleration limit", "minimum brake vibration", "maximum
    brake vibration", "minimum throttle vibration", "maximum
    throttle vibration", "minimum brake resistance", "maximum
    brake resistance" "minimum throttle resistance", "maximum
    throttle resistance"],
    "old_value": [60, 60, 20, 20, 80, 80, 40, 60, 40, 60, 20, 40, 20, 40],
    "new_value": [90, 90, 90, 90, 30, 30, 10, 20, 10, 20, 70, 100, 70,
    100],
    "justification": "To make the car feel heavy and unresponsive, the
    force feedback on the controller is increased. To achieve
    this, both the brake effect intensity and throttle effect
    intensity are increased from 60% to 90%, and the minimum brake
    resistance, maximum brake resistance, minimum throttle
    resistance and maximum throttle resistance values are increased
    from 20%, 40%, 20% and 40% to 70%, 100%, 70% and 100% respectively.
    To make the car feel more unresponsive, a lot of vibrations
    should not be felt while turning the car, especially when grip
    loss is encountered. Towards this, the minimum and maximum
    throttle and brake vibrations are reduced for low frequency
    vibrations, and the inverse grip loss sensitivity values are
    also increased for both the brake and throttle to delay the sensation
    of losing grip. To further make the car feel heavy, the turning
    acceleration scale is reduced to 30% to reduce the vibrations
    generated due to turning. Lastly, the acceleration limit is
    reduced to 30% to make the throttle response bottom out early during
    acceleration instead of ramping up further."
}
‘‘‘

User: I want the car to feel grippy and responsive, like a formula one racing
car.

Initial Settings:
brake trigger mode: 2
throttle trigger mode: 2
brake effect intensity: 60
throttle effect intensity: 60
inverse brake grip loss sensitivity: 20
inverse throttle grip loss sensitivity: 20
turning acceleration scale: 80
forward acceleration scale: 100

```
acceleration limit: 80
minimum brake vibration: 40
maximum brake vibration: 60
minimum throttle vibration: 40
maximum throttle vibration: 60
minimum brake resistance: 20
maximum brake resistance: 40
minimum throttle resistance: 20
maximum throttle resistance: 40

Response:
‘‘‘json
{
    "settings_list": ["brake effect intensity", "throttle
    effect intensity", "inverse brake grip loss sensitivity",
    "inverse throttle grip loss sensitivity", "turning acceleration
    scale", "acceleration limit", "minimum brake vibration", "maximum
    brake vibration", "minimum throttle vibration", "maximum
    throttle vibration", "minimum brake resistance", "maximum
    brake resistance" "minimum throttle resistance", "maximum
    throttle resistance"],
    "old_value": [60, 60, 20, 20, 80, 80, 40, 60, 40, 60, 20, 40, 20, 40],
    "new_value": [90, 90, 0, 0, 100, 100, 60, 80, 60, 80, 70, 100, 70,
    100],
    "justification": "To make the car feel responsive, the force
    feedback and vibrations on the controller are increased. To achieve
    this, both the brake effect intensity and throttle effect
    intensity are increased from 60% to 90%, and the minimum brake
    resistance, maximum brake resistance, minimum throttle
    resistance and maximum throttle resistance values are increased
    from 20%, 40%, 20% and 40% to 70%, 100%, 70% and 100% respectively.
    To make the car feel more grippy, a lot of vibrations should be
    felt while turning the car, especially since a formula one car rides
    low and close to the road, particularly when grip loss is
    encountered. Towards this, the minimum and maximum throttle and
    brake vibrations are increased for strong, high frequency
    vibrations, and the inverse grip loss sensitivity values are
    also decreased for both the brake and throttle to deliver the
    sensation of losing grip early. To further make the car feel light
    and nimble, the turning acceleration scale is increased to 100% to
    increase the vibrations generated due to turning. Lastly, the
    acceleration limit is increased to 100% to make the throttle
    response ramp up as the car reaches its top speed."
}
‘‘‘

User: I want to feel some vibrations when I accelerate, especially when
```

```
I'm slipping, but I don't want to feel any vibrations on the brake. Oh also, both
pedals should be stiff, with the brake being stiffer than the
accelerator.

Initial Settings:
brake trigger mode: 2
throttle trigger mode: 2
brake effect intensity: 60
throttle effect intensity: 60
inverse brake grip loss sensitivity: 100
inverse throttle grip loss sensitivity: 100
turning acceleration scale: 50
forward acceleration scale: 50
acceleration limit: 255
minimum brake vibration: 0
maximum brake vibration: 0
minimum throttle vibration: 0
maximum throttle vibration: 0
minimum throttle resistance: 0
maximum throttle resistance: 0
minimum brake resistance: 0
maximum brake resistance: 0


Response:
'''json
{
    "settings_list": ["brake effect intensity", "throttle
    effect intensity", "inverse brake grip loss sensitivity",
    "inverse throttle grip loss sensitivity", "turning acceleration
    scale", "acceleration limit", "minimum brake vibration", "maximum
    brake vibration", "minimum throttle vibration", "maximum
    throttle vibration", "minimum brake resistance", "maximum
    brake resistance" "minimum throttle resistance", "maximum
    throttle resistance"],
    "old_value": [60, 60, 20, 20, 80, 80, 40, 60, 40, 60, 20, 40, 20, 40],
    "new_value": [90, 90, 0, 0, 100, 100, 60, 80, 60, 80, 70, 100, 70,
    100],
    "justification": "To make the car feel responsive, the force
    feedback and vibrations on the controller are increased. To achieve
    this, both the brake effect intensity and throttle effect
    intensity are increased from 60% to 90%, and the minimum brake
    resistance, maximum brake resistance, minimum throttle
    resistance and maximum throttle resistance values are increased
    from 20%, 40%, 20% and 40% to 70%, 100%, 70% and 100% respectively.
    To make the car feel more grippy, a lot of vibrations should be
    felt while turning the car, especially since a formula one car rides
```

```
    low and close to the road, particularly when grip loss is
    encountered. Towards this, the minimum and maximum throttle and
    brake vibrations are increased for strong, high frequency
    vibrations, and the inverse grip loss sensitivity values are
    also decreased for both the brake and throttle to deliver the
    sensation of losing grip early. To further make the car feel light
    and nimble, the turning acceleration scale is increased to 100% to
    increase the vibrations generated due to turning. Lastly, the
    acceleration limit is increased to 100% to make the throttle
    response ramp up as the car reaches its top speed."
}
'''

# The existing values of the settings listed above are shown in the following

<Settings list appended here>

<User preference profile appended here>
```

A.2.2 *Stage 2: Updating User Preference Profile*

```
You are a helpful racing engineer analyst who processes
conversation data between a racing driver and their engineer to
summarize your understanding of the driver's haptic
feedback preferences for their car.
In addition, you are given an existing summary of the driver's preferences to
begin with. The summary you generate should add to this existing document, WITHOUT
EXCEEDING 500 WORDS TOTAL.

# Guidelines to follow
- You should pay attention to the driver's queries to the race engineer, and keep
track of the specific words they use to describe the haptic feedback. For
example, adjectives such as 'bumpy', 'slipping', 'sliding', 'strong', 'weak'
etc. are important to understand the words that the driver uses to describe
different haptic effects.
- You should ensure that the summary generated DOES
NOT contain any information that cannot directly be inferred from the
conversation's transcript.
- Make sure that the generated summary also contains as much
information from the existing summary so as to not lose information
unnecessarily.
- If the existing summary contains something that conflicts with the
conversation data, prioritize the more recent data (i.e., the
conversation data) when forming the final summary.
- Make sure that the generated summary is LESS THAN 500 words.
```

```
- If the existing summary provided to you is empty, ignore it and rely only on
the conversation data to generate your summary.
- ONLY return the summary as a JSON object with one field titled 'summary'.
- DO NOT include any paragraph breaks in the summary. Make sure it is
all in one paragraph.

# The following is the existing summary for the user's profile, along with
the most recent chat transcript.

<User preference profile appended here>

<Past 3 chat interactions appended here>
```

B

INITIAL SETTINGS CONFIGURATION

The following initial settings were used with TacTalk. The initial seed was an important hyperparameter, since it influenced TacTalk's parameter changes in the long run, and could potentially cause deadlocks in cases where TacTalk modified a redundant parameter repeatedly without looking at other parameters. For instance, if the effect intensity was set to 0 in the beginning, TacTalk would not increase it before modifying attributes such as vibration and resistance.

```
brake trigger mode: 2
throttle trigger mode: 2
brake effect intensity: 60
throttle effect intensity: 60
inverse brake grip loss sensitivity: 100
inverse throttle grip loss sensitivity: 100
turning acceleration scale: 50
forward acceleration scale: 50
acceleration limit: 100
minimum brake vibration: 0
maximum brake vibration: 0
minimum throttle vibration: 0
maximum throttle vibration: 0
minimum throttle resistance: 0
maximum throttle resistance: 0
minimum brake resistance: 0
maximum brake resistance: 0
```

C

TECHNICAL EVALUATION DATASET

C.1 PROMPTS USED FOR TECHNICAL EVALUATION

The following prompts were used for TacTalk's technical evaluation. All of these prompts follow a 'I would like the car to feel like a ...' structure, followed by the name or defining traits of a car. These names/traits were grouped into high, average and low performance categories as shown in the following:

```
# High Performance Cars
Ferrari LaFerrari
Formula 1 Car
Red Bull RB20 F1 Car
Lamborghini GT3 Car
Toyota Le Mans Hyper Car
Super Formula Car
Nissan GTR Skyline
Formula 1 McLaren MCL38
Ferrari 458 Italia
Toyota Supra
Porsche 911
Koenigsegg Jesko
McLaren P1
Lamborghini Gallardo
Mercedes AMG One
Bugatti Veyron
Pagani Huayra

# Average Performance Cars
Hyundai Elantra
Honda Civic
Toyota Corolla
Honda Accord
Toyota Camry
Volkswagen Jetta
Volkswagen Golf
Mid-Size Family Sedan
Small Hatchback
Fiat Panda
Chevrolet Spark
Ford Focus
Mini Cooper
Suzuki Swift
```

```
Kia Seltos
Nissan Rogue
Honda CRV

# Low Performance Cars
Garbage Truck
Semi Truck
Hearse
Slow, Old Car
Tacoma Pickup Truck
Chevrolet Silverado
Schoolbus
Sambabus
Limousine
Soapbox Racer
Fiat 500
Ford Model T
Renault Dauphine
Subaru 360
Ford Pinto
City Rover
Underpowered sluggish car
```

D

SCREENING QUESTIONNAIRE

The following screening questionnaire was used to select participants for the TacTalk user study. Participants were required to 1) be above 18 years of age, 2) know at least 4 terms, and 3) have played at least 1 game not including Mario Kart.

```
Q1. Enter your email:
Q2: Please state whether your age is 18 or above:
Q3: Please select the driving-related words/phrases that you are familiar with:
- Traction/Grip
- Throttle
- Brake
- Drifting
- Burnout
- Trail Braking
- Differential
- Brake Bias
- Centring Force
- RPM (Revs Per Minute)
- Aquaplaning
Q4: Please select the racing games that you have played:
- Mario Kart
- Need For Speed (series)
- Burnout (series)
- Forza Horizon (series)
- Driveclub
- Forza Motorsport (series)
- Gran Turismo (series)
- F1 (series)
- Assetto Corsa (series)
- BeamNG
```