Assessing Effectiveness and Interpretability of Light Behaviors in Smart Speakers

Sahiti Kunchay
sahiti@psu.edu
College of Information Sciences and Technology
Pennsylvania State University
University Park, PA, USA

Saeed Abdullah
saeed@psu.edu
College of Information Sciences and Technology
Pennsylvania State University
University Park, PA, USA

ABSTRACT

Light is an important communication modality. Smart speakers leverage light behaviors with different colors and shapes to deliver a wide range of information. However, there has not been sufficient work to assess the effectiveness of these light behaviors. That is, can users correctly interpret light behaviors from smart speakers? How effective are these light behaviors across different cultures? To answer these questions, we conducted a survey with 1,006 smart speaker users from seven countries. On average, they were able to correctly identify only 37% of light behaviors used in smart speakers. Overall, Amazon Echo users accurately identified a higher proportion of light behaviors than Google Home users. Furthermore, the perceived usability varied considerably across different countries. We also interviewed six experts to design more effective light behaviors for smart speakers. Our findings have important implications for future smart speaker design and the use of light as a communication modality in HCI.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Interaction techniques**.

KEYWORDS

Smart-speakers, Light behavior, Expressive Light, Notifications

ACM Reference Format:
Sahiti Kunchay and Saeed Abdullah. 2021. Assessing Effectiveness and Interpretability of Light Behaviors in Smart Speakers. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3469595.3469610

1 INTRODUCTION

Digital devices rely on light to communicate a wide range of information to users. Light is used to indicate current status, convey errors, show progress, deliver spatial information, and provide event notifications [24]. One of the key advantages of leveraging light behaviors is their ability to communicate unobtrusively. They can communicate useful information about device status and interactions without disrupting primary tasks. In other words, light as a communication modality can lead to "calm technology" [43] — it supports user interactions as needed and disappears into the periphery when users are not paying attention.

In recent years, a significant number of consumer products have leveraged light behaviors to support a wide range of use cases. Smart speakers also rely on light behaviors to support user interactions and convey complex informational states. For example, they use light to communicate information about errors during device set up, WiFi connectivity, and notifications for incoming calls and messages. All mainstream smart speakers have light apparatus with varying abilities and physical shapes. For example, Amazon Echo devices use light rings while Google Home devices use dots (see Figure 1). Consequently, all smart speakers have developed a substantial set of light behaviors to support a wide range of tasks and interactions.

The integration of light behaviors in smart speakers pose unique design challenges. These behaviors are not limited to the use of point light — "a small, single color light emitting element" with varying intensity [16]. Instead, smart speakers use multiple colors and patterns for communication (e.g., pulsing yellow circle for reminders in Echo devices). This can potentially allow smart speakers to be more "expressive" [16] and support a larger number of use cases compared to devices using one-dimensional point lights (e.g., smartphones). However, the resulting complexity can also hinder user interpretation. Harrison et al. [16] noted the necessity of a light behavior to be *iconic* — that can consistently produce single interpretation in a given context. Iconic light behaviors require little memorization from users. That is, iconic light behaviors support an important usability heuristic: *recognition rather than recall* [27].

The use of complex patterns and colors might lead to many non-iconic light behaviors in smart speakers and subsequently, reduce their overall usability. However, there has not been any systematic study to assess the effectiveness and expressivity of light behaviors in smart speakers. Do smart speaker users correctly understand different light behaviors? Are these light behaviors iconic [16] with one dominant interpretation?

Smart speakers have become a global trend. For example, Amazon Echo devices are available in more than 80 countries [10] supporting multiple languages. The number of smart speakers worldwide is estimated to be 163 million by the end of 2021 [1], which makes them the fastest growing global trend among connected devices. As a result, light behaviors in smart speakers are now used to support user tasks and interactions across different countries and cultures. However, color interpretation varies across cultures

CUI '21, July 27–29, 2021, Bilbao (online), Spain

ACM ISBN 978-1-4503-8998-3/21/07…$15.00
https://doi.org/10.1145/3469595.3469610

[13, 23, 37, 42]. This has important implications for light behaviors in smart speakers given their global reach. Given the emerging importance of smart speakers as a global consumer device, our current lack of understanding regarding the effectiveness and expressivity of the light behaviors in smart speakers represents a serious knowledge gap with both theoretical and practical consequences.

In this paper, we aim to address this gap. Specifically, we have conducted an online survey with 1,006 smart speaker users from seven countries: Australia, France, Germany, India, Japan, the UK, and the US. We focused on light behaviors used in Amazon Echo and Google Home devices given their wide cross-cultural adoption [19, 20, 30]. Furthermore, We also interviewed six experts to explore design challenges and opportunities for light behaviors in smart speakers. Overall, our findings indicate that users do not understand the light behaviors — we find that users are able to accurately identify only 37% of the light behaviors presented to them.

There are a number of findings that might be useful for smart speaker interaction designers including associations of accuracy and usability with device type, demographics, and device usage characteristics. Some of our key findings include significant associations of usability with a user's country of residence, indicating cultural differences in how users interpret the light behavior system. Furthermore, a number of important device characteristics serve as significant predictors of accuracy and usability. Importantly, there is a significant difference in accuracy among different smart speaker types, with Amazon Echo device users reporting higher percentage of accurately identified light behaviors than Google Home device users.

Our findings have important implications for the design of light behaviors in smart speakers. We also discuss the design recommendations derived from our analysis of the survey and interview sessions in our discussion section. This paper makes knowledge contributions in two important research directions:

- *Smart speaker interaction design*: This study addresses an important research gap regarding interaction design. Light behaviors play an important role in shaping user experience for smart speakers. However, prior work on smart speakers have mostly focused on voice interactions [6, 36]. This study, thus, complements prior work and advances the current state of knowledge regarding the user interface and interaction design of smart speakers. As a result, the findings here have implications for both smart speaker developers and UX researchers aiming to advance multi-modal user interfaces.
- *Light as an HCI communication modality*: While almost all digital devices use light to communicate information, there has been a lack of systematic study to assess their efficacy across different contexts within the HCI community. For example, the work by Harrison et al. [16] was conducted almost a decade ago (2012) and it considered only a limited subset of light behaviors (monochromatic "point lights"). This represents a significant HCI knowledge gap given the increasingly pervasive use of complex and multi-colored light behaviors in modern devices. Our study is an important step to address this knowledge gap. Specifically, this study i) identified significant usability issues with light behaviors used globally; ii) established accuracy data for a range of light behaviors across different countries, which can serve as a baseline for future work; and iii) provided actionable guidelines for designing complex light behaviors. These findings are important to advance our current state of knowledge regarding a widely used but under-explored HCI communication modality.

2 RELATED WORK

Light has been used as a communication modality in numerous pervasive devices in creative and highly effective ways. In this section, we present the growing body of work in light communication research as well as how the recent work in this domain makes a strong case for examining the efficacy of the light behavior system in smart speakers. We also aim to establish the necessity of this work through the exploration of existing smart speaker research and build the discourse around why our study furthers efforts in both HCI and UX research pertaining to smart speakers.

2.1 Light as a communication media

Light as a medium of communication has gained significant popularity in the recent past, with everyday devices using a range of light behaviors, or representations of light forms, to convey various informational states. In recent years, usage of light to convey these information states has grown beyond the everyday appliance to smartphones, smartwatches, and smart displays. Harrison et al. [16] pioneered the effort of introducing light communication research to HCI through their design vocabulary of light behaviors for smartphones. This work characterized various light behaviors in smartphones and provided a rich discussion on the use of point lights in ubiquitous devices. They defined expressive lights, or point lights, as small LED lights that have traditionally been used in a wide variety of devices and appliances to convey low-dimensional information. Pohl et al. [31] presented a novel notification system that used light scattering on the users' wrists to convey information. Kerber et al. [17] used expressive lights on a low-resolution display over the smartwatch screen to convey different informational states related to emails, instant messages, and alarms. Similarly, Freeman et al. [14] presented an interactive light feedback system to accompany smartphones in order to provide feedback for different user gestures. Moreover, A number of recent studies (e.g., [25, 26]) have explored the design space of ambient light displays to derive guidelines for a system that encodes various kinds of information within these displays.

There has also been work in the expressive light domain to design and develop light vocabularies in Human-robot Interaction (HRI). For instance, Baraka et al. [4, 5] analyzed how expressive lights could be useful in conveying informational states from robots to humans. Song et al. [39] investigated the use of light to express emotion in social robots. In a subsequent study, Song et al. [40] used bioluminescence inspired expressive lights to examine their effect on human willingness to interact with the robot.

These studies have explored the design space of conveying information through novel light behaviors designed for specific domains. Prior work found that successful light behaviors are distinguishable [31], able to attract users' attention, and interpretable [26], thus qualifying some of the important characteristics of effective light behavior systems. These studies also hypothesized that other devices would significantly benefit from studying and designing light behavior sets [16], specifically multi-modal configurations in which lights are used in conjunction with speech [5, 33]. However, no prior work has explored the use and effectiveness of light behaviors in smart speakers. We posit that smart speakers would significantly benefit from this investigation of their light behavior systems' efficacy and usability, given that many popular smart speakers already use these light apparatus to convey critical information.

However, smart speaker light behaviors cannot simply be ported from light behaviors in other devices because they significantly differ from the device domains in which these light behavior studies have been traditionally conducted (i.e., smartphones and robots). There is a different interaction dynamic present with smart speakers that stems from the use of a different interface — the voice user interface (VUI). This dynamic determines what is essential to the voice-based communication (such as conversation context, fillers, speech commands), and what is considered information that might interrupt the natural flow of conversation (such as informational states pertaining to volume, Wi-Fi, updates, device state). These latter kinds of information are considered complementary to the conversation, and hence are communicated through the light apparatus. Further, the kinds of information delivered through light vary from smartphones and robots as the core and associated functionalities of the device vary as well. For instance, "low power" is a significant informational state that needs to be conveyed through smartphone light behaviors, but has no meaning within a smart speaker context given their constant access to power. Hence, although light vocabularies have been established in the past, they cannot directly be adapted for smart speaker light behaviors. We can, however, use similar methodology to study smart speaker light behaviors, which is what this study does. In this work, we adapted Harrison et al.'s methodology [16] to examine and analyze the light behavior system in smart speakers.

2.2 Smart speaker usage and light

The growth of smart-speakers such as Amazon Echo and Google Home has been exponential in the past few years [20], and is forecasted to continue to do so [19]. Recent work within the HCI and ubiquitous computing communities examining the usage of smart-speakers in household settings [34, 36] demonstrates a high integration of the devices into the everyday lives and routines of users. Given their ubiquity, smart speakers are poised to deal with unique, and often personal, interactions that require designers of such technology to take into account the myriad of social situations and concerns that accompany such devices.

Smart speakers operate on the premise of being capable of carrying out coherent, uninterrupted conversation. Such conversations would require that the smart speaker seamlessly pick up on implicit conversational cues — both verbal and nonverbal. However, capturing and processing these cues is a challenging task that has not been perfected by current popular smart speakers [41]. This arises due to the structured nature of the conversation that strictly follows the <trigger word, question, answer> paradigm (<"*Alexa, what is the weather like today?*", "*It's sunny with highs of 28°C and lows of 17°C in Sunnyvale, GA.*">). Any deviation from this structure by the user (e.g., adding verbal cues that are neither questions or commands such as *"Okay"* and *"Sounds Good"*) [32, 41] causes communication breakdowns. Beneteau et al. [6] presented the different forms in which smart speaker communication breakdowns occur within households, with one the most prevalent infractions being the agent misunderstanding what the user is trying to elucidate, creating a gap between what the user expects as a response and the ensuing interaction. Luger et al. [22] investigated this 'gulf between user expectation and experience of conversational agents', and elucidated several cases where the input and response mapping was not adequate.

The aforementioned work demonstrates that users often require visual confirmation before performing sensitive, important, and impactful tasks [22] such as sending texts and making calls. Moreover, Wu et al.'s work comparing Intelligent Personal Assistant (IPA) use among native and non-native English speakers [44] emphasized a need to tailor visual feedback to enhance the IPA experience for non-native speakers. Smart speakers leverage this preference for visual confirmation to mitigate user concerns and improve the overall interaction with the smart speaker by providing complementary information through the visual apparatus. Hence, communication through the light apparatus supports efficient dialogue with a smart speaker. However, these light behaviors must be clear and coherent to the user to facilitate effective interactions between smart speakers and their users. But, there has been no prior work that has examined how effective the existing light behaviors are in terms of conveying these critical informational states. Our work aims to investigate the efficacy of existing light behaviors and critically analyze the various components that make up the light vocabulary that these behaviors subscribe to.

Furthermore, investigating the light behaviors could have far-reaching consequences for novel applications and often overlooked smart speaker interactions. For instance, Blair et al. [8, 9] highlight the needs and challenges of deaf older adults using conversational agents. The work recommends not only relying on auditory feedback, but to use dual-feedback mechanisms to improve upon the device's accessibility. Using effective light communication to convey various device states accurately and efficiently may make a huge difference in the interaction dynamics between this population and their smart speakers. However, given the current gap in the knowledge of smart speaker light behaviors, there is a pressing need to look into how users interact with these light behaviors, and whether they understand these behaviors.

3 METHODOLOGY

We adapted the methodology from Harrisson et al. [16] to assess the interpretability and efficacy of smart speaker light behaviors. More specifically, a key goal in this study was to collect and analyze data from a large number of smart speaker users across geographical boundaries and understand culture-specific device usage and interpretation of light behaviors. That is, we wanted to establish a baseline of user interaction data regarding light behaviors across countries and devices. Given this goal, we carefully considered the feasibility of different options (such as interviews, observational studies/case studies, diaries) and decided on the current survey method as it afforded us the ability to analyze a large number of users' reported interactions across geographical boundaries, and was the most feasible and cost-effective approach. Before deploying the survey, we validated it in a pilot study with 243 users.

Furthermore, we selected different countries to ensure cultural diversity and user representation (i.e., regions supported by both Alexa and Google devices). Our primary criteria for choosing countries to recruit participants were: i) countries with a high adoption rate of both Google Home and Amazon Echo devices allowing comparison across devices; and ii) maximizing cultural diversity in our dataset (e.g., population size and language), instead of solely focusing on users from Europe and North America. Towards this, we conducted an online survey with participants from seven countries: Australia, France, Germany, India, Japan, the UK, and the US. We specifically aimed to understand the differences in interaction dynamics and investigate whether these differences impact their ability to accurately identify light behaviors. Furthermore, we also conducted interviews with six design experts to understand the challenges and opportunities to design light behaviors in smart speakers.

3.1 Survey

The participants in our survey were recruited using a professional service — Qualtrics Panel — to facilitate high quality data collection across different countries. We aimed to recruit 150 participants from each country. Our inclusion criteria were: i) the participant must be a user or owner of an Amazon Echo and/or a Google Home device, and ii) at least be 18 years of age. To ensure data quality, we used "attention checks" (e.g., a question asking participants to select a specific answer) in the survey. We also incorporated "speed checks" to filter out surveys completed within unreasonably short time, as a number of studies have identified completion speed as an important quality indicator for survey responses (e.g., [21]). We also filtered out participants with inappropriate responses to text-field based questions such as "When was your first interaction with a smart speaker?". We had a total of 1006 participants who met these criteria.

In the survey, we focused on three key areas of interest: (a) ownership and usage, wherein we asked participants about their smart speaker ownership duration, their first interaction with the device, number of devices they own, and how often they use the device; (b) interaction characteristics such as orientation to the device, how often they physically interact with the device and how much attention they pay to the light while interacting with the device; (c) accuracy and usability, in which we ask participants to identify different light behaviors as well as assessing perceived usability. The survey items described below have been tested internally and in a pilot study with 243 smart speaker owners to ensure clarity and avoidance of bias.

The survey begins with general questions about participant demographics and smart speaker ownership, including the number and type of devices they own. If participants had two or more different kinds of devices, they were asked to pick the device that they use most frequently to streamline the accuracy module. In terms of the sub-types of devices, Echo Dot, Echo, Echo Spot, and Echo Plus were the options provided for Amazon Echo Devices, while Home, Home Mini, and Home Max were the options provided for Google Home devices. We selected these devices given their reliance on light behaviors as well as their wide adoption [19, 20].

In order to understand the interaction dynamic users have with their smart speakers, we asked the participants a set of questions that established use of and reliance on light behaviors, including: how often they look at the light behaviors for information, their orientation with respect to the device while interacting with it, changes in their orientation with respect to the device if other people are in the room, how often they use the device, how much they rely on the light for information, and light behaviors they find particularly useful. We describe these questions in detail below.

Participants were asked to rate their usage frequency on a six item Likert scale (from "Never" to "Multiple times a day"). We then asked participants how often they physically interact with the device (e.g., by tapping on it or using the buttons on top of the device) using a seven item Likert scale (from "Never" to "Always"). We have defined this metric as physical interaction frequency (PIF) in this paper. We used a similar seven item Likert scale for four other questions: (a) "*Do you typically face your smart-speaker while interacting with it?*" (interaction orientation), (b) "*Does your orientation (facing towards or away from the smart-speaker) change when multiple people are in the room interacting with the smart-speaker?*" (people orientation), (c) "*When using the smart-speaker do you pay attention to the light on the top of the smart-speaker?*" (light attention), and (d) "*To what extent, do you believe you rely on the light patterns to obtain information?*" (reliance). We also asked participants their physical proximity to smart speakers during usual interaction using a five item Likert scale (from " 3 feet" to "Usually use the smart speaker from a different room").

In the subsequent section, we asked participants to identify light behaviors from smart speakers. *Each participant was presented with light behaviors only from the devices they owned.* We included all light behaviors available in a given device for the survey. These resulted in 16 light behaviors for Google Home Mini/Max, 17 for Google Home, and 11 for Amazon Echo devices [1]. The list of light behaviors is presented in Table 1 of the supplementary material. Each of the light behaviors was presented as a GIF along with a contextual clue intended to help the participants identify the behavior. An example of an accuracy question is presented in Figure 1. Participants were given five different options and were asked to choose the option that represented the light behavior[2]. For each light behavior, four options were selected randomly from the pool of existing light behaviors for that particular device, along with the correct option. These other four options were feasible alternatives for the light behavior.

We were also interested in determining how individuals interact with light behaviors from smart speakers. Toward this, a survey question asked participants about their frequency of light use ("*How often did you look at the light apparatus on your smart speaker in the last 7 days?*"). To understand the circumstances under which people looked towards the light for information, we had two questions: (a) "*Do you look at the smart-speaker to obtain the information about the following? Check all that apply.*" with the following options: (i)

[1]Light behavior numbers are at the time of conducting the survey; Since then, 1-3 light behaviors have been added to the devices, some exclusively to newer devices.
[2]We have provided all GIFs and associated contextual clues in the supplementary material (see Tables 1 & 2).

Smart-speaker Resetting, (ii) Smart-speaker Booting up, (iii) Volume Changes, (iv) Wifi Connections and Errors with Connections, (v) Smart-speaker Processing Your Commands, (vi) Alarm, Timer or Reminder, (vii) Notifications, (viii) Smart-speaker Updates, (ix) Smart-speaker is Muted; and (b) "*Under what circumstances do you look at the light apparatus? Check all that apply.*" with the following options: (i) Each time I talk to the smart-speaker (ii) When the smart-speaker is not responding/acknowledging, (iii) When the smart speaker misunderstands my commands, (iv) When smart-speaker requires clarification (e.g., "Sorry, I don't know that"), (v) When the smart-speaker takes longer than usual to process my request, (vi) Other (Specify). Lastly, we deployed the System Usability Scale (SUS) [11] to assess perceived usability of light behaviors in smart speakers. We specifically chose SUS given its wide adoption in prior studies on smart speakers and applicability for the light behavior system.

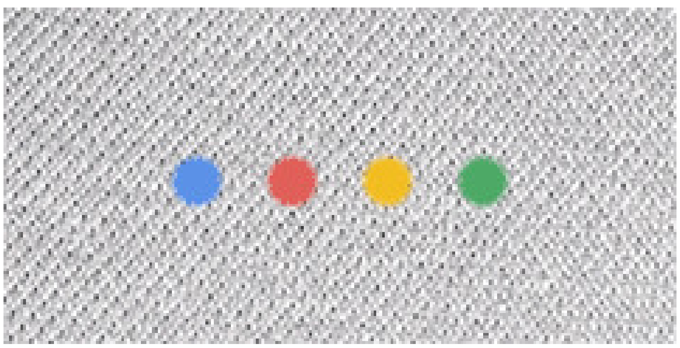

Figure 1: Example of an image (in the form of an animated GIF) presented to the survey participants. This GIF was accompanied by the following contextual clue: "*Scenario: After connecting your google home device to a power source, you observe the following light pattern. What information do you think the device is trying to convey?*". We used GIFs from official documentation provided by Amazon Echo [2] and Google Home [15].

3.2 Design Sessions

We also conducted design sessions to understand different expert perspectives on the existing light behaviors, and to determine the steps necessary to develop a light vocabulary for smart speakers. For the design sessions, we identified six individuals in the field of design and interaction. The recruitment process involved compiling a repository of developers, technological evangelists, researchers and practitioners in the field of either voice communication design or lighting design. We then proceeded to contact each of these individuals through email and Twitter, asking whether they would be willing to participate in a 30 minute interview to offer their expert opinions and assessments of the light behavior systems in Amazon Echo and Google Home devices. We placed no constraints on regions while recruiting experts leading to different cultural backgrounds. We concluded recruitment with six design experts: four individuals from the Women in Voice[3] — a group specifically focusing on voice interactions, one engineer from the Amazon Alexa team, and one assistant professor in a public university specializing in lighting design. The participants come from a range of backgrounds, including UI/UX design, engineering, lighting research, and voice interaction development. Each interview participant has had extensive experience and knowledge in one of the two fields: voice communication design or lighting design. These sessions were conducted virtually after survey deployment and had no impact on the questions asked in the survey.

The design sessions are conducted as follows: each participant is first given an online list of Amazon Echo and Google Home light behaviors and their corresponding informational states to help inform their responses to the session. Then, we asked them about useful and effectively communicated light behaviors. We also asked them to identify common characteristics of effectively designed light behaviors, and whether there is a design rationale behind the light behaviors in smart speakers. A brief overview of the script of the design session is included in the supplementary material (see Table 4).

4 RESULTS

4.1 Survey Results

4.1.1 Participant Demographics and Usage Characteristics. We had a total of 1,006 valid participants in our survey dataset after cleaning the dataset and removing outliers. Outliers were removed using two procedures: values beyond 1.5 times the interquartile range were removed from the dataset, and Smirnov-Grubbs Test was used to remove outliers. We applied these two procedures on our two main variables of interest: accuracy (proportion of accurately identified light behaviors) and usability (SUS score). We had 642 users of Amazon Echo devices and 364 users of Google Home devices. The participants were from 7 different countries: Australia (n = 145), France (n = 145), Germany (n = 144), India (n = 148), Japan (n=135), UK (n = 150), and the US (n = 139). The split in terms of devices in each country is given in Figure 2. A chi-square test of independence showed that there was a significant association between device type and country, ($\chi^2(6, N = 1006) = 224.7$, $P < 0.01$).

On average, each user had 1.68 ($SD = 1.08$) devices for 15.18 ($SD = 10.53$) months. The average age of our participants was 36 ($SD = 12.52$), and 26.04% held bachelor's degrees (n = 262), with 22.96% having Master's degrees (n = 231). In terms of household composition, 45.33% live with a family of 1-3 members (n = 456), and 27.93% live with a family of 4 or more members (n = 281). 50.3% of the users were female (n = 506), with 49.30% male (n = 496), 0.3% non-binary (n = 3), and 0.1% prefer not to disclose their gender (n = 1). In terms of impairments, 3.2% reported some form of vision impairment (n = 32), and 4% reported either partial or complete hearing impairment (n = 40). Table 5 in the supplementary material describes these variables in further detail. The device usage characteristics reported by the users is depicted in Figure 3.

In our dataset, the top three circumstances under which users look at the smart-speaker are: "each time I talk to the smart-speaker", "when the smart-speaker is not responding/acknowledging", and

[3]Women in Voice. https://womeninvoice.org/

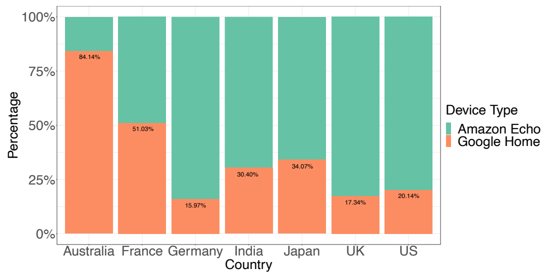

Figure 2: Country Based Amazon Echo and Google Home Split.

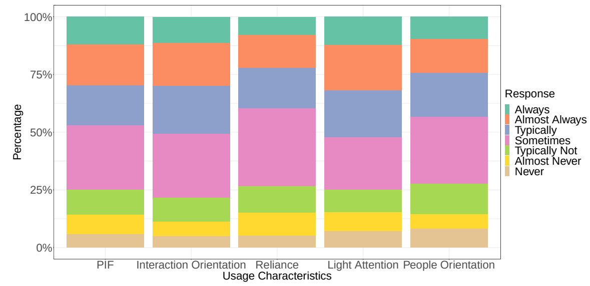

Figure 3: Users' responses to device usage characteristics.

"when the smart speaker misunderstands my commands". Also, the top three information seeking categories when users look at their smart speakers are: volume changes, WiFi connections and errors with connections, and smart-speaker processing commands.

The relationships between demographics and usage characteristics are summarized in Table 1. A Shapiro-Wilk test showed that age, $W = 0.93$, $p < 0.01$, was significantly non-normal. This requires non-parametric analyses. As such, we conducted Kruskal-Wallis rank sum tests to analyze the relationships between age and usage characteristics. Further, we used Pearson's Chi-squared test of independence to analyze the relationships between the device usage characteristics and country, education, household, and gender. All of the usage characteristics significantly varied by country of origin, education levels, and household composition.

4.1.2 Accuracy and Device Type. Overall, the accuracy is quite low — **on average, each user accurately identified only 36.61% of the light behaviors presented to them**. An examination of the accuracy distribution reveals that it exhibits positive skew and negative kurtosis. A Shapiro-Wilk test showed that accuracy, $W = 0.98$, $p < 0.01$, was significantly non-normal. Hence, we employ non-parametric tests for analyses involving accuracy.

The accuracy is significantly different among devices: ($H(1) = 7.16$, $p < 0.01$). Focused comparisons of the mean ranks between the devices showed that **Amazon Echo users accurately identified a higher proportion of light behaviors than Google Home users** (*difference* = 50.93, Mean Accuracy $\text{Proportion}_{AmazonEchoUsers}$ = 0.37, Mean Accuracy $\text{Proportion}_{GoogleHomeUsers}$ = 0.34) (see Figure 4). The critical difference (α =.05 corrected for the number of tests) was 37.36. We hypothesize that this may be due to a number of factors: (a) Amazon Echo devices use a lower number of light behaviors making it easier for users to remember, (b) Amazon Echo devices employ clear color-based distinctions between light behaviors, which makes them more distinguishable, and (c) Amazon Echo's light apparatus takes the form of a ring, affording the display of a wider spectrum of patterns compared to the LED dots on top on Google Home.

4.1.3 Accuracy and Demographics. We used beta regression for this analysis as the dependent variable — accuracy — is a proportion (no. of accurately identified light behaviors / total no. of light behaviors), and regular regression methods cannot be applied to compositional data such as proportions. Our analysis shows that **age (β = -0.01, p < 0.01) is a significant negative predictor of the proportion of accurately identified behaviors**. We hypothesize that this negative relationship between age and accuracy arises due to the fact that smart speaker usage patterns differ across different age groups, with younger users interacting with smart speakers in distinctly different ways compared to older users of smart speakers [7]. These specific domains of interaction might influence the accuracy with which users can identify light behaviors. Accuracy does not vary significantly across education levels ($H(10) = 11.75$, $p = 0.30$), household compositions ($H(3) = 0.14$, $p = 0.99$), or gender ($H(3) = 6.78$, $p = 0.08$).

A Kruskal-Wallis test analyzing the relationship between proportion of accurately identified light behaviors and the user's country of residence showed that **accuracy does not significantly vary based on the user's country of residence** ($H(6) = 12.43$, $p = 0.05$). Moreover, a beta regression analysis of the relationship between proportion of accurately identified light behaviors and the duration of device ownership showed that **duration of device ownership is not a significant predictor of accuracy** ($p = 0.05$). Furthermore, **the number of devices is also not a significant predictor of accuracy** ($p = 0.05$).

4.1.4 Accuracy and Device Usage Characteristics. In our dataset, individuals who mentioned never looking at the smart speaker while interacting with it accurately identified 33% of light behaviors on average (i.e., mean proportion of accurately identified light behaviors = 0.33). This is significantly lower than those who mentioned looking at the smart speakers "sometimes" (mean = 0.39, $\beta = 0.31$, $p < 0.01$) or "typically" (mean = 0.37, $\beta = 0.23$, $p < 0.01$) during interactions. In other words, individuals who never look at the visual features of smart speakers are on average worse in correctly identifying light behaviors, which makes sense. However, users who indicated that they "almost always" look at smart speakers during interaction had equal accuracy in identifying light behaviors (mean = 0.33, $\beta = 0.15$, $p = 0.04$) as individuals who never look at the smart speaker while interacting with it. We speculate this might be due to information overload (i.e., these users being exposed to non-distinguishable light patterns for a much wide range of informational states). However, future studies should aim to replicate this finding and further explore the potential causes.

On the other hand, we found that **reliance on light behaviors is a significant predictor of accuracy** ($p = 0.02$). Participants who

	Age			Country			Education			Household			Gender		
	H	df	P	χ^2	df	P	χ^2	df	P	χ^2	df	P	χ^2	df	P
Light Attention	52.36	60	0.75	213.94	36	**<0.01**	92.43	60	**<0.01**	50.06	18	**<0.01**	17.02	18	0.52
Reliance	87.41	60	**0.01**	205.64	36	**<0.01**	170.63	60	**<0.01**	50.97	18	**<0.01**	21.36	18	0.26
Usage Frequency	71.92	60	0.14	250.44	30	**<0.01**	108.35	50	**<0.01**	36.83	15	**<0.01**	16.09	15	0.38
Interaction Orientation	76.68	60	0.07	200.27	36	**<0.01**	88.32	60	**0.01**	33.24	18	**0.02**	26.34	18	0.09
People Orientation	73.17	60	0.12	202.29	36	**<0.01**	97.31	60	**<0.01**	33.08	18	**0.02**	15.95	18	0.60
PIF	74.67	60	0.10	236.17	36	**<0.01**	155.18	60	**<0.01**	38	18	**<0.01**	22.61	18	0.21

Table 1: Relationships between demographics and device usage characteristics. Significant relationships denoting associations between variables are presented in bold. Importantly, all of the usage characteristics significantly vary by country of origin, education levels, and household composition.

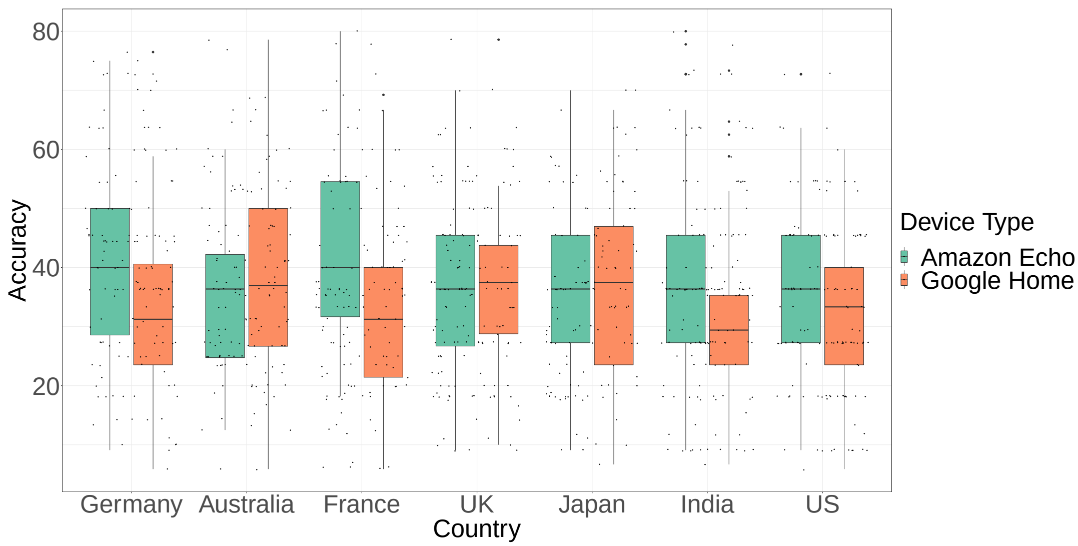

Figure 4: Device-based variation in accuracy by Country. Overall, Amazon Echo device users exhibit significantly higher accuracy than Google Home device users. Users from Germany exhibit highest average accuracy, while US exhibits the lowest; Users from France exhibit highest variation of mean accuracy across devices.

indicated that they never rely on light patterns in smart speakers to obtain information had a lower accuracy in identifying light behaviors (mean = 0.34) compared to other groups. Participants who "sometimes" (mean = 0.37, $\beta = 0.20$, $p = 0.03$), "typically" (mean = 0.39, $\beta = 0.23$, $p < 0.01$), and "almost always" (mean = 0.35, $\beta = 0.29$, $p < 0.01$) relied on light patterns for information had much better accuracy.

4.1.5 Perceived Usability and Device Type. Overall, the **perceived usability of the light behavior system across the users is quite low, with an average SUS score of 49.05**. As a comparison point, Bangor et al. [3] found that the mean SUS score from 964 usability tests across various interface types was 70 (less than 5% of all studies had a SUS score of below 50). The perceived usability distribution exhibited positive skew and kurtosis. A Shapiro-Wilk test showed that perceived usability, $W = 0.97$, $p < 0.01$, was significantly non-normal. Hence, we employ non-parametric tests for analyses involving perceived usability. **There is no significant difference in the perceived usability among the devices**, $H(1) = 0.69$, $p = 0.4$ (see Figure 6).

4.1.6 Perceived Usability and Demographics. In our dataset, perceived usability did not vary with age ($p = 0.22$), gender ($H(3) = 3.77$, $p = 0.29$), education levels ($H(10) = 3.42$, $p = 0.97$), or household composition ($H(3) = 7.12$, $p = 0.07$). It also did not vary by number of devices owned ($p = 0.7$) or ownership duration ($p = 0.55$). However, **there was a significant difference in perceived usability among users from different countries** ($H(6) = 24.54$, $p < 0.01$). We conducted pairwise comparisons using Wilcoxon rank sum test (also known as Mann-Whitney's U test) with continuity correction, and found that **there was a significant difference between the average perceived usability of users from the US and every other country** (see Figure 6). We hypothesize that these significant differences in usability arise primarily because usability testing of these devices and the light behaviors apparatus might mostly be conducted in the North American regions, and might not take into consideration the opinions and perceptions of people from

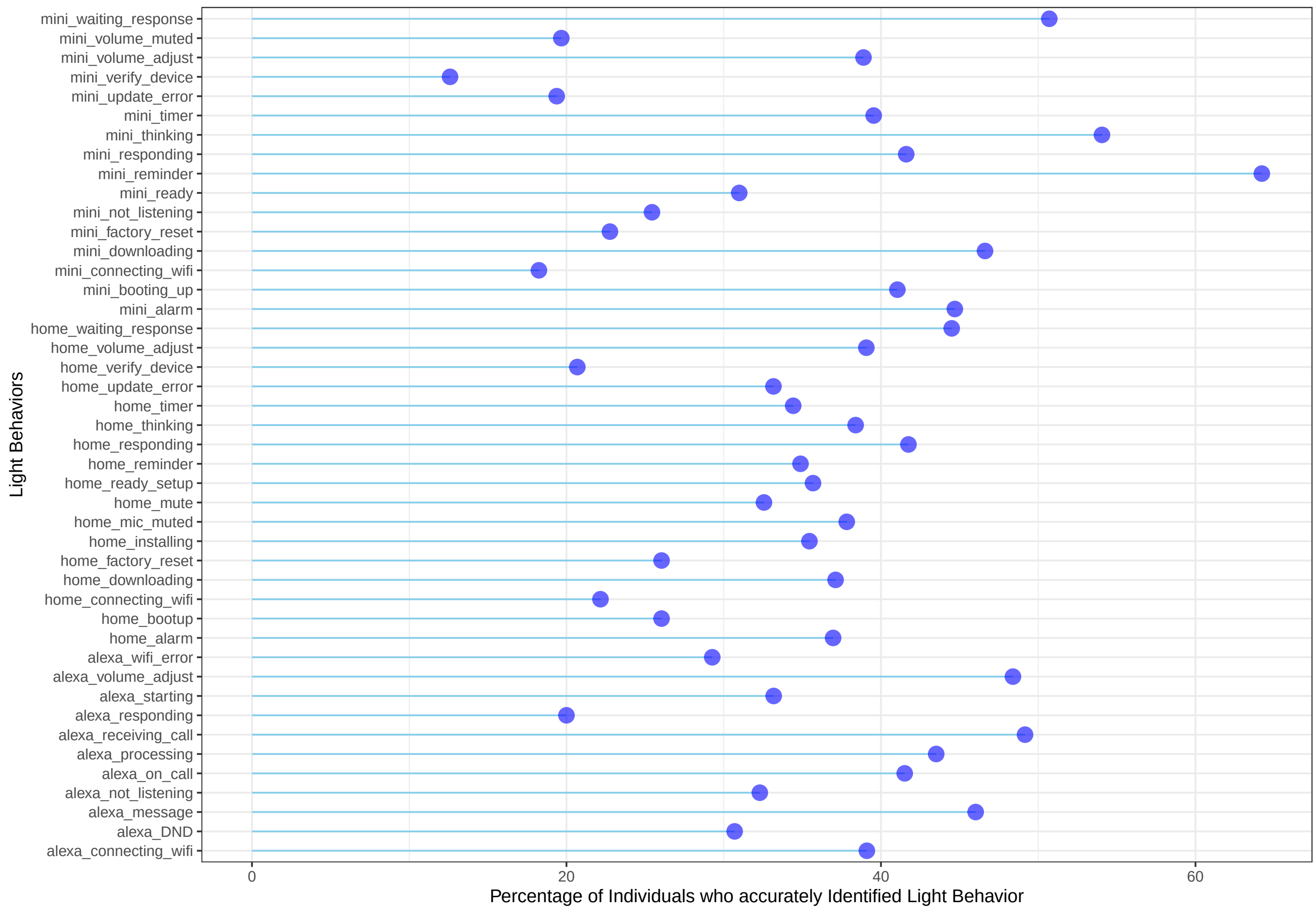

Figure 5: Percentage of individuals who accurately identified various light behaviors. The top three light behaviors that are most identifiable are: Reminder notification in Google Mini/Max, Timer in Google Mini/Max and Google Mini/Max waiting for a response. The three least identifiable behaviors are: Google Mini/Max verifying device, Amazon Echo responding and Google Mini/Max connecting to WiFi. The mapping and exact values of the percentages of individuals who accurately identified the various light behaviors is given in Table 6 of the Supplementary Material.

diverse geographical regions. There has been strong preliminary work to support this hypothesis, with cultural differences in interacting with VUIs and IPAs between native and non-native English speakers being shown to have a significant impact on the perceived usability of these interfaces [35, 44].

4.1.7 Perceived Usability and Device Usage Characteristics. We also found that the **usage frequency of a smart speaker is a significant predictor of the usability of the smart-speaker (p = 0.04)** in our dataset. Users who did not use smart speakers frequently had a better perceived usability score (mean SUS = 51.63) than those who use smart speakers at least once a day (mean = 49.42, β = 2.52, p < 0.01). We speculate this might be due to the fact that frequent use exposed users to more usability issues.

Further, **attention paid to the light on top of a smart-speaker is a significant predictor or the reported usability of the smart-speaker** (p < 0.01). In our dataset, users who never look at light behaviors indicated better perceived usability (mean score = 51.57) than users who sometimes (mean = 47.52, β = 3.05, p = 0.03) look at the lights in smart speakers. We hypothesize that these findings result from the inherent confusion and frustration that arises due to low interpretability of current light behaviors in smart speakers. As stated in 4.1.4, users who sometimes or typically look at the smart speakers during interaction are able to accurately identify 37-39% of the light behaviors, but this indicates that approximately 61-63% of the light behaviors are still not interpretable, leading to users perceiving the system as cumbersome and complex.

In our dataset, accuracy and perceived usability are not correlated (Pearson's r(1004) = -0.01, p = 0.7). However, we did find that **reliance on light behaviors to obtain information is a significant predictor of perceived usability** (p = 0.02). Compared to those who never rely on the light behavior system for information (mean usability = 47.17), those who "always" rely on these light behaviors report a higher perceived usability score (mean = 52.42, β = -3.91, p < 0.01). This result makes sense when placed in the context

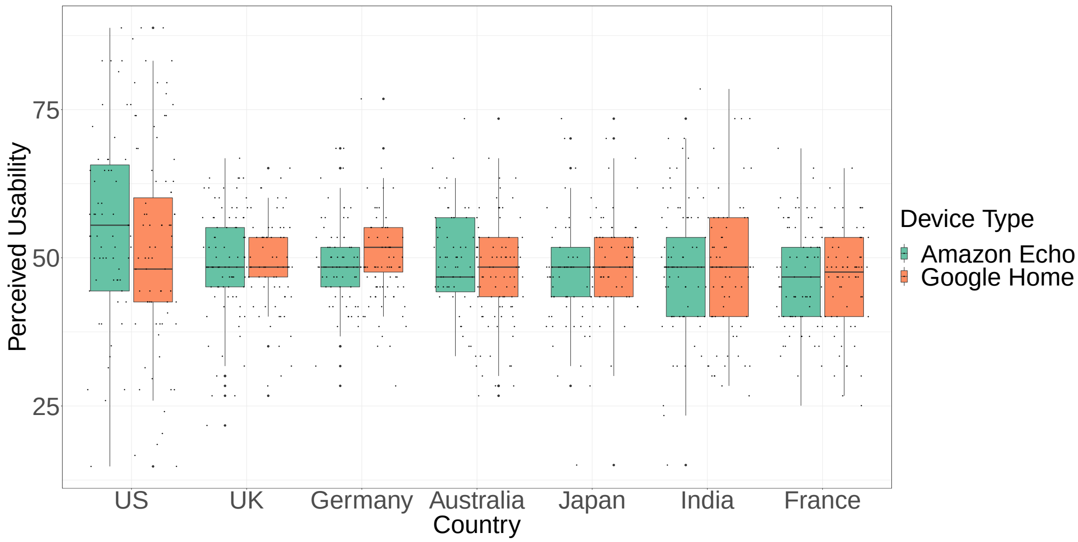

Figure 6: Device-based variation in Perceived Usability by Country. Overall, there is no significant difference in perceived usability across Amazon Echo's and Google Home's light behaviors. There was a significant difference between the mean perceived usability of users from US and every other country, with the US also exhibiting the highest variation in perceived usability across devices.

of the relationship between reliance and accuracy (4.1.4) – users who always rely on the light behaviors for information are both able to accurately identify a higher proportion of light behaviors as well as find the light behavior system more usable.

4.2 Interview Results

In the following section, we will present major themes from our interviews with six designers who are experts in either voice interfaces or light behaviors. We recorded the interviews and then transcribed the recordings. One author performed a bottom-up thematic analysis to identify common themes using a qualitative interpretivist approach. Another author checked the generated themes against data to ensure consistency and avoid any potential bias.

All interview participants agreed that light behaviors in smart speakers help to improve user experience in a number of ways.

4.2.1 Avoiding interruptions. P4 pointed out the importance of light behaviors given the unique affordances of smart speakers: "*Alexa and Google Home are [designed to be] voice first. They really want to avoid cluttering the audio experience with random noise. First of all, you wouldn't want Alexa just repeating 'Hey, I need WiFi, Hey I need WiFi' till she gets WiFi. [Light behaviors] are just very convenient ways of saying 'Hey I need your attention — I am in a certain state and I might need your attention to resolve that'*". That is, light behaviors can provide useful information without demanding a user's immediate attention. P4 also noted that unwanted interruptions from smart speaker audio can lead to serious user annoyance: " *When [my co-worker's smart speaker] was muted, every so often, it would chime in saying: 'Hey, I am muted by the way. So if you want to talk to me please unmute me'. And for me, that was incredibly irritating*". Light behaviors in smart speakers help to improve the usability of smart speakers by minimizing such task and user interruptions. These behaviors essentially redirect information between the center and periphery of the user's attention as needed, embodying the principles of calm technology [43].

4.2.2 Complementing audio interactions. Light behaviors can complement audio interactions from smart speakers. Some participants commented that the use of light behaviors can "*humanize*" the smart speaker interfaces. P5 noted that: "*Alexa's tone does not go up or down [during interactions]. There's a little bit of the intonation in the speech, but she's not getting louder or softer [...] it's not as nuanced as human interactions can be. But, I think the inclusion of the light — sometimes spinning, sometimes stationary — [can make Alexa's interactions] more nuanced, like human interactions can be nuanced*". P4 had similar comments: "*[when interacting with Alexa,] the cyan pointing in the direction of the person speaking, that's very important in making it seem more humane. That [is,] she is listening to you and not just listening to the world*'

However, experts also identified a number of usability issues and challenges with current light behaviors in smart speakers.

4.2.3 Current light behaviors in smart speakers can be difficult to interpret. The light behaviors in smart speakers often are not iconic [16] — lacking consistent interpretation. P4 noted the difficulty of interpreting light behaviors for reminder notifications and muted device status in Google Home devices: " *I don't necessarily associate orange with a microphone being muted [muted and the reminder notification] both appear to be the same light pattern, as far as I can tell. [...] I think, in that case, if I saw the reminder notification, I might think that the device was muted. I might try to turn it up and be confused there*".

The lack of iconic light behaviors requires additional contexts from the device to ensure successful interactions. P6 mentioned: " *I don't think I would understand the visual feedback by itself, I would definitely need the voice message from the device to tell me what is wrong [...] I'm familiar with Google Home Mini. So I can tell if the light is popping [...] something is wrong with that. But I would think that voice message from the device is very important for me to understand the feedback.*". P2 specifically noted that "*[when WiFi errors happen], I would rather like to listen to a command or a prompt*" instead of just relying on light behaviors.

Being forced to interpret light behaviors can be particularly difficult for new users. P5 mentioned: " *if I'm completely new to the device, I would have no idea what's going on. It just takes a little bit of reading the documentation, especially during setup. I remember when I first got the device, it was like 'What am I doing?'*". Some of these challenges are also due to the relative novelty of complex light behaviors. P1 pointed out the difficulty of successfully designing light behaviors in smart speakers: "*if you don't have any past experience with those signals, you wouldn't know what that is. And if you're trying to introduce a new signal to users, it's really hard for people to take in. It takes time and experience.*"

Participants suggested to use familiarity as a strategy to improve understanding and interpretability. This means sourcing light behaviors from already known and learnt behaviors that are present in ubiquitous devices. "Tying it to an action they already know" was a simple yet effective suggestion that might considerably increase the ease of using the light behavior system. A good example here is the light behavior for the volume change — both Amazon Echo and Google Home devices use light behaviors that mimic a progression "bar" used across other devices including phones and TVs. The resultant light behavior is highly effective — volume changes is one of the mostly highly recognized light behaviors in our dataset (see Figure 5).

4.2.4 *Too many light behaviors can lower usability.*

A consistent theme in our interviews was that smart speakers use too many light behaviors to convey different states and information. P2 noted "*it could probably be a bit overwhelming. It's just a lot that is being communicated over the lights, and I think that it could be a bit hard to learn all these different light behaviors*". P3 also pointed out that requiring users to memorize different light patterns can reduce usability of these devices: "*I think it's too many! I think especially with our lives online, we have so many things bombarding us for our attention, in the home, on the computer, on our phones, that we don't have a lot of attention for things [...] I'm not going to spend the time to memorize what a certain thing means. If I have to go look something up, it has failed me in a certain way. Because, I don't have the time for that. So I think having to memorize the pattern or the color palette or other stuff doesn't really work anymore*".

P2 suggested to reduce the number of different light behaviors: "*I would concentrate on the most important ones. For example, listening, processing, alarm, download, calling functions, and notifications. And then I would dismiss the others, because from my perspective they're not too important to understand [...] because you won't have a WiFi issue every day. [By] reducing the number of light behaviors, people would actually learn those [behaviors] better*".

4.2.5 *Culture differences.*

Experts also noted the necessity for culture specific design of light behaviors given the global use of smart speakers. P5, who grew up in India but currently lives in the USA, pointed out "*we have a very culture-centric take on designing things or working, which doesn't necessarily translate well across the sea [...] I doubt the whole light thing would go down well with the average Indian user, who's speaking to Alexa in Hindi. Why is it red, why is it orange, why is it blue? I don't know if the average Indian user would have the same kind of empathy and response [to these light behaviors], and that's probably just a cultural thing*".

P3 also expressed similar concerns: "*someone from a non-western, non-European culture might interpret a color differently than from a western, European mindset. So that's something to consider. I don't know if other cultures might find red not to mean stop. The muted [state in Alexa and Google Home is represented by] either red or orange color. That could be interpreted differently and may take a bit of a learning curve for those folks.*"

4.2.6 *Accessibility challenges.*

Colors in light behaviors can also lead to accessibility challenges. P3 noted that: "*I think Alexa having so many colors instead of using patterns is easier for me to understand, but I also don't have an accessibility issue. If I was a person who couldn't see color this might have been a problem. So I think in that regard, [Google Home] devices having a rotating or a combination pattern might be a little bit better for somebody who can't see [these colors]*". P5 suggested that: "*it's better to be on the blue and cyan side because some people have trouble with green. My father used to be color-blind, so he couldn't really quite see green*". Overall, the experts recommended to test different colors and patterns to ensure that accessibility needs are addressed adequately by smart speakers.

4.2.7 *Gulf of evaluation.*

The lack of iconic and interpretable light behaviors can result in widening the "gulf of evaluation" [28] for users. In other words, it is particularly difficult for users to recover from errors and handle unexpected outcomes when they can't interpret light behaviors from smart speakers. P4 mentioned that " *If I saw a white light on my Google Home, and I didn't know what the notification was for, which happens occasionally on my Alexa devices — the notifications pop up, but I don't know what they're for — I would not ask, 'What are my notifications?', or 'What's up?'. I would not really know what to ask*". One potential way to address this gulf of evaluation is to enable users to directly query the interpretation of current light behavior (e.g., "Alexa, why are you showing this color?"). This can help users to better understand device state and recover from errors.

4.2.8 *Lack of customizability.*

There is currently a lack of options for customizing and personalizing light behaviors in smart speakers. P3 noted that "*[setting customized colors] was kind of a big thing with android phones — changing the color of the notification light, colors to whatever you personally prefer. So maybe that would be an option [for smart speakers] as well*". Being able to customize colors can help to address culture specific needs. It will also allow users to choose light behaviors that might reflect their own use cases (e.g., more interpretable patterns to support their own frequent use cases).

5 DISCUSSION

Smart-speakers are increasingly important consumer devices with the fastest growing user base of all emerging connected devices globally [20]. However, a core aspect of its interface — light behaviors — has been mostly overlooked in prior HCI work. This study addresses this gap by evaluating light behaviors in smart speakers across seven different countries. In this section, we compare and contrast our findings from our two datasets — the online survey and the design expert interviews. We also provide actionable design recommendations for a more understandable and usable light behavior in smart speakers.

In this study, we examined the current state of the light behaviors from two sets of devices with high global adoption [20] — Amazon Echo and Google Home devices. In our dataset, Amazon Echo device users exhibit better understanding of the light behaviors than Google Home device users. Overall, the number of correctly identified light behaviors was low (37% on average) even with contextual cues. Current smart speakers have too many light behaviors (4.2.4) — Google Home Mini/Max use 16 light behaviors and Google Home has 17 light behaviors in comparison to Amazon Echo devices with 11 light behaviors. Furthermore, these light behaviors often do not lead to consistent interpretation (i.e., lack of *iconic* behaviors). Our design experts pointed out a number of light behaviors with potentially ambiguous interpretation (e.g., reminders vs muted). The number of light behaviors along with the difficulty of consistent interpretation mean smart speaker users need to memorize light behaviors in their devices. In other words, light behaviors in smart speakers implicitly lead users to *recall* instead of *recognize*, which can considerably lower their usability [27].

On the other hand, the design experts noted that systems with simplified light vocabulary, such as traffic lights (a popular example among our design experts) work so well because of the few simple yet consistent uses of design elements such as colors and patterns. This schema of simple and few design elements lowers the burden on users across a variety of constructs — time (users spending lesser time to interpret these behaviors), mental burden (users not having to recall from a catalogue of behaviors), and ease of use. Overall, a number of design experts suggested reducing the number of different light behaviors in smart speakers to support most important user tasks and interactions. A simplified light behavior system can lower user confusion and improve overall usability.

Another potential reason for low interpretability of smart speaker light behaviors might be due to the need for training. When light behavior systems, or any systems for that matter, are not intuitive to the users, designers must rely on training users to use the system appropriately and effectively. The inherent lack of familiarity in current smart speaker light behaviors that several design experts brought up justifies a need for training smart speaker users to better understand these light behaviors. However, this training must happen in tandem with how users already use the system. One way of achieving this goal might be to accompany light behaviors in initial stages of smart speaker use with voice prompts that help users associate light behaviors with informational states. A number of design experts suggested that additional contexts (i.e., voice prompts) accompanying light behaviors can considerably enhance the learning experience of smart speaker users, especially at the beginning phase.

Our design experts also pointed out the lack of discoverability [28] to be a serious usability concern for current light behaviors in smart speakers. Discoverability enables users to easily determine what actions are available in a given product, as well as significantly improving task performance and usability scores in smart speakers [18]. The current light behaviors, however, are not easily discoverable (e.g., how would a new user know which light behavior indicates a reminder notification?). Furthermore, light behaviors, when designed effectively, can be a powerful tool to support discoverability — indicating system state and possible actions. However, the current poor understanding of the light behaviors can impede the discoverability of smart speakers. Not being able to accurately interpret light behaviors means that users miss out on critical informational states and useful feedback, which invariably degrades users' experience with smart speakers.

There were also discrepancies between our survey data and perceptions of design experts. The design experts indicated that light behaviors for error states (such as update errors, Wi-Fi errors, muted states) are communicated clearly and effectively. However, according to the survey results, error states such as *"mini_update_error"*, *"alexa_not_listening"*, and *"mini_not_listening"* are not highly identifiable behaviors (all < 40% identifiability — see Figure 5). This demonstrates a gap between the everyday users of these devices and those who are trained in design thinking and engineering of the product. The error states that appear logical, clear, and well-communicated to the design experts might not translate well in the real world settings. On the other hand, simple, intuitive light behaviors that draw from prior experience (e.g., volume changes; Amazon Echo pointing in the device of the speaker) are recognized by both experts and users. This clearly demonstrates that leveraging users' prior experiences and interactions with other devices and agents can result in successful and effective design of light behaviors. This calls for a greater emphasis on understanding what counts as intuitive for people from diverse backgrounds and experiences, and tailoring interface elements accordingly. Furthermore, this also supports the experts' recommendation to strategically use familiarity as a tactic to improve understanding and usability of light behaviors.

Cultural differences in understanding was a major point that emerged from the interviews. This is also consistent with our survey findings. In our dataset, all usage characteristics differed by country of origin. This indicates that behaviors such as paying attention to the light, looking at the speaker while interacting with it, frequency of usage, and frequency of physical interaction with the device vary across different countries. This provides a potential explanation as to why the usability differs across countries — it might be due to the fact that interaction with light behaviors and smart speakers fundamentally differ across cultures. Designers should aim to ensure light behaviors have consistent meaning with local cultural values and enable customization to improve understanding and usability of the light behavior system across different regions and cultures. Specifically, smart-speaker designers can benefit from an expanded analysis of diverse interpretations of color and light in non-western countries and cultures, as highlighted, for example, in Yokosawa et al.'s [45] and Yu's [46] work in cross cultural interpretations of color.

Further, future work can also be guided by principles governing the role of color and light in affective computing [38].

Moreover, the design experts note that users might run into accessibility issues while using smart speakers and their light behaviors. Currently, the accessibility features of Alexa [4] do not enable customizations for individuals with color blindness. There is also no provision to have information that is traditionally delivered through light to be communicated over audio only, which could better support individuals with vision impairments. One of our design experts hypothesized that the Google Home devices largely use a neutral color (white) to aid users with vision impairments. However, we have found no documentation to validate this claim [5]. We believe future iterations of smart speakers should specifically focus on making the light behaviors more accessible to a wider population.

Given these various considerations, and based on our survey and interview findings, we make the following design recommendations:

- **Simplify the light behavior system.** Reducing the number of light behaviors will enable the user to focus their attention on important and urgent information that they might need to know to interact effectively with the smart speaker. It will considerably reduce the burden on the user to learn and memorize different light behaviors that they might not need or use often. Limiting the light behaviors to a few key behaviors will improve the understandability and interpretability. In order to understand which behaviors to keep and which ones to dismiss, designers should run observational user studies to identify key use cases and light behaviors. Further, our analysis shows that "volume changes" (*mini_volume_adjust*, *home_volume_adjust*, and *alexa_volume_adjust* light behaviors), "Wifi connections and errors with connections", and "smart speaker processing your commands" are the top three categories of information users seek when they look at the smart speaker. Moreover, "Each time I talk to the smart-speaker", "When the smart-speaker is not responding or acknowledging", and "When the smart speaker misunderstands my commands" are the top three circumstances under which users look at the smart-speaker. We hope that designers can leverage our data and findings to develop a simpler, more intuitive light behavior system that is considerate of the kinds of information that users want and need.
- **Provide additional context.** Using voice prompts to provide additional information accompanying the light behaviors has two benefits: i) it allows new users to get accustomed to the meanings and contexts of the light behaviors and the information that they convey, enabling users to understand what the light behaviors mean, rather than having to commit meanings to memory; ii) it enables users with vision impairments to effectively interact with the device. For example: expert P6 in our interviews noted that voice prompts when errors happen (such as WiFi failure) would be helpful to ensure that users are informed of the breakdown, so the onus of extracting that information through a light behavior they may or may not understand is not completely on the user.
- **Allow for customizability.** Customization enables users to choose how the device behaves, and has numerous benefits: there is lesser burden on the designers to choose and develop universally understandable behaviors; it enables people with different impairments to successfully interact with the device in meaningful ways; and it accommodates users with diverse preferences. Further, customization has been shown to improve usability and trust among smart speaker users [12]. To support better customizability in smart speakers, designers can explore options provided in some Android phones, where users can select notification light patterns in their phones. This enables users to craft their own interactions from prior experiences they might have with devices. This also allows for users from different device ecosystems to quickly and efficiently adapt to a new smart speaker by customizing their preferences.
- **Improve accessibility.** As our experts noted, the use of color and different patterns in current light behaviors might lower accessibility for individuals with vision impairments. Future work should focus on providing guidelines and tools to better accommodate the needs of individuals with vision impairments. Furthermore, visual communication through light in smart speakers is important to accommodate a wide range of users with diverse needs. Better light communication in smart speakers will specifically benefit the deaf and hard of hearing (DHH) community. In the US, 20% of individuals (48 million) report some degree of hearing loss [29]. For them, relying on just audio communication can be a significant barrier to use smart speakers [8]. Effective light behaviors in smart speakers can help to address these accessibility issues. Through this work, we hope to guide designers through making design choices towards more accommodating and usable light behaviors.
- **Ground light behaviors in previously learned experiences.** Designers of smart speaker interaction should also aim to develop iconic light behaviors by leveraging users' experiences with other devices. That is, future work should explore how to better integrate existing knowledge and behaviors in a user-centered strategy to help users intuitively understand light behaviors, rather than having to rely on memory to interpret them. Behaviors such as volume changes and Echo responding to the user are good examples of interpretable behaviors that do not rely on memory, rather, they rely on existing behaviors that the users already know.

The wide adoption of smart speakers and their use of complex light behaviors provides a unique opportunity to develop an effective, interpretable, and accessible light vocabulary guided by these recommendations. In other words, designers of smart speakers can help to develop consistent guidance and rules regarding colors and patterns for different light behaviors, similar to a design vocabulary suggested by Harrison et al. [16] for point lights in smartphones. They examined different kinds of information conveyed through light in smartphones and leveraged expert opinions to iteratively

[4]Accessibility Features for Alexa https://www.amazon.com/gp/help/customer/display.html?nodeId=202158280
[5]Accessibility features on Google Nest speakers and displays https://support.google.com/googlenest/answer/9286728?hl=en&ref_topic=7195017

design, refine and test different light behaviors. However, they focused their design on "point lights" — light behaviors with only a single color and point sources. Given that smart speaker light behaviors are not constrained by these parameters, a rich and comprehensive light vocabulary based on color, speed, pattern, and intensity can potentially be used to convey increasingly complex information states in a consistent and interpretable way. Such a light vocabulary will be useful not only for the future design of smart speakers but also for light as a communication modality in general.

6 LIMITATIONS

The study has a number of limitations. First, the device usage characteristics are based on self-report, and hence they are susceptible to reporting bias. Users' actual interactions with smart speakers might be different from what they perceive, and subsequently, report. This might have an impact on the associations we have established between accuracy, usability, and device usage characteristics. Second, we have not explored culture-specific light behavior associations in great detail. Evaluating the impact of color and pattern interpretations in greater depth might provide insights into interpretability and perceived usability. Third, our design experts have mostly had experience with either Amazon Echo devices or Google Home devices, but not both. This might bias their assessment of useful light behaviors. Finally, we did not explore usage patterns of smart speakers with displays (e.g., Amazon Echo Show). User interactions and interpretation of light behaviors might be different for such smart speakers.

7 CONCLUSION

This is the first study to conduct an empirical analysis of light behaviors in smart speakers across different cultures. For this, we collected data from an online survey across seven countries. We also conducted six design sessions with experts. Our findings show that current light behaviors in smart speakers are difficult for users to interpret. Our data also indicate that users from various countries might interact with light behaviors and smart speakers differently. Our findings call for rethinking of the existing practices and honing down on culture specific differences to design better, more coherent light behaviors. We have also provided design recommendations towards more cohesive and comprehensible light behaviors in smart speakers. Given the ubiquity of smart speakers, these light behaviors impact a global user population. Our findings, thus, can potentially help the next iterations of smart speakers to have more consistent and easier to understand light patterns, significantly benefiting their large user base.

REFERENCES

[1] 2020. Global smart speaker market 2021 forecast. https://www.canalys.com/newsroom/canalys-global-smart-speaker-market-2021-forecast

[2] Amazon. 2019. What Do the Lights on Your Echo Device Mean? https://www.amazon.com/gp/help/customer/display.html?nodeId=201601790. [Accessed 14 August, 2019].

[3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[4] Kim Baraka, Ana Paiva, and Manuela Veloso. 2016. Expressive lights for revealing mobile service robot state. In *Robot 2015: Second Iberian Robotics Conference*. Springer, 107–119.

[5] Kim Baraka, Stephanie Rosenthal, and Manuela Veloso. 2016. Enhancing human understanding of a mobile robot's state and actions using expressive lights. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 652–657.

[6] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[7] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.

[8] Johnna Blair and Saeed Abdullah. 2019. Understanding the Needs and Challenges of Using Conversational Agents for Deaf Older Adults. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 161–165.

[9] Johnna Blair and Saeed Abdullah. 2020. It Didn't Sound Good with My Cochlear Implants: Understanding the Challenges of Using Smart Assistants for Deaf and Hard of Hearing Users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–27.

[10] Jeff Blankenburg. 2017. Amazon Echo Devices Rolling Out to Customers in Over 80 Countries. https://developer.amazon.com/blogs/alexa/post/3d8b2e1e-d8c5-41f6-a193-5456a590ef9a/amazon-echo-now-available-to-customers-in-over-60-countries.

[11] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[12] Eugene Cho, S Shyam Sundar, Saeed Abdullah, and Nasim Motalebi. 2020. Will deleting history make alexa more trustworthy? effects of privacy and content customization on user experience of smart speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] Dianne Cyr, Milena Head, and Hector Larios. 2010. Colour Appeal in Website Design within and across Cultures: A Multi-Method Evaluation. *International Journal of Human-Computer Studies* 68, 1 (Jan. 2010), 1–21. https://doi.org/10.1016/j.ijhcs.2009.08.005

[14] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2015. Interactive light feedback: Illuminating Above-Device gesture interfaces. In *IFIP Conference on Human-Computer Interaction*. Springer, 478–481.

[15] Google. 2019. Learn about the lights on your speaker. https://support.google.com/googlenest/answer/7073219?hl=en. [Accessed 14 August, 2019].

[16] Chris Harrison, John Horstman, Gary Hsieh, and Scott Hudson. 2012. Unlocking the expressivity of point lights. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1683–1692.

[17] Frederic Kerber, Christoph Hirtz, Sven Gehring, Markus Löchtefeld, and Antonio Krüger. 2016. Managing smartwatch notifications through filtering and ambient illumination. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct*. 918–923.

[18] Philipp Kirschthaler, Martin Porcheron, and Joel E Fischer. 2020. What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.

[19] Ilker Koksal. 2020. The Sales Of Smart Speakers Skyrocketed. https://www.forbes.com/sites/ilkerkoksal/2020/03/10/the-sales-of-smart-speakers-skyrocketed/.

[20] Paul Lee. 2018. Smart Speaker Sales Set to Soar. https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/smart-speaker-voice-computing.html.

[21] Dominik Johannes Leiner. 2019. Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. In *Survey Research Methods*, Vol. 13. 229–248.

[22] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5286–5297.

[23] Thomas J. Madden, Kelly Hewett, and Martin S. Roth. 2000. Managing Images in Different Cultures: A Cross-National Study of Color Meanings and Preferences. *Journal of International Marketing* 8, 4 (Dec. 2000), 90–107. https://doi.org/10.1509/jimk.8.4.90.19795

[24] Andrii Matviienko, Wilko Heuten, Vanessa Cobus, Heiko Müller, Jutta Fortmann, Andreas Löcken, Susanne Boll, Maria Rauschenberger, Janko Timmermann, and Christoph Trappe. 2015. Deriving Design Guidelines for Ambient Light Systems. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia - MUM '15*. ACM Press, Linz, Austria, 267–277. https://doi.org/10.1145/2836041.2836069

[25] Andrii Matviienko, Maria Rauschenberger, Vanessa Cobus, Janko Timmermann, Heiko Müller, Jutta Fortmann, Andreas Löcken, Christoph Trappe, Wilko Heuten, and Susanne Boll. 2015. Deriving design guidelines for ambient light systems. In *Proceedings of the 14th international conference on mobile and ubiquitous multimedia*. 267–277.

[26] Heiko Müller, Andreas Löcken, Wilko Heuten, and Susanne Boll. 2014. Sparkle: an ambient light display for dynamic off-screen points of interest. In *Proceedings of*

the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. 51–60.

[27] Jakob Nielsen. 1994. Enhancing the Explanatory Power of Usability Heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*. ACM Press, Boston, Massachusetts, United States, 152–158. https://doi.org/10.1145/191666.191729

[28] Donald A. Norman. 2013. *The Design of Everyday Things* (revised and expanded edition ed.). Basic Books, New York, NY.

[29] National Institute on Deafness and Other Communication Disorders. 2019. Quick Statistics About Hearing. (2019). https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

[30] Sarah Perez. 2020. Nearly 70% of US Smart Speaker Owners Use Amazon Echo Devices.

[31] Henning Pohl, Justyna Medrek, and Michael Rohs. 2016. ScatterWatch: subtle notifications via indirect illumination scattered in the skin. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services.* 7–16.

[32] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems.* 1–12.

[33] Aljoscha Pörtner, Lilian Schröder, Robin Rasch, Dennis Sprute, Martin Hoffmann, and Matthias König. 2018. The Power of Color: A Study on the Effective Use of Colored Light in Human-Robot Interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, 3395–3402.

[34] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. " Alexa is my new BFF" Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* 2853–2859.

[35] 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.* 548–553.

[36] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference.* 857–868.

[37] Mario Silic, Dianne Cyr, Andrea Back, and Adrian Holzer. 2017. *Effects of Color Appeal, Perceived Risk and Culture on User's Decision in Presence of Warning Banner Message.* SSRN Scholarly Paper ID 2932982. Social Science Research Network, Rochester, NY.

[38] Marina V Sokolova and Antonio Fernández-Caballero. 2015. A review on the role of color and light in affective computing. *Applied Sciences* 5, 3 (2015), 275–293.

[39] Sichao Song and Seiji Yamada. 2017. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI.* IEEE, 2–11.

[40] Sichao Song and Seiji Yamada. 2018. Bioluminescence-inspired human-robot interaction: designing expressive lights that affect human's willingness to interact with a robot. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.* 224–232.

[41] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the role of conversational cues in guided task support with virtual assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–7.

[42] Ran Wei, Xiaoxia Wan, Qiang Liu, Guo Cao, and Haiwen Wang. 2017. Regional Culture Preferences to LED Light Color Rendering. In *Advanced Graphic Communications and Media Technologies (Lecture Notes in Electrical Engineering)*, Pengfei Zhao, Yun Ouyang, Min Xu, Li Yang, and Yujie Ouyang (Eds.). Springer, Singapore, 33–40. https://doi.org/10.1007/978-981-10-3530-2_5

[43] Mark Weiser and John Seely Brown. 1996. Designing calm technology. *PowerGrid Journal* 1, 1 (1996), 75–85.

[44] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services.* 1–9.

[45] Kazuhiko Yokosawa, Karen B Schloss, Michiko Asano, and Stephen E Palmer. 2016. Ecological effects in cross-cultural differences between US and Japanese color preferences. *Cognitive Science* 40, 7 (2016), 1590–1616.

[46] Hui-Chih Yu. 2014. A cross-cultural analysis of symbolic meanings of color. *Chang Gung Journal of Humanities and Social Sciences* 7, 1 (2014), 49–74.