The Bot on Speaking Terms: The Effects of Conversation Architecture on Perceptions of Conversational Agents

Christina Wei
christina.wei@mail.utoronto.ca
University of Toronto
Toronto, Canada

Young-Ho Kim
yghokim@younghokim.net
NAVER AI Lab
Republic of Korea

Anastasia Kuzminykh
anastasia.kuzminykh@utoronto.ca
University of Toronto
Toronto, Canada

ABSTRACT

Conversational agents mimic natural conversation to interact with users. Since the effectiveness of interactions strongly depends on users' perception of agents, it is crucial to design agents' behaviors to provide the intended user perceptions. Research on human-agent and human-human communication suggests that speech specifics are associated with perceptions of communicating parties, but there is a lack of systematic understanding of how speech specifics of agents affect users' perceptions. To address this gap, we present a framework outlining the relationships between elements of agents' conversation architecture (dialog strategy, content affectiveness, content style and speech format) and aspects of users' perception (interaction, ability, sociability and humanness). Synthesized based on literature reviewed from the domains of HCI, NLP and linguistics (n=57), this framework demonstrates both the identified relationships and the areas lacking empirical evidence. We discuss the implications of the framework for conversation design and highlight the inconsistencies with terminology and measurements.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **HCI design and evaluation methods**; **Natural language interfaces**.

KEYWORDS

conversational agents, natural language interface, chatbots, virtual assistants, user perceptions, anthropomorphized perceptions, conversation architecture, speech variations

ACM Reference Format:
Christina Wei, Young-Ho Kim, and Anastasia Kuzminykh. 2023. The Bot on Speaking Terms: The Effects of Conversation Architecture on Perceptions of Conversational Agents. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3571884.3597139

1 INTRODUCTION

Conversational agents (CAs), such as Amazon Alexa or Google Assistant, are designed to interact with humans using natural language through voice or text modalities. Since these agents communicate in natural language, the effectiveness of interactions between users and conversational agents in achieving users' goals depend heavily on the *conversation architecture*, which refers to the content and manner of communication utilized by the agents [47, 77]. Previous research suggests that there is a relationship between CAs' conversation architecture and how users perceive these agents (e.g., [11, 23, 24]). While various studies investigated how some elements of conversation architecture affect aspects of users' perceptions or discussed the speech specifics of agents as part of a broader analysis of the state of research [11, 23], there is no comprehensive analysis of conversation architecture elements that are relevant to user perceptions. As such, this paper aims to provide a systematic understanding of the effect of conversation architecture elements on users' perceptions of conversational agents. Such an understanding is essential for guiding the design of conversation architecture, which would allow us to both predict and develop users' appropriate perceptions of agents.

To gain this systematic understanding, we explored the following research questions in this paper: **(RQ1) what aspects of users' perceptions of agents have been explored in relation to the effects of conversation architecture elements?**, **(RQ2) what elements of conversation architecture are relevant to users' perceptions of agents?**, and **(RQ3) what is known about the specifics of these relationships between conversation architecture elements and the perceptions of CAs?** In order to answer these questions, we performed a comprehensive review of the existing literature published from 2010 to 2022 and synthesized the effect of conversation architecture elements on the perceptions of CAs based on the final dataset of 57 relevant papers (Figure 3). Based on this review, we grouped the perceptions of agents that were shown to be affected by conversation architecture into four categories (interaction, ability, sociability and humanness) and eleven aspects (Table 1), and grouped the speech variations related to perceptions of agents into four categories (dialog strategy, content affectiveness, content style and speech format) and eleven elements (Table 2). We found that perceptions of interaction with CAs (usability, engagement, satisfaction) have been most explored in relation to speech variations, while the perceptions of CAs' ability (intelligence, competence, credibility) have been explored the least. Our results also show that the conversation architecture elements of response delay and agent-initiated dialog related to perceptions of CAs are under-explored in literature. Thus, our work contributes to the CUI community by presenting a framework that synthesizes the findings on the effect of conversation architecture on perceptions of conversational agents while also identifying and clustering perception aspects and conversation architecture elements.

CUI '23, July 19–21, 2023, Eindhoven, Netherlands

ACM ISBN 979-8-4007-0014-9/23/07…$15.00
https://doi.org/10.1145/3571884.3597139

In the remainder of the paper, we first outline related background work, then describe the literature review process and the framework synthesis method, followed by a detailed description of the taxonomies of the aspects of the perception of agents and conversation architecture elements, as well as the synthesized framework for their relationships. We conclude by discussing the implications of this work, including the associated research challenges and opportunities, as well as ethical considerations for designing the conversational architecture to affect users' perceptions of conversational agents.

2 RELATED WORK

Similar to human-human communication, users utilize natural languages when interacting with conversational agents. However, while it is tempting to leverage components from human-human conversation in designing interactions with conversational agents, studies have demonstrated that there are differences between human-human communications and human-agent communications [12, 34]. Currently, CAs are considered as a user-controlled tool rather than a social companion, with a focus on the utilitarian aspects of the conversation [12]. When conversing with agents, users tend to use shorter messages with limited vocabulary, adapting to the style of the agent [34]. Also, some studies found that small talk and humour are considered unnecessary or even inauthentic in conversations with CAs, while these elements are important in scaffolding communications with human partners [12, 17]. There are also machine-like traits that users preferred in a CA, such as the ability to interact through multimodal media, and a machine's perceived ability to be objective and non-judgemental [17, 44]. Given human-agent communications are different from human-human communications, it is important to look specifically at the design of user experiences for conversational agents.

There are a number of studies investigating factors affecting user experiences for conversational agents, which includes exploring the relationship between conversation architecture elements and agents' perceptions. For instance, the use of affective language is commonly explored, with studies finding that conversational agents are perceived as more socially present and emotionally intelligent if they use sentiment-adaptive responses based on user's utterances [16, 91]. Another commonly explored element is the use of prosody in voice-based agents, such as the use of express prosody contributing to higher perceived enjoyment and intimacy with the agent [45]. However, there is a paucity of research to synthesize these findings on user experiences with agents across literature.

In recent years, several papers have synthesized the current state of research for particular types of conversational agents (e.g. text-based, voice-based, polyadic) highlighting major trends, topics, methods and evaluating metrics [11, 73, 93]. Other papers looked at specific dimensions of user experiences or domains of usage, such as Van Pinxteren et al. [81] examining the effects of agents' human-like communicative behaviours on their relationships with users, and Kocaballi et al. [48] exploring the challenges and opportunities of CAs in healthcare. There are also publications aimed to unify concepts used in the research for CA user experiences, for instance Feine et al. [23] creating a taxonomy of social cues, and Finch and Choi [24] analyzing evaluation protocols for dialogue systems. To the best of our knowledge, there is currently no synthesis on the speech specifics of CAs, the perceptions of agents, or the relationship between them.

Conversation architecture has been demonstrated to play an important role in users' perceptions of agents [47, 68, 77]. Yet, little is known about the overall landscape on the how the specifics of speech variations affect these perceptions. Given the paucity of research in this area, this paper aims to synthesize the identified relationships between conversation architecture elements and users' perceptions of agents.

3 METHOD

3.1 Data Collection

To perform a literature analysis of past findings on the relationship between the specifics of conversation architecture and perceptions of CAs, we identified, reviewed, and selected relevant literature using PRISMA guidelines [65]. In accordance with accepted practices, we defined the scope of our literature review by considering two factors: (1) time—specifically publications from 2010 to 2022—and (2) source—using the ACM Digital Library [60, 79, 93]. Through our analysis, we didn't find any common terms used to refer to either conversation architecture elements or aspects of perceptions. As such, we generalized our search term to capture articles related to conversational agents. Based on search terms used in previously published literature reviews [11, 73], the following keywords are used to search for publications related to conversational user interfaces:

"conversational agent" OR "natural language interface" OR "IPA" OR "intelligent personal assistant" OR "chatbot" OR "speech interface" OR "voice assistant" OR "intelligent agent" OR "human-chatbot communication" OR "virtual agent" OR "dialog system" OR "voice user interface" OR "human computer dialog*"*

The following selection criteria are applied to identify literature related to the effect of conversational architecture on the perception of agents:

- The paper was published between 2010 and 2022.
- The paper is peer-reviewed and written in English.
- The paper contains the use of voice-based or text-based conversational agents.
- The paper contains studies on the effect of conversation architecture elements on the perceptions of agents.
- The paper contains effects of conversation architecture elements can be separated from other effects (e.g. embodiment).

The initial query retrieved 2901 unique publications. We screened the titles and abstracts of the papers based on the selection criteria above, resulting in 221 papers. The fully body of these papers were then reviewed, selecting 49 papers that met the criteria. A search for additional literature based on citation tracing was also performed to supplement the corpus, adding 8 more papers to the selection. In total, we identified 57 relevant articles for analysis (see Figure 1).

Literature Corpus Characteristics. The papers reviewed (n=57) were published between May 2011 and November 2022. Most of the papers were published in or after 2019, with a slight trend upwards

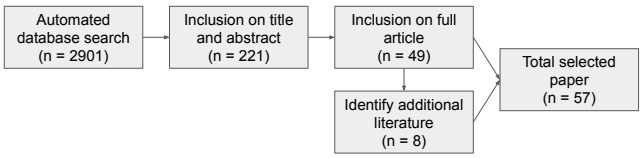


Figure 1: Literature search process

over the years (see Figure 2). The vast majority of the papers in our corpus were published in conference proceedings (n=50), while the remaining papers were published in journals. Out of the papers published in conferences, the top conferences were The ACM Conference on Human Factors in Computing Systems[1] (n=16), Conversational User Interfaces[2] (n=7), Human-Agent Interaction[3] (n=4), and Intelligent Virtual Agents[4] (n=4).

On the modality characteristics of the conversational agents in our corpus, there is a roughly even split between voice and text modalities, with slightly more papers exploring voice-based CAs (n=30) compared to text-based CAs (n=27).

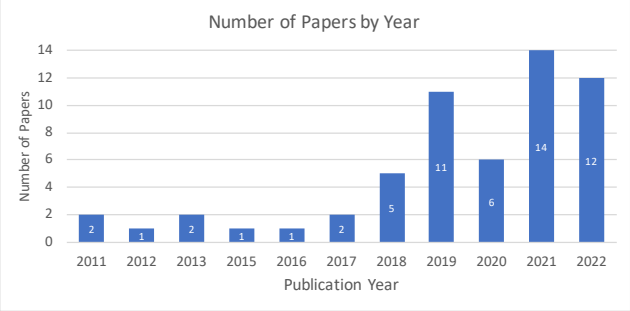


Figure 2: Number of papers by year

3.2 Data Analysis

We analyzed the data by conducting reflexive thematic analysis approach with open and axial coding procedures [4]. First, we performed open coding on each article to extract labels related to perceptions of agents and conversation architecture. These labels were then grouped into conceptual categories through axial coding. The first author was responsible for performing the data analysis, with the second and third authors cross-checking the codes. We resolved disagreements through regular discussions and multiple rounds of revisions.

Perceptions of Agents. To understand the perceptions of agents related to conversation architecture (RQ1), assessments for the perceptions of agents were extracted from each paper. For example, likeability of the agent contains Godspeed questionnaire [3]'s set of questions on likeability used by [55], Subjective Assessment of Speech System Interface (SASSI) questionnaire [37]'s set of questions on likeability used by [7, 10] and questionnaire items like "*this voice agent was likeable*" used by [14]. This step resulted in 83 unique codes. These codes are then organized based on similarity of their meanings, resulting in 11 aspects of perceptions. For instance, the aspect of personality traits of the agent contains codes such as likeability, funniness, and kindness. Lastly, the 11 aspects are grouped into 4 categories: perception of interaction with agent, perception of agent's ability, perception of sociability with agent, and perception of agent's humanness (Table 1).

Conversation Architecture Elements. As for the elements of conversation architecture that are relevant to the perception of agents (RQ2), we extracted the speech variation features used in each paper. Some of the papers used multiple features in their CA design, such as the anthropomorphic agent used in Seeger et al's study [77]. In these composite situations, we broke the CA design down into its individual features. For example, the design of the anthropomorphic agent [77] was divided into features of emotional expressions, is-typing indicator, emoticons, and response delay and captured as separate codes. 58 unique codes were created as part of this process, capturing features like sentiment-adaptive responses [16], lexical alignment [78], and typos [86]. We then organized these codes into elements based on similarity, resulting in 11 elements of conversation architecture. For example, the element of disfluency contains fillers [41, 85], interjections [5, 38], repetitions [90] and typos [86]. These elements are then grouped into 4 categories: dialog strategy, content affectiveness, content style, and speech format (Table 2).

Relationship Between Perceptions of Agents and Conversation Architecture. To study the effect of conversation architecture on users' perceptions (RQ3), we extracted the connections explored in each paper, using the perception aspects and conversation architecture elements developed in the previous data analysis steps. Each connection's effect was also recorded based on whether an association was found in the study and the nature of that association. Out of our review corpus, 265 connections between perceptions of agents and conversation architecture were explored in literature. To analyze the specifics of conversation architecture that affect the perceptions of CAs, 72 connections that did not result in observed relationships were discarded, resulting in 193 relationships for analysis. For example, the connection between the matching style of a CA and user satisfaction was removed because Hoegen et al. [35] did not find a significant difference between the style matching agent and the non-style matching agent for overall interaction satisfaction. Out of the 193 relationships, 10 were based on comparisons between CAs with multiple speech element variations (e.g. [77, 83]). They were not included in the final framework as we could not attribute the perceptions of agents to the individual elements. Overall, 183 relationships between individual conversation architecture elements and perceptions of agents are incorporated into our synthesized framework, visualized as a heatmap to demonstrate their relationships with each other (Figure 3).

4 FINDINGS

In this section, we present our findings on the perception of agents, conversation architecture elements, and their relationship with each other.

[1] https://dl.acm.org/conference/chi
[2] https://dl.acm.org/conference/cui
[3] https://dl.acm.org/conference/hai
[4] https://dl.acm.org/conference/iva

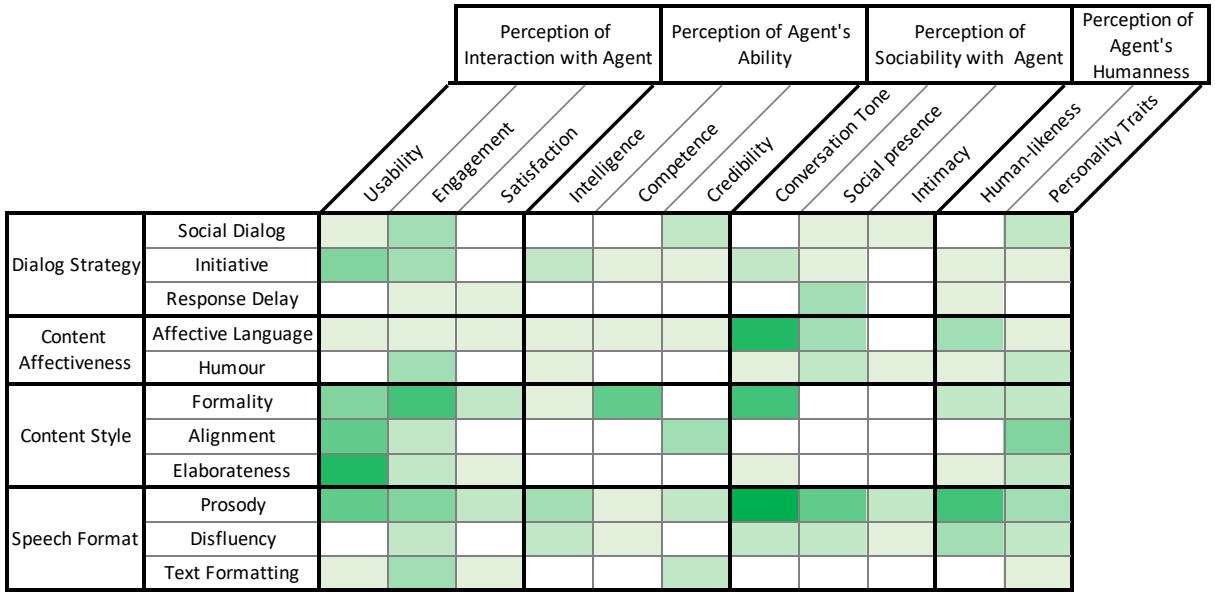


Figure 3: Synthesized framework on the identified relationship between perception of agents and conversation architecture elements. The heatmap visualizes the number of relationships identified from the literature. Note that a single paper may describe more than one relationship.

4.1 Perception of Agents

As shown in Table 1, we defined four categories for the perceptions of conversational agents covering 11 different aspects of perception. The details of each category and aspect are discussed below.

Perception of interaction with agent assesses the overall interaction quality between users and conversational agents. The three aspects in this perception category are: usability, engagement and satisfaction. *Usability* captures the utilitarian component of the interaction, whether it was accurate, easy to use, efficient or helpful. Some commonly used methods to evaluate usability include the response accuracy portion of the SASSI questionnaire [37], or the NASA Task Load Index (NASA-TLX) [32] for cognitive workload. Questions such as "*the system is easy to use*" or "*it is easy to understand the agent*" are also used to evaluate usability. *Engagement* on the other hand captures users' emotional reactions to the CA, including whether they have enjoyed the conversation, or felt annoyed or frustrated with the interaction. Some commonly used methods to evaluate engagement include the annoyance portion of the SASSI questionnaire [37], and the Use Engagement Scale (UES) [71]. Questions such as "*I enjoyed using the system*" or "*I felt frustrated with the agent*" are also used to evaluate engagement. Lastly, *Satisfaction* captures users' overall satisfaction interacting with the agent. Questions such as "*the overall assessment of conversing with the CA was satisfactory*" are used to evaluate this aspect of perception.

Perception of agent's ability assesses the perceived capabilities of the agent. Unlike system capabilities that affect agent's performance, this perceived ability category captures perceptions of agents with equivalent system capabilities but varied conversation architecture elements. Specifically, *intelligence* captures the agent's perceived expertise and knowledge, and it is commonly administered through the Godspeed questionnaire [3] using the set of questions related to perceived intelligence. Also survey questions such as asking users about the agent's intelligence and domain knowledge are used to evaluate perceived intelligence. *Competence* takes intelligence one step further by examining the agent's perceived ability to put its intelligence into practice. Surveys or qualitative feedback are usually used to assess if the agent is capable and competent, and whether users have confidence in the agent's ability to get the job done. Lastly, *credibility* captures agent's perceived truthfulness, benevolence, and user's confidence in CAs. Some commonly used methods to evaluate trust include the Trust Propensity Scale [59] and the Individualized Trust Scale (ITS) [87]. Questions such as "*is the agent honest*" and "*can I trust the agent with sensitive information*" are also used to evaluate trust.

Perception of sociability with agent assesses the emotional connections that users have with conversational agents. This category includes the perception aspects of conversation tone, social presence and intimacy. *Conversation tone* captures the affective impression of the agent's tone, such as empathy and expressiveness. Also, this aspect includes whether the tone used by the agent was perceived as friendly, polite or warm. *Social presence* captures the sense of connectedness and psychological distance users have with an agent. This perception aspect also includes the sense of familiarity or similarity with the agent, and whether users feel the agent behaves like them or has similar attitudes to them. The aspect of *intimacy* extends social presence into the realm of the quality of relationships with an agent. Some commonly used methods to assess intimacy include the set of social attraction questions from the interpersonal attraction questionnaire [61] and the quality of relationship inventory (QRI) [72]. These questionnaires include

Table 1: Categories of the perception aspects with examples and related papers identified from our review.

Perception Aspects		Examples	No. Papers	Papers
Perception of Interaction with Agent	Usability	accuracy, ease of use, efficiency, helpfulness	23	[2][5][7][10][14][19][20][30][31][33][39][40][42][46][45][51][55][58][62][63][75][78][88]
	Engagement	enjoyment, annoyance, desirable, intention to use	28	[1][5][6][7][13][14][20][21][27][28][33][40][42][45][46][54][55][58][62][64][66][75][78][82][84][89][90][94]
	Satisfaction	service satisfaction, quality of interaction	12	[5][10][16][20][31][25][35][39][58][75][88][91]
Perception of Agent's Ability	Intelligence	knowledgeable, intelligent, expertise	11	[2][6][7][14][19][22][38][41][78][84][91]
	Competence	competent, appropriate	6	[13][51][42][53][63][86]
	Credibility	trustworthy, truthfulness, confidence	15	[1][7][19][21][33][35][40][42][51][54][55][58][77][80][88]
Perception of Sociability with Agent	Conversation Tone	friendly, warm, empathetic, persuasive	20	[2][7][13][14][15][16][19][33][38][39][42][43][46][53][62][66][80][90][91][94]
	Social Presence	connectedness, familiarity, psychological distance	18	[5][6][7][10][14][16][25][26][27][28][43][45][53][56][57][58][69][86]
	Intimacy	intimate, rapport, quality of relationship	7	[10][43][45][54][55][57][86]
Perception of Agent's Humanness	Human-likeness	human-like, natural, artificial, machine-like	20	[2][6][7][10][13][16][25][30][39][41][42][56][58][63][69][70][77][85][86][94]
	Personality Traits	likeable, kind, witty, funny, creepy	20	[1][5][6][7][14][30][31][33][39][40][41][46][54][55][62][70][83][84][85][94]

questions like "*I think the agent could be a friend of mine*", or "*I feel we could establish a personal relationship with each other*".

Perception of agent's humanness assesses the specifics of anthropomorphized perceptions of conversational agents. *Human-likeness* captures whether the agent presented itself as natural and human-like, or artificial and machine-like. The Godspeed questionnaire [3] set of questions related to anthropomorphism and the Ascent of Man scale [52] are commonly used methods to evaluate human-likeness. Survey questions with semantic scales such as "*human-like / machine-like*" and "*artificial / natural*" are also used frequently to assess this aspect of perception. *Personality traits* captures the human characteristics that are attributed to the agent. This is commonly collected as qualitative feedback from users, commenting on whether the agent is extroverted or introverted, or its perceived personality such as likeable, funny or witty. Sometimes the Big-5 personality traits questionnaire [29] is used to map an agent's disposition on various personality dimensions.

There is good coverage across the four perception categories based on our reviewed corpus. The perception category of interaction with agent is most commonly explored in literature, followed by the perception category of sociability with agent. Some aspects of perception are under explored in literature, such as the perceived intimacy with a CA or the perceived competence of an agent. This may be due to the controlled lab settings for experiments, where participants are given defined scenarios for interaction. This type of environment is not conducive to forming relationships with a conversational partner, as noted by Linnerman et al. in their discussions [55]. The same factor could impact the assessment of an agent's perceived competence, as users may not feel comfortable assessing the expertise of their conversational partner.

4.2 Conversation Architecture Elements

As shown in Table 2, we defined four categories for conversation architecture across 11 different elements. The details of each category and element are discussed below.

The category of **dialog strategy** refers to the approach used by a conversational agent to engage in a dialogue with a user, which includes the use of social dialogs, agent-initiated content, and adding delays to responses. Specifically, the element of *social dialogue* captures the use of non-task related conversations with users to build social connections, such as using self-disclosure [54] and small talk [57, 83]. The element of *initiative* captures utterances that are initiated by an agent without direct prompts from users. For example, this element includes CAs designed to proactively initiate conversation repair [2, 14], or actively elicit feedback from users [89]. Lastly, *response delay* refers to the tactic of deliberately delaying for a certain period of time before an agent responds to users [25, 27]. It is mainly used in text-based CAs alongside visual displays of typing indicators [26].

Content affectiveness category refers to the conversational agent's use of language to convey emotions or to elicit emotions from users. The element of *affective language* captures injections of emotional words or phrases into an agent's utterances, such as the use of affective expressions [77, 91, 94], sentiment-adaptive responses [16], and encouraging words [33]. The other element in

Table 2: Categories of the conversation architecture elements with examples and related papers identified from our review.

Conversation Architecture Elements		Examples	No. Papers	Papers
Dialog Strategy	Social Dialog	social talk, self-disclosure	7	[1][54][57][66][75][83][84]
	Initiative	proactive dialogue, conversation repair	4	[2][14][51][89]
	Response Delay	response delay, typing indicator	4	[25][26][27][77]
Content Affectiveness	Affective Language	emotional expressions, sentiment-adaptive responses	11	[15][16][33][39][53][56][66][77][84][91][94]
	Humour	humorous content, jokes	6	[6][28][43][64][69][83]
Content Style	Formality	formal vs. casual, honorific expressions	9	[13][20][31][42][46][58][66][70][84]
	Alignment	content matching, lexical alignment, agreeableness	7	[33][35][40][55][78][82][83]
	Elaborateness	elaborate, concise, sentence structure	6	[30][62][66][75][83][84]
Speech Format	Prosody	pitch, intonation, speech rate, spoken accent	12	[7][10][19][22][31][38][42][45][57][63][80][94]
	Disfluency	fillers, interjections, repetitions, typos	6	[5][38][41][85][86][90]
	Text Formatting	capitalization, emoticons	6	[21][46][77][84][86][88]

this category, *humour*, captures an agent's attempt to include jokes in its dialog. Several studies in our reviewed corpus have explored the effect of humour on various perceptions of agents (e.g. [6, 43]).

The category of **content style** refers to the variations of language used in a message aside from the content meaning of the message (i.e. how something is said). The element of *formality* describes the linguistic style used by a conversational agent, such as using formal language like honorific expressions to address users [70]. *Alignment* captures the degree that an agent matches its utterances to users, such as being lexical aligned with the content and structure of users' sentences [40, 55], as well as agreement with users [82]. Lastly, the element of *elaborateness* captures the sentence complexity and length of agents' utterances. For example, Roy et al. [75] explored the differences in perceptions for elaborateness variations such as "*Cloudy, possibility of snow, high: 4, low: -10*" vs. "*Today will be cloudy, with a high of 4 and a low of -10. Snow is predicted*".

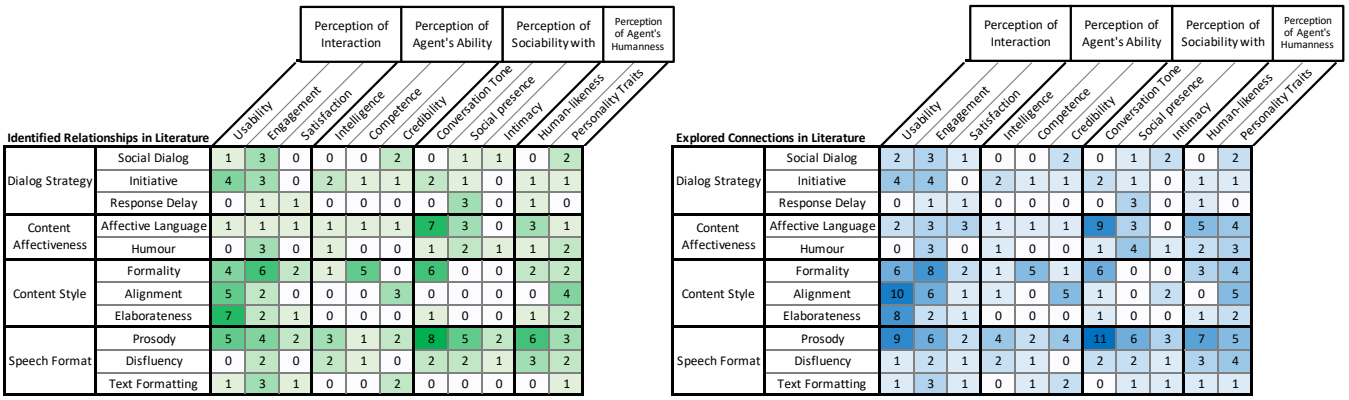
Lastly, the category of **speech format** refers to the non-verbal component of a conversational agent's utterances for both text and voice modalities. *Disfluency* captures the use of non-lexical utterances like filler words ("*um, uh*") [38, 41] or repetition of words within a sentence [90]. It also includes the use of typos [86] for text-based agents. For voice-based agents specifically, the element of *prosody* encompasses vocal qualities like pitch [31, 42], speech rate [10], and spoken accent [22]. For text-based CAs, the *text formatting* element includes different formats agents uses to present information to users, such as using capitalization [86] or emoticons [46, 88].

Overall there are similar number of conversation architecture elements explored across the four categories of dialog strategy, content affectiveness, content style, and speech format. Looking across the 11 different elements of conversation architecture, there is a good coverage of studies on the effect of affective language and prosody cues on the perceptions of agents. However, there are less studies exploring the effect of agent initiated content and response delay on the perception of agents. Specifically for the element of agent initiated content, we found various research on conversation repairs [50, 74] and agents proactively sharing content with users [18, 92], but many of them did not not include assessments of users' perceptions of CAs. This may be due to the fact that the usage of agent-initiated content is relatively new, as CA interactions have historically been driven by users. As such, current research is focused on the functional aspects of agents initiating contents instead of exploring the perceptions of agents. As for response delay, our research suggests this element is under-explored in literature, as only 4 studies in our reviewed corpus examined agents using response delay, with majority of these publications published by the same authors.

4.3 Relationship Between Perceptions of Agents and Conversation Architecture

Our synthesized framework (Figure 3) includes 183 identified relationships out of the 265 explored connections between perceptions of agents and conversation architecture elements. The differences between explored connections and identified relationships are shown as a side-by-side comparison in Figure 4, with detailed discussions in the subsections below.

4.3.1 Perception of Interaction with Agent. The effects of conversation architecture elements on the perception of interaction with agent is the most explored connections in the corpus (n=97). Out of



Identified Relationships in Literature		Perception of Interaction			Perception of Agent's Ability			Perception of Sociability with			Perception of Agent's Humanness	
		Usability	Engagement	Satisfaction	Intelligence	Competence	Credibility	Conversation Tone	Social presence	Intimacy	Human-likeness	Personality Traits
Dialog Strategy	Social Dialog	1	3	0	0	0	2	0	1	1	0	2
	Initiative	4	3	0	2	1	1	2	1	0	1	1
	Response Delay	0	1	1	0	0	0	0	3	0	1	0
Content Affectiveness	Affective Language	1	1	1	1	1	1	7	3	0	3	1
	Humour	0	3	0	1	0	0	1	2	1	1	2
Content Style	Formality	4	6	2	1	5	0	6	0	0	2	2
	Alignment	5	2	0	0	0	3	0	0	0	0	4
	Elaborateness	7	2	1	0	0	0	1	0	0	1	2
Speech Format	Prosody	5	4	2	3	1	2	8	5	2	6	3
	Disfluency	0	2	0	2	1	0	2	2	1	3	2
	Text Formatting	1	3	1	0	0	2	0	0	0	0	1

Explored Connections in Literature		Perception of Interaction			Perception of Agent's Ability			Perception of Sociability with			Perception of Agent's Humanness	
		Usability	Engagement	Satisfaction	Intelligence	Competence	Credibility	Conversation Tone	Social presence	Intimacy	Human-likeness	Personality Traits
Dialog Strategy	Social Dialog	2	3	1	0	0	2	0	1	2	0	2
	Initiative	4	4	0	2	1	1	2	1	0	1	1
	Response Delay	0	1	1	0	0	0	0	3	0	1	0
Content Affectiveness	Affective Language	2	3	3	1	1	1	9	3	0	5	4
	Humour	0	3	0	1	0	0	1	4	1	2	3
Content Style	Formality	6	8	2	1	5	1	6	0	0	3	4
	Alignment	10	6	1	1	0	5	1	0	2	0	5
	Elaborateness	8	2	1	0	0	0	1	0	0	1	2
Speech Format	Prosody	9	6	2	4	2	4	11	6	3	7	5
	Disfluency	1	2	1	2	1	0	2	2	1	3	4
	Text Formatting	1	3	1	0	1	2	0	1	1	1	1

Figure 4: Number of past work that identified relationships between conversation architecture and the perception of agents (on the left) and number of past work that explored possible connections (on the right).

these explored connections, the majority of them (n=66) discovered relationships between the perceptions of interaction with agent related to the studied architecture element. Looking at the difference between modalities, we noticed more voice-based agents (n=25) resulted in null relationships compared to text-based agents (n=6). The conversation architecture element of *alignment* contributed the most to the null results for voice-based agents, with 9 out of 11 explored connections not finding any relationships related the perception of interaction with agents.

Across the aspects for the perception of interaction, both usability and engagement were assessed frequently in our reviewed corpus, with speech variations having effects on both perception aspects across all conversation architecture element categories. The perception aspect of user satisfaction is commonly used to evaluate conversational agents within the customer service domain answering transactional inquiries (e.g. [16, 20, 25]).

The category of conversation architecture with the most identified relationships to perceptions of interaction is content style. Specifically for the element of *elaborateness*, users found the use of full sentences more useful than keyword only [30, 75]. Otherwise, the effect of an agent's elaborateness on user's perception of interaction depends on the user's preference [62], as well as the topic of discussion [30]. As for *formality*, various studies reported significant differences in the perceptions of interaction between CAs using casual vs. formal styles. However, there are mixed results on the effects on formality, as some users experienced higher engagement interacting with agent using casual style of conversation [13], while in another study users found the CA using formal language style as less engaging as it is boring [46].

There are also a number of papers in our reviewed corpus exploring the effect of speech format elements on the perception of interaction with CAs, with many of them finding significant relationships between them. The element with the most identified relationships is *prosody*. For instance, studies found that a CA's expressiveness in vocal cues increased participants' engagement ratings [94], and different pitches of voice affect users' perception of engagement and usability [7, 31].

Dialog strategy and content affectiveness categories of conversation architecture did not have as many identified relationships with perceptions of interaction. It is worth noting the effect of using the *initiative* element, as CAs that use self-initiated content are perceived as more efficient and higher quality of interaction [14]. Also, some studies have discovered conflicting effects between different perception aspects of interaction for CAs using *social dialog*. One such example is users enjoying of the conversation with CAs using social talk [54, 75], but perceiving the agent as less efficient [75].

4.3.2 Perception of Agent's Ability. The effects of conversation architecture elements on the perception of agent's ability is the least explored connection in the corpus (n=39). Out of these explored connections, the majority of them (n=30) found relationships between the studied speech element and the perceptions of agent's ability. There are no notable differences between the text and voice modalities of CAs for either explored connections or the identified relationships.

Perception aspects of agent's ability had similar number of identified relationships with speech elements across intelligence, competence and credibility. Our review revealed that the perceived ability of a CA is also dependent on other influencing factors in addition to conversation architecture elements. For instance, Kraus et al. [51] found that the perception of competence of a proactive CA is dependent on task difficulty. Also, the perception of intelligence for an agent using fillers depended on the context of the conversation, as the filler-speaking agent was perceived as less intelligent in task-oriented conditions, but was seen as slightly more intelligent in social-oriented conditions [41].

The conversation architecture element of *prosody* has relationships identified with each perception aspect of an agent's ability. For example, Chan et al. [7] found that agents using kin's voices are rated as more intelligent and credible than generic voices. Also, the style of speech used by an agent affects the perception of appropriateness of tone, which is an aspect of the perceived competence of an agent [63]. Different content styles of *formality* used by a CA also impacts its perceived competence [13, 42]. Related to the perception aspect of credibility, being *lexically aligned* with the user improved the rating of trustworthiness of an agent [35, 55]. There are some mixed results on the use of emoticons within the

text formatting element. One study found that chatbots using emoticons as less trustworthy [88], while another study found that users assigned higher scores of confidence to the agent using emojis [21].

There are several conversation architecture elements with little to no explorations related to the perception of agent's ability. Specifically, both categories of dialog strategy and content affectiveness are minimally explored in our reviewed corpus, with no explored connections with *response delay* and only one or two connections explored for *social dialog* and *humour*. While some speech elements within the content style category have been evaluated for agents' perceived abilities, it is worth noting that the element of *elaborateness* did not have any explored connections with perceived ability of agents. Further research in these areas is needed to close these knowledge gaps.

4.3.3 Perception of Sociability with Agent. The perception of agent's sociability is the second most explored category within our reviewed corpus (n=64), with majority of the connections found relationships between the studied conversation architecture elements and perceptions of agent (n=49). Interestingly, studies explored more perceptions of sociability with agent related to voice-based CAs (n=40) compared to text-based CAs (n=24), potentially due to the assumption that interactions with text-based agents are perceived as less personal and more formal [49]. There are some mixed results for voice-based CAs, as 12 out of the 28 explored connections did not find any significant results. Drilling down to the specific elements, we noticed that the use of *alignment* in voice-based agents did not effect the perception of sociability with agent [33, 55]. As there are no papers in our reviewed corpus studying the effect of text-based agents using alignment on the perceptions of sociability, we were not able to compare the effect of alignment between modalities.

Across the aspects in this perception category, perceived conversation tone has the most number of identified relationship (n=27), followed by perceived social presence (n=17). There are only 5 relationships found between speech elements and perceived intimacy. One reason for this is half of the explored connections with speech elements did not have any effect on the perception of sociability with agent, such as the use of capitalization [86], lexical alignment [55], and social talk [57]. For the perception aspect of social presence, we noticed a gap exploring elements in the content style category (formality, alignment, elaborateness), as there are no connections explored at this intersection.

Looking at the different categories of conversation architecture elements, the heatmap (Figure 3) shows that there are not many relationships in content style category of conversation architecture that are related to the aspects of perceived sociability with CAs. Most of the identified effects are concentrated at the intersection of *formality* and perceived conversation tone. Studies have found that CAs using casual style of conversation are perceived as warm, empathetic and friendly [42, 46], while CAs using formal style of conversation are perceived as polite but lacks empathy [13]. The rest of the speech elements related to the perceptions of sociability are minimally explored in literature, indicating the need for more research in this area.

There are a few specific conversation architecture elements that has more identified relationships with perceptions of sociability with agent. Specifically, CAs using *affective language* are perceived to be more empathetic [15, 16, 91] and emotionally expressive [94], as well as being more emotional connected with their users [53, 56]. Also, different variations of a voice-based CA's *prosody* have effects on the perceived sociability with agent, such as an agent using expressive prosody is assessed as more intimate and more similar with the user [45]. Lastly, the conversation architecture element of *humour* has identified relationships across aspects of perceived sociability with CAs. Our review found that humor has effects on human-agent relationships, as humorous agents are rated as more friendly, intimate, and similar by users compared to non-humorous agents [28, 43]. This is contrary to Clark et al. [12]'s findings that while humour is an important conversational characteristic for human-human interactions, it is viewed as a novelty feature for human-agent conversations.

4.3.4 Perception of Agent's Humanness. There are 55 explored connections between conversation architecture elements and the perception of sociability with agents, with majority of them (n=38) discovering relationships in literature. There are significantly more explorations for voice-based agents (n=39) as compared to text-based agents (n=16). Previous studies have found that modality may have an effect on the perception of humanness, as voice-based agents are perceived as more human-like as compared to text-based agents [9]. Out of the 39 explored connections for voice-based agents, 14 of them did not result in any relationships. This is especially evident in the speech element of *affective language* for voice-based agents, as most of the connections resulted in null relationships. For example, when comparing the ratings of human-likeness or likeability for a speech agent employing expressive words to the one not using any, Zhu et al [94] were unable to detect any statistically significant differences in both of the observation study and interaction study.

Across the aspects in this perception category, both human-likeness and personality traits have relationships with almost all the conversation architecture elements. Specifically for the use of *prosody* in CAs, our review found some opposing effects on the perception of the agent's humanness. In the case of Chan et al's study [7], participants rated the agent using kin voices as significantly more likeable compared to the generic voices, but it was perceived as eerie. This may be a warning indicator to beware of the uncanny valley effect [67] when designing conversation architecture elements to elicit anthropomorphized perceptions from users.

There are a few notable conversation architecture elements related to the perceived humanness of conversational agents. The element of *prosody* such as varying pitch, intonation and speech rate has more identified relationships with perceived personality traits of an agent as compared to perceived human-likeness, especially on the perception of likeability for an agent [10, 42, 63]. For the element of *disfluency*, the effect on perceived humanness depended on the context of the conversation. Studies have found that participants perceived the filler-condition agent as more likeable in the social-oriented situation, but did not find the same effect in task-oriented situations [41, 85]. The use of *alignment* in an agent has a number of identified relationships with the perception of

personality traits of an agent, with some studies finding CAs that are lexically aligned with a user are more likeable [40, 55].

5 DISCUSSION

5.1 Research Challenges and Opportunities

5.1.1 The Need for Consistent Evaluation Protocols for Perceptions of Agents. We found a diversity of approaches used to assess perceptions of agents in our reviewed corpus. While we made our best efforts to group perceptions used in literature based on similarity with each other, the inconsistent and composite nature of these perception assessments makes it challenging to synthesize them into a framework that contains homogeneous components in each perception aspect that are comparable with each other.

Based on the literature reviewed in this paper, it is unclear whether perception aspects using similar labels are being evaluated *consistently* through different approaches. For example, we found several methods used to assess the perceived human-likeness of an agent. A commonly used survey is adapted from the Godspeed questionnaire [3], which evaluates human-likeness based on users' impression of the agent as fake / natural, machinelike / humanlike, unconscious / conscious, and artificial / lifelike (used by [35], [41] and [70]). Another method used to evaluate human-likeness is adapted from Holtgraves et al.'s [36] questionnaire, which asks users questions related to the agent's perceived human-likeness, skillfulness, thoughtfulness, politeness, responsiveness and engagement (used by [16] and [25]). One study [86] used the Ascend of Man scale [52] with pictures showing the evolution from ape to man, asking users to choose a depiction that best represents the agent's perceived human-likeness. It is unclear whether these different methods are capturing assessments of similar perception aspects that can be compared with each other. Another example is the evaluation of the perception of agent's empathy. In Diedrech et al.'s study [16], empathy is assessed by asking users whether the CA is giving users individual or personal attention. In another study that also assesses the perception of empathy [15], the RoPE Scale [8] is used with questions like "*the robot cares about my feelings*" or "*the robot comforts me when I am upset.*" These evaluations of empathy seem to have different underlying meanings, one assessing the personalization aspect of CAs, while the other is assessing the emotional aspect of CAs.

In addition the problem of consistency in evaluating perceptions, there is the issue of *composite measures* being used in the assessment of users' perceptions. These composite measures collapse multiple aspects of perception into one measurement, making it impossible to break down perceptions into more granular aspects for analysis. One such example is Ma et al's study [58] to evaluate different approaches used by CAs to reply to users' uncertain queries. A single UX score is used to evaluate users' perceptions, which composes of questions on whether the user thinks the CA's response is pleasing / trustworthy / natural / acceptable / shorten the distance between CA and the user. This UX score encompasses multiple aspects across several perception categories. While the study has found significant effect for the use of formal language on the user-rated UX score, it is not possible to understand how the details of formality is related to the perception categories of interaction (pleasing, acceptable), sociability (shorten the distance between CA and user), and humanness (natural, trustworthy). The humanness questionnaire from Holtgraves et al. [36] has a similar problem, evaluating across perception categories of interaction (responsiveness, engagement), ability (skillfulness), sociability (politeness), and humanness (human-likeness, thoughtfulness).

There has been some effort recently towards unifying the evaluation of conversational agents, such as the work by Finch et al. [24] presenting a comprehensive analysis of current evaluation protocols. More research is needed in this area to standardize the assessment of perceptions to make them consistent, granular, and comparable across literature. In the interim, we urge researchers to maintain the granularity of measurements to ensure that the subtleties among different perceptions of agents are preserved in their analysis.

5.1.2 Investigate the Relationship Across Perception Aspects of Agents. There is evidence in literature that perception aspects of agents have effects on each other. For instance, Moussawi et al. [68] conducted a study to understand the correlations between different perceptions related to users' intention to adopt a conversational agent. Specifically, they found a correlation within the category of perceived ability, where an agent's perceived intelligence is positively correlated to the perceived initial trust of the agent. This study also found that users with higher perceived intelligence of an agent are more likely to attribute higher ratings for perceived human-likeness, as well as for the usability and engagement aspects in the perception category of interaction. Lastly, the authors found that perceived humanness have a positive impact on perceived enjoyment, which lead to higher intention to adopt the CA. This study indicates that the perception aspects identified in this paper are not independent of each other.

Correlations between perception aspects are discussed in a few of the papers we reviewed. One of the studies showed that the CA's perceived anthropomorphized personality trait of agreeableness has an influence on the users' perception of credibility [1]. Seeger and Heinzl [77]'s study found a similar correlation, where higher perceived anthropomorphism led to lower loss of perceived trust. There are also correlations between aspects within the same perception category, such as a speech agent that is rated higher in perceived human-likeness is associated with higher likability ratings [94]. Also, within the perception category of social connection, one study found that perceived social distance is positively related to perceived social attraction [86].

These correlations demonstrate that some perception aspects have effects on each other, either within the same category or across different categories. While there are limited research into this area, further investigations are needed to understand the relationships between various perception aspects.

5.1.3 Research Directions for the Effects of Conversation Architecture on Users' Perceptions. Our synthesized framework (Figure 3) lays the foundation in understanding the effect of conversation architecture on the perception of agents. It showcases the density of explored relationships between them, as well as highlighting areas that are under-explored. In order to continue building our knowledge, it is important to investigate further into the effect of contextual factors as well as multiple conversation architecture elements on users' perceptions.

There are a few *under-explored areas* in our synthesized framework. The perception category of agent's ability has the least number of explored connections with conversation architecture compared to the other perception categories. Specifically, more studies are needed to understand the effects of social dialog, response delay, humour and elaborateness on users' perceptions of agents' ability. For the perception category of sociability with agents, there is a general lack of explored connections with conversation architecture elements in the content style category, especially for alignment and elaborateness. Lastly, more studies on the effect of using agent-initiated content and response delay on the perceptions of agents are needed.

Some studies have discussed how differences in *contextual factors* resulted in variations in the perceptions of agents while using similar conversation architecture elements. One of the factors is users' prior experiences with CAs, as Gnewuch et al. [25] found that experienced users perceived the agent using response delay as lower in social presence because it is seen as inefficient to wait for the CA to respond, but novice users perceived higher social presence conversing with the CAs using response delays because it is more similar to conversations with human partners [25]. Another factor is users' characteristics, as the perceived trustworthiness of a style-matching agent depended on users' own conversational style [35]. Some other contextual factors that resulted in differences in perceptions of an agent include the purpose of conversation (e.g. transactional vs. social) [41], anonymity of the conversation [54], and the sensitivity of information discussed in the interaction [13]. A comprehensive review to identify these contextual factors as well as to understand their effects on the perceptions of agents across various conversation architecture elements would be useful to tailor CA design for specific situations.

As we gain a better understanding of the effects of a single conversation architecture element on the perceptions of agents, we can extend the research to using *multiple conversation architecture elements* in an agent. Several papers in our reviewed corpus incorporated composite elements in an agent, such as the design of an anthropomorphized chatbot using elements such as affective language, emoticons, and response delays to assess users' perceptions compared to a non-anthropomorphized chatbot [77]. In addition to exploring the combined effects of conversation architecture elements on users' perceptions, it would also be interesting to understand the relative importance of each element on users' perceptions. This can be studied by analyzing the effect of modifying conversation architecture elements individually, as well as their combined effect on user's perceptions (e.g. [31, 57, 94]). For instance, Habler et al. [31] found that the effect of social dialog is larger than the effect of prosody on the perception of agents.

5.2 Design Implications and Ethical Considerations

Although there are existing research on the relationship between conversation architecture used by CAs and how users perceive them, there is currently a lack of a comprehensive understanding of these connections. Our work proposes a framework that establishes a foundational understanding of the current state of research and highlights under-explored areas that require further investigation. This framework also serves as a guide for designers by summarizing the various conversation architecture elements that can be utilized in design, along with their corresponding influences on user perception of agents. Designers can use this framework to understand how their choice of conversation architecture affects user perception of agents, or to design for specific perceptions of agents using conversation architecture elements. For instance, designers seeking to enhance the perception of competence in an agent may investigate the effects of initiative strategies, affective language, formal language, and different prosody settings. However, designers of conversational agents need to be aware of and consider ethical implications and potential negative impacts for users. There are three main areas of concern related to the effect of conversation architecture on the perceptions of agents: gender stereotypes, influencing users' actions, and privacy concerns.

For the element of prosody, various studies explored the effect of different pitches on the perceptions of agents. Even in studies that are not explicitly analyzing the effect of different gendered voices in an agent, possible stereotypes may still exist in the study. Based on our reviewed corpus, studies found that lower pitches commonly associated with men are considered to be more desirable and authoritative but less friendly [42, 80]. For Dubiel et al. [19]'s study, the agent with a lower mean pitch was selected as the more persuasive voice. This result may be demonstrating users' unconscious bias to select a male-sounding voice as more persuasive and authoritative over female-sounding voices that are commonly associated with higher pitches. While some guidelines recommend designing agents to be androgynous to avoid gender stereotypes [76], there are limitations in creating gender-ambiguous voices. Currently, there are no defined guidelines on what is perceived as a gender-neutral voice [42]. There is also a lack of voice generators available to generate voices that are perceived as androgynous [80]. To address the issue of gender stereotype, CA designers should give users easy access to choose from a variety of voices.

Studies have demonstrated that conversation architecture elements can be used to design perceptions of agents to make CAs more persuasive [19]. This opens up the ethical issue of influencing users' attitudes and behaviours through these persuasion techniques. In a study by Chan et al [7], they found that CAs using kin's voices are perceived as more credible and likable, with a higher perceived social presence with the agent. These perceptions contribute to the agent being more engaging and persuasive, therefore people are more likely to comply with its requests. In another study, Andrews [1] found that tailoring the personality of the CA to users will positively impact an agent's persuasiveness. These persuasion techniques can be beneficial to help users achieve their goals, but they can also be used for harmful actions, such as trying to get users to believe in false information. Designers who use persuasive techniques in CAs should be transparent about their goal of influencing human perceptions and behaviors, and ensure that their actions are in the best interest of the users.

Another key area of concern is privacy. Given the natural language format of conversational agents, users may be disclosing more sensitive and personal information than needed for the interaction. In relationship to conversation architecture, using elements like self-disclosure and persuasive voice prosody settings could result in users perceiving higher trust with the agent, leading them

to disclose more sensitive information [19, 54]. This can expose users to attacks, such as CAs using voice impersonation to ask for personal information for malicious intents [7]. As such, CAs ought to be constructed in a manner that ensures complete clarity regarding the storage and usage of user data. Moreover, designers need to protect the privacy of users by taking into account the sensitivity of the data, determining who can access it, and devising measures to prevent malicious users from exploiting it.

5.3 Limitations and Future Directions

Our systematic literature review was limited to papers published in the ACM Digital Library between 2010 and 2022. We may have missed literature published outside this time period, as well as in other libraries. Also, most of the studies in our corpus are based on lab experiments using short interactions with users. The generalizability of these findings needs to be verified through longer-term engagements with conversational agents deployed in real-world situations. Lastly, we needed to rely on our own interpretations to categorize some of the perception measures into the framework due to inconsistent protocols used in the evaluation of perceptions of agents.

There are several future research directions that could enhance the proposed framework outlined in this paper. Firstly, establishing consistent measures for evaluating users' perceptions of agents is crucial to increase validity and consistency. Additionally, exploring under-examined aspects of the framework, such as how an agent's elaborateness affects its perceived ability (intelligence, competence, credibility), is necessary to further contribute to the body of knowledge. Moreover, delving deeper into the nuances that affect the relationship between conversation architecture and users' perceptions of agents is essential. This includes exploring contextual factors, such as users' prior experiences with conversational agents, as well as understanding how the combination of multiple conversation architecture elements can influence users' perceptions.

6 CONCLUSION

In this paper, we share our findings of a systematic review of existing literature published in the ACM Digital Library on the effect of conversation architecture elements on the perceptions of CAs. Through the synthesis of 57 papers in our corpus, we found that current literature has explored users' perceptions of agents related to interaction, ability, sociability and humanness in relation to the effects of conversation architecture (RQ1). Also, we observed that conversation architecture elements related to dialog strategy, content affectiveness, content style, and speech format are relevant to users' perceptions of agents (RQ2). Based on our in-depth analysis, we present a framework outlining the identified relationships between elements of agents' conversation architecture and aspects of users' perception (RQ3). Our investigation also revealed the need for consistent protocols in evaluating perceptions of agents, as measurements are inconsistent across studies. Also, more research is needed to investigate the under-explored areas in the framework, the relationship across perception aspects, the influence of contextual factors, and the effect of composite conversation architecture elements on users' perceptions. While our research contribute to the design of conversation architecture to orchestrate specific perceptions of agents, we urge designers to incorporate ethical perspectives into their design considerations, including potential gender stereotypes, the use of persuasive techniques to influence users, and privacy issues related to users disclosing sensitive information to agents.

ACKNOWLEDGMENTS

This work was supported by NAVER Corp.

REFERENCES

[1] Pierre Y Andrews. 2012. System personality and persuasion in human-computer dialogue. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 2 (2012), 1–27.
[2] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
[3] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1 (2009), 71–81.
[4] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
[5] Jessy Ceha and Edith Law. 2022. Expressive Auditory Gestures in a Voice-Based Pedagogical Agent. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.
[6] Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law. 2021. Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[7] Sam WT Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. 2021. KinVoices: Using Voices of Friends and Family in Voice Interfaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
[8] Laurianne Charrier, Alisa Rieger, Alexandre Galdeano, Amélie Cordier, Mathieu Lefort, and Salima Hassas. 2019. The rope scale: a measure of how empathic a robot is perceived. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 656–657.
[9] Eugene Cho, Maria D Molina, and Jinping Wang. 2019. The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants. *Cyberpsychology, Behavior, and Social Networking* 22, 8 (2019), 515–520.
[10] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. " Nobody speaks that fast!" An empirical study of speech rate in conversational agents for people with vision impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[11] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers* 31, 4 (2019), 349–371.
[12] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[13] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–13.
[14] Andrea Cuadra, Shuran Li, Hansol Lee, Jason Cho, and Wendy Ju. 2021. My bad! repairing intelligent voice assistant errors improves interaction. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–24.
[15] Karl Daher, Jacky Casas, Omar Abou Khaled, and Elena Mugellini. 2020. Empathic chatbot response for medical assistance. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–3.
[16] Stephan Diederich, Max Janssen-Müller, Alfred Benedikt Brendel, and Stefan Morana. 2019. Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: insights from an experiment with a conversational agent. (2019).
[17] Philip R Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R Cowan. 2019. Mapping perceptions of humanness in intelligent personal assistant interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[18] Mateusz Dubiel, Alessandra Cervone, and Giuseppe Riccardi. 2019. Inquisitive mind: a conversational news companion. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.

[19] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.

[20] Ela Elsholz, Jon Chamberlain, and Udo Kruschwitz. 2019. Exploring language style in chatbots to increase perceived product value and user engagement. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 301–305.

[21] Ahmed Fadhil, Gianluca Schiavo, Yunlong Wang, and Bereket A Yilma. 2018. The effect of emojis when interacting with conversational interface assisted health coaching system. In *Proceedings of the 12th EAI international conference on pervasive computing technologies for healthcare*. 378–383.

[22] Pedro Guillermo Feijóo-García, Mohan Zalake, Alexandre Gomes de Siqueira, Benjamin Lok, and Felix Hamza-Lup. 2021. Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*. 68–75.

[23] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161.

[24] Sarah E Finch and Jinho D Choi. 2020. Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. *arXiv preprint arXiv:2006.06110* (2020).

[25] Ulrich Gnewuch, Stefan Morana, Marc Adam, and Alexander Maedche. 2018. Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. (2018).

[26] Ulrich Gnewuch, Stefan Morana, Marc TP Adam, and Alexander Maedche. 2018. 'The Chatbot is typing…'–The Role of Typing Indicators in Human-Chatbot Interaction. In *Proceedings of the 17th Annual Pre-ICIS Workshop on HCI Research in MIS*. 0–5.

[27] Ulrich Gnewuch, Stefan Morana, Marc TP Adam, and Alexander Maedche. 2022. Opposing Effects of Response Time in Human–Chatbot Interaction. *Business & Information Systems Engineering* (2022), 1–19.

[28] Kenro Go, Toshiki Onishi, Asahi Ogushi, and Akihiro Miyata. 2021. Conversational Agents Replying with a Manzai-style Joke. In *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*. 221–230.

[29] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[30] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. 2022. Keep it Short: A Comparison of Voice Assistants' Response Behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–12.

[31] Florian Habler, Valentin Schwind, and Niels Henze. 2019. Effects of Smart Virtual Assistants' Gender and Language. In *Proceedings of Mensch und Computer 2019*. 469–473.

[32] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[33] Jennifer Healey and Dalila Szostak. 2013. Relating to speech evoked car personalities. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1653–1658.

[34] Jennifer Hill, W Randolph Ford, and Ingrid G Farreras. 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in human behavior* 49 (2015), 245–250.

[35] Rens Hoegen, Deepali Aneja, Daniel McDuff, and Mary Czerwinski. 2019. An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 111–118.

[36] TM Holtgraves, Stephen J Ross, CR Weywadt, and TL Han. 2007. Perceiving artificial social agents. *Computers in human behavior* 23, 5 (2007), 2163–2174.

[37] Kate S Hone and Robert Graham. 2000. Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering* 6, 3-4 (2000), 287–303.

[38] Jiaxiong Hu, Yun Huang, Xiaozhu Hu, and Yingqing Xu. 2021. Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[39] Yaxin Hu, Yuxiao Qu, Adam Maus, and Bilge Mutlu. 2022. Polite or Direct? Conversation Design of a Smart Display for Older Adults Based on Politeness Theory. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[40] Shen Huiyang and Wang Min. 2022. Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions. *International Journal of Human–Computer Interaction* 38, 1 (2022), 28–41.

[41] Yuin Jeong, Juho Lee, and Younah Kang. 2019. Exploring effects of conversational fillers on user perception of conversational agents. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[42] Iris Jestin, Joel Fischer, Maria Jose Galvez Trigo, David Large, and Gary Burnett. 2022. Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–11.

[43] Peter Khooshabeh, Cade McCall, Sudeep Gandhe, Jonathan Gratch, and James Blascovich. 2011. Does it matter if a computer jokes. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 77–86.

[44] Hyeji Kim, Inchan Jung, and Youn-kyung Lim. 2022. Understanding the Negative Aspects of User Experience in Human-likeness of Voice-based Conversational Agents. In *Designing Interactive Systems Conference*. 1418–1427.

[45] Jieun Kim, Woochan Kim, Jungwoo Nam, and Hayeon Song. 2020. " I Can Feel Your Empathic Voice": Effects of Nonverbal Vocal Cues in Voice User Interface. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[46] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[47] Bart P Knijnenburg and Martijn C Willemsen. 2016. Inferring capabilities of intelligent agents from their external traits. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 1–25.

[48] Ahmet Baki Kocaballi, Emre Sezgin, Leigh Clark, John M Carroll, Yungui Huang, Jina Huh-Yoo, Junhan Kim, Rafal Kocielnik, Yi-Chieh Lee, Lena Mamykina, et al. 2022. Design and Evaluation Challenges of Conversational Agents in Health Care and Well-being: Selective Review Study. *Journal of medical Internet research* 24, 11 (2022), e38525.

[49] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for workplace reflection: a chat and voice-based conversational agent. In *Proceedings of the 2018 designing interactive systems conference*. 881–894.

[50] Kazunori Komatani and Hiroshi G Okuno. 2010. Online error detection of barge-in utterances by using individual users' utterance histories in spoken dialogue system. In *Proceedings of the SIGDIAL 2010 Conference*. 289–296.

[51] Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. 2020. Effects of proactive dialogue strategies on human-computer trust. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 107–116.

[52] Nour Kteily, Emile Bruneau, Adam Waytz, and Sarah Cotterill. 2015. The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of personality and social psychology* 109, 5 (2015), 901.

[53] Minha Lee, Gale Lucas, Johnathan Mell, Emmanuel Johnson, and Jonathan Gratch. 2019. What's on Your Virtual Mind? Mind Perception in Human-Agent Negotiations. In *Proceedings of the 19th ACM international conference on intelligent virtual agents*. 38–45.

[54] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. " I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[55] Gesa Alena Linnemann and Regina Jucks. 2018. 'Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?'—Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction. *Interacting with Computers* 30, 3 (2018), 173–186.

[56] Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura. 2019. Positive emotion elicitation in chat-based dialogue systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27, 4 (2019), 866–877.

[57] Nichola Lubold, Erin Walker, and Heather Pon-Barry. 2016. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 255–262.

[58] Qianli Ma, Yaping Zhang, Wenti Xu, and Ronggang Zhou. 2022. Ask a Further Question or Give a List? How Should Conversational Agents Reply to Users' Uncertain Queries. *International Journal of Human–Computer Interaction* (2022), 1–15.

[59] Roger C Mayer and James H Davis. 1999. The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of applied psychology* 84, 1 (1999), 123.

[60] Kelly McConvey, Shion Guha, and Anastasia Kuzminykh. 2023. A Human-Centered Review of Algorithms in Decision-Making in Higher Education. *arXiv preprint arXiv:2302.05839* (2023).

[61] James C McCroskey, Virginia P Richmond, and John A Daly. 1975. The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research* 1, 4 (1975), 323–332.

[62] Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 2018. Exploring the impact of elaborateness and indirectness on user satisfaction in a spoken dialogue system. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 165–172.

[63] Teruhisa Misu, Etsuo Mizukami, Yoshinori Shiga, Shinichi Kawamoto, Hisashi Kawai, and Satoshi Nakamura. 2011. Toward construction of spoken dialogue system that evokes users' spontaneous backchannels. In *Proceedings of the SIGDIAL 2011 Conference*. 259–265.

[64] Tomoki Miyamoto, Daisuke Katagami, and Yuka Shigemitsu. 2017. Improving relationships based on positive politeness between humans and life-like agents. In *Proceedings of the 5th International Conference on Human Agent Interaction*. 451–455.

[65] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.

[66] Joonas Moilanen, Aku Visuri, Sharadhi Alape Suryanarayana, Andy Alorwu, Koji Yatani, and Simo Hosio. 2022. Measuring the Effect of Mental Health Chatbot Personality on User Engagement. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*. 138–150.

[67] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.

[68] Sara Moussawi, Marios Koufaris, and Raquel Benbunan-Fich. 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets* 31, 2 (2021), 343–364.

[69] Radosław Niewiadomski, Jennifer Hofmann, Jérôme Urbain, Tracey Platt, Johannes Wagner, Piot Bilal, T Ito, C Jonker, M Gini, and O Shehory. 2013. Laugh-aware virtual agent and its impact on user amusement. (2013).

[70] Subaru Ouchi, Kazuki Mizumaru, Daisuke Sakamoto, and Tetsuo Ono. 2019. Should speech dialogue system use honorific expression?. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. 232–233.

[71] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[72] Gregory R Pierce, Irwin G Sarason, Barbara R Sarason, Jessica A Solky-Butzel, and Lauren C Nagle. 1997. Assessing the quality of personal relationships. *Journal of Social and Personal Relationships* 14, 3 (1997), 339–356.

[73] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630.

[74] Fabian Reinkemeier and Ulrich Gnewuch. 2022. Designing Effective Conversational Repair Strategies for Chatbots. (2022).

[75] Quentin Roy, Moojan Ghafurian, Wei Li, and Jesse Hoey. 2021. Users, Tasks, and Conversational Agents: A Personality Study. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. 174–182.

[76] Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations.. In *AICS*. 104–115.

[77] Anna-Maria Seeger and Armin Heinzl. 2021. Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service?. In *ECIS*.

[78] Laura Spillner and Nina Wenig. 2021. Talk to Me on My Level–Linguistic Alignment for Chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–12.

[79] Brodrick Stigall, Jenny Waycott, Steven Baker, and Kelly Caine. 2019. Older adults' perception and use of voice user interfaces: a preliminary review of the computing literature. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. 423–427.

[80] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by default?–exploring the effect of voice assistant gender and pitch on trait and trust attribution. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–7.

[81] Michelle ME Van Pinxteren, Mark Pluymaekers, and Jos GAM Lemmink. 2020. Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management* 31, 2 (2020), 203–225.

[82] Sarah Theres Völkel and Lale Kaya. 2021. Examining user preference for agreeableness in chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–6.

[83] Sarah Theres Völkel, Samantha Meindl, and Heinrich Hussmann. 2021. Manipulating and evaluating levels of personality perceptions of voice assistants through enactment-based dialogue design. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–12.

[84] Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User perceptions of extraversion in chatbots after repeated use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[85] Mirjam Wester, Matthew Aylett, Marcus Tomalin, and Rasmus Dall. 2015. Artificial personality and disfluency. In *Sixteenth Annual Conference of the International Speech Communication Association*.

[86] David Westerman, Aaron C Cross, and Peter G Lindmark. 2019. I believe in a thing called bot: Perceptions of the humanness of "chatbots". *Communication Studies* 70, 3 (2019), 295–312.

[87] Lawrence R Wheeless and Janis Grotz. 1977. The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3, 3 (1977), 250–257.

[88] Marilena Wilhelm, Tabea Otten, Eva Schwaetzer, and Kinga Schumacher. 2022. Keep on Smiling: An Investigation of the Influence of the Use of Emoticons by Chatbots on User Satisfaction. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–6.

[89] Ziang Xiao, Sarah Mennicken, Bernd Huber, Adam Shonkoff, and Jennifer Thom. 2021. Let Me Ask You This: How Can a Voice Assistant Elicit Explicit User Feedback? *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–24.

[90] Jie Yang, Hirofumi Kikuchi, Takatsugu Uegaki, and Hideaki Kikuchi. 2021. The effect of the repetitive utterances complexity on user's perceived empathy and desire to continue dialogue by a chat-oriented dialogue system. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. 241–244.

[91] Yang Yang, Xiaojuan Ma, and Pascale Fung. 2017. Perceived emotional intelligence in virtual agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2255–2262.

[92] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Völkel, Johannes Schöning, Rainer Malaka, and Yvonne Rogers. 2022. Understanding circumstances for desirable proactive behaviour of voice assistants: The proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–14.

[93] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX research on conversational human-AI interaction: A literature review of the ACM digital library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–24.

[94] Qingxiaoyang Zhu, Austin Chau, Michelle Cohn, Kai-Hui Liang, Hao-Chuan Wang, Georgia Zellou, and Zhou Yu. 2022. Effects of Emotional Expressiveness on Voice Chatbot Interactions. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–11.

A APPENDIX

Table 3: Articles included in literature review

Authors	Title	Year	Venue
Andrews [1]	System personality and persuasion in human-computer dialogue	2012	ACM Transactions on Interactive Intelligent Systems (TiiS)
Ashktorab et al. [2]	Resilient chatbots: Repair strategy preferences for conversational breakdowns	2019	ACM Conference on Human Factors in Computing Systems (CHI)
Ceha and Law [5]	Expressive Auditory Gestures in a Voice-Based Pedagogical Agent	2022	ACM Conference on Human Factors in Computing Systems (CHI)
Ceha et al. [6]	Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?	2021	ACM Conference on Human Factors in Computing Systems (CHI)
Chan et al. [7]	KinVoices: Using Voices of Friends and Family in Voice Interfaces	2021	ACM Conference on Computer Supported Cooperative Work (CSCW)
Choi et al. [10]	"Nobody speaks that fast!" An empirical study of speech rate in conversational agents for people with vision impairments	2020	ACM Conference on Human Factors in Computing Systems (CHI)
Cox and Ooi [13]	Does Chatbot Language Formality Affect Users' Self-Disclosure?	2022	ACM Conference Conversational User Interfaces (CUI)
Cuadra et al. [14]	My bad! repairing intelligent voice assistant errors improves interaction	2021	ACM Conference on Computer Supported Cooperative Work (CSCW)
Daher et al. [15]	Empathic chatbot response for medical assistance	2020	ACM International Conference on Intelligent Virtual Agents (IVA)
Diederich et al. [16]	Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: insights from an experiment with a conversational agent	2019	International Conference on Computers and Information Systems (ICIS)
Dubiel et al. [19]	Persuasive synthetic speech: Voice perception and user behaviour	2020	ACM Conference Conversational User Interfaces (CUI)
Elsholz et al. [20]	Exploring Language Style in Chatbots to Increase Perceived Product Value and User Engagement	2019	Conference on Human Information Interaction and Retrieval (CHIIR)
Fadhil et al. [21]	The effect of emojis when interacting with conversational interface assisted health coaching system	2018	EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)
Feijóo-García et al. [22]	Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations	2021	ACM International Conference on Intelligent Virtual Agents (IVA)
Gnewuch et al. [25]	Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction	2018	European Conference on Information Systems (ECIS)
Gnewuch et al. [26]	The Chatbot is typing…'–The Role of Typing Indicators in Human-Chatbot Interaction	2018	International Conference on Computers and Information Systems (ICIS)
Gnewuch et al. [27]	Opposing Effects of Response Time in Human-Chatbot Interaction	2022	Business & Information Systems Engineering (BISE)
Go et al. [28]	Conversational Agents Replying with a Manzai-style Joke	2021	Australian Conference on Human-Computer Interaction (OzCHI)
Haas et al. [30]	Keep it Short: A Comparison of Voice Assistants' Response Behavior	2022	ACM Conference on Human Factors in Computing Systems (CHI)
Habler et al. [31]	Effects of Smart Virtual Assistants' Gender and Language	2019	Conference on Message understanding (MuC)
Healey and Szostak [33]	Relating to speech evoked car personalities	2013	ACM Conference on Human Factors in Computing Systems (CHI)

Table 3 – continued from previous page

Authors	Title	Year	Venue
Hoegen et al. [35]	An End-to-End Conversational Style Matching Agent	2019	ACM International Conference on Intelligent Virtual Agents (IVA)
Hu et al. [38]	Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues	2021	ACM Conference on Human Factors in Computing Systems (CHI)
Hu et al. [39]	Polite or Direct? Conversation Design of a Smart Display for Older Adults Based on Politeness Theory	2022	ACM Conference on Human Factors in Computing Systems (CHI)
Huiyang and Min [40]	Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions	2022	International Journal of Human-Computer Interaction (IJHCI)
Jeong et al. [41]	Exploring effects of conversational fillers on user perception of conversational agents	2019	ACM Conference on Human Factors in Computing Systems (CHI)
Jestin et al. [42]	Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles	2022	ACM Conference Conversational User Interfaces (CUI)
Khooshabeh et al. [43]	Does it matter if a computer jokes	2011	ACM Conference on Human Factors in Computing Systems (CHI)
Kim et al. [45]	"I Can Feel Your Empathic Voice": Effects of Nonverbal Vocal Cues in Voice User Interface	2020	ACM Conference on Human Factors in Computing Systems (CHI)
Kim et al. [46]	Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality	2019	ACM Conference on Human Factors in Computing Systems (CHI)
Kraus et al. [51]	Effects of proactive dialogue strategies on human-computer trust	2020	ACM Conference on User Modeling, Adaptation and Personalization (UMAP)
Lee et al. [53]	What's on Your Virtual Mind? Mind Perception in Human-Agent Negotiations	2019	ACM International Conference on Intelligent Virtual Agents (IVA)
Lee et al. [54]	"I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot	2020	ACM Conference on Human Factors in Computing Systems (CHI)
Linnemann and Jucks [55]	'Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?'—Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction	2018	Interacting with Computers (IWC)
Lubis et al. [56]	Positive emotion elicitation in chat-based dialogue systems	2019	IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)
Lubold et al. [57]	Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion	2016	ACM/IEEE International Conference on Human-Robot Interaction (HRI)
Ma et al. [58]	Ask a Further Question or Give a List? How Should Conversational Agents Reply to Users' Uncertain Queries	2022	International Journal of Human-Computer Interaction (IJHCI)
Miehle et al. [62]	Exploring the impact of elaborateness and indirectness on user satisfaction in a spoken dialogue system	2018	ACM Conference on User Modeling, Adaptation and Personalization (UMAP)
Misu et al. [63]	Toward construction of spoken dialogue system that evokes users' spontaneous backchannels	2011	SIGdial Workshop on Discourse and Dialogue (SIGDIAL)
Miyamoto et al. [64]	Improving Relationships Based on Positive Politeness Between Humans and Life-Like Agents	2017	International Conference on Human-Agent Interaction (HAI)
Moilanen et al. [66]	Measuring the Effect of Mental Health Chatbot Personality on User Engagement	2022	International Conference on Mobile and Ubiquitous Multimedia (MUM)

Table 3 – continued from previous page

Authors	Title	Year	Venue
Niewiadomski et al. [69]	Laugh-aware virtual agent and its impact on user amusement	2013	International Conference on Autonomous Agents and Multiagent Systems (AAMAS)
Ouchi et al. [70]	Should speech dialogue system use honorific expression?	2019	International Conference on Human-Agent Interaction (HAI)
Roy et al. [75]	Users, Tasks, and Conversational Agents: A Personality Study	2021	International Conference on Human-Agent Interaction (HAI)
Seeger and Heinzl [77]	Chatbots often Fail! Can Anthropomorphic Design Mitigate Trust Loss in Conversational Agents for Customer Service?	2021	European Conference on Information Systems (ECIS)
Spillner and Wenig [78]	Talk to Me on My Level-Linguistic Alignment for Chatbots	2021	International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)
Tolmeijer et al. [80]	Female by default?–exploring the effect of voice assistant gender and pitch on trait and trust attribution	2021	ACM Conference on Human Factors in Computing Systems (CHI)
Völkel and Kaya [82]	Examining User Preference for Agreeableness in Chatbots	2021	ACM Conference Conversational User Interfaces (CUI)
Völkel et al. [83]	Manipulating and Evaluating Levels of Personality Perceptions of Voice Assistants through Enactment-Based Dialogue Design	2021	ACM Conference Conversational User Interfaces (CUI)
Völkel et al. [84]	User perceptions of extraversion in chatbots after repeated use	2022	ACM Conference on Human Factors in Computing Systems (CHI)
Wester et al. [85]	Artificial personality and disfluency	2015	INTERSPEECH
Westerman et al. [86]	I Believe in a Thing Called Bot: Perceptions of the Humanness of "Chatbots"	2019	Communication Studies
Wilhelm et al. [88]	Keep on Smiling: An Investigation of the Influence of the Use of Emoticons by Chatbots on User Satisfaction	2022	ACM Conference Conversational User Interfaces (CUI)
Xiao et al. [89]	Let Me Ask You This: How Can a Voice Assistant Elicit Explicit User Feedback?	2021	ACM Conference on Computer Supported Cooperative Work (CSCW)
Yang et al. [90]	The Effect of the Repetitive Utterances Complexity on User's Perceived Empathy and Desire to Continue Dialogue by a Chat-oriented Dialogue System	2021	International Conference on Human-Agent Interaction (HAI)
Yang et al. [91]	Perceived Emotional Intelligence in Virtual Agents	2017	ACM Conference on Human Factors in Computing Systems (CHI)
Zhu et al. [94]	Effects of Emotional Expressiveness on Voice Chatbot Interactions	2022	ACM Conference Conversational User Interfaces (CUI)