Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Impact of auditory sense on trust and brand affect through auditory social interaction and control

Atieh Poushneh

California State University-Bakersfield, School of Business and Public Administration, Management and Marketing Department, United States

A R T I C L E   I N F O

Keywords:
Voice assistant
Social interaction
Surprise
Trust

A B S T R A C T

Voice assistants—or voice-enabled artificial intelligence—have changed the way people interact with their surroundings dramatically. Utilizing an enactive view of social cognition theory, this study demonstrates how voice assistants can act as [semi] autonomous agents to hold instantaneous social interactions with consumers. This research employed two experimental studies. Study 1 used two voice assistant mobile applications, Microsoft Cortana and Google Assistant, and Study 2 used Amazon Alexa and Microsoft Cortana. The contributions this paper makes are two-fold. First, the results illustrate how perceived auditory sense drives perceived auditory control through auditory social interactions with a voice assistant that lead to brand affect and consumers’ trust in the voice assistant. Second, results shed light on the role of surprise as a repelling drive that attenuates the effect of perceived auditory control on brand affect.

1. Introduction

Voice assistants are promoting “online mentality” (Rauschnabel et al., 2018) and becoming a significant part of consumers’ daily lives (Kim et al., 2018). Examples of voice assistant mobile applications are Apple’s Siri, Google Assistant, and in-home smart speakers are Amazon’s Echo, Google’s Home, and Apple’s Home.

A voice assistant or a [semi] autonomous agent refers to the auditory ability to sense the environment and act upon it. An autonomous agent can sustain its identity and differentiate itself from its environment (Di Paolo and Thompson, 2014). Voice assistants use an algorithm developed by computer scientists to be a [semi] “autonomous,” “intelligent,” or “emotionally intelligent” agent. Voice assistant technology has been criticized for not being a fully autonomous or a conscious agent acting upon the environment (Damasio, 2019) to express emotion in voice (Shuller and Schuller, 2018) internally as sadness, disappointment, or joy.

Nonetheless, developers are working on algorithms to give voice assistants, social characteristics. A recent study documented people reacting to a robot that was asking them not to shut it off (Hortmann et al., 2018), showing that people respond socially to robots demonstrating human-like behavior. It is hard to switch off a robot begging you not to. Such a robot demonstrates some identities and can act [semi] autonomously. Voice assistants can also act as [semi] autonomous agents to sense and act upon the environment and exchange information with consumers (e.g., Kyselo, 2019) in real time (Spaulding, 2014) and transforms the dynamics of their information processing and decision-making (Brill et al., 2019).

Yet, there is no agreement on adopting an enactive view of cognition to examine how voice assistant as a [semi] autonomous agent can enact consumers’ cognition (e.g., Kyselo, 2019; Metta et al., 2010). Previous literature has scrutinized the embedded and embodied view of cognition to address how voice assistants influence consumers’ decision-making (Davenport et al., 2020; Pagni et al., 2019; Strayer et al., 2017; Morituchi, 2019). Yet, it is still unclear how the auditory sense of voice assistant influences their auditory embodiment and consumers’ responses (Deng et al., 2019; Schuetzler et al., 2018; Kramer et al., 2009). For example, Kim et al. (2018)’ study showed that nonverbal cues (e.g., visual and gesture) and verbal embodiment (e.g., speech) influence consumers’ trust. In contrast, Schuetzler et al. (2018) demonstrated that verbal and nonverbal (e.g., gaze, body movement) embodiments do not influence consumers’ perception.

Previous research studies have shown that embedding communication ability (Guzman, 2019)—such as perceived ease of use and usefulness (Morituchi, 2019)—through interactions with voice assistants enhances consumer trust (Kim et al., 2018) and social image (McLean and Osei-Frimpong, 2019) by engaging consumers to (Pagni et al., 2019; Kachouie et al., 2014). Embedding emotion in voice (Schuller, 2018) further enhances social desirability (Schuetzler et al., 2018) and consumers’ believability (Demeur et al., 2011). Consumers can feel the
intensity of emotion of an agent in conversation with the agent, such as disappointment and enthusiasm in Alexa (Haselton, 2019). According to Amazon, the new version of Alexa improved overall customer satisfaction by 30% when it responded with emotions (Schwartz, 2019). Along with embedding verbal cues of voice (e.g., speech) in voice assistants, nonverbal cues including visual, gesture, and locomotion drive consumers’ trust and social behavior with voice assistants (Kim et al., 2018).

There is scant research from a marketing perspective investigating the underlying mechanism describing how auditory sense drives consumers’ trust in voice assistant and brand response (e.g., Klaus and Zaichkowski, 2020; Kim et al., 2018). This study intends to answer two questions: 1) How do voice assistants enhance consumers’ trust in voice assistant and brand affect? And 2) How does surprise leverage the above-mentioned effect on brand affect?

Unlike the early understanding of embedded cognition as mental processes taking place solely in the consumer’s mind, independent of external factors, and embodied cognition as mental and bodily processes that occur in the exchange of information between a consumer and an agent (e.g., Gibson, 1979; Wilson and Golonka, 2013), the enactive view of cognition posits that cognitive process does not necessarily require an interaction of two biological or living organisms (Di Paolo and Thompson, 2014; Metta et al., 2010). Cognition can be enacted as an exchange of information between a consumer and either a biological or [semi] autonomous non-biological agent (e.g., Kyselo, 2014); for example, the icub autonomous social robot (Metta et al., 2010).

Using an enactive perspective, this study applies social cognition theory (De Jaegher, Di Paolo, and Gallagher, 2010; De Jaegher and Di Paolo, 2008; Clark, 1998; Varela et al., 2016; Wilson, 2002) to examine the extent to which the perceived auditory sense of voice assistants drive consumers’ trust and brand affect and how surprise plays a role. Surprise plays a critical role in consumers’ information processing, and consumers try to avoid or minimize it (Parr and Friston, 2019). Consumers do not expect to be surprised when they interact with a voice assistant.

The remainder of this paper is organized as follows. First, a brief literature review of the main concepts drawn on for the study will be presented. Next, the hypotheses, methodology, and results will be explained. Finally, a discussion and conclusion will be presented, followed by an exploration of the finding’s implications, the study’s limitations, and suggestions for future research.

### 1.1. Theoretical framework

The embodied view of cognition refers to the fact that consumers’ decisions are not based solely on what is in their minds but also their bodies and environment (Chiel and Beer, 1997; Clark, 1998; Steels and Brooks, 1995). The enactive view of social cognition posits that a [semi] autonomous agent can produce and maintain its identity and actively participate in social interaction rather than being passive to be a subordinate. Autonomy denotes self-maintaining identity (De Jaegher, Di Paolo, and Gallagher, 2010). Table 1 demonstrates the relevant theories employed in past literature.

This study postulates that voice assistants, as situated agents, are also [semi] autonomous to some extent to regulate their identity (De Jaegher and Di Paolo, 2008). Yet, even as voice assistants become dead if they run out of batteries (Hoffmann and Pfeifer, 2018) or disconnect from the Internet or Cloud Computing, they manifest social and emotional traits; therefore, aligning with the proponents of the enactive view of cognition, this study argues that they can act as a [semi] autonomous agents to maintain their identity.

Social cognition theory suggests how people think and perceive other living or non-living autonomous agents’ behaviors (Abramova and Slors, 2018). Social cognition can be shaped through social interactions between at least two biological or non-biological autonomous agents (Di Paolo, 2020; De Jaegher, Di Paolo, and Gallagher, 2010). Social interaction facilitates social cognition and requires regulated coupling and non-accidental coordination between at least two agents. In other words, social interaction underlies the concept of “coordination” (De Jaegher and Di Paolo, 2008).

Coupling refers to the influence of one variable in a system on another. In other words, the state of one agent influences the state of another agent. For example, a person (an agent) who walks with his/her dog (another agent) held by a leash (coupled) (De Jaegher, Di Paolo, and Gallagher, 2010). Not all coupling meets the criteria of social interaction. The coupling of a voice assistant with other artificial intelligence agents can improve consumer experience (Koeher, 2016). Coupled agents should be coordinated to be considered as having a social interaction. For example, running to a friend on a sidewalk in downtown L.A. is not a social interaction; however, it might lead to initiate a conversation. In this study, coupling refers to the coexistence of a voice assistant and a consumer, and coordination refers to the auditory sense (recognizing the consumer’s voice and responding to the consumer’s inquiry) between the voice assistant and the consumer.

To initiate auditory social interaction, consumers are coupled (via Blue tooth, WiFi) and coordinated through auditory sense with a voice assistant. The level of auditory sense (coordination) influences the level of depth and length of the social interaction between consumers and voice assistants (De Jaegher and Di Paolo, 2008). This study hypothesizes that perceived auditory sense leads to perceived auditory social interaction. Perceived auditory sense refers to the ability of voice assistant to recognize (sense) the consumer’s voice and acts upon it. That is, a consumer initiates a voice interaction with a voice assistant that senses the environment and ready to act upon the consumer’s inquiry. The auditory abilities embedded in voice assistants facilitate social interactions (Deng et al., 2019) and enhances social desirability (Schuetzler et al., 2018) and consumers’ perception (Araujo, 2018). Table 2 demonstrates the previous studies related to voice assistants.

Voice assistants rely on cloud-based computing, and they can learn from consumers through auditory interaction (Davenport et al., 2020)
and update their auditory response. As a result, auditory social interaction will be smoother, and consumers will feel in control during the interaction. For example, when consumers say, “Hey Siri, what is the best running shoe brand?” and she listens and responds to their inquiries instantaneously, they feel in control during the voice interaction and are more likely to trust Siri and be happy with the brand. Fig. 1 depicts the theoretical framework.

### 1.2. Hypotheses development

The pivotal role of auditory sense has been acknowledged in previous studies (e.g., Javornik et al., 2020). Voice, as a sensory input, conveys crucial information such as the credibility, reliability, and other personality factors of a speaker (Nass et al., 1997; Till and Busler, 1998).

Voice assistants can be socially desirable (McLean and Osei-Frimpong, 2019; Heerink et al., 2010). Similar to the human voice, non-human voice conveyed by voice assistants can convey reliable information, positive emotion that enhances consumer engagement (Davenport et al., 2020; Pagani et al., 2019). The auditory ability of voice assistants motivates consumers to interact with such interfaces (McLean and Osei-Frimpong, 2019). It thereby smooths social interaction (Pagani et al., 2019; Kachouie et al., 2014) and prompts consumers to hold deep conversations that are emotionally pleasing with the voice assistants (Klaus and Zaichkowsky, 2020; Wilson, 2002).

This study proposes that the perceived auditory sense of the voice assistant enhances perceived auditory social interaction with consumers. The concept of perceived auditory sense refers to the coordination of the consumer’s inquiry with the auditory ability of the voice assistant to recognize the consumer’s voice and act upon it. Perceived auditory social interaction, meanwhile, refers to the exchange of auditory information (Liu and Shrum, 2002) between the consumer and the voice assistant.

Consumers feel a social presence when they verbally interact with such a device (Chattaraman et al., 2019) as a voice assistant (McLean and Osei-Frimpong, 2019) when coupled and coordinated with the consumers’ inquiry.

Auditory ability influences consumers’ perception (Araujo, 2018) and facilitates social interaction between a consumer and a voice assistant (Deng et al., 2019). However, a lack of auditory sense, such as an inability to understand the language or patterns of consumers’ speech, makes a voice assistant unable to interact with the consumer. As the voice assistant’s perceived auditory sense improves, so will consumers’ perception of their ability to respond to it and effectively engage in enhanced social interaction. Therefore, this study hypothesizes the following:

**H1.** Perceived auditory sense with voice assistant leads to perceived auditory social interaction.

![Fig. 1. Conceptual framework.](image-url)
Previous research studies have studied the effect of perceived control on consumers’ responses. Perceived control is one of the benefits of interaction with such interactive media (Kim, 2011) as voice assistants (Klaus and Zaichkowsky, 2020). Consumers feel in control when a voice assistant listens and responds to consumer’s inquiries in real time (Liu, 2003). It allows consumers to interact and manipulate their environment and sense embodiment. It is a critical variable in increasing consumers’ confidence (Hoffman and Novak, 1996), and it encourages consumers to interact with a voice assistant if it does not threaten the consumers’ safety and security. Otherwise, a voice assistant can inhibit consumers from interacting with it (Evans and Brown, 1988) because it decreases consumers’ confidence (Moller et al., 2006).

In this study, perceived control refers to perceived (embodied) auditory control. Perceived auditory control refers to the extent to which consumers feel in control during auditory interactions with a voice assistant. Social interaction facilitates consumers’ learning, imagination, and attention (Velestinos and Doering, 2010) and builds social relationships through verbal interactions with an artificial agent (Chattaraman et al., 2019). Perceived social interaction enhances perceived interactivity (Chattaraman et al., 2019). As fully autonomous agents, consumers feel in control in social interactions with such a [semi] autonomous agent such as voice assistant. Therefore, this study hypothesizes the following:

H2. Perceived auditory social interaction leads to consumers’ perceived auditory control.

Trust has shifted from the product brands to the brands conveying auditory experience (Klaus and Zaichkowsky, 2020). Consumers’ trust in voice assistants refers to the degree to which consumers feel that their voice assistant is trustworthy and sincere and that they can rely on the assistant to perform tasks. Consumers are more likely to trust voice assistants delivering charming, pleasing, and trustworthy voice (Klaus and Zaichkowsky, 2020). Previous literature has also shown the effects of embodiment—such as gestures, posture, eye movements, and auditory ability—on trust (Cassell et al., 2001) and believability (De Rosis et al., 2003).

Perceived social presence and social interaction enhance trust in a voice assistant (Chattaraman et al., 2019), satisfaction (Verhagen et al., 2014), and the likelihood of an online purchase (Lu et al., 2016) through perceived auditory control. Perceived trust motivates consumers to take risks, although they expect positive outcomes of trust (Rousseau et al., 1998). When consumers feel in control during auditory social interaction with a voice assistant, they can rely on the information it provides; thus they trust their voice assistant. Therefore, this study postulates the following:

H3. Perceived auditory control enhances consumer’s trust in a voice assistant.

Brand affect refers to a positive emotional response toward a brand (Chaudhuri and Holbrook, 2001). Perceived auditory sense can be a brand clue brand clues similar to other sensory brand clues, such as logos, sounds, and smells (Bettingen and Lueddicke, 2009). Previous literature has shed light on the effect of perceived control over emotional effect on a pleasant customer service experience and consumers’ responses (Jin and Bateos, 1991), and social role brands play to facilitate social interaction (Swaminathan et al., 2020). Consumers who can perceive control during voice interaction, they feel happiness (Owens and Dierer, 1995). This study postulates that perceived auditory control enhances brand affect. Yet, this study expects an inverse result when surprise intervenes.

For instance, let us assume that you have an Alexa at home, and you wake her up by saying, “Hey Alexa.” Then, you continue interacting with her and extend your inquiries to play a song, find the weather forecast, and to place a toothpaste and a chocolate bar to your shopping cart. In this process, she played the right song and responded to your inquiry about the weather forecast; however, you get surprised, then you noted that something went wrong and she placed an incorrect item in your shopping cart. Will you still have a positive emotional response with Amazon’s Alexa?

Emotion is an essential source of information that could affect consumers’ evaluation of products and brands (Hasford, Hardesty, and Kidwell, 2015). Surprise, as an emotion, is an essential source of information that could affect consumers’ evaluation of products and brands (Hasford, Hardesty, and Kidwell, 2015). Surprise is elicited when there is a “schema discrepancy” (e.g., Meyer et al., 1997). A schema is a type of private and personal information about the nature of products, events, or situations (Rumelhart, 1984). Consumers continuously check whether their schemas match the responses of their voice assistants. When consumers sense a discrepancy between what they expect to hear and what they actually hear, they will be surprised. For example, surprise might occur when a consumer requests that Alexa play a song by Lady Gaga and Alexa plays a song from a different singer. Surprise can also be elicited when a voice assistant’s response is not what consumers expect, which inversely transforms the consumer’s brand affect experience. When an unexpected response elicits surprise from the voice assistant, the effect of perceived auditory control on brand affect is diluted. Therefore, this study expects that perceived auditory control will be inversely related to brand affect when surprise is evoked during an interaction with a voice assistant. Therefore, this study hypothesizes that:

H4. Surprise attenuates the effect of perceived control on brand affect.

2. Empirical studies

Two experiments were designed to test the hypotheses. Study 1 examines the effect of perceived auditory social interaction on consumers’ trust in voice assistant through perceived auditory control. Study 2 examines the effect of perceived auditory sense on perceived auditory social interaction and perceived auditory control and thereby brand affect and consumers’ trust in voice assistant. In addition, it highlights the interaction effect of surprise and perceived auditory control on brand affect.

2.1. Study 1: design, stimuli selection and procedure

Two commonly used voice assistant mobile applications were utilized for this study, and to answer the research questions, a lab experiment was employed (Kerling and Lee, 2000) in which participants were randomly assigned to one of two voice assistance mobile applications, Google Assistant mobile application (n1 = 59), and Microsoft Cortana (n2 = 68).

The sample consisted of 127 young consumers (70 males and 56 females) between the ages of 21 and 41. This study was conducted in a laboratory environment located at a large, Southwestern university in the U.S. in 2018. All voice assistant mobile applications were installed on the researcher’s smart phone, which participants used to interact with one of the two applications.

Convenience sampling was used to recruit the participants, and no incentive for participation was offered. Participants were informed that they would interact with a voice assistant. After interacting with the mobile application for about 7 min, they closed the application and completed a questionnaire. Participants were instructed to ask the mobile application questions related to their daily activities such as playing music, finding weather forecasts, searching for information, and like questions. The survey’s questions focused on perceived auditory social interaction, perceived auditory control and consumers’ trust in voice assistant.

Before being exposed to the treatments, participants answered prescreening questions to ensure that they were familiar with internet use and technology (Olsson et al., 2012). The prescreening questions were: “I am familiar with using mobile applications,” “I frequently use mobile
applications to shop,” “I usually use voice assistants for my daily work”.

2.2. Measures

To measure consumers’ trust in voice assistants, three items from Lee (2005), Jarvenpaa et al. (1999) and Doney and Cannon (1997) were used. To operationalize perceived auditory social interaction, two items from Liu (2003) were used. To measure perceived auditory control, three items from Nel et al. (1999) were borrowed. All items were measured using a 7-point Likert scale with the anchors being “strongly disagree” and “strongly agree.”

2.3. Results: prescreening, manipulation check, reliability analysis, confirmatory factor analysis

SPSS 24.0 software was used to obtain descriptive statistics and run reliability analyses. ANOVA was applied to ensure participants were not significantly different in terms of technology use and familiarity ($F = 0.144, p > .1$). Next, ANOVA was used to examine whether the manipulation was successful. Results indicate that perceived auditory social interaction ($F = 23.74; p < .001$), perceived auditory control ($F = 39.791; p < .001$) and consumer’s trust ($F = 8.896, p < .003$) were significantly different between the two groups.

To examine reliability, Cronbach Alphas were checked. Cronbach Alphas ranged from 0.59 to 0.84 thus demonstrating construct internal consistencies (Nunnally and Bernstein, 1994). Next, exploratory factor analysis (EFA) was conducted with the eight items, using the maximum likelihood method (MLE) and Varimax rotation to explore the underlying constructs (KMO = 0.863, $X^2 = 472.530, df = 28, p < .001$). Three constructs emerged with acceptable default eigenvalues of 1. Additionally, based on Harman’s single factor test, this study also examined common method bias variance to ensure a single construct explained 48.43% of total variance, which is less than 50%.

Smart PLS was used to examine confirmatory factor analysis, check convergent and discriminant validity, obtain the average variances extracted (AVE), and composite reliabilities (CR). One item from perceived auditory control was removed due to its low level of factor loading. CR ranged from 0.77 to 0.90. The AVE ranged from 0.57 to 0.76, thus meeting the recommended threshold value of 0.5 for convergent validity (McDonald and Ho, 2002). AVEs above 0.5, and the square roots of AVEs above inter-factor correlations show discriminant validity (Fornell and Larcker, 1981). All constructs satisfied the requirements for convergent validity and discriminant validity. Table 3 shows Cronbach Alphas ($\alpha$), average variance extracted (AVE), construct reliabilities (CR), and confirmatory factor loadings (CFA). Table 4 demonstrates the correlations and discriminant validity.

The structural model was tested using Multi-Group Analysis with Smart PLS 3.0 because Smart PLS works well with small samples. Moreover, it also works well for exploratory or theory development research (Hair et al., 2012). The results of Smart PLS 3 demonstrated that perceived voice interaction significantly influences consumers’ perceived control in two groups ($\beta_{\text{Cortana}} = .531, t_{\text{Cortana}} = 6.538, p_{\text{Cortana}} < .001; \beta_{\text{Google}} = .673, t_{\text{Google}} = 7.096, p_{\text{Google}} < .001$). Perceived voice interaction significantly enhances consumers’ trust in voice assistant in two groups ($\beta_{\text{Cortana}} = .496, t_{\text{Cortana}} = 4.664, p_{\text{Cortana}} = .001; \beta_{\text{Google}} = .798, t_{\text{Google}} = 18.848, p_{\text{Google}} < .001$). Therefore, H2 and H3 were supported. Table 5 displays the results of hypotheses testing.

3. Study 2

This study was conducted in a laboratory located in a western U.S. city. Convenience sampling was used to recruit participants. The sample consisted of 116 adult consumers (65 males and 60 females). The age of the participants ranged from 20 to 65 years. All questionnaires were used in the analyses. Missing data was a minor issue and did not exert any undue effect.

Participants were randomly assigned to one of two mobile voice assistants: Amazon’s Alexa and Microsoft Cortana. Participants were familiar with the voice assistant. Of the sample, 75% had already interacted with voice interface mobile applications through their smartphones.

Participants used their own smart phones and were informed they would speak with the mobile application using only voice interaction. They were guided to talk about their daily activities for 15 min. Participants could initiate either a shallow or deep conversation with their voice interface. After interacting with the mobile application, they closed it and then completed a questionnaire.

3.1. Measures

One item was developed to capture perceived auditory sense (“This voice assistant recognizes my voice when I call her name”). Similar to Study 1, to measure consumers’ trust in voice assistants, three items from Lee (2005), Jarvenpaa et al. (1999) and Doney and Cannon (1997) were used. To measure perceived auditory social interaction, two items from Liu (2003) were used. To measure perceived auditory control, three items from Nel et al. (1999), and for surprise, three items from Richins (1997) were borrowed. To measure brand affect, three items from Chaudhuri and Holbrook (2001) were used. All items were measured using a 7-point Likert scale with the anchors being “strongly disagree” and “strongly agree.”

Table 3

| Correlations and discriminant validity for studies 1 & 2. |
|----------------------------------------------------------|
| Control | Trust | Brand affect | Surprise | Surprise*control |
| Auditory sense | Auditory social interaction | |
| Auditory sense | 1 | |
| Auditory social interaction | .46** | .865* |
| Auditory control | .49** | .681* | .752* |
| Trust | .38** | .696* | .649* | .870* |
| Brand affect | .42** | .53** | .680** | .47** | .89** |
| Surprise | .27** | .57** | .54** | .48** | .55** | .90** |
| Surprise*control | .011** | .06** | .06** | .01** | .03** | .11** | 1 |

Note:
Study 1: **
Study 2: *
Table 4

Constructs: Average variance extracted (AVE), construct reliability (CR), CFA for studies 1&2.

| Constructs                                      | Definition of construct                                           | CFA Study 1 | CFA Study 2 |
|-------------------------------------------------|-------------------------------------------------------------------|-------------|-------------|
| Perceived auditory sense (Study 2)              | Refers to the ability of voice assistant to recognize (sense) the consumer’s voice and acts upon it. |             |             |
| Perceived auditory social interaction (Study 1: α = .83, AVE = .75, CR = .90) | Refers to exchange of auditory information between the consumer and the voice assistant. | .856        | .900        |
| Perceived auditory social interaction (Study 2: α = .88, AVE = .73, CR = .85) | This voice assistant makes me feel like it wants to listen to me. | .869        | .810        |
| This voice assistant encourages me to offer feedback. | This voice assistant recognizes my voice when I call her name. |             |             |
| Consumers’ trust in voice assistant (Study 1: α = .84, AVE = .76, CR = .90, R² = .42) | Refers to the degree to which consumers feel their voice assistant is trustworthy and sincere, and they feel they can rely on the assistant to perform tasks. | .801        | .731        |
| Consumers’ trust in voice assistant (Study 2: α = .78, AVE = .70, CR = .87, R² = .26) | I trust the voice assistant because it keeps my best interests in mind. | .912        | .910        |
| This voice assistant is trustworthy.             | I think that the information offered by this voice assistant was sincere and honest. | .901        | .840        |
| Consumers’ perceived auditory control (Study 1: α = .58, AVE = .57, CR = .77, R² = .46) | Refers to the extent to which consumers feel their voice control during auditory interaction with a voice assistant. |             |             |
| Consumers’ perceived auditory control (Study 2: α = .70, AVE = .63, CR = .84, R² = .27) | When I used this voice assistant, I felt in control. | .912        | .641        |
| This voice assistant allowed me to control the voice interaction. | This voice assistant allowed me to control the voice assistant. | .890        | .871        |
| I felt I had (no) control while interacting with the voice assistant. | I felt I had (no) control while interacting with the voice assistant. | RC          | .861        |
| (RC) Deleted.                                   |                                                                  |             |             |
| Surprise—Study 2 (α = .88, AVE = .80, CR = .92) | Indicates an emotion which is evoked during an interaction with a voice assistant. |             |             |
| Surprised                                       |                                                                  | –           | .841        |
| Amazed                                          |                                                                  | –           | .942        |
| Astonished                                      |                                                                  | –           | .910        |
| Brand affect—Study 2: (α = .87, AVE = .80, CR = .92, R² = .51) | Refers to a positive emotional response toward a brand. | –           | .890        |
| I feel good when I use this brand.               |                                                                  |             |             |
| This brand makes me happy.                      |                                                                  |             | .841        |
| This brand gives me pleasure.                   |                                                                  |             | .942        |

Note: – Perceived auditory sense, surprise and brand affect were measured in Study 2.
RC: Reverse coded. This item was reverse coded for Study 1 and was removed.

Table 5

Results of hypotheses testing.

| Hypotheses                                                      | Studies | Results |
|-----------------------------------------------------------------|---------|---------|
| Perceived auditory sense—Perceived auditory social interaction  | Study 2 | Supported |
| Perceived auditory social interaction → perceived auditory control | Study 1 and Study 2 | Supported |
| Perceived auditory control—Consumers’ trust in voice assistant | Study 1 and Study 2 | Supported |
| Perceived auditory control—Brand affect                        | Study 2 | Supported |
| Surprise* perceived auditory control—Brand affect               | Study 2 | Supported |

3.2. Results: manipulation check, reliability analysis, confirmatory factor analysis

ANOVA was used to examine whether the manipulation was successful. ANOVA was also applied to ensure that participants were not significantly different in terms of voice assistant use, technology use, and familiarity (F = 0.010, p > .1). Perceived auditory sense (F = 4.403, p < .03) and perceived auditory social interaction (F = 6.852, p < .01), perceived auditory control (F = 6.161, p < .02), brand affect (F = 6.571, p < .02), and surprise (F = 9.891, p < .003) were all significantly different across the two groups except consumers’ trust (F = 1.215, p > .1).

Structural test. To examine reliability, Cronbach Alphas were checked, Cronbach Alphas ranged from 0.68 to 0.88, thus demonstrating construct internal consistencies (Nunnally and Bernstein, 1994). Next, exploratory factor analysis (EFA) was conducted with the fifteen items using the maximum likelihood method (MLE) and Varimax rotation to explore the underlying constructs (KMO = 0.833, χ² = 1023.58, df = 105, p < .001). Five constructs emerged with acceptable default eigenvalues of 1. Additionally, based on Harman’s single factor test, this study also examined common method bias variance to ensure a single construct explained 39.91% of total variance, which is less than 50%.

Smart PLS was used to examine confirmatory factor analysis, check convergent and discriminant validity, obtain the average variances extracted (AVE), and composite reliabilities (CR). Composite reliabilities (CR) ranged from 0.84 to 0.92. The AVE ranged from 0.63 to 0.80, thus meeting the recommended threshold value of 0.5 for convergent validity (McDonald and Ho, 2002). Average variances extracted (AVE) above 0.5, and the square roots of AVEs above inter-factor correlations show discriminant validity (Fornell and Larcker, 1981). Table 3 demonstrates the correlations and discriminant validity. All constructs satisfied the requirements for convergent validity and discriminant validity. Table 4 shows constructs, items, and confirmatory factor loadings (CFA).

The structural model was tested using Multi-Group Analysis with Smart PLS 3.0. The results demonstrated that perceived auditory sense enhances perceived auditory social interaction across two groups (βAlexa = .58, tAlexa = 6.27, pAlexa < .001; β Cortana = .30, tCortana = 1.98, pCortana < .05). In addition, the effect of perceived auditory social interaction on perceived auditory control was also significant across two groups (β Alexa = .58, tAlexa = 6.67, pAlexa < .001; β Cortana = .47, t Cortana = 3.48, pCortana < .001). H1 and H2 were supported. In relation to H3, the effect of perceived auditory control on consumers’ trust in voice assistant was also significant across both groups (β Alexa = .49, t Alexα = 5.33, p Alexa < .001; β Cortana = .55, t Cortana = 4.54, p Cortana < .001) and thereby H3 was supported. In relation to H4, results indicated that surprise attenuated the effect of perceived auditory control on brand affect across two groups (β Alexa = -0.16, tAlexa = 2.35, pAlexa < .02; β Cortana = -.19, t Cortana = 2.06, p Cortana < .04).
will be with voice assistants such as Amazon’s Alexa, Echo, or Google Assistant (Alsinn, 2018).

The results of this study indicate that perceived auditory sense initiates perceived auditory social interactions that shape brand affect and consumers’ trust in their assistants through perceived auditory control. The results also highlight the influence of surprise on brand affect. Surprise is inversely related to brand affect.

Voice assistants mainly rely on voice commands to verbally interact with consumers. Voice assistants utilize natural language processing (NLP) and machine learning to 1) analyze and interpret consumers’ language, accent, patterns of speech, and 2) respond to requests in real time (Pickell, 2019). Embedded auditory ability can motivate consumers to verbally interact with the voice assistant and feel a sense of control over the auditory interaction. As consumers feel in control during their interactions with a voice assistant, their trust in and happiness with the brand are enhanced. Consumers’ trust is a significant construct for building a relationship with customers-technology (Cassell et al., 2001) and shaping their behavior (Schurr and Ozanne, 1985) toward online shopping (Lu et al., 2016).

Voice assistants such as Amazon’s Alexa use Neural TTS text-to-speech technology that responds to users with emotions. Perceived verbal and nonverbal auditory embodiment enhance consumers’ trust in voice assistant (Cassell et al., 2001). For example, when a user asks a question properly, Alexa can respond happily, or when the consumer asks a question about which team won or lost a game, Alexa can sound disappointed when the consumer’s favorite team has lost (Gao, 2019). Emotion is contagious and is transferred to consumers during interactions with the voice assistant.

The results of this study also highlight the importance of surprise. A discrepancy between what consumers predict and what they are exposed to is what evokes surprise. Surprise is used for emotion-based relationship marketing because of its major benefits, which are customer satisfaction, customer retention, and consumer trust (e.g., Vanhamme and Lindgreen, 2001). When surprise is elicited, consumers are less likely to be able to predict the response of the voice assistant, and the effect of perceived auditory control on brand affect is attenuated.

5. Theoretical implications

Today, consumers interact with real time technologies such as voice assistant mobile applications or smart speakers. Drawing on an enactive view of social cognition theory, this study deepens our understanding of how cognition is enacted during auditory interaction with voice assistants. This study highlights the role of semi-autonomous of voice assistants to maintain consumer-initiated interactions. Proponents of radical embodiment have criticized the enactive view of cognition (Hoffmann and Pfeifer, 2018) to scrutinize the effect of voice assistants on consumers’ cognition. To them, voice assistants are not fully autonomous due to their lack of organic body and metabolism to fully become self-maintained or self-regulated.

This study demonstrated that situated voice assistants that are kept in consumers’ houses or installed on smartphones are not fully autonomous to self-regulate (e.g., Di Paolo, 2005; 2009) however yet, such voice assistants are semi-autonomous to act upon the environment. When voice assistance is tightly coupled or coordinated with consumers’ requests, it can sense and perceive those requests. To that end, consumers are more likely to maintain an auditory social interaction when voice assistant can detect and recognize their voices and act upon their inquiries.

Perceived auditory control enables consumers to interact with a voice assistant that leads to brand affect and trust in voice assistant. However, when surprise is elicited as the result of an unexpected response (Friston, 2009, 2010) by voice assistants, the effect of perceived auditory control on brand affect is diluted.

6. Managerial implications

Gigantic companies, such as Walmart and Target are using voice assistants such as Google Home to remind their customers when to reorder most frequently used products for instance milk or coffee (Sun et al., 2019). Yet, voice interface would be a helpful replacement of touch interface option for consumers who are more favorably to avoid contamination (Argo et al., 2006). For example, consumers who seriously avoid touching self-service technology interface due to being infected or spreading contagious viruses including SARS, or COVID-19.

There is an increasing desire among consumers to adopt a voice assistant in their daily lives. Voice assistants allow people to multitask, and therefore reduce cognitive load (Strayer et al., 2017). People use this technology voice to shop online or search for information such as news, music, weather, and so on. It is estimated that 53 million Americans own at least one smart speaker, 74% of smart speaker owners play music, 66% ask weather-related questions, more than 50% of people ask funny questions (Handley, 2019), 27% conduct searches (Juniper Research, 2018), and 24% used their smart speakers to shop online (Handley, 2019).

In addition to functional communication quality (Guzman, 2019), technology developers are focusing on humanizing voice assistants to display human traits, emotions, and social attributes. For example, socially interactive robots such as Replika (emotional companion robot), Pepper (customer service robot), Stich Fix (a fashionable mobile social application) display social and emotional characteristics to interact with consumers socially.

Socially interactive artificial agents can be programmed to adjust or modify their tone of voice to express excitement, disappointment, or other emotions when responding to users. Amazon’s Alexa, for example, can respond to consumers with happiness, excitement, disappointment, or empathy (Schlosser, 2019). Alexa uses a suitable speech style when she speaks and delivers news and music (Schlosser, 2019). Consumers can feel the intensity of embodied emotions, such as disappointment and enthusiasm.

Companies should pay attention to the design of voice assistants to minimize surprise when consumers are willing to purchase a product. When companies try to build a relationship with their customers, they should avoid surprise in product development, packaging, and delivery. Customers would not like to encounter a surprise when they are serious about taking a purchase action. Companies may adopt surprise when they want to draw customers’ attention (Lindgreen and Vanhamme, 2003). For example, Siri provides surprising but funny responses to users. In this case, a surprise builds memorable and pleasant moments for users.

Further, the use of machine learning also helps voice assistants learn from consumers’ responses and predict what users expect in future interactions; therefore, social interaction improves. To assist customers in searching for product information, companies need to develop voice assistants that can recognize the customers’ patterns of speech, language, and accent so that they can respond to requests promptly. Designing faultless voice assistant engages consumers in social interaction, that leads to perceived control and trust in voice assistant and brand affect. For example, consumers tell Alexa, “I do not like this music, play the next one,” Alexa responds: “Okay, I will not play this music.” Alexa learns quickly from consumer desires.

7. Limitation and suggestions for future research

This study has some limitations. First, only voice assistant mobile applications were used. Second, the sample size was limited. Larger sample size in different contexts is required to generalize the results. In addition, in-home smart speakers such as Amazon’s Echo or Apple’s Home are necessary to understand how perceived auditory sense drives consumer responses. For future research, the effect of physical, social, and emotional embodiment on consumers’ responses...
Nunnally, J.C., Bernstein, L.H., 1994. Psychometric theory McGraw-hill series. In: Psychology, vol. 3. McGraw-Hill, New York.
Pagani, M., Recat, M., Hofacker, C.F., 2019. Adding voice to the omnichannel and how that affects brand trust. J. Interact. Market. 48, 89–105.
Parr, T., Friston, K.J., 2019. Generalized free energy and active inference. Biol. Cybern. 1135.6, 495–513.
Pickel, D., 2019. https://learn.g2.com/machine-learning,
Rauschnabel, P.A., He, J., Ro, Y.K., 2018. Antecedents to the adoption of augmented reality smart glasses: a closer look at privacy risks. J. Bus. Res. 92, 374–384.
Richins, M.L., 1997. Measuring emotions in the consumption experience. J. Con. Res. 242, 127–146.
Rousseau, D.M., Sitkin, S.B., Burt, R.S., Camerer, C., 1998. Not so different after all: a cross-discipline view of trust. Acad. Manag. Rev. 233, 393–404.
Rumelhart, D.E., 1984. Schemata and the Cognitive System.
Schlosser, Kurt, 2019. https://www.geekwire.com/2019/alexa-emotional-amazon-created-realistic-new-voice-tones-assistant/.
Schuetzler, R.M., Giboney, J.S., Grimes, G.M., Nunamaker Jr., J.F., 2018. The influence of conversational agent embodiment and conversational relevance on socially desirable responding. Decis. Support Syst. 114, 93–102.
Schuller, D., Schuller, B.W., 2018. The age of artificial emotional intelligence. Computer 51 (9), 38–46.
Schurr, P.H., Ozanne, J.L., 1985. Influences on exchange processes: buyers’ preconceptions of a seller’s trustworthiness and bargaining toughness. J. Con. Res. 114, 939–953.
Schwartz, E.H., 2019. https://voicebot.ai/2019/11/27/alexa-is-learning-to-speak-emotionally/.
Spaulding, S., 2014. In: Shapiro, Lawrence (Ed.), Embodied cognition and theory of mind in hand book of Embodied Cognition. Routledge Press,
Strayer, D.L., Cooper, J.M., Turzill, J., Coleman, J.R., Hopman, R.J., 2017. The smartphone and the driver’s cognitive workload: a comparison of Apple, Google, and Microsoft’s intelligent personal assistants. Canadian J. Exp. Psych./Revue canadienne de psychologie expérimentale 712, 93.
Steele, L., Brooks, R., 1995. The“ Artificial Life” Route to“ Artificial Intelligence”: Building Situated Embodied Agents.
Swaminathan, V., Sorescu, A., Steenkamp, J.B.E., O’Guinn, T.C.G., Schmitt, B., 2020. Branding in a hyperconnected world: refocusing theories and rethinking boundaries. J. Mark. 842, 24–46.
Sun, C., Shi, Z.J., Liu, X., Ghose, A., Li, X., Xiong, P., 2019. The Effect of Voice AI on Consumer Purchase and Search Behavior.
Till, B.D., Busler, M., 1998. Matching products with endorsers: Attractiveness versus expertise. J. Con. Mark. 15 (5), 576–586.
Vanhamme, J., Snelders, D., 2003. What if you surprise your customers ... Will they be more satisfied? Findings from a pilot experiment. In: Keller, P.A., Rook, D.W. (Eds.), NA - Advances in Consumer Research, 30. Association for Consumer Research, Valdosta, GA, pp. 48–55.
Varela, F.J., Thompson, E., Rosch, E., 2016. The Embodied Mind: Cognitive Science and Human Experience. MIT press.
Veletzianos, G., Doering, A., 2010. Long-term student experiences in a hybrid, open-ended and problem based Adventure Learning program. Australas. J. Educ. Technol. 262.
Verhagen, T., Van Nes, J., Feldberg, F., Van Dolen, W., 2014. Virtual customer service agents: using social presence and personalization to shape online service encounters. J. Computer-Mediated Commun. 193, 529–545.
Wilson, M., 2002. Six views of embodied cognition. Psychon. Bull. Rev. 94, 625–636.
Wilson, A.D., Golonka, S., 2013. Embodied cognition is not what you think it is. Front. Psychol. 4, 58.