Easy to Hear, Easy to Select?: Investigating Speech Rate as a Potential Deceptive Pattern in Non-Native Users' Decision-Making

Yuichiro Kinoshita
Meiji University
Nakano, Tokyo, Japan
kinoshita@meiji.ac.jp

Satoshi Nakamura
Meiji University
Nakano, Tokyo, Japan
satoshi@snakamura.org

Abstract

Voice user interfaces (VUIs) increasingly mediate everyday decision-making, raising concerns about deceptive patterns (dark patterns) in voice-based interactions. Prior work has shown that voice features can influence users' choices, but such effects have mainly been examined in users' native languages. In practice, however, users may need to interact with VUIs in a non-native language, where subtle differences in speech presentation may be harder to resist. This study examined speech rate as a potential deceptive pattern in non-native voice interactions. We conducted three experiments with Japanese non-native English speakers to investigate whether manipulating the speech rate of synthetic speech when presenting options could bias users' selections. The results showed that speech rate manipulation did not clearly bias selections. However, they also indicated that further investigation is needed when the difference in speech rate is more pronounced.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Auditory feedback**.

Keywords

Deceptive Patterns, Dark Patterns, Voice Interaction, Non-Native Language

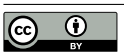
ACM Reference Format:
Yuichiro Kinoshita and Satoshi Nakamura. 2026. Easy to Hear, Easy to Select?: Investigating Speech Rate as a Potential Deceptive Pattern in Non-Native Users' Decision-Making. In *Proceedings of the 14th Nordic Conference on Human-Computer Interaction (NordiCHI '26), October 03–07, 2026, Vaasa, Finland*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3829807.3829876

1 Introduction

Voice user interfaces (VUIs) are becoming increasingly prevalent. Beyond household devices such as smart speakers, the potential of voice-operated automated teller machines [27] and voice-interactive robots in shopping malls [28] is being explored, suggesting that VUIs will be used in an even wider range of settings. As VUIs become more prevalent, there is a concern that deceptive patterns[1] (DPs) may emerge in VUIs [43, 51]. DPs are user interface designs that manipulate users' decision-making or behavior to benefit service providers. For example, obstruction of subscription cancellation processes [60] and high-demand messages on shopping sites [41, 67] are designs that constitute DPs.

Research on DPs has primarily focused on graphical user interfaces (GUIs), revealing that DPs are widely used in websites [41, 46] and mobile applications [13, 26, 44]. Studies have also addressed the detection [10, 33, 39, 47, 77] of and countermeasures [58, 59, 67] against DPs in GUIs. While extensive research has been conducted on DPs in screen-based interfaces, few studies have examined DPs in VUIs, despite the possibility that they may also be employed in VUIs [43, 51]. Owens et al. [51] discussed how DPs may be employed in voice assistants, illustrating the possibility of DPs such as presenting information favorable to the company in a louder voice while delivering unfavorable information in a quieter voice. Dubiel et al. [15] demonstrated that the fidelity of synthetic speech could potentially manipulate users' decision-making. Furthermore, Leschanowsky et al. [36] indicated that variations in speech rate when presenting options in a consent choice scenario could influence users' selections.

Although speech rate has been shown to potentially influence users' decision-making [36], sufficient investigation has not yet been conducted. Moreover, to the best of our knowledge, research on DPs in VUIs has only examined the effects on decision-making when information is presented in users' native language [15, 36, 82]. However, VUIs still do not adequately accommodate diverse users [29, 73], and there are situations in which users interact with VUIs in a non-native language. In addition, there is a possibility that DPs which hinder users' understanding of important content or functions by using a language different from their native language [26, 63] could also be employed in VUIs, potentially creating situations in which users are compelled to interact with VUIs in a non-native language. Since listening comprehension and understanding of meaning are more difficult in a non-native language [8, 17], the deliberate manipulation of speech rate may have a greater impact on users' decision-making in such situations. For example, a service provider could slowly present the option they want users to select, while presenting the other options at a speech



NordiCHI '26, Vaasa, Finland

ACM ISBN 979-8-4007-2373-5/26/10
https://doi.org/10.1145/3829807.3829876

[1] Brignull [5] originally coined the term "Dark Patterns" in 2010. However, due to concerns about racial connotations, Brignull later changed the terminology to "Deceptive Patterns" (or "Deceptive Design"). Accordingly, this paper uses the term "Deceptive Patterns."

rate that is extremely fast for non-native users, thereby potentially steering users toward the slowly presented option.

VUIs are interfaces that support both voice input from users and voice output using synthetic speech [62]. This paper focused specifically on voice output and examined the influence of speech rate manipulation on decision-making as a potential DP that may be employed in VUIs in the future.

The research questions (RQs) of this paper are as follows:

RQ1: Can users' selections be biased by deliberately manipulating speech rate when options are presented via synthetic speech in a non-native language?

RQ2: Does the effect on users' selections change when speech rate differences between options are varied?

To answer RQ1, we examined whether deliberately manipulating speech rate can bias users' selections when options are presented via synthetic speech in a non-native language. To address RQ2, we conducted the experiment using a wider variety of speech rates than in a previous study [36], aiming to deepen the understanding of how speech rate affects users' selections. Specifically, we conducted experiments with Japanese participants who were non-native speakers of English, using two options presented at different speech rates. After a preliminary investigation of how speech rate affects users' selections, we further examined this effect in greater detail.

The main contributions of this paper are as follows:

- We investigated whether manipulating the speech rate of synthetic speech in a non-native language could bias users' selections, and found that such manipulation did not clearly bias users' selections.
- We examined whether varying the difference in speech rate produces different effects on users' selections, and showed that while no significant differences in effects were observed across conditions, further investigation is needed for cases where the speech rate difference is more pronounced.
- We highlighted the need for VUIs to accommodate diverse users and discussed directions for future research on DPs in VUIs.

2 Related Work

2.1 Deceptive Patterns

Since Brignull [5] introduced the term deceptive patterns (dark patterns) in 2010, many DPs have been identified and categorized. Mathur et al. [41] identified 1,818 DPs across more than 11,000 shopping websites, and Di Geronimo et al. [13] revealed that over 95% of 240 popular mobile applications employed DPs. A weak correlation between a service's popularity and the use of DPs has been reported [24]. In addition, DPs in specific services such as games [48, 80, 81], video streaming services [9, 61], and social networking services [44, 66] have also been investigated. Furthermore, DPs are also present in privacy-related consent choices [4, 49, 68]. Gray et al. [21] proposed an ontology that hierarchically organizes DP types based on previously identified DPs and existing classification schemes.

While users feel frustrated when encountering DPs [19, 38], it has been shown that mild, non-aggressive DPs do not provoke emotional backlash [37]. Additionally, users find it difficult to recognize DPs [13], and even when they are aware of the presence of DPs, their behavior can still be manipulated by DPs [3]. Therefore, Bongard-Blanchy et al. [3] highlighted the need for interventions against them. In response, several approaches have been explored. Chen et al. [10] developed a system that can automatically detect them in mobile applications with 93% accuracy. In addition, methods using deep learning to automatically detect DPs on e-commerce websites have also been proposed [11]. Furthermore, research has been conducted on intervention methods to mitigate the effects of DPs [67], visual countermeasures against DPs [59], and educational platforms to enhance users' resistance to DPs [78].

Research on DPs has primarily focused on screen-based interfaces. However, the need to broaden the scope beyond screen-based interfaces has been discussed. Lacey and Caudwell [35] pointed out that the cuteness of robots could function as a DP, and Kowalczyk et al. [34] revealed that DPs also exist in Internet-of-Things devices. In addition, Greenberg et al. [22] explored the possibility of DPs in physical space design. Moreover, DPs exist in VUIs, and their use in VUIs may expand in the future [51]. This paper focused on voice output in VUIs and examined whether users' selections could be biased by manipulating speech rate when options were presented in a non-native language.

2.2 Influencing User Decision-Making in Voice User Interfaces

Because VUIs rely on voice-based interaction, a different modality from GUIs, they may influence users' behavior in ways not previously seen in GUIs. Buchta et al. [6] showed that in their experiment, paid items tended to be selected more frequently when participants operated through a VUI than through a GUI. On the other hand, it has been shown that proactive feedback from a conversational voice agent may slow down decision-making and encourage users to deliberate [14]. Dubiel et al. [15] found that when options were presented using two different voices, options presented with a voice that had greater pitch variation and a faster speech rate were more likely to be selected. Furthermore, Pias et al. [53] showed that manipulating characteristics of synthetic speech, such as tone, age, and gender, can increase persuasiveness and influence users' purchasing decisions. Zhang et al. [82] also demonstrated that manipulating voice features could steer people's behavior.

Regarding verbal consent in voice assistants, Seymour et al. [65] proposed recommendations for voice-based consent, taking into account factors such as time constraints on responses and the inherent difficulty of comprehending spoken information. Leschanowsky et al. [36] compared screen-based and voice-based permission requests and showed that voice-based permission requests were less likely to be accepted. They also examined the effect on decision-making by varying the speech rate when saying "Accept" and "Decline" in voice-based permission requests, and reported a tendency for participants to be more likely to accept when either Accept or Decline was spoken faster, regardless of which one was sped up. Our study builds on the work of Leschanowsky et al. [36]. While they examined the effects of speech rate variations in a native language on decision-making, our study differs in that we investigated the effects in a non-native language. In addition, whereas

Leschanowsky et al. [36] used two voices at different speeds, we conducted experiments using a total of four different speech rates, aiming to further the understanding of how speech rate variations affect decision-making.

2.3 Interacting with Voice User Interfaces in a Non-Native Language

A key challenge is that VUIs do not adequately accommodate users from diverse cultural backgrounds [29, 73]. Users find it difficult to use VUIs in a non-native language [54] and experience a high mental workload [75]. Song et al. [69] compared English speech recognition accuracy on Amazon's Alexa between native English speakers and non-native English speakers whose native language is Korean. They found that the recognition accuracy was 98% for native speakers and only 55% for non-native speakers. In addition, non-native speakers require longer thinking time than native speakers when giving commands to VUIs and feel anxious about their pronunciation [76]. These are issues related to users' speech directed at VUIs. However, problems also exist on the listening side, when users receive voice output from VUIs.

When listening in a non-native language, users may fail to catch words they already know or may be unable to understand the intended message [8, 17]. It is also known that some people have difficulty distinguishing certain sounds (e.g., /l/ and /r/) [18, 23]. Based on these findings, when users listen to voice output from VUIs in a non-native language, their decision-making may be more susceptible to manipulation through voice features, due to the human tendency to prefer more fluent information (e.g., information presented at an easily comprehensible speech rate) over less fluent information (e.g., information presented at a speech rate that is too fast to comprehend) [55]. In this paper, we focused on speech rate as one of the voice features and investigated whether varying the speech rate of VUI voice output in a non-native language could bias users' selections.

3 Method

3.1 Overview of the Three Experiments

We conducted three experiments to answer *RQ1: Can users' selections be biased by deliberately manipulating speech rate when options are presented via synthetic speech in a non-native language?*, and *RQ2: Does the effect on users' selections change when speech rate differences between options are varied?* An overview of the experiments is as follows:

Experiment 1 (Section 4) aimed to identify question and option sets for use in subsequent experiments examining whether speech rate manipulation biases users' selections. This screening process was intended to minimize the influence of factors other than speech rate (e.g., substantial differences in popularity between options) on decision-making. In Experiment 1, we used a total of 15 questions (Table 1), each with two options, and presented options via speech in a non-native language without manipulating speech rate to identify question and option sets that showed no selection bias.

Experiment 2 (Section 5) examined whether participants' selections could be biased by presenting two options at different speech rates in a non-native language, using the questions and options identified in Experiment 1. This experiment served as a preliminary investigation, using speech at 113 words per minute (WPM) and 139 WPM.

Experiment 3 (Section 6) added speech at 165 WPM and 191 WPM to the 113 WPM and 139 WPM used in Experiment 2. Two options were presented at different speeds in a non-native language across a total of six combinations. We used the same question and option sets as in Experiments 1 and 2 and examined the differences in effects on users' selections when speech rate differences were varied.

This research was approved by our institutional ethics board. The following sections describe the experimental design common to all three experiments.

3.2 Scenario and Task

Compared with decisions that are considered economically high involvement, such as buying a car, users may be more susceptible to influence by VUIs in low involvement decisions that do not require careful deliberation, such as purchasing food [15, 45, 71]. Therefore, we adopted a scenario involving low involvement choices made while traveling abroad. By setting this scenario, we could create diverse situations requiring choices and made it easier for participants to imagine situations in which they would face interactions in a non-native language. We instructed participants:

> Imagine that you are going on an overseas trip alone. During the trip, you will encounter 15 events, and for each event, a question will be presented. Please answer each question by selecting one of two options.

We presented participants with a total of 15 questions, each with two options. As two options have been shown to be optimal for selection tasks involving multiple attributes [2, 74], we used two options. Participants were native Japanese speakers with English as a non-native language. We presented questions on screen in Japanese to ensure that participants could reliably understand their meaning, while we presented options via English speech to examine the effect of speech rate manipulation in a non-native language on selections. Participants responded to each question by clicking to select one of the two options.

3.3 Design of Questions and Options

Table 1 shows the list of questions and options used in the experiments. The 15 questions were created by the authors through discussion, arranged in chronological order of travel experiences. Since vocabulary knowledge strongly influences listening performance [70], most of the English words used in the options were selected from vocabulary learned by the end of high school in Japan. Each option consisted of five to seven words, and both options within the same question contained the same number of words. To minimize differences in popularity between options and to prevent participants from making judgments based on only part of the audio, options within the same question were designed to differ in only a few words.

Table 1: Questions and options used in the experiments. Questions were presented in Japanese on screen, and options were presented via English synthetic speech.

Questions (English translation, originally in Japanese)	Options (originally in English)
1. You are on the plane. You can choose a snack and drink set. Which do you choose?	Chocolate cookies with a cup of tea Chocolate cookies with a cup of coffee
2. You have arrived at your destination. You enter a restaurant for lunch. Which dish do you choose?	Grilled chicken with rice and salad Grilled fish with rice and salad
3. After lunch, you are going to visit a natural heritage site. Which do you choose?	Beautiful natural heritage with forests Beautiful natural heritage with waterfalls
4. You are checking in at the hotel. You are told you can choose the room location. Which view do you choose?	Room with a city view Room with a garden view
5. After dinner, you return to the hotel. You decide to attend a short concert at the hotel. Which do you choose?	Short concert with jazz music Short concert with classical music
6. It is the next day. You have breakfast at the hotel. Which bread do you choose?	Bread with butter and jam Croissant with butter and jam
7. After breakfast, you are going to visit a museum. Which do you choose?	Local museum to learn about history Local museum to learn about culture
8. You are having pasta for lunch. Which do you choose?	Pasta with cream sauce and bacon Pasta with tomato sauce and bacon
9. After lunch, you are going to visit a garden. Which do you choose?	Garden with roses and trees Garden with lavender and trees
10. You are choosing a restaurant for dinner. Which do you choose?	Traditional restaurant serving Western foods Modern restaurant serving Western dishes
11. After dinner, you are going to see the night view. Which view do you choose?	City view from a quiet seaside City view from a tall tower
12. It is the last day. You choose a breakfast menu at the hotel. Which do you choose?	Breakfast with eggs and toast Breakfast with fruits and yogurt
13. After breakfast, you are going to see an opera. Which do you choose?	Opera set in ancient times Opera set in modern times
14. You have arrived at the airport. You have some time before boarding, so you visit a cafe in the airport. Which do you choose?	Popular cafe with delicious parfaits Popular cafe with delicious cakes
15. You are on the plane. You are going to watch a movie. Which do you choose?	Comedy movie with a touching story Comedy movie with a thrilling story

3.4 Audio Generation

We generated the English synthetic speech used for presenting options using Google Cloud Text-to-Speech[2] (TTS). We chose Google Cloud TTS because it allows speech rate to be controlled by specifying a numerical value for the speaking rate parameter. Among the available voices, we used "en-US-Wavenet-H,"[3] an American English female voice. We adopted a female voice because female voices are widely used in commercial voice assistants [7]. The average fundamental frequency (F_0) of the voice was 213.3 Hz (SD = 37.1 Hz). All audio was structured so that the first option was spoken after the word "Option 1," followed by the second option spoken after the word "Option 2." For example, the audio was presented as: *"Option 1, Room with a city view, Option 2, Room with a garden view."*

3.5 Experimental System and Procedure

We implemented a web system for the experiment (Figure 1). Participants accessed our system via PC. The system first displayed a page explaining the overview of the experiment, the participation requirements, and consent to participate. The system also provided a sample audio clip to confirm that participants could play audio. After participants agreed to participate, the system presented the experimental scenario, and on the following pages, displayed each question one at a time in Japanese text on screen. Options were displayed on screen as "Option 1" and "Option 2." When participants pressed the audio playback button, the system presented

[2]https://cloud.google.com/text-to-speech
[3]https://docs.cloud.google.com/text-to-speech/docs/audio/en-US-Wavenet-H.wav

two options via synthetic speech. The audio could be replayed any number of times. The system did not allow participants to select an option until the audio playback was completed. After listening to the speech, participants selected one of the options as their answer to the question. The order in which the options were presented was randomized each time.

After participants made selections for all 15 questions, the system transitioned to a questionnaire page. The questionnaire included attention check question to verify that participants were reading carefully, and participants' gender and age were collected at the end of the questionnaire. Other questionnaire items are described in the respective experiment sections. After participants completed the questionnaire, the system notified them that the experiment had ended.

3.6 Participant Recruitment

We conducted online experiments and limited participation to PC users. We used Yahoo! Crowdsourcing[4] (YCS) for participant recruitment. YCS is one of the widely used crowdsourcing services in Japan, and its high reliability and usefulness have been demonstrated [64]. Since YCS allows recruitment to be restricted to male or female participants[5], we recruited participants in two separate rounds for each experiment to ensure gender balance. In one round, recruitment was limited to male participants, and in the other, to female participants, with an equal number recruited in each round. All experiments were expected to take five minutes, and based on local wage standards, participants were compensated with 100 yen worth of points credited to their YCS accounts. Participants were required to be adults who were native speakers of Japanese and non-native speakers of English.

4 Experiment 1: Identifying Question and Option Sets with No Significant Selection Bias

To examine the potential for biasing users' selections through speech rate manipulation, our experiments needed to be designed so that factors other than speech rate had as little influence on users' selections as possible. Therefore, we first conducted a selection experiment in which options were presented via speech at the same speech rate, and identified question and option sets that showed no significant selection bias. In subsequent experiments (Section 5 and Section 6), we used the question and option sets identified in this section to examine whether speech rate manipulation could bias users' selections.

4.1 Experimental Setup

The experiment used the scenario and system described in Section 3. As the purpose of this experiment was to identify question and option sets in which selection bias was not caused by factors other than speech rate, the two options presented via speech were delivered at the same speech rate.

Although the average speech rate of native English speakers is 152–170 WPM [79], this may feel fast for non-native speakers. Therefore, we conducted a pilot test with 13 Japanese students who were non-native speakers of English in our laboratory, and generated speech at an average speech rate of 139 WPM so that non-native speakers could understand the spoken content.

For each of the 15 questions, participants selected one of two options presented at the same speech rate. In the post-experiment questionnaire, participants rated their comprehension of the content of the spoken options on a 7-point Likert scale (1 = did not understand at all, 7 = understood everything) to verify that they could understand the speech.

4.2 Results

4.2.1 Participants. The first experiment was conducted on November 19, 2025, with 200 participants recruited. After excluding those who failed the attention check, 192 participants (95 males, 96 females, and 1 undisclosed) remained for analysis. Five participants were aged 20–29, 28 were aged 30–39, 49 were aged 40–49, 73 were aged 50–59, 27 were aged 60–69, and 10 were aged 70 or older. The average experiment completion time was 4 minutes and 33 seconds.

4.2.2 Selection Results and Questionnaire. Table 2 shows the selection results for each of the 15 questions. In this study, we set the significance level at 0.05 and conducted a binomial test to examine whether the observed proportions differed from the chance level (50%) in each question. We applied the Holm method for multiple comparisons. The binomial tests showed no significant deviations from chance level in 10 questions (Q3, 5, 6, 7, 8, 10, 11, 12, 13, and 15).

The mean rating for the questionnaire item asking the extent to which participants understood the content of the options presented via speech was 4.33 (SD = 1.44) on a 7-point scale (1 = did not understand at all, 7 = understood everything), suggesting that participants had a moderate level of comprehension and that the speech rate and vocabulary difficulty were generally manageable. In addition, the average number of audio playbacks was 1.27 (SD = 0.51), indicating that participants did not tend to replay the audio multiple times.

5 Experiment 2: Preliminary Investigation of Speech Rate Effects in Non-Native Synthetic Speech

In this section, we preliminarily examined whether users' selections can be biased by manipulating speech rate in non-native synthetic speech, using the question and option sets with no significant selection bias identified in Section 4.

5.1 Experimental Setup

We conducted the experiment using the scenario and experimental system described in Section 3. Following the selection experiment design of Kinoshita et al. [32], we applied speech rate manipulation to only 5 of the 15 questions to prevent participants from noticing the experimental purpose. The five questions were randomly selected from the 10 questions that showed no significant selection bias in the previous experiment (Section 4.2), and speech rate manipulation was applied to Q5, 7, 8, 10, and 11 (Table 1). For the remaining questions, both options were presented at the same speech rate (139 WPM).

[4]https://crowdsourcing.yahoo.co.jp/
[5]At the time of this study, YCS did not support limiting recruitment to other gender categories.

Figure 1: Screenshot of the experimental system. The same system was used across all experiments, with only the audio files changed between experiments.

Since the effect of presenting one option 20% faster than another has been examined [36], in this experiment, we instead presented one option approximately 20% slower than the other option (139 WPM), at 113 WPM. Because Google Cloud TTS allows the speech rate to be changed partway through a single utterance, we configured the speech rate to differ between the two options. For example, for the pair "Option 1, Short concert with jazz music, Option 2, Short concert with classical music," "Short concert with jazz music" was spoken at the faster speech rate (139 WPM), while "Short concert with classical music" was spoken at the slower speech rate (113 WPM). We randomized the order in which the two options were presented for every question. For questions with speech rate manipulation, which option was presented at the slower speech rate was also randomized.

In the post-experiment questionnaire, in addition to the items described in Section 3.5, participants rated their sense of control over their selections on a 7-point Likert scale (1 = no control at all, 7 = complete control), based on the questionnaire by Rohden and Espartel [56]. We included this measure because users may not notice that their decision-making is being steered by DPs in VUIs [15, 36, 82].

We used YCS for participant recruitment. Participants were required to be adults who were native speakers of Japanese and non-native speakers of English, and those who had participated in the previous experiment (Section 4) were prevented from participating in this experiment.

5.2 Results

5.2.1 Participants. We ran the second experiment on November 25, 2025, recruiting 200 participants. Following the exclusion of participants who did not pass the attention check, the final sample consisted of 177 participants (91 males, 85 females, and 1 undisclosed). Eight participants were aged 20–29, 18 were aged 30–39, 49 were aged 40–49, 66 were aged 50–59, 26 were aged 60–69, 9 were aged 70 or older, and 1 was undisclosed. The average experiment completion time was 4 minutes and 44 seconds.

5.2.2 Selection Results and Questionnaire. Table 3 shows the total number of selections for each option in the five questions in which the speech rate was manipulated. Binomial tests revealed significant deviations from chance level (50%) in Q5 and Q10. The bias observed in these two questions may have been influenced by syllable structure, as discussed by Leschanowsky et al. [36]. On the other hand, it cannot be ruled out that the absence of selection bias in the previous experiment (Section 4) was coincidental. Therefore, we conducted subsequent analyses using only Q7, 8, and 11, which showed no significant deviation from chance level in this experiment.

The proportion of selections for the option presented at the faster speech rate (139 WPM) was 52.54% (279/531, 95% CI [48.29%, 56.76%]), and the proportion for the option presented at the slower speech rate (113 WPM) was 47.46% (252/531, 95% CI [43.24%, 51.71%]). A binomial test revealed no significant deviation from chance level (p = .259).

Figure 2 shows the number of questions in which each participant selected the option presented at the faster speech rate and the option presented at the slower speech rate, for the three analyzed questions with speech rate manipulation. Some participants (N = 32) selected the option presented at the faster speech rate in all three questions, while others (N = 27) selected the option presented at the slower speech rate in all three. The distribution showed a slight tendency toward selecting the faster option in two of the three questions.

The average number of audio playbacks was 1.20 (SD = 0.46), similar to that in the previous experiment (Section 4.2). The mean rating for perceived control over one's own decisions (1 = no control at all, 7 = complete control) was 5.12 (SD = 1.31), suggesting that participants generally felt they were in control of their selections.

6 Experiment 3: Effects of Varying Speech Rate Differences in Non-Native Synthetic Speech

In this section, we prepared a wider variety of speech rates and examined whether users' selections can be biased when options are presented at different speech rates via non-native synthetic speech,

Table 2: Selection results for each question in Experiment 1 (Section 4), in which both options were presented at the same speech rate (139 WPM). p-values are the results of binomial tests for selection bias, adjusted using the Holm method.

	Options	p-value	Adj. p-value
Q1	Chocolate cookies with a cup of tea: 64.06% (123/192) Chocolate cookies with a cup of coffee: 35.94% (69/192)	< .001	.002
Q2	Grilled chicken with rice and salad: 77.08% (148/192) Grilled fish with rice and salad: 22.92% (44/192)	< .001	< .001
Q3	Beautiful natural heritage with forests: 53.13% (102/192) Beautiful natural heritage with waterfalls: 46.88% (90/192)	.427	1.000
Q4	Room with a city view: 63.02% (121/192) Room with a garden view: 36.98% (71/192)	< .001	.005
Q5	Short concert with jazz music: 54.69% (105/192) Short concert with classical music: 45.31% (87/192)	.220	1.000
Q6	Bread with butter and jam: 52.08% (100/192) Croissant with butter and jam: 47.92% (92/192)	.614	1.000
Q7	Local museum to learn about history: 51.56% (99/192) Local museum to learn about culture: 48.44% (93/192)	.718	1.000
Q8	Pasta with cream sauce and bacon: 57.29% (110/192) Pasta with tomato sauce and bacon: 42.71% (82/192)	.051	.460
Q9	Garden with roses and trees: 61.46% (118/192) Garden with lavender and trees: 38.54% (74/192)	.002	.020
Q10	Traditional restaurant serving Western foods: 56.77% (109/192) Modern restaurant serving Western dishes: 43.23% (83/192)	.071	.496
Q11	City view from a quiet seaside: 57.29% (110/192) City view from a tall tower: 42.71% (82/192)	.051	.460
Q12	Breakfast with eggs and toast: 53.13% (102/192) Breakfast with fruits and yogurt: 46.88% (90/192)	.427	1.000
Q13	Opera set in ancient times: 58.33% (112/192) Opera set in modern times: 41.67% (80/192)	.025	.250
Q14	Popular cafe with delicious parfaits: 65.63% (126/192) Popular cafe with delicious cakes: 34.38% (66/192)	< .001	< .001
Q15	Comedy movie with a touching story: 56.25% (108/192) Comedy movie with a thrilling story: 43.75% (84/192)	.097	.580

and whether the magnitude of speech rate differences affects the degree of influence on selections.

6.1 Experimental Setup

We conducted the experiment using the experimental scenario, questions, and options described in Section 3. Using the implemented system (Figure 1), questions were presented in Japanese on screen, and options were presented via English speech.

To prevent participants from noticing the purpose of the experiment, we applied speech rate manipulation to 5 of the 15 questions. The five questions (Q5, 7, 8, 10, and 11) were the same as those used in the experiment in Section 5. For the remaining 10 questions, both options were presented at the same speech rate (139 WPM), and the data from these 10 questions were not used in the analysis.

In the experiment in Section 5, we examined the effects of speech rate manipulation on users' selections using the combination of 113 WPM and 139 WPM. In this section, we additionally generated speech at 165 WPM and 191 WPM. With a total of four different speech rates, there were six possible combinations for presenting two options. The speech rate difference between the two options was set to either 26, 52, or 78 WPM, depending on the condition. We randomly assigned one of these six conditions to each participant. For all questions, the order in which the two options were presented was randomized each time, and for questions with speech rate manipulation, which option was presented at the slower speech rate was also randomized each time.

In the post-experiment questionnaire, in addition to the items described in Section 3.5, participants rated their perceived control

Table 3: Proportion of selections for each option in the five speed-manipulated questions, and binomial test results in Experiment 2 (Section 5).

	Options	p-value	Adj. p-value
Q5	Short concert with jazz music: 64.97% (115/177) Short concert with classical music: 35.03% (62/177)	< .001	.001
Q7	Local museum to learn about history: 58.76% (104/177) Local museum to learn about culture: 41.24% (73/177)	.024	.191
Q8	Pasta with cream sauce and bacon: 41.24% (73/177) Pasta with tomato sauce and bacon: 58.76% (104/177)	.024	.191
Q10	Traditional restaurant serving Western foods: 62.71% (111/177) Modern restaurant serving Western dishes: 37.29% (66/177)	< .001	.008
Q11	City view from a quiet seaside: 49.72% (88/177) City view from a tall tower: 50.28% (89/177)	1.000	1.000

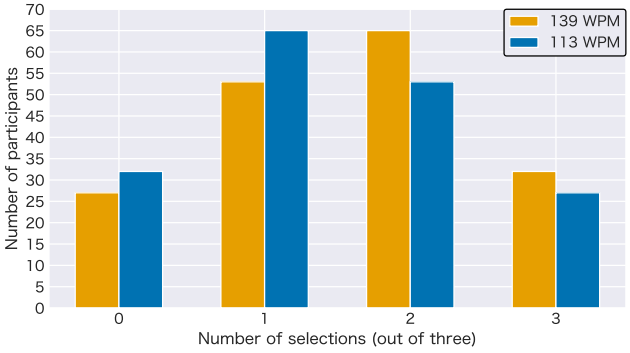


Figure 2: Distribution of the number of faster (139 WPM) and slower (113 WPM) option selections per participant for the three analyzed speed-manipulated questions in Experiment 2 (Section 5).

over their selections using a 7-point Likert scale (1 = no control at all, 7 = complete control). In addition, to assess participants' English proficiency, we used a subset of the LEAP-Q [40], in which participants self-rated their speaking, listening, and reading abilities on an 11-point scale from 0 (none) to 10 (perfect). The LEAP-Q is a questionnaire for measuring language proficiency whose validity has been confirmed [30, 40], and it is used in various fields such as psychology and linguistics.

We used YCS for recruiting participants. Participants were required to be adults who were native speakers of Japanese and non-native speakers of English, and those who had participated in the previous experiments (Section 4 and Section 5) were prevented from participating.

6.2 Results

6.2.1 Participants and Questionnaire. The third experiment was carried out on February 20, 2026. Of the 400 participants recruited, those failing the attention check were removed, leaving 353 participants (175 males, 176 females, 1 transgender, and 1 undisclosed) for analysis. The average experiment completion time was 4 minutes and 47 seconds. Ten participants were aged 20–29, 36 were aged 30–39, 101 were aged 40–49, 122 were aged 50–59, 56 were aged 60–69, 27 were aged 70 or older, and 1 was undisclosed. Using the self-ratings of speaking, listening, and reading from a subset of the LEAP-Q [40], we calculated the average of the three ability ratings for each participant. We classified participants with an average of 7 or above as high proficiency [52], those with 4 or above but below 7 as moderate proficiency, and those below 4 as low proficiency [1]. Of the participants, 6 were classified as high proficiency, 25 as moderate proficiency, and 322 as low proficiency.

Table 4 shows the number of participants assigned to each condition, their ratings of perceived control over their selections (1 = no control at all, 7 = complete control), and the number of audio playbacks. We conducted a Kruskal-Wallis test to examine whether there were differences in perceived control ratings across conditions, and a significant difference was found ($H = 15.349, p = .009$). Post-hoc Wilcoxon rank sum tests (Shaffer-corrected) showed that perceived control ratings in the *113 vs. 165 WPM* condition were significantly higher than those in the *113 vs. 139 WPM* condition ($Adj.p = .007, r = .321$). No significant differences were found between the other conditions. In addition, a Kruskal-Wallis test revealed no significant difference in the number of audio playbacks across conditions ($H = 10.086, p = .073$).

6.2.2 Selection Results. We counted the total number of selections for each option across the speed-manipulated questions (Q5, 7, 8, 10, and 11) and conducted binomial tests to examine selection bias. Q10 showed a significant deviation from chance level ($Adj.p < .001$), while the other four questions did not (Q5: $Adj.p = 1.000$, Q7: $Adj.p = .152$, Q8: $Adj.p = 1.000$, Q11: $Adj.p = 1.000$). Subsequent analyses used data from these four questions.

Table 5 shows the proportion of selections for the option presented at the faster speech rate and the option presented at the slower speech rate for each condition. In all conditions, the proportion of slower option selections was higher than that of faster option selections. However, no significant deviation from chance level was found in any of the conditions. The condition with the largest speech rate difference (113 vs. 191 WPM) showed a slightly

Table 4: Number of participants assigned to each condition, perceived control ratings over their selections (1 = no control at all, 7 = complete control), and number of audio playbacks in Experiment 3 (Section 6).

	113 vs. 139	139 vs. 165	165 vs. 191	113 vs. 165	139 vs. 191	113 vs. 191
N of participants	53	55	59	66	57	63
Perceived control	M = 4.34, SD = 1.57	M = 5.18, SD = 1.35	M = 4.85, SD = 1.28	M = 5.27, SD = 1.41	M = 4.68, SD = 1.65	M = 4.86, SD = 1.38
N of playbacks	M = 1.21, SD = 0.54	M = 1.24, SD = 0.49	M = 1.24, SD = 0.56	M = 1.20, SD = 0.47	M = 1.23, SD = 0.56	M = 1.19, SD = 0.51

higher proportion of slower option selections compared to the other conditions. However, a chi-square test showed no significant difference in the proportion of slower and faster option selections across conditions ($\chi^2(5, N = 1,412) = 1.927, p = .859$). We also conducted a binomial generalized linear mixed model (GLMM) with speech rate difference as a fixed effect and participant as a random effect to analyze slower option selections. The results revealed no significant effect of speech rate difference (Table 6).

Figure 3 shows the distribution of the number of questions (out of the four analyzed questions with speed manipulation) in which each participant selected the option presented at the faster speech rate and the option presented at the slower speech rate, for each condition. In all conditions, many participants selected the slower speech rate option in two or more of the four questions. In addition, some participants selected the slower option in all four questions. In the *113 vs. 191 WPM* condition, the modes of the number of slower option selections were 2 and 3 out of four, while those of faster option selections were 1 and 2 (Figure 3f). In the other conditions, the mode was 2 for both faster and slower option selections.

7 Discussion

We conducted three experiments to examine how manipulating speech rate when presenting options via synthetic speech in a non-native language affects users' selections. Specifically, in the first experiment (Section 4), we identified appropriate question and option sets for the investigation. In the second experiment (Section 5), we conducted a preliminary investigation of the effects of speech rate manipulation in non-native speech on users' selections. In the third experiment (Section 6), we used various speed combinations to examine how varying speech rate differences affect users' selections. Below, we answer the research questions of this study and discuss design implications.

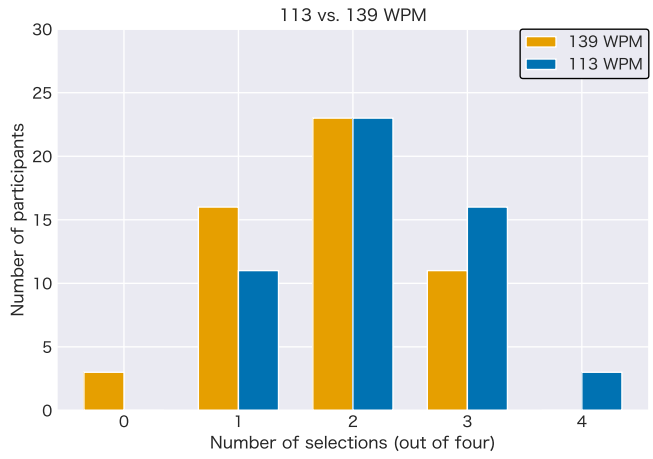
7.1 RQ1: Can users' selections be biased by deliberately manipulating speech rate when options are presented via synthetic speech in a non-native language?

Our results suggest that manipulating speech rate when presenting options via synthetic speech in a non-native language may not bias users' selections. This was contrary to our expectation that users would be more vulnerable to the manipulation of voice features when interacting in a non-native language. Leschanowsky et al. [36] examined the effects of speech rate manipulation in voice-based permission requests in users' native language. They reported that participants tended to select "Accept" regardless of whether "Accept" or "Decline" was spoken faster, indicating that participants did not tend to select the option presented at a faster or slower speech rate. Our results are similar in that participants did not clearly tend to select either the faster or slower option. In other words, the effect of speech rate manipulation on users' selections may not differ substantially between native- and non-native-language voice interactions. However, since options within the same question in this study were designed to differ in only a few words to prevent participants from making judgments based on only part of the audio (Section 3.3), the similarity between options may have affected the results.

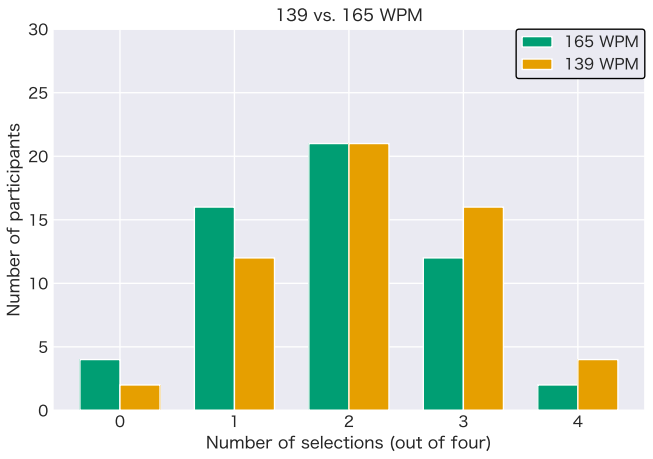
Leschanowsky et al. [36] also suggested that disrupting the balance of speech rate may cause users to rely more on their existing heuristics. Therefore, in situations such as cookie consent or permission requests, where users tend to accept [25, 72], speech rate manipulation may be effective. However, in decision-making scenarios like those in our experiments, or in situations where users have a clear intention (e.g., wanting to cancel a subscription), speech rate manipulation may not be deceptive.

Overall, although no selection bias was observed through speech rate manipulation, some participants selected the option presented at the faster speech rate in all speed-manipulated questions, while others selected the option presented at the slower speech rate in all of them. In an exploratory analysis, we found that participants who reported low perceived control over their selections (3 or below on the 7-point scale) did not consistently select options presented at either the faster or slower speech rate across all speed-manipulated questions. Just as some users are more susceptible to being steered by DPs in GUIs [57], there may also be users who are more susceptible to being steered by speech rate manipulation. Future research should investigate which user attributes make users more susceptible to influence from voice feature manipulation.

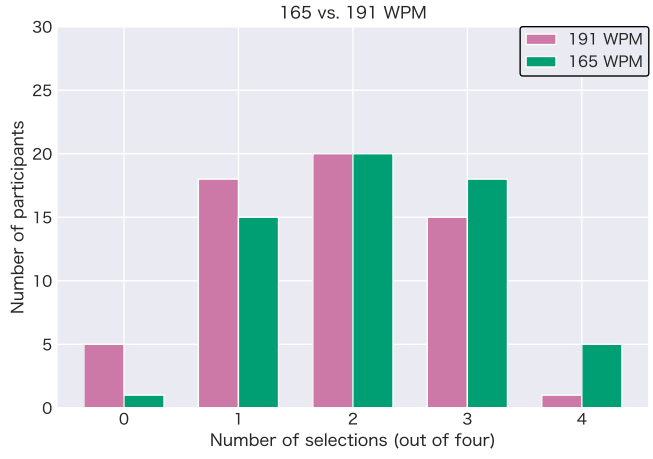
Furthermore, research on DPs in VUIs should investigate which voice features contribute to steering users' decision-making and which do not. Our results suggest that speech rate alone may not bias users' selections. However, since manipulating other voice features such as pitch in combination may steer users' decision-making [15], it is also necessary to investigate the combined effects of manipulating multiple voice features simultaneously.



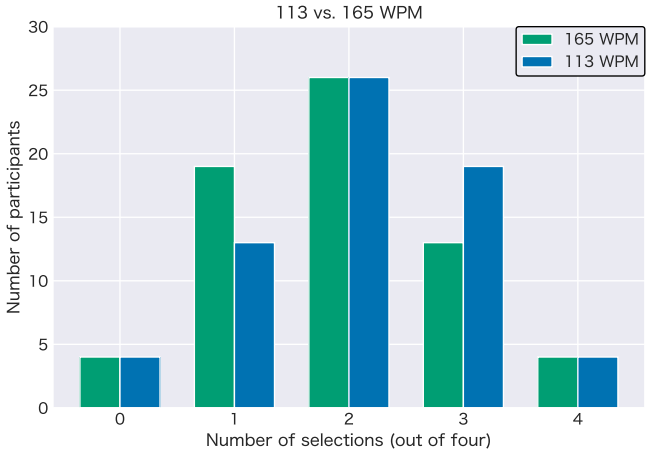
(a) 113 vs. 139 WPM



(b) 139 vs. 165 WPM

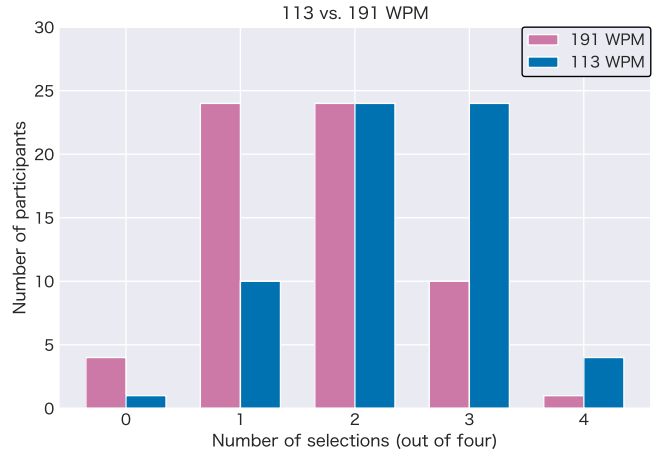


(c) 165 vs. 191 WPM



(d) 113 vs. 165 WPM

(e) 139 vs. 191 WPM



(f) 113 vs. 191 WPM

Figure 3: Distribution of the number of faster and slower option selections per participant in four analyzed speed-manipulated questions for each condition.

Table 5: Proportion of slower and faster option selections for each condition, and binomial test results in Experiment 3 (Section 6).

Condition (WPM)	Speed Diff. (WPM)	Slower option selections	Faster option selections	p-value	Adj. p-value
113 vs. 139	26	55.19% (117/212), 95% CI [48.46, 61.73]	44.81% (95/212), 95% CI [38.27, 51.54]	.149	.745
139 vs. 165	26	53.64% (118/220), 95% CI [47.04, 60.11]	46.36% (102/220), 95% CI [39.89, 52.96]	.312	.936
165 vs. 191	26	54.36% (129/236), 95% CI [48.29, 60.89]	45.34% (107/236), 95% CI [39.11, 51.71]	.171	.745
113 vs. 165	52	52.27% (138/264), 95% CI [46.26, 58.22]	47.73% (126/264), 95% CI [41.78, 53.74]	.498	.936
139 vs. 191	52	53.51% (122/228), 95% CI [47.03, 59.87]	46.49% (106/228), 95% CI [40.13, 52.97]	.321	.936
113 vs. 191	78	57.94% (146/252), 95% CI [51.77, 63.87]	42.06% (106/252), 95% CI [36.13, 48.23]	.014	.083

Table 6: Binomial generalized linear mixed model output for the probability of selecting the slower speech option.

Factors	Estimate	Std. Error	Z	p-value
Intercept	0.105	0.133	0.795	.426
Speed Diff.	0.002	0.003	0.630	.529

7.2 RQ2: Does the effect on users' selections change when speech rate differences between options are varied?

We prepared synthetic speech at four different speech rates for presenting options, and examined how varying speech rate differences affect users' selections. The experimental results showed no significant selection bias in any speech rate combination and the binomial GLMM also revealed no significant effect of speech rate difference. However, since humans have a simplicity bias [50], a tendency to reduce cognitive effort, further research should investigate whether users become more likely to select the option presented at a slower speech rate as the speech rate difference becomes larger than those used in our experiments.

In the experiment, no significant difference in the number of audio playbacks was observed across conditions with different speech rate combinations, with the average number of playbacks being approximately 1.2 across all conditions. This suggests that participants did not attempt to listen again even when options were presented at a fast speech rate. Consequently, even if options were presented at an even faster speech rate, users may be reluctant to replay the audio and may either choose randomly or select the option they were able to comprehend.

Regarding perceived control over selections, a significant difference was found between the *113 vs. 139 WPM* condition and the *113 vs. 165 WPM* condition, but no differences were found among the other conditions. Since speech rate was manipulated in only 5 of the 15 questions and the remaining 10 were presented at the same speech rate, participants' perceived control ratings may have remained relatively high, which may explain the lack of differences among the other conditions. Therefore, the results regarding perceived control should be interpreted with caution.

The strategy of steering decision-making by increasing the speech rate difference can be considered conceptually similar to the Interface Interference DP in GUIs. Interface Interference refers to the manipulation of user interfaces to privilege certain actions over others, confusing users or making it difficult for them to find important action possibilities [20]. Just as users feel aversion toward such DPs in GUIs [19, 38], an extremely large speech rate difference may cause users to feel distrust or discomfort. Future research should investigate at what point the speech rate difference becomes large enough to evoke such negative feelings in users and how this affects their decision-making.

7.3 Design Implications

VUIs should be designed so that users can make decisions based on their own intentions. In this study, participants' comprehension of the content presented via English synthetic speech was not sufficiently high, and most participants did not have high proficiency in English, their non-native language. As prior research [73] has pointed out, multilingual support in VUIs is essential. In addition to multilingual support, using simple and accessible vocabulary is also important to ensure that users do not struggle with comprehension.

Designers should also enable users to adjust the speech rate of VUI output. Currently, Apple's Siri and Amazon's Alexa already allow users to adjust the speech rate. When designing VUIs, the speech rate should be adjustable in multiple levels, and information should be presented at a consistent speed. Although this study focused solely on speech rate, voice tone and pitch also affect decision-making [82], and designers should enable users to control these features as well. In addition, VUI designs should be dynamically adjustable according to context, taking into account factors such as

the setting in which the VUI is used and situations such as emergencies [16, 31, 42].

In our experiments, the average number of audio playbacks was approximately 1.2, indicating that users tended not to replay the audio. Therefore, it is crucial to design VUIs in a way that reduces users' reluctance to listen again. Since users show frustration with lengthy responses from VUIs [51], it is necessary to convey information concisely. For example, when users ask to hear something again, VUIs should replay only the relevant portion rather than repeating from the beginning. Such design approaches may be useful for reducing users' reluctance to listen again.

It has been noted that standardized guidelines for VUI design do not exist [43]. Given that the specific nature of voice interaction has received insufficient attention [12], it is necessary to establish design guidelines for VUIs that enable users to achieve their intended decision-making and behavior. To this end, further research is needed on DPs that may be present or may emerge in VUIs. Researchers also need to establish a definition of DPs in VUIs so that regulatory authorities can identify and address them.

8 Limitations and Future Work

We acknowledge several limitations of this study. First, since we recruited Japanese adults who were non-native speakers of English as participants, the results obtained in this study may differ for people from different cultural backgrounds or for different languages. In addition, since the experiment adopted an overseas trip scenario, the results may not be applicable to decision-making in different contexts, limiting the generalizability of the findings. Although we selected most of the words used in the options from vocabulary learned by the end of high school in Japan to minimize the influence of word knowledge on selections, its influence on participants' selections cannot be completely ruled out. Furthermore, since most participants had low English language proficiency, we were unable to analyze differences in the effects of speech rate manipulation on selections based on language proficiency levels.

In the future, we plan to recruit participants with varying levels of language proficiency and investigate how differences in proficiency affect the impact of speech rate manipulation on decision-making. In addition, while this study presented options via speech and required participants to make selections by clicking, we would like to investigate the effects when participants make selections by speaking, thereby completing the interaction entirely through voice. Moreover, we plan to explore the generalizability of our findings by conducting experiments in more realistic decision-making contexts using VUIs (e.g., shopping), and examining situations involving a larger number of options.

9 Conclusion

Although previous research has examined how manipulation of voice features affects users' decision-making in voice user interfaces, to the best of our knowledge, this study is the first to examine such effects on non-native users in voice interactions in the context of deceptive patterns. We conducted three experiments with Japanese participants who were non-native speakers of English to examine whether manipulating speech rate when presenting options via synthetic speech in a non-native language can bias users' selections. Our study found no clear evidence that manipulating speech rate biases users' selections. Future research should investigate how a wider range of speech-rate differences affects users' selections and examine the impact of voice features beyond speech rate.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP26K02962 and JP26KJ2028.

References

[1] Ana Inés Ansaldo, Ladan Ghazi-Saidi, and Daniel Adrover-Roig. 2015. Interference Control In Elderly Bilinguals: Appearances Can Be Misleading. *Journal of Clinical and Experimental Neuropsychology* 37, 5 (2015), 455–470. doi:10.1080/13803395.2014.990359

[2] Alan Baddeley. 2010. Working memory. *Current Biology* 20, 4 (2010), R136–R140. doi:10.1016/j.cub.2009.12.014

[3] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 763–776. doi:10.1145/3461778.3462086

[4] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. In *Proceedings on Privacy Enhancing Technologies*, Vol. 2016. 237–254. doi:10.1515/popets-2016-0038

[5] Harry Brignull. 2023. Deceptive patterns – user interfaces designed to trick you. Retrieved January 28, 2026 from https://www.deceptive.design/

[6] Karolina Buchta, Piotr Wójcik, Konrad Nakonieczny, Justyna Janicka, Damian Gałuszka, Radosław Sterna, and Magdalena Igras-Cybulska. 2022. Microtransactions in VR. A qualitative comparison between voice user interface and graphical user interface. In *2022 15th International Conference on Human System Interaction (HSI)*. 1–5. doi:10.1109/HSI55341.2022.9869475

[7] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 223 (Nov. 2019), 19 pages. doi:10.1145/3359325

[8] Xun Cao, Naomi Yamashita, and Toru Ishida. 2016. How Non-native Speakers Perceive Listening Comprehension Problems: Implications for Adaptive Support Technologies. In *Collaboration Technologies and Social Computing*. Springer Singapore, Singapore, 89–104.

[9] Akash Chaudhary, Jaivrat Saroha, Kyzyl Monteiro, Angus G. Forbes, and Aman Parnami. 2022. "Are You Still Watching?": Exploring Unintended User Behaviors and Dark Patterns on Video Streaming Platforms. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 776–791. doi:10.1145/3532106.3533562

[10] Jieshan Chen, Jiamou Sun, Sidong Feng, Zhenchang Xing, Qinghua Lu, Xiwei Xu, and Chunyang Chen. 2023. Unveiling the Tricks: Automated Detection of Dark Patterns in Mobile Applications. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 114, 20 pages. doi:10.1145/3586183.3606783

[11] Qishuo Cheng, Jinxin Xu, Mengxia Qiu, Jiajian Zheng, and Ruisi Li. 2025. Detection and Evaluation of Dark Patterns in E-business Websites Utilizing Deep Learning. In *Proceedings of the 2024 International Conference on Big Data Mining and Information Processing (BDMIP '24)*. Association for Computing Machinery, New York, NY, USA, 194–198. doi:10.1145/3735014.3735898

[12] Silvia De Conca. 2023. The present looks nothing like the Jetsons: Deceptive design in virtual assistants and the protection of the rights of users. *Computer Law & Security Review* 51 (2023), 105866. doi:10.1016/j.clsr.2023.105866

[13] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376600

[14] Mateusz Dubiel, Luis A. Leiva, Kerstin Bongard-Blanchy, and Anastasia Sergeeva. 2024. "Hey Genie, You Got Me Thinking about My Menu Choices!" Impact of Proactive Feedback on User Perception and Reflection in Decision-making Tasks. *ACM Trans. Comput.-Hum. Interact.* 31, 5, Article 56 (Nov. 2024), 30 pages. doi:10.1145/3685274

[15] Mateusz Dubiel, Anastasia Sergeeva, and Luis A. Leiva. 2024. Impact of Voice Fidelity on Decision Making: A Potential Dark Pattern?. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 181–194. doi:10.1145/3640543.3645202

[16] Jonas Fritsch, Pelin Karaturhan, Stina Marie Hasse Jørgensen, Ada Ada Ada, and Søren Lyngsø Knudsen. 2025. Towards a Framework for Exploring Synthetic Voices in VUI Design. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 724–739. doi:10.1145/3715336.3735729

[17] Christine C.M Goh. 2000. A cognitive perspective on language learners' listening comprehension problems. *System* 28, 1 (2000), 55–75. doi:10.1016/S0346-251X(99)00060-3

[18] Hiromu Goto. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9, 3 (1971), 317–323. doi:10.1016/0028-3932(71)90027-3

[19] Colin M. Gray, Jingle Chen, Shruthi Sai Chivukula, and Liyang Qu. 2021. End User Accounts of Dark Patterns as Felt Manipulation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 372 (Oct. 2021), 25 pages. doi:10.1145/3479516

[20] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108

[21] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner. 2024. An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 289, 22 pages. doi:10.1145/3613904.3642436

[22] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. Association for Computing Machinery, New York, NY, USA, 523–532. doi:10.1145/2598510.2598541

[23] Susan G. Guion, James E. Flege, Reiko Akahane-Yamada, and Jesica C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America* 107, 5 (05 2000), 2711–2724. doi:10.1121/1.428657

[24] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 377 (Oct. 2021), 29 pages. doi:10.1145/3479521

[25] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 621, 27 pages. doi:10.1145/3491102.3501985

[26] Shun Hidaka, Sota Kobuki, Mizuki Watanabe, and Katie Seaborn. 2023. Linguistic Dead-Ends and Alphabet Soup: Finding Dark Patterns in Japanese Apps. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 3, 13 pages. doi:10.1145/3544548.3580942

[27] Graham I. Johnson and Lynne Coventry. 2001. "You Talking to Me?" Exploring Voice in Self-Service User Interfaces. *International Journal of Human–Computer Interaction* 13, 2 (2001), 161–186. doi:10.1207/S15327590IJHC1302_5

[28] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2009. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (HRI '09)*. Association for Computing Machinery, New York, NY, USA, 173–180. doi:10.1145/1514095.1514127

[29] Amanda Kann. 2022. Voice Assistants Have a Plurilingualism Problem. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 16, 5 pages. doi:10.1145/3543829.3544526

[30] Margarita Kaushanskaya, Henrike K. Blumenfeld, and Viorica Marian. 2020. The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition* 23, 5 (2020), 945–950. doi:10.1017/S1366728919000038

[31] Jieun Kim and Susan R. Fussell. 2025. Should Voice Agents Be Polite in an Emergency? Investigating Effects of Speech Style and Voice Tone in Emergency Simulation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 61, 17 pages. doi:10.1145/3706598.3714203

[32] Yuichiro Kinoshita, Yuto Sekiguchi, Riho Ueki, Kouta Yokoyama, and Satoshi Nakamura. 2023. Do People Tend to Select a Delayed Item?. In *Human-Computer Interaction*. Springer Nature Switzerland, Cham, 397–407.

[33] Daniel Kirkman, Kami Vaniea, and Daniel W. Woods. 2023. DarkDialogs: Automated detection of 10 dark patterns on cookie dialogs. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. 847–867. doi:10.1109/EuroSP57164.2023.00055

[34] Monica Kowalczyk, Johanna T. Gunawan, David Choffnes, Daniel J Dubois, Woodrow Hartzog, and Christo Wilson. 2023. Understanding Dark Patterns in Home IoT Devices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 179, 27 pages. doi:10.1145/3544548.3581432

[35] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'Dark Pattern' in Home Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 374–381. doi:10.1109/HRI.2019.8673274

[36] Anna Leschanowsky, Anastasia Sergeeva, Judith Bauer, Sheetal Vijapurapu, and Mateusz Dubiel. 2025. Exploring the Impact of Modality and Speech Rate Manipulation in Voice Permission Requests—Limits of Applicability and Potential for Influencing Decision-Making. *International Journal of Human-Computer Studies* 205 (Nov. 2025), 103590. doi:10.1016/j.ijhcs.2025.103590

[37] Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a Light on Dark Patterns. *Journal of Legal Analysis* 13, 1 (03 2021), 43–109. doi:10.1093/jla/laaa006

[38] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. 2021. Towards the Identification of Dark Patterns: An Analysis Based on End-User Reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction (IndiaHCI '20)*. Association for Computing Machinery, New York, NY, USA, 24–33. doi:10.1145/3429290.3429293

[39] S M Hasan Mansur, Sabiha Salma, Damilola Awofisayo, and Kevin Moran. 2023. AidUI: Toward Automated Recognition of Dark Patterns in User Interfaces. In *Proceedings of the 45th International Conference on Software Engineering (ICSE '23)*. IEEE Press, 1958–1970. doi:10.1109/ICSE48619.2023.00166

[40] Viorica Marian, Henrike K. Blumenfeld, and Margarita Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research* 50, 4 (2007), 940–967. doi:10.1044/1092-4388(2007/067)

[41] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 81 (Nov. 2019), 32 pages. doi:10.1145/3359183

[42] Niharika Mathur, Hasibur Rahman, and Smit Desai. 2026. "Who wants to be nagged by AI?": Investigating the Effects of Agreeableness on Older Adults' Perception of LLM-Based Voice Assistants' Explanations. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, Article 33, 6 pages. doi:10.1145/3772363.3798685

[43] Thomas Mildner, Orla Cooney, Anna-Maria Meck, Marion Bartl, Gian-Luca Savino, Philip R Doyle, Diego Garaialde, Leigh Clark, John Sloan, Nina Wenig, Rainer Malaka, and Jasmin Niess. 2024. Listening to the Voices: Describing Ethical Caveats of Conversational User Interfaces According to Experts and Frequent Users. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 307, 18 pages. doi:10.1145/3613904.3642542

[44] Thomas Mildner, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 192, 15 pages. doi:10.1145/3544548.3580695

[45] Emi Moriuchi. 2019. Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing* 36, 5 (2019), 489–501. doi:10.1002/mar.21192

[46] Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300472

[47] Asmit Nayak, Yash Wani, Shirley Zhang, Rishabh Khandelwal, and Kassem Fawaz. 2025. Automatically Detecting Online Deceptive Patterns. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25)*. Association for Computing Machinery, New York, NY, USA, 96–110. doi:10.1145/3719027.3765191

[48] Sam Niknejad, Thomas Mildner, Nima Zargham, Susanne Putze, and Rainer Malaka. 2024. Level Up or Game Over: Exploring How Dark Patterns Shape Mobile Games. In *Proceedings of the 23rd International Conference on Mobile and Ubiquitous Multimedia (MUM '24)*. Association for Computing Machinery, New York, NY, USA, 148–156. doi:10.1145/3701571.3701604

[49] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376321

[50] Shaul Oreg and Mahmut Bayazit. 2009. Prone to Bias: Development of a Bias Taxonomy from an Individual Differences Perspective. *Review of General Psychology* 13, 3 (2009), 175–193. doi:10.1037/a0015656

[51] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In *Proceedings of the 2022 European Symposium on Usable Security (EuroUSEC '22)*. Association for Computing Machinery, New York, NY, USA, 64–78. doi:10.1145/3549015.3554213

[52] Sabra D. Pelham and Lise Abrams. 2014. Cognitive advantages and disadvantages in early and late bilinguals. *Journal of experimental psychology. Learning, memory, and cognition* 40, 2 (2014), 313–325. doi:10.1037/a0035224

[53] Sabid Bin Habib Pias, Ran Huang, Donald S. Williamson, Minjeong Kim, and Apu Kapadia. 2024. The Impact of Perceived Tone, Age, and Gender on Voice Assistant Persuasiveness in the Context of Product Recommendations. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 20, 15 pages. doi:10.1145/3640794.3665545

[54] Aung Pyae and Paul Scifleet. 2018. Investigating differences between native english and non-native english speakers in interacting with a voice user interface: a case of google home. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 548–553. doi:10.1145/3292147.3292236

[55] Rolf Reber, Piotr Winkielman, and Norbert Schwarz. 1998. Effects of Perceptual Fluency on Affective Judgments. *Psychological Science* 9, 1 (1998), 45–48. doi:10.1111/1467-9280.00008

[56] Simoni F. Rohden and Lélis Balestrin Espartel. 2024. Consumer reactions to technology in retail: choice uncertainty and reduced perceived control in decisions assisted by recommendation agents. *Electronic Commerce Research* 24, 2 (Feb. 2024), 901–923. doi:10.1007/s10660-024-09808-7

[57] Arianna Rossi, Rachele Carli, Marietjie W. Botes, Angelica Fernandez, Anastasia Sergeeva, and Lorena Sánchez Chamorro. 2024. Who is vulnerable to deceptive design patterns? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability. *Computer Law & Security Review* 55 (2024), 106031. doi:10.1016/j.clsr.2024.106031

[58] René Schäfer, Paul Miles Preuschoff, and Jan Borchers. 2023. Investigating Visual Countermeasures Against Dark Patterns in User Interfaces. In *Proceedings of Mensch Und Computer 2023 (MuC '23)*. Association for Computing Machinery, New York, NY, USA, 161–172. doi:10.1145/3603555.3603563

[59] René Schäfer, Paul Miles Preuschoff, René Röpke, Sarah Sahabi, and Jan Borchers. 2024. Fighting Malicious Designs: Towards Visual Countermeasures Against Dark Patterns. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 296, 13 pages. doi:10.1145/3613904.3642661

[60] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. 2022. Understanding Account Deletion and Relevant Dark Patterns on Social Media. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 417 (Nov. 2022), 43 pages. doi:10.1145/3555142

[61] Brennan Schaffner, Yaretzi Ulloa, Riya Sahni, Jiatong Li, Ava Kim Cohen, Natasha Messier, Lan Gao, and Marshini Chetty. 2025. An Experimental Study Of Netflix Use and the Effects of Autoplay on Watching Behaviors. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW030 (May 2025), 22 pages. doi:10.1145/3710928

[62] Dirk Schnelle-Walka. 2011. I tell you something. In *Proceedings of the 16th European Conference on Pattern Languages of Programs (EuroPLoP '11)*. Association for Computing Machinery, New York, NY, USA, Article 10, 26 pages. doi:10.1145/2396716.2396726

[63] Katie Seaborn, Tatsuya Itagaki, Mizuki Watanabe, Yijia Wang, Ping Geng, Takao Fujii, Yuto Mandai, Miu Kojima, and Suzuka Yoshida. 2024. Deceptive, Disruptive, No Big Deal: Japanese People React to Simulated Dark Commercial Patterns. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 95, 8 pages. doi:10.1145/3613905.3651099

[64] Katie Seaborn and Satoshi Nakamura. 2025. Quality and representativeness of research online with Yahoo! Crowdsourcing. *Frontiers in Psychology* Volume 16 - 2025 (2025). doi:10.3389/fpsyg.2025.1588579

[65] William Seymour, Mark Cote, and Jose Such. 2023. Legal Obligation and Ethical Best Practice: Towards Meaningful Verbal Consent for Voice Assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 166, 16 pages. doi:10.1145/3544548.3580967

[66] Shamim Seyson and Wesley Willett. 2025. Exploring the Evolution of Dark Patterns and Manipulative Design on Instagram. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 263, 8 pages. doi:10.1145/3706599.3719771

[67] Ray Sin, Ted Harris, Simon Nilsson, and Talia Beck. 2025. Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? *Behavioural Public Policy* 9, 1 (2025), 61–87. doi:10.1017/bpp.2022.11

[68] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by design - dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 19, 12 pages. doi:10.1145/3419249.3420132

[69] Jae Yung Song, Anne Pycha, and Tessa Culleton. 2022. Interactions between voice-activated AI assistants and human speakers and their implications for second-language acquisition. *Frontiers in Communication* Volume 7 - 2022 (2022). doi:10.3389/fcomm.2022.995475

[70] Lars Stenius Stæhr. 2009. VOCABULARY KNOWLEDGE AND ADVANCED LISTENING COMPREHENSION IN ENGLISH AS A FOREIGN LANGUAGE. *Studies in Second Language Acquisition* 31, 4 (2009), 577–607. doi:10.1017/S0272263109990039

[71] Vito Tassiello, Jack S. Tillotson, and Alexandra S. Rome. 2021. "Alexa, order me a pizza!": The mediating role of psychological power in the consumer–voice assistant interaction. *Psychology & Marketing* 38, 7 (2021), 1069–1080. doi:10.1002/mar.21488

[72] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. Association for Computing Machinery, New York, NY, USA, 973–990. doi:10.1145/3319535.3354212

[73] Kimi Wenzel and Geoff Kaufman. 2024. Designing for Harm Reduction: Communication Repair for Multicultural Users' Voice Interactions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 879, 17 pages. doi:10.1145/3613904.3642900

[74] Andi Winterboer and Johanna D. Moore. 2007. Evaluating information presentation strategies for spoken recommendations. In *Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys '07)*. Association for Computing Machinery, New York, NY, USA, 157–160. doi:10.1145/1297231.1297260

[75] Yunhan Wu, Justin Edwards, Orla Cooney, Anna Bleakley, Philip R. Doyle, Leigh Clark, Daniel Rough, and Benjamin R. Cowan. 2020. Mental Workload and Language Production in Non-Native Speaker IPA Interaction. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. doi:10.1145/3405755.3406118

[76] Yunhan Wu, Martin Porcheron, Philip Doyle, Justin Edwards, Daniel Rough, Orla Cooney, Anna Bleakley, Leigh Clark, and Benjamin Cowan. 2022. Comparing Command Construction in Native and Non-Native Speaker IPA Interaction through Conversation Analysis. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. doi:10.1145/3543829.3543839

[77] Yuki Yada, Tsuneo Matsumoto, Fuyuko Kido, and Hayato Yamana. 2023. Why is the User Interface a Dark Pattern?: Explainable Auto-Detection and its Analysis. In *2023 IEEE International Conference on Big Data (BigData)*. 6308–6310. doi:10.1109/BigData59044.2023.10386308

[78] Jingzhou Ye, Yao Li, Wenting Zou, and Xueqiang Wang. 2025. From Awareness to Action: The Effects of Experiential Learning on Educating Users about Dark Patterns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 224, 22 pages. doi:10.1145/3706598.3713493

[79] Jiahong Yuan, Mark Y. Liberman, and Christopher Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. In *Interspeech*. doi:10.21437/Interspeech.2006-204

[80] José Pablo Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *International Conference on Foundations of Digital Games*. https://api.semanticscholar.org/CorpusID:17683705

[81] Gloria Xiaodan Zhang, Yijia Wang, Taro Leo Nakajima, and Katie Seaborn. 2025. First Contact with Dark Patterns and Deceptive Designs in Chinese and Japanese Free-to-Play Mobile Games. *Proc. ACM Hum.-Comput. Interact.* 9, 6, Article GAMES025 (Oct. 2025), 26 pages. doi:10.1145/3748620

[82] Shuning Zhang, Han Chen, Yabo Wang, Yiqun Xu, Jiaqi Bai, Yuanyuan Wu, Shixuan Li, Xin Yi, Chunhui Wang, and Hewu Li. 2025. The Manipulative Power of Voice Characteristics: Investigating Deceptive Patterns in Mandarin Chinese Female Synthetic Speech. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 150 (Sept. 2025), 32 pages. doi:10.1145/3749510