



Stepped Warm-Up—The Progressive Interaction Approach for Human-Robot Interaction in Public

Min Zhao¹(✉), Dan Li¹, Zhun Wu¹, Shiyan Li¹, Xiaonan Zhang²,
Lu Ye², Guangfu Zhou², and Daisong Guan¹

¹ Baidu AI Interaction Design Lab, Beijing, China
zhaomin04@baidu.com

² Baidu Natural Language Processing Department, Beijing, China

Abstract. Initiating interaction is one of the most basic functions of service robots, and it has a vital impact on the subsequent interaction process. In the present study, we proposed a brand-new approach to initiate interaction—the progressive interaction approach. Specifically, robots actively send social cues to potential users in a progressively enhancing manner. Based on the concept of this approach, we modeled the behavior of a robot named Xiaodu, and further validated the practical benefits of this approach in an experimental study, in which participants were asked to rate their experience after interacting with Xiaodu with different initiating strategies. The findings suggested that compared to the reactive approach, the progressive interaction approach led to stronger positive emotions (self-reported) and was perceived to be more natural and friendly. Participants also reported higher affection and higher interaction intention towards the progressive interaction approach. The study has some implications for designing robots' behavior in the interaction initiating process.

Keywords: Human-Robot Interaction (HRI) · Initiating interaction · Progressive interaction approach · Reactive approach · Facial expression · Face recognition · Attracting attention

1 Introduction

There are increasing amount of service robots being developed and applied in various public places, such as in museums [1, 2], airports [3], train stations [4], and shopping malls [5]. These service robots are expected to interact with people and act as receptionists or information staff. Initiating interaction is one of the most basic functions of these service robots [6], and it has a vital impact on the subsequent interaction process [7].

However, there are still many remaining issues existing for robots to initiate interaction in public places. Currently, there are mainly two approaches for service robots to initiate interaction. The first is called “the reactive approach”, in which a robot waits until a user to initiate interaction [8–10]. Robots who adopt this initiating strategy usually use certain behavior to exhibit their availability and reciprocity [11], and to encourage users to initiate interactions [12, 13, 15]. This approach is resistant to many complex issues [14], nevertheless, it will miss out on potential users who are hesitating

or uncertain about how to interact with robots. The other is “the proactive approach”, in which a robot proactively seeks for people who need help [6, 8]. Compared to the reactive approach, this approach is more initiative [16] and thus more likely to help those with potential needs but do not know they can turn to robots [6]. Besides, this approach is also useful for advertisement purposes. Yet, being too proactive, it may be perceived as “rude”, “disturbing”, or even “annoying” [8, 15]. In summarize, neither of the two approaches are natural enough in practical applications: the reactive approach is too passive, while the proactive approach may seem intrusive and cause annoyance.

The aim of this study was to design a more natural approach for robots to initiate interaction. We expected that with this approach, even those who had never encountered robots before, could interact with them in a natural, efficient, and pleasant way. Our approach was mainly based on behavior patterns revealed in human-human interactions, as they were established through long social practices and could be accepted by the majority of people without further explanations. Specifically, we designed a range of behaviors of a robot under the guidance of the approach. Moreover, an experimental study was conducted to validate the practical benefits of the approach.

2 Related Work

2.1 Initiating Interaction Between Human-Human in Public Places

The interaction initiating process is extremely delicate, and requires cooperation [17–19]. Moreover, it is a dynamic process. In the process of initiating interaction, people will use various social cues, which are understandable to all participants of the interaction [20].

Hall [21] suggests a series of steps in the process of initiating interaction: (a) getting a person’s attention; (b) assessing that person’s willingness to interact; (c) creating physical proximity to enable interaction.

One key issue in initiating interaction in public places is to recognize the intention to interact. Goffman [20] points out that in public places, an “encounter” is normally initiated by one person, using specific eye contact (a brief glance) or certain body language and posture. The interaction is not considered as officially started until the cues sent by the initiator are recognized by the receiver. In return, the receiver usually sends back eye contact (e.g., gaze) or body language as recognition. Specifically, gaze, among other kinds of eye contact, is a vital social cue for initiating social contact in public places, as it is one of the most directional social cues to express the intention to interact [22, 23].

As suggested by Hall [21], once the interaction intention is recognized by mutual parties, the next step is to create physical proximity to enable interaction. Hall [24] further proposes the concept of “proxemics”, which refers to the physical distance and/or closeness between people. He defines four kinds of personal spatial zones from a relatively close to far distance: “Intimate Zone”, “Personal Zone”, “Social Zone”, “Public Zone”. The widely accepted personal spatial zone was summarized by Lambert [25] in Table 1. Moreover, it appears that distance itself is not only a common social

cue that can reflect and influence social relationships and attitudes [26, 27], but it can also affect one’s ability to see or touch the other person, and thus play an important role in the use of other social cues. For instance, Kendon [28] suggests that friends usually exchange greetings twice, first using body language at a far distance and again by smiling at a closer distance.

Table 1. Human-human personal spatial zones (cf. Lambert)

Personal spatial zone	Range	Situation
Close intimate	0 to 0.15 m	Lover or close friend touching
Intimate zone	0.15 m to 0.45 m	Lover or close friend only
Personal zone	0.45 m to 1.2 m	Conversation between friends
Social zone	1.2 m to 3.6 m	Conversation to non-friends
Public zone	3.6 m +	Public speech making

2.2 Initiate Interaction Between Human-Robot in Public Places

In general, robots are expected to obey common social norms when they are starting and maintaining communication with humans [29]. Most people are willing to try to interact with robots if the robot showed appropriate social behavior [30].

“How to attract users’ attention” is a popular research topic. Studies find that robots with humanoid bodies are more likely to attract users’ attention [31], and robots with face can attract significantly more users to stop [32]. Robots can also catch users’ eyes by moving their heads or blinking their eyes. However, when users are engaged in other things (e.g., watching news on television), speech turns out to be the best way to attract their attention, followed by waving gesture and eye LED blinking gesture, while attempting to build eye-contact works the worst in this kind of scenarios [33]. Gaze behavior is only valid when the robot has users’ attention already [33]. Some researchers designed different possible behavior including head motions (e.g., looking at the nearest person), different facial expressions (e.g., happy, sad, and angry) and different language acts (e.g., come here, do you like robots?). After self-learning from interactions with humans, the robot developed a “positive attitude” to attract attention, that is, saying nice things, looking at people, and smiling [34].

Gaze behavior is still an effective social cue in human-robot interaction, given that the robot has attracted users’ attention already. Mutlu and colleagues [35] applied key gaze mechanisms in human conversations on robots, and found that humanlike gaze mechanisms successfully helped the robot signal different participant roles, manage turn-exchanges, and shaped how participants perceive the robot and the conversation. Besides, other studies suggest that prolonged eye-contact from the robot while approaching or being approached by users can increase the user’s affection towards the robot [36].

“Human-Robot Proxemics” has been another focus of research. It is suggested that when approaching robots, users are likely to keep robots at personal or social zones, same as what they will behave when they are approaching another human-being [30, 37–39]. Similarly, when being approached by robots, users will feel more

comfortable when robots stopped at the personal and social zones [40, 41]. If robots were too close, users will even step back to make themselves feel more comfortable. Besides, users prefer to be approached from the front instead of from the back, and front-left or front-right is preferred than the direct front-on direction [42–44].

Although there have been abundant findings accumulating in the field of robots' social behavior, few studies have focused on the whole dynamic process of "initiating interaction". So far, studies towards initiating interaction between human-robot have been mainly concentrated on building algorithms, such as how to detect humans and track their positions, how to determine intention and interests of humans, and how to recognize humans etc., or simply on one part of the initiating process, such as how to attract users' attention. The present study focused on the whole process of "initiating interaction", and further explored how these designed behaviors based on the progressive interaction approach would be perceived by users.

3 The Progressive Interaction

3.1 The Approach of "the Progressive Interaction"

Based on analyses of human behavior and our previous investigations, we suggest that during the process of human-robot interaction in public places, robots are expected to send interaction signals more actively, and in a progressively enhancing manner. This is what we call "the progressive interaction".

Specifically, as humans and robots getting closer, humans will have different expectations towards robots. We divided the process into three stages, which are named as "far field", "mid field", and "near field", based on the order they will appear in users' mental world. Every stage is corresponding to certain distance ranges in physical world, and there are certain behaviors that will fit in users' expectations for every stage:

- **Far Field:** The aim of robots in this stage is to gain users' attention, and to make users' aware that "I'm noticed by HIM/HER". This is a critical stage, as if it failed, the following human-robot interaction would seem to be abrupt or even impossible to make. This field is corresponding to a distance ranging approximately from 2.7 to 4.2 m. Robots are expected to use facial expressions and body movements to attract attention. Such as smiling, friendly eye contact, waving, tilting head to one side, nodding, and so on;
- **Mid Field:** Robots are supposed to further express "the intention to interact", and make users clearly aware that "I'm the only one in HIS/HER eyes" and HE/SHE is intended to further interact. This will also encourage users to approach the robot spontaneously. The distance corresponding to this field is about 1.2–2.7 m. Within this distance, users expect the robot to send out interactive signals using a variety of combinations, so that the user can be certain that he/she is the target. For example, using voice to send greetings (such as good morning, hello), at the same time, smiling, or waving, etc.;
- **Near Field:** Robots need to "start a dialogue" first, and will be perceived as initiative and friendly. "HE/SHE is 'Liao' (hitting on) me", and with this impression, the dialogue between human and robot develops naturally. The distance

corresponding to the near field is within 1.2 m. Within this field, the role of language is highlighted. Users expect the robot to “initiate dialogue” first, such as introducing himself/herself and asking if any help was needed. At the same time, users expect more enthusiastic smiles and body movements (such as shaking hands, hugging, etc.).

The distance corresponding to each field was established in a previous study we carried out, in which participants ($N = 32$) were asked to approach a robot from certain direction and with certain pace that he/she was comfortable with. While approaching, they were asked to evaluate where were the appropriate points for the robot to attract attention (far field), express interaction intention (mid field), and start a dialogue (far field). The specific distance was measured and calculated afterwards.

3.2 Designing the Progressive Interaction

We applied the approach, the progressive interaction, on a robot called Xiaodu. The focus of the present study was listed in Table 2.

Table 2. Designing the progressive interaction

Stages	Purpose	Behavior
Far field	Attracting attention	Facial expressions
Mid field	Initiating an interactive request	Voice, facial expressions
Near filed	Starting a dialogue	Voice, facial expressions

The Robot “Xiaodu”. Xiaodu is a “formal employee” in Baidu company, 160 cm tall, 110 cm width, and works as a receptionist in the company hall (see Fig. 1). Xiaodu is benefited from the AI techniques (e.g., NLP, dialogue system, speech recognition) of Baidu, and is able to communicate smoothly with users in multiple aspects, such as communicating emotions, providing information and other services.







Fig. 1. Xiaodu at work

Facial Expressions in Far Field. The aim of this stage is to attract users’ attention. Facial expressions of Xiaodu need to be perceived and understood by users. In return, users will be willing to pay attention to Xiaodu, and even further interact.

Six groups of facial expressions are designed by our UX designers and researchers, based on the features of facial expressions in human-human interactions, facial expression database of Xiaodu, and popular stickers used in online chatting. Participants were invited to evaluate these expressions from far field, regarding their understanding, affection, at what degree they were attracted, and the intention to further interact. The expression of “Raised eyebrows” was selected (as shown in Table 3).

Table 3. Facial expressions of Xiaodu for different fields

The default expression: expressionless	Far field: raised eyebrows	Mid field: smiling eyes	Near field: smiling eyes with heart-shaped blush
			

*Face expressions were animated

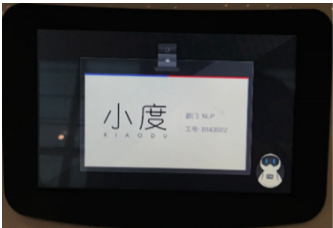

Facial Expression and Voice in Mid Field. The core in mid field is to “initiate an interaction request”. For this stage, users need to be able to perceive the robot’s intention to interact. After the same process as in far field stage, the expression “smiling eyes” was selected for the mid field stage (see Table 3). As for voice, the classic way of “greeting” was selected—“Hi, how are you”.

In addition, the dual-modality design (“smiling eyes + greeting”) was favored by participants, comparing to merely smiling eyes. The former design left participants an impression that the robot was more enthusiastic, more directional, and more interactive. We further investigated whether the effect of voice was also significant after the whole interaction process in later validation experiment.

Facial Expressions, Voice and Face Recognition in Near Field. The aim for this stage is to “start a dialogue”. “Smiling eyes with heart-shaped blush” was designed and selected for this stage (as shown in Table 3). As for voice, Xiaodu introduced himself/herself and then the rules for subsequent interaction—“I’m Xiaodu. If you want to chat you can just speak to me, I’m listening”.

During the process of user evaluation, we found that participants paid a lot of attention to the screen in the front of Xiaodu and would spontaneously keep eyes on the screen. Xiaodu was capable of face recognition, thus, we investigated whether displaying the face of the user would improve his/her experience in later experiment (as shown in Table 4).

Table 4. The screen of Xiaodu in far field

The default display	Face recognition
	

4 Experiment Validation of the Progressive Interaction Approach

4.1 Objectives

The study aimed to investigate whether there was any difference in user experience between the progressive interaction approach and traditional initiating interaction approach (the reactive approach); and the second aim was to explore the difference in user experience among different designs in the progressive interaction approach.

4.2 Design

The experiment adopted a within-subject design, and all participants were required to experience all 5 trials. After each trial, participants were asked to fill in questionnaires regarding their emotions and attitudes.

One of the trials was the control condition, in which Xiaodu adopted the reactive approach. Specifically, Xiaodu kept the default facial expression, and participants could walk in the front of Xiaodu, touch the screen, and he/she would be reminded to pick up the microphone to start a conversation.

The other four trials were the experiment condition. They were based on the progressive interaction approach. Specifically, a 2 (*Mid field: with/without voice*) \times 2 (*Near field: with/without face recognition*) design was used for the experiment condition (as shown in Fig. 2).

To minimize the influence of learning effect and fatigue effect, we randomized the order of trials for each participant.

The present study only focused on the initiating process, and later human-robot dialogue was not included in this study. Thus, to avoid the impacts of the robot's responses on participants' emotions and attitudes, all participants were required to ask the same question "How's the weather today".

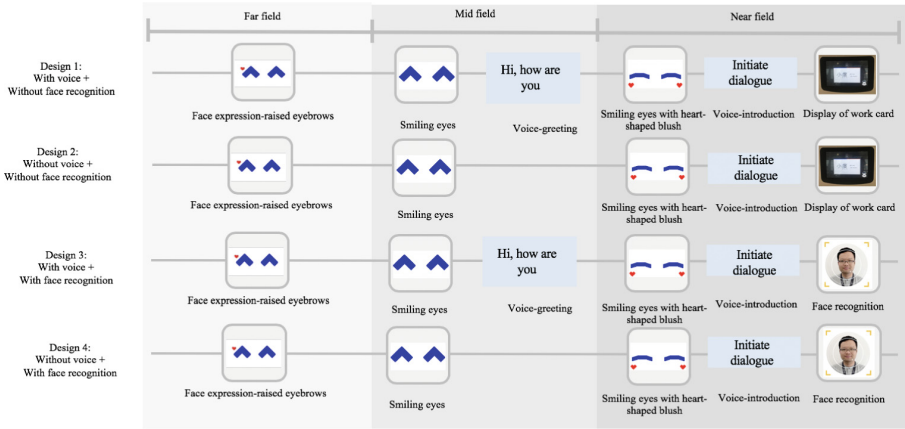


Fig. 2. Designs for the progressive interaction approach. Introduction in near field: I’m Xiaodu. If you want to chat you can just speak to me, I’m listening.

4.3 Participants

25 participants were recruited, including 12 males and 13 females. Their age ranged from 20 to 45 years old. 11 of them reported previous experience with service robots in public places. All participants reported normal or corrected to normal vision, and normal hearing.

All participants volunteered to participate in the study and agreed to make audio and video recordings of the research process. At the end of the study, all participants were given appropriate compensation.

4.4 Experiment Environment and Equipment

Experimental Set-Up. The experiment took place in the open office area of Baidu company. The whole experiment area was approximately 8 m in length, and 2.5 m in width. A white curtain was set up behind Xiaodu, and one experimenter recorded the facial expression of users’ behind the curtain (as shown in Fig. 3).

The Robot Xiaodu. As mentioned before, Xiaodu is a humanoid robot in Baidu company. In different trials, it would display different kinds of behavior.

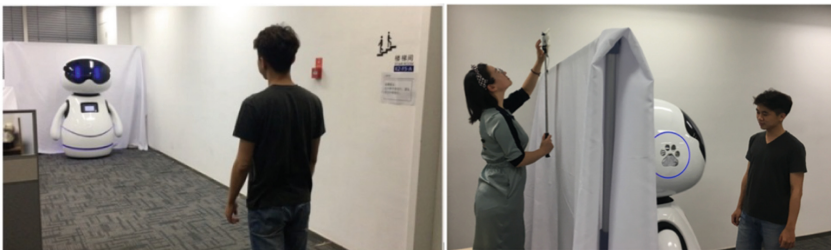


Fig. 3. The experimental set-up

To save the time and other resources needed to develop a fully autonomous robot, the experiment used the Wizard of Oz (WoZ) paradigm [45]. The facial expressions and voice of Xiaodu were actually controlled by an experimenter, who practiced multiple times to ensure consistency among different trials.

4.5 Measures

Emotions. Initiating interaction is a dynamic, subtle and transient process. In order to better evaluate user experience in this process, we included “emotion” as an important dependent variable, and it was measured with two criteria: objective face expression analysis and subjective self-report.

Objective Emotions. Noldus FaceReader [46] offers a relatively objective method to measure emotions through analyzing facial expressions, and it can automatically detect various emotions such as happiness, surprise and anger. For a given time period, face expressions in recordings will be analyzed, and different criteria will be generated, such as emotional arousal level, emotional valence.

Considering the emotions that participants might experience in the experiment, we included four kinds of emotions: happiness, surprise, neutral, and negative emotion. Specifically, negative emotion was the combination of sadness, anger, fear, and disgust. We adopted the criterion—emotional arousal level. Values for emotional arousal level range from 0–1, and higher values indicate stronger emotions.

Subjective Emotions. Participants were asked to rate their emotional intensity in an 8-point Likert questionnaire (from 0–7, higher scores indicated stronger emotional intensity), in which they were required to answer the question, “How did you feel in the process of interacting with Xiaodu just now”, repeatedly for different kinds of emotions. Five kinds of emotions were rated in the questionnaire, which is, happiness, surprise, confusion, disgust, and neutral. The questionnaire was filled in after every trial.

Attitudes. Aside from emotions, we were also interested in participants’ perceptions and cognitions towards Xiaodu. Thus, we also asked participants to evaluate the following attitudes, with one single question each. Participants were asked to rate in a 7-point Likert questionnaire (from 1–7, higher scores indicated stronger agreement):

- **Naturalness:** How natural do you think this interaction approach is?
- **Friendliness:** How friendly do you think this interaction approach is?
- **Affection:** How much do you like this interaction approach?
- **Interaction Intention:** If Xiaodu initiated interaction with you in this way, how much would you like to interact with him?

4.6 Procedure

Participants were first welcomed, and were informed about the aim and general process of the study. After introduction, there was a practice phase, in which participants could get familiar with the experimental set-up, the robot (those who encountered Xiaodu for the first time were usually excited), and find out the comfortable pace for him/her to

approach Xiaodu. Then, the experiment began. Participants were asked to imagine that this was the first time they encountered Xiaodu, and they were intended to check the weather through Xiaodu. All participants went through all five trials (one for reactive approach and four for the progressive interaction approach). Participants’ facial expressions during the experiment were recorded by one experimenter. After every trial, they filled out questionnaires about their emotions and attitudes during the trial. Finally, participants were interviewed for the reasons of their ratings after all trials.

4.7 Data Analysis

SPSS 23.0 was used for data analysis. First, we conducted descriptive analysis for all dependent variables. To evaluate the difference between the progressive interaction approach and the reactive approach, one-way repeated measures ANOVA was used. Further post hoc analysis was adopted if the ANOVA showed significant differences among the five trials.

Furthermore, we were interested in the effects of the two factors (“voice in mid field” and “face recognition in near field”) in the progressive interaction designs. Thus, we conducted the two-way repeated measures ANOVA to see whether voice and face recognition were preferred by participants.

5 Results

5.1 Objective Emotions

The objective emotions of all participants in different trials are shown in Table 5. Because of technique issues, only 21 of participants were included in this analysis. In general, participants exhibited a neural or happy state in their facial expressions of all trials.

One-way repeated measures ANOVA was used to investigate the differences among different trials. Mauchly’s test indicated that the assumption of sphericity had been violated except for the emotion happiness. Thus, we looked at the effects of trials after Green-house Geisser corrected (except for happiness). The findings suggested that there was no significant difference among different trials in all kinds of emotions (as shown in Table 6).

Table 5. Descriptive statistics of objective emotions

	N	Neutral		Happiness		Surprise		Negative emotion	
		M	SD	M	SD	M	SD	M	SD
Design 1	21	0.50	0.27	0.31	0.22	0.03	0.06	0.02	0.03
Design 2	21	0.48	0.28	0.30	0.23	0.05	0.07	0.03	0.02
Design 3	21	0.45	0.29	0.34	0.21	0.05	0.10	0.02	0.03
Design 4	21	0.48	0.26	0.32	0.21	0.06	0.11	0.02	0.02
The reactive approach	21	0.49	0.24	0.31	0.18	0.03	0.04	0.03	0.03

Design 1: With voice + Without face recognition; Design 2: Without voice + Without face recognition; Design 3: With voice + With face recognition; Design 4: Without voice + With face recognition

Table 6. The effects of trials on objective emotions

Emotions	Type	<i>df</i>	F	<i>p</i>
Neutral	Greece-house Geisser	2.776	0.753	0.516
Happiness	Sphericity assumed	4.000	0.483	0.748
Surprise	Greece-house Geisser	2.255	1.039	0.369
Negative emotion	Greece-house Geisser	2.336	0.595	0.581

Besides, we conducted two-way repeated measures ANOVA to investigate the different among different designs of the progressive interaction. Since both of the factors had two levels (with/without), the assumption of sphericity was considered as automatically meet. The effects of voice and face recognition, and their interaction effects are listed in Table 7. It turned out that there was no significant difference in four designs, nor is the interaction effect of the two factors significant.

Table 7. The effects of voice and face recognition on objective emotions

	Emotions	<i>df</i>	F	<i>p</i>
Voice	Neutral	1	0.049	0.827
	Happiness	1	0.596	0.449
	Surprise	1	0.937	0.345
	Negative emotion	1	0.059	0.811
Face recognition	Neutral	1	1.055	0.317
	Happiness	1	1.051	0.318
	Surprise	1	1.400	0.251
	Negative emotion	1	0.289	0.597
Voice * Face interaction	Neutral	1	1.350	0.259
	Happiness	1	0.032	0.859
	Surprise	1	0.227	0.639
	Negative emotion	1	0.700	0.413

5.2 Subjective Emotions

The ratings of different kinds of self-reported emotions in different trials are shown in Table 8. In general, participants reported more positive emotions (e.g., happiness) and less negative emotions (e.g., confusion) after the trials of the progressive interaction than after the trial of the reactive approach.

As with the objective emotions, one-way repeated measures ANOVA was used and the results are listed in Table 9. Trials had significant effects on self-reported happiness, surprise, and confusion. Thus, post hoc analysis was conducted to see the difference between any two of the five trials. Compared to the reactive approach, participants reported significantly higher scores in happiness, surprise, but lower scores in confusion for the progressive interaction approach (as shown in Table 10).

Table 8. Descriptive statistics of subjective emotions

	N	Happiness		Surprise		Confusion		Disgust		Neutral	
		M	SD	M	SD	M	SD	M	SD	M	SD
Design 1	25	4.12	2.60	2.60	2.80	0.64	1.52	0.00	0.00	0.40	1.38
Design 2	25	3.92	2.43	1.96	2.51	0.16	0.80	0.16	0.80	0.68	1.62
Design 3	25	4.32	2.38	2.48	2.80	0.88	1.17	0.16	0.80	0.00	0.00
Design 4	25	4.32	2.45	2.40	2.56	0.48	1.36	0.00	0.00	0.12	0.60
The reactive approach	25	0.36	1.32	0.76	1.93	5.00	2.14	0.44	1.58	0.72	1.72

Design 1: With voice + Without face recognition; Design 2: Without voice + Without face recognition; Design 3: With voice + With face recognition; Design 4: Without voice + With face recognition

Table 9. The effects of trials on subjective emotions

Emotions	Type	df	F	p
Happiness	Sphericity assumed	4.000	25.962***	0.000
Surprise	Greece-house Geisser	3.101	3.505*	0.018
Confusion	Sphericity assumed	4.000	50.030***	0.000
Disgust	Sphericity assumed	4.000	1.141	0.342
Neutral	Greece-house Geisser	2.424	1.668	0.192

*** Significant at the .001 level; * Significant at the .05 level.

Table 10. Post Hoc analysis. Paired comparisons between the reactive approach and other four designs respectively.

	With the reactive approach trial	Mean difference	p
Happiness	Design 1	-3.760***	0.000
	Design 2	-3.560***	0.000
	Design 3	-3.960***	0.000
	Design 4	-3.960***	0.000
Surprise	Design 1	-1.840**	0.006
	Design 2	-1.200	0.059
	Design 3	-1.720**	0.009
	Design 4	-1.640*	0.019
Confusion	Design 1	4.360***	0.000
	Design 2	4.840***	0.000
	Design 3	4.120***	0.000
	Design 4	4.520***	0.000

Design 1: With voice + Without face recognition; Design 2: Without voice + Without face recognition; Design 3: With voice + With face recognition; Design 4: Without voice + With face recognition

*** Significant at the .001 level; ** Significant at the .01 level;

* Significant at the .050 level.

Findings of two-way repeated measures ANOVA suggested that there was no significant difference in different designs of the progressive interaction approach (see Table 11).

Table 11. The effects of voice and face recognition on subjective emotions

	Emotions	df	F	p
Voice	Happiness	1	0.115	0.737
	Surprise	1	0.728	0.402
	Confusion	1	2.966	0.098
	Disgust	1	/	/
	Neutral	1	0.980	0.332
Face recognition	Happiness	1	0.651	0.428
	Surprise	1	0.204	0.656
	Confusion	1	1.205	0.283
	Disgust	1	/	/
	Neutral	1	3.868	0.061
Voice * Face interaction	Happiness	1	0.150	0.702
	Surprise	1	0.611	0.442
	Confusion	1	0.030	0.864
	Disgust	1	1.000	0.327
	Neutral	1	0.235	0.632

5.3 Attitudes: Naturalness, Friendliness, Affection and Interaction Intention

Descriptive statistics are listed in Table 12. Compared to the reactive approach, Xiaodu with the progressive interaction approach were generally considered as more natural and friendly, and participants also reported higher affection, higher interaction intention with.

Table 12. Descriptive statistics of attitudes

	N	Naturalness		Friendliness		Affection		Interaction intention	
		M	SD	M	SD	M	SD	M	SD
Design 1	25	4.71	1.37	5.75	1.07	5.42	0.97	5.67	1.01
Design 2	25	4.50	1.32	5.58	1.10	5.25	1.07	5.67	1.01
Design 3	25	4.79	1.18	5.63	1.06	5.46	1.02	5.63	1.06
Design 4	25	4.88	1.12	5.54	1.14	5.54	0.93	5.63	1.21
The reactive approach	25	2.76	1.33	3.24	1.59	2.84	1.21	2.92	1.66

Design 1: With voice + Without face recognition; Design 2: Without voice + Without face recognition; Design 3: With voice + With face recognition; Design 4: Without voice + With face recognition

Again, we conducted one-way repeated measures ANOVA to see if the differences of different approaches were significant, and found that there was significant difference in all four aspects of attitudes (see Table 13). Thus, post hoc analysis was conducted, and the results are listed in Table 14.

Table 13. The effects of trials on attitudes

Attitudes	Type	df	F	p
Naturalness	Sphericity Assumed	4.000	21.239***	0.000
Friendliness	Greece-house Geisser	1.590	39.338***	0.000
Affection	Greece-house Geisser	2.295	54.649***	0.000
Interaction intention	Greece-house Geisser	2.108	38.107***	0.000

*** Significant at the .001 level.

Table 14. Post Hoc analysis. Paired comparisons between the reactive condition and other four designs respectively.

	With the reactive trial	Mean difference	p
Naturalness	Design 1	-1.917***	0.000
	Design 2	-1.708***	0.000
	Design 3	-2.000***	0.000
	Design 4	-2.083***	0.000
Friendliness	Design 1	-2.583***	0.000
	Design 2	-2.417***	0.000
	Design 3	-2.458***	0.000
	Design 4	-2.375***	0.000
Affection	Design 1	-2.583***	0.000
	Design 2	-2.417***	0.000
	Design 3	-2.625***	0.000
	Design 4	-2.708***	0.000
Interaction intention	Design 1	-2.667***	0.000
	Design 2	-2.667***	0.000
	Design 3	-2.625***	0.000
	Design 4	-2.625***	0.000

Design 1: With voice + Without face recognition; Design 2: Without voice + Without face recognition; Design 3: With voice + With face recognition; Design 4: Without voice + With face recognition

*** Significant at the .001 level.

In addition, two-factor repeated measures ANOVA was used to see the effect of two factors (voice/face recognition) in different progressive interaction approach designs. Results showed no significant difference of the two factors in all four aspects (see Table 15).

Table 15. The effects of voice and face recognition on attitudes

	Emotions	<i>df</i>	F	<i>p</i>
Voice	Naturalness	1	0.185	0.671
	Friendliness	1	1.131	0.299
	Affection	1	0.193	0.664
	Interaction intention	1	0.000	1.000
Face recognition	Naturalness	1	1.184	0.288
	Friendliness	1	0.561	0.461
	Affection	1	1.484	0.236
	Interaction intention	1	0.058	0.811
Voice * Face interaction	Naturalness	1	0.723	0.404
	Friendliness	1	0.489	0.491
	Affection	1	1.000	0.328
	Interaction intention	1	0.000	1.000

6 Discussion

In the present study, we compared two approaches to initiate interaction: the progressive interaction approach and the reactive approach. Our main findings were that the progressive interaction approach resulted in more positive self-reported emotions, and was perceived to be more natural and friendlier. Moreover, participants reported higher affection and higher interaction intention towards the progressive interaction approach. However, no difference was found in objective emotions of the two approaches, nor in the four designs of the progressive interaction approach.

Participants reported more positive emotions and attitudes towards the progressive interaction approach than the reactive approach. The findings were also confirmed in the interview after the experiment. We found that participants successfully received the robot's intention to interact through signals the robot sent, such as facial expressions, greetings, and self-introductions. With these signals, participants started interacting with the robot naturally. On the contrary, participants were confused about "how to initiate dialogue" with the robot in the reactive approach condition.

Interestingly, no significant difference was found in the objective emotions between the progressive interaction approach and the reactive approach, and in all trials, participants mainly exhibited neutral and positive facial expressions. Culture might be one of the influencing factors of this finding. Studies suggested that compared to Westerners, Easterners were more likely to control their facial expressions, since they paid more attention to the appropriateness of expressing emotions [47]. Moreover, it was common for Chinese to use smiles or laughter to cover up negative emotions, unless in front of close others [48]. Thus, it was possible that during the experiment, participants used awkward smiles to cover up their confusion and awkwardness, which were recognized as positive emotions by the Noldus FaceReader. The assumption was consistent with experimenters' observations. Moreover, participants also reported that they were confused about "how to initiate dialogue" with the robot in reactive approach condition.

There was no significant difference in four designs of the progressive interaction approach in all subjective evaluations. In general, participants had positive attitudes and emotions towards the progressive interaction designs, which suggested that these four designs were all relatively natural and friendly. However, it was worth noting that some participants expressed their concerns towards face recognition. 48% of participants reported that they did not like that their faces were shown on the screen, while only 16% of participants clearly indicated the opposite attitude. Reasons for dislikes were mainly: (1) Privacy issues, they fear that the image information might be recorded, analyzed and stored, or even be used for other purposes; (2) The image itself, the image captured by Xiaodu was not ideal because of wrong angles, looked fat or ugly etc.; (3) Interaction was not natural, as participants needed to switch between Xiaodu's face and the screen back and forth, participants could not focus on Xiaodu's face as in daily interaction. Besides, it felt awkward to communicate with his/her own image presented on the screen. On the other hand, those who liked face recognition reported that (1) High-tech feeling, face recognition made Xiaodu seemed more intelligent; (2) More directional; participants felt being seen by Xiaodu and was certain that him/herself was the target for further interaction. Thus, we suggested that face recognition should be used with caution, especially in the process of initiating interaction.

There were several limitations of this study. One was that the experiment was conducted in the constrained open office area, which was not necessarily considered as public places. A more ideal environment would be in the hall of Baidu building. Moreover, this was a very targeted study, and thus its' generalizability needs to be considered when applying the findings. Specifically, a humanoid robot was used who was often perceived to be adorable. The characters of the robot itself may have some positive impacts on users' emotions and attitudes. Moreover, our definitions of the far/mid/near fields and expectations corresponded were targeted at Chinese users. Not to mention the role Xiaodu played when we were putting the progressive interaction into concrete behavior patterns. Thus, the advantages of the progressive interaction need to be validated in other types of robots and in other cultural backgrounds. Another limitation was that we didn't include the proactive approach into comparisons in the present study, which we would be interested to investigate in the future.

There are still many interesting topics that worth investigating in the field, which we would also be dedicated to in the future. For example, how to design body language in different fields; how to make the greetings and introductions more natural and diverse.

7 Conclusions

The present study focused on service robots in public places, and proposed a brand-new approach—the progressive interaction approach, for robots to initiate interaction. Furthermore, the approach was preliminarily validated by an experimental study, in which it was compared to a relatively traditional approach, the reactive approach. Specifically, we found that: (1) compared to the reactive approach, the progressive interaction approach led to more positive emotions, and was perceived to be more

natural and friendly. Participants also reported higher affection, higher interaction intention towards the progressive interaction approach; (2) There was no significant difference among the four designs of the progressive interaction approach; (3) During the process of initiating interaction, face recognition did not cause more positive experience but revealed many concerns from users. Thus, whether to use face recognition or not in application should be considered cautiously. In conclusion, our study enriched the understanding of the human-robot interaction, and made a step forward in designing a natural and friendly human-robot interaction process.

References

1. Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N.: Interactive humanoid robots for a science museum. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006, pp. 305–312. IEEE, New York (2006)
2. Thrun, S., et al.: MINERVA: a second-generation museum tour-guide robot. In: Proceedings of 1999 IEEE International Conference on Robotics and Automation, ICRA 1999, pp. 1999–2005. IEEE, New York (1999)
3. Ramírez, O., Khambhaita, H., Chatila, R., Chetouani, M., Alami, R.: Robots learning how and where to approach people. In: Proceedings of 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 347–353. IEEE, New York (2016)
4. Hayashi, K., et al.: Humanoid robots as a passive-social medium: a field experiment at a train station. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, HRI 2007, pp. 137–144. IEEE, New York (2007)
5. Kanda, T., Glas, D.F., Shiomi, M., Ishiguro, H., Hagita, N.: Who will be the customer?: A social robot that anticipates people’s behavior from their trajectories. In: Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp 2008, pp. 380–389. ACM, New York (2008)
6. Satake, S., Kanda, T., Glas, D., Imai, M., Ishiguro, H., Hagita, N.: How to approach humans? Strategies for social robots to initiate interaction. In: Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2009, pp. 109–116. IEEE, New York (2009)
7. Mead, R., Atrash, A., Matarić, M.J.: Automated proxemic feature extraction and behavior recognition: applications in human-robot interaction. *Int. J. Soc. Robot.* **5**(3), 367–378 (2013)
8. Kato, Y., Kanda, T., Ishiguro, H.: May I help you?: Design of human-like polite approaching behavior. In: Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2015, pp. 35–42. ACM, New York (2015)
9. Rashed, M.G.: Observing people’s behaviors in public spaces for initiating proactive human-robot interaction by social robots. Ph.D. Dissertation (2016)
10. Rashed, M.G., Das, D., Kobayashi, Y., Kuno, Y.: A study on proactive methods for initiating interaction with human by social robots. *Asian J. Conver. Technol.* **4**(2) (2018). ISSN: 2350-1146
11. Yamazaki, K., Kawashima, M.: Prior-to-request and request behaviors within elderly day care: implications for developing service robots for use in multiparty settings. In: Bannon, L. J., Wagner, I., Gutwin, C., Harper, R.H.R., Schmidt, K. (eds.) ECSCW 2007, pp. 61–78. Springer, Heidelberg (2007). https://doi.org/10.1007/978-1-84800-031-5_4

12. Schulte, J., Rosenberg, C., Thrun, S.: Spontaneous, short-term interaction with mobile robots. In: Proceedings of 1999 IEEE International Conference on Robotics and Automation, ICRA 1999, pp. 658–663. IEEE, New York (1999)
13. Hayashi, K., et al.: Humanoid robots as a passive-social medium - a field experiment at a train station. In: Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, HRI 2007, pp. 137–144. IEEE, New York (2007)
14. Peters, C.: Evaluating perception of interaction initiation in virtual environments using humanoid agents. In: Proceedings of the 17th European Conference on Artificial Intelligence, pp. 46–50. IOS Press, Amsterdam (2006)
15. Finke, M., Koay, K., Dautenhahn, K., Nehaniv, C.: Hey, I'm over here – How can a robot attract people's attention? In: Proceedings of 2005 IEEE International Workshop on Robot and Human Interactive Communication, pp. 7–12. IEEE, New York (2005)
16. Buss, M., et al.: Towards proactive human-robot interaction in human environments. In: Proceedings of the 2nd International Conference on Cognitive Infocommunications, pp. 1–6. IEEE, New York (2011)
17. Schefflen, A.E.: *Body Language and Social Order*. Prentice-Hall, Englewood Cliffs (1972)
18. Kendon, A.: *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, Cambridge (1990)
19. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)
20. Goffman, E.: *Behavior in Public Place: Notes on the Social Organization of Gatherings*. The Free Press, New York (1963)
21. Hall, E.T.: *The Hidden Dimension*. Anchor Books, New York (1969)
22. Argyle, M., Cook, M.: *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge (1976)
23. Kampe, K.K.W., Frith, C.D., Frith, U.: “Hey John”: signals conveying communicative intention toward the self activate brain regions associated with “mentalizing”, regardless of modality. *J. Neurosci.* **23**(12), 5258–5263 (2003)
24. Hall, E.T.: A system for the notation of proxemic behavior. *Am. Anthropol.* **65**(5), 1003–1026 (1963)
25. Lambert, D.: *Body Language*. Harper Collins, London (2004)
26. Hall, E.T.: Proxemics. *Curr. Anthropol.* **9**(2–3), 83–108 (1968)
27. Sommer, R.: *Personal Space: The Behavioral Basis of Design*. Prentice Hall, Englewood Cliffs (1969)
28. Kendon, A.: Features of the structural analysis of human communication behavior. In: von Raffler-Engel, W. (ed.) *Aspects of Nonverbal Communication*, pp. 29–43 (1980)
29. Dautenhahn, K.: Socially intelligent robots: dimensions of human robot interaction. *Philos. Trans. R. Soc. B Biol. Sci.* **362**(1480), 679–704 (2007)
30. Breazeal, C.: Social interactions in HRI: the robot view. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **34**(2), 181–186 (2004)
31. Imai, M., Ono T., Ishiguro, H.: Physical relation and expression: joint attention for human-robot interaction. In: Proceedings of the 10th IEEE International Workshop on Robot and Human Communication, RO-MAN 2001, pp. 512–517. IEEE, New York (2001)
32. Bruce, A., Nourbakhsh, I., Simmons, R.: The role of expressiveness and attention in human-robot interaction. In: Proceedings of the 2002 IEEE International Conference on Robotics and Automation, pp. 4138–4142. IEEE, New York (2011)
33. Torta, E., van Heumen, J., Cuijpers, Raymond H., Juola, J.F.: How can a robot attract the attention of its human partner? A comparative study over different modalities for attracting attention. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) *ICSR 2012. LNCS (LNAI)*, vol. 7621, pp. 288–297. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34103-8_29

34. Thrum, S., et al.: Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *Int. J. Robot. Res.* **19**(11), 972–999 (2000)
35. Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., Ishiguro, H.: Conversational gaze mechanisms for humanlike robots. *J. ACM Trans. Interact. Intell. Syst.* **1**(2), 12:1–12:33 (2012)
36. Mead, R., Mataric, M.J.: Robot have need too: how and why people adapt their proxemic behavior to improve robot social signal understanding. *J. Hum. Robot. Interact.* **5**(2), 48–68 (2016)
37. Huttenrauch, H., Eklundh, K.S., Green, A., Topp, E.: Investigating spatial relationships in human-robot Interaction. In: *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5052–5059. IEEE, New York (2006)
38. Silvera-Tawil, D., Rye, D., Velonaki, M.: Human-robot interaction with humanoid Diamandini using an open experimentation method. In: *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pp. 425–430. IEEE, New York (2015)
39. Walters, M.L., et al.: Close encounters: spatial distances between people and a robot of mechanistic appearance. In: *Proceedings of 5th IEEE-RAS International Conference on Humanoid Robots*, pp. 450–455. IEEE, New York (2005)
40. Walters, M.L., Dautenhahn, K., Woods, S.N., Koay, K.L., Te Boekhorst, R., Lee, D.: Exploratory studies on social spaces between humans and a mechanical-looking robot. *Connect. Sci.* **18**(4), 429–439 (2006)
41. Sardar, A., Joosse, M., Weiss, A., Evers, V.: Don't stand so close to me: users' attitudinal and behavioral responses to personal space invasion by robots. In: *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2012*, pp. 229–230. IEEE, New York (2012)
42. Dautenhahn, K., et al.: How may I serve you?: A robot companion approaching a seated person in a helping context. In: *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006*, pp. 172–179. IEEE, New York (2006)
43. Walters, M.L., Dautenhahn, K., Woods, S.N., Koay, K.L.: Robotic etiquette: results from user studies involving a fetch and carry task. In: *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, HRI 2007*, pp. 317–324. IEEE, New York (2007)
44. Ball, A., Rye, D., Silvera-Tawil, D., Velonaki, M.: Group vs. individual comfort when a robot approaches. In: Tapus, A., André, E., Martin, J.C., Ferland, F., Ammi, M. (eds.) *ICSR 2015. LNCS (LNAI)*, vol. 9388, pp. 41–50. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25554-5_5
45. Steinfeld, A., Jenkins O.C., Scassellati, B.: The Oz of Wizard: simulating the human for interaction research. In: *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2009*, pp. 101–107. IEEE, New York (2009)
46. Noldus Official Website. <https://www.noldus.com/human-behavior-research/products/facereader>
47. Hutchison, A., Gerstein, L.: Emotion recognition, emotion expression, and cultural display rules: implications for counseling. *J. Asia Pac. Couns.* **7**(1), 19–35 (2017)
48. Yan, W.: Facial expressions and differences between Chinese and Western cultures. *Wen Shi Zi Liao* **9**(1), 97–98 (2010)