Effects of Language Familiarity in Simulated Natural Dialogue with a Virtual Crowd of Digital Humans on Emotion Contagion in Virtual Reality

Matias Volonte*
Clemson University

Chang-Chun Wang †
National Chiao Tung University

Elham Ebrahimi‡
University of North Carolina Wilmington

Yu-Chun Hsu §
National Chiao Tung University

Kuan-Yu Liu ¶
National Chiao Tung University

Sai-Keung Wong ‖
National Chiao Tung University
National Yang-Ming Chiao Tung University

Sabarish V. Babu **
Clemson University

Abstract

This investigation compared the emotional impact caused by a crowd of affective virtual humans (VHs) that communicated in the users' native or foreign language. We evaluated the users' affective reactions to a crowd of VHs that exhibited distinct emotional expressions. A total of four emotions were presented, which were Positive, Negative, Neutral, and Mixed. The VHs performed verbal and non-behaviors accordingly. Under the Mixed condition, the VHs were divided into three groups equally and each group was uniquely assigned one of the three emotions (i.e., positive, negative, and neutral). Users collected ten items from a virtual reality market. To complete the tasks, they interacted using natural speech with the emotional VHs. Three language conditions were investigated: one condition in USA and another two conditions in Taiwan. The group of participants in USA interacted with the VHs in English; and the two groups of participants in Taiwan interacted with the VHs using a foreign (English) language and a native (Mandarin) language respectively. We discovered that the medium of communication or language familiarity had a strong influence on participants' emotional reactions. When participants interacted in a foreign language with VHs with a positive emotional disposition, we found their positive emotional reactions were subdued and negative reactions were elevated. However, this was not the case when participants interacted with VHs in their native language, as their emotional reactions were contingent on the emotional disposition of the VHs.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Animations, Evaluation/methodology; I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality;

1 Introduction

Virtual crowd simulations refer to the process of recreating behaviors (e.g., gathering and interaction) of crowds of agents (or virtual humans) in digital scenarios. They are commonly used in films and video games [37], evacuation [49], urban planning [4], and training systems [61]. In addition, crowd simulations are used for reconstructing emotionally stressful environments, such as fire emergencies [40] and evacuation drills [25]. Typically these specific types of crowd simulations are commonly used to study how emotional contagion or emotion propagation increases the behavioral realism of agents or to study human factors [74].

Emotion contagion refers to a person's affect and related behaviors that influence others to trigger similar emotions and behaviors and vice versa [24]. Emotion contagion research in crowds has mostly examined aspects such as crowd dynamics, performance, and effectiveness. These models simulate emotions and personality traits of the virtual humans (VHs) to produce realistic behavior. Popular emotion contagion models are the OCC (Ortony, Clore, and Collins) model, [56], OCEAN personality model [3], and the emotional reciprocal velocity obstacles model [73]. However, some models incorporate virtual agents that exhibit low-fidelity verbal and non-verbal behaviors. The affective behaviors missing in these agents are positive or negative facial expressions, specific body postures, tone of voice (emotional prosody), or gaze [54]. Studying emotional contagion is important, especially to understand human reactions and decision making in highly emotional scenarios. Furthermore, the models mentioned above do not support agents with natural language interaction capabilities. Thus, these models cannot convey emotion through speech. Users cannot communicate through natural language with these agents so the emotion contagion effect is limited to only facial expressions and body postures. The lack of access to simulated conversations hinders the emotional propagation effect since natural language is a strong channel to convey emotion and cause emotion contagion [33]. There are not many research studies in virtual crowds that investigate emotion contagion in which users are able to engage in natural conversations through speech based dialogue with a crowd of affective VHs.

Language has a direct impact on emotional contagion and varies depending on the users' level of proficiency. Emotional processing is enhanced in a first, more proficient language and reduced in a later learned, less proficient language. Pavlenko et al. [47] stated that the cause of this effect is due to the *affective processing* theory. Affective processing is defined as the responses triggered by the automatic appraisal of verbal stimuli, which triggers an increase in arousal levels. Additionally, there is evidence that bilinguals speak longer about emotional topics in their second language because of reduced embarrassment [6]. The second language speakers commonly acknowledge that taboo terms can be uttered with greater ease in their second language than in their first language [23]. The effects of interacting with virtual agents have not been deeply examined yet in a first or second language in interaction with crowds of emotional VHs in immersive virtual environments.

A study examined the users' emotional contagion effect during interaction with a crowd of VHs that showed opposite-valenced emotional expressions [65]. The agent-based crowd model had rich properties, including eye gaze, facial expression, body motion, and verbal and non-verbal behaviors. However, users interacted with the

*e-mail: mvolont@clemson.edu.
†e-mail: ccwang6612.cs07g@nctu.edu.tw. The first and second authors contributed equally.
‡e-mail: ebrahimie@uncw.edu
§e-mail: i_always_smile123@yahoo.com.tw
¶e-mail: liu6352609@gmail.com
‖e-mail: cswingo@cs.nctu.edu.tw
**e-mail: sbabu@clemson.edu

2642-5254/21/$31.00 ©2021 IEEE
DOI 10.1109/VR50410.2021.00040

agents through natural speech only in the English language. The emotional contagion effect was weak in the study, but users showed more interest in engaging with virtual agents expressing positive emotions than with agents expressing negative emotional verbal and non-verbal behaviors.

In this paper, we studied the extent to which the emotional contagion effect was mediated by language familiarity in a crowd simulation with natural speech-based dialogue capabilities. In a between-subject design, participants interacted through natural language with a crowd of VHs to collect objects from a marketplace. This study presented four conditions of emotional crowds, namely Positive (POS), Negative (NEG), Neutral (NEU), and Mixed (MIX). The VHs in Positive and Negative conditions exhibit opposite-valenced emotional facial expressions for easy recognition while the VHs in the neutral condition do not show any emotional expression; in the Mixed condition, the crowd is composed of VHs with the the three different kinds of emotions. Furthermore, we presented three language familiarity levels: one condition in the USA and two conditions in Taiwan. Participants in the USA interacted with the VHs in English, and the other two groups in Taiwan interacted with the VHs using a foreign (English) language or a native (Mandarin) language. English and Mandarin were selected because they are one of the most popular in the world. This investigation provides novel and original data on how language mediates the emotion contagion effects with interactive affective virtual crowds. We also compared the emotional contagion across different groups of language modes of communication. This provides more general and robust results. The contribution of this investigation is valuable to the academic community as the effects of language familiarity on emotional contagion with interactive crowds have not been thoroughly investigated. Our results have implications to the design of virtual crowd simulations in entertainment, education and training scenarios.

2 Related Work

Non-Verbal communication.

Non-verbal communication is the transmission of information through facial expressions, gestures, body posture, and the social interaction distance between individuals [19]. These non-verbal behaviors in interpersonal relationships provide information about the expresser's emotional state. There is strong evidence that the recognition of emotions composed by non-verbal behaviors are inherited and shared by all humans and recognized cross-culturally [17]. Paul Ekman provides evidence that facial emotional expressions are inherited biologically to humans and are universally recognized [17]. He classified seven facial expressions that correspond to distinct universal emotions, namely disgust, sadness, happiness, fear, anger, surprise, and contempt. Furthermore, Sauter et al. studied nonverbal emotional vocalization recognition between different cultural groups [57]. Results showed that several primarily negative emotions have vocalizations that can be recognized across cultures, while most positive emotions are communicated with culture-specific signals. Moreover, there is evidence that emotional bodily movements can also be recognized universally. A study presented different body postures represented by point lights to western and eastern participants, and then the participants were asked to label the conveyed emotions. Results showed that movement kinematics alone can effectively convey emotions across disparate cultures [46].

Emotional Contagion and Virtual Humans.

Emotional contagion is the *"process in which a person or group influences the emotions or behaviors of another person or group through the conscious or unconscious induction of emotion states and behavioral attitudes"* [60]. Emotional contagion attracted interest from many fields not limited to computer animation, simulation, and virtual reality. Studies are, but not limited to, the effects of emotion contagion on social networks [14], how it affects relationships in work environments [35], or in the medical field [45]. Emotional contagion also attracted interest in the field of virtual humans researchers since human-virtual human interaction evokes similar social responses to that of human-human interactions [55]. Studies show that emotional contagion during interaction with virtual agents is caused by animation [71], appearance [63], and verbal and non-verbal behaviors [66]. Qu et al. [52] asked participants to have a conversation with a virtual woman who displayed either positive or negative facial expressions when speaking and listening to the participants. Results showed that the two emotions displayed by the virtual woman during the interaction, and especially in the speaking phase, evoked emotional contagion by producing a congruent emotional state in the participants.

Crowd modeling and emotion contagion.

Traditional crowd simulation techniques focus on crowd movement and collision avoidance, such as the social force model [26], continuum crowd simulation [34], space colonization [15], reciprocal collision avoidance [62], radial-view based method [42], and the collision probability fields [70]. Agent-based emotional contagion models simulate agents' emotions that affect their decision making [7,41]. Li et al. [41] proposed an approach to simulate emotional contagion in crowds with emotional and stressful scenarios, such as violent riots. This method involves modeling antagonistic roles within the crowd using two components, namely a mental emotion and an external emotion. Studies have been investigated to evaluate user task performance [30] and user behaviors [65] in immersive environments. The system [65] supports a variety of verbal and non-verbal behaviors, including gaze, walking styles (gaits), facial expressions, and voice emotional prosody (e.g., speech rate, vocal range). Such behaviors are essential in social interaction [27].

Language and emotion contagion.

Language familiarity affects users' emotional responses. Evidence suggests that affective processing may be weaker in a learned language than in a naturally acquired language. Thinking in a foreign language may reduce decision-making biases because the foreign language provides a greater emotional and cognitive distance compared to the native language [39]. A study analyzed the users' reaction time in a visual processing task between native English speakers and Chinese speakers for whom English was a second language. Results revealed that native English speakers showed a large attentional blink following a taboo distractor while the attentional blink time for Chinese speakers was shorter. Additionally, electrodermal monitoring revealed that some emotional phrases elicited larger skin-conductance responses (SCRs) when presented in the native language than that in a second learned language [13].

A study compared the recognition accuracy between Chinese and non-Chinese participants about the emotional expressions of a virtual human [51]. The experiment presented a within-subject design that included listening and speaking phases. Several conversation clips between the VH and a person in Chinese were presented to participants. In each clip, the VH spoke 10 sentences in total and kept silent between each sentence, listening to her conversational partner's speech. The study found that emotion was recognized by people from different cultural backgrounds as the VH because both Chinese and non-Chinese participants correctly identified the valence of the virtual human's emotional state.

3 System Design and Architecture

The experiments were conducted in an HTC Vive Pro Head-Mounted Displays. The simulation was executed in a Windows PC with Intel (R) Core (TM) Processors with an NVIDIA GeForce RTX Graphics card and 32 GB of memory RAM.

3.1 Virtual Market Simulation

The environment consisted of a virtual market with a street in the center and shops on both sides of the street. Two access points (entrance/exit) are located at the ends of the street. The items for

Figure 1: The English interface. A panel shows the recognized words (top), task items (left), and the word tips (right). The Mandarin interface shows them in Chinese characters.

collection are fruits and vegetables that are commonly found in real markets. The system has three components which are crowd simulation, user interaction, and task manager. The component of crowd simulation updates agents in the virtual market. There are two kinds of agents which are pedestrians and vendors. The pedestrians walk along the street. There are vendors and each of them stands at one shop which has a unique item for the user to collect. The interaction component allows the user to interact with pedestrians and vendors. To collect a specific item, the user must talk in a natural language to the pedestrians to gather a piece of guidance information about the location of the vendor. The interaction component supports English and Mandarin languages. Only one of the languages is adopted under an experiment condition. The system has an interface that displays the recognized words, names of the task items, and word tips for conversation. A tick symbol is displayed next to an item that has been collected. Figure 1 shows the English interface. The task manager manages the task items.

The user can talk to a vendor or a pedestrian within the distance of social interaction [21, 69]. The system motivates users to interact with agents based on two mechanisms. First, the pedestrians have a probability that they know the vendor locations for the acquired items. The probability is moderate (e.g., 0.8) and otherwise the users would drop interest in interacting with the pedestrians. Second, if the pedestrian knows the location of the acquired item and the vendor is only one shop away from the current location, the pedestrian responds briefly in a clear direction. However, the pedestrian uses a hand to point to a location in the case that the vendor is far away from the current location. Thus, the user may ask one or more pedestrians to locate the vendor location.

3.2 Computation of Crowd Movement

The pedestrians enter and leave the market periodically at the access points. After a pedestrian leaves the market, a new pedestrian is generated at an entrance. We adopt the social force model to drive the pedestrians to walk [26] because of its efficiency. Mainly, the social force model uses attractive forces and repulsive forces to make pedestrians move towards goals and avoid collision, respectively. The idea is to compute the net force $\mathbf{f}_i$ for each pedestrian i to achieve collision avoidance. The velocity of pedestrian i is updated as $\mathbf{v}_i^{(t+1)} = \mathbf{v}_i^{(t)} + \frac{\mathbf{f}_i}{M_i}\Delta t$, where M_i is the mass of pedestrian i, t is the simulation time, and Δt is the simulation time step. To control the speed of pedestrian i, $\mathbf{v}_i^{(t+1)}$ is modified as $\mathbf{v}_i^{(t+1)} = min(s_i, \|\mathbf{v}_i^{(t+1)}\|)\frac{\mathbf{v}_i^{(t+1)}}{\|\mathbf{v}_i^{(t+1)}\|}$, where s_i is the maximum walking speed (e.g., $1.4ms^{-1}$). After that the position of pedestrian i is updated by $\mathbf{p}_i^{(t+1)} = \mathbf{p}_i^{(t)} + \mathbf{v}_i^{(t+1)}\Delta t$. We adopt the radial view-based method [42] to drive the pedestrians to walk away from the user and an interacting pedestrian.

3.3 Behaviors of agents

The agents perform verbal and non-verbal behaviors to convey positive, negative, or neutral emotional states. The behaviors included facial expressions [16], body postures [59], gaze behavior [2], the speed of the response, intensity, voice pitch, vocal range, and speech rate [8, 50], head pose [5], hand motion, and hand pointing. Hand pointing is achieved based on inverse kinematic with factors such as the shoulder joint position, the vendor location, and the hand direction. Furthermore, facial expressions (e.g., sad and happy) are modeled based on Facial Action Coding System (FACS) which defines specific action units that can be adopted to construct a facial expression [16]. In eye gaze modeling, eyeballs, head pose, and body postures are coordinated while a target is being watched [48]. During the social interaction, gaze and body direction may or may not be congruent with each other [1], resulting in a positive or negative impression. Eyes look towards a target in a direct gaze while they look downwards or directions other than the target in an inverted gaze [1, 68]. We adopt a direct gaze and an inverted gaze to model agents with positive and negative emotions, respectively.

3.4 Interaction with Pedestrians and Vendors

We designed a minimal set of sentence structures for the user to interact with pedestrians and vendors for the English and the Mandarin language versions. For example, the responses from a pedestrian are *Walk straight and the shop is on your left hand*, *Go in this direction*, and *I don't know where it is*. The users' voice is converted to text and the speech domain associated with the text is analyzed. Thus, the pedestrians and vendors can respond to the user based on the type of the speech domains (e.g., greetings, appreciation, leave-taking).

A pilot study (around 3 participants) showed that participants were frustrated when they spoke some words that were recognized as other words (especially in Mandarin). Therefore, the recognized keywords of the item were shown in the dialogue interface. Furthermore, we selected the items which had distinct pronunciations so that a high recognition accuracy could be achieved.

A user initiates a conversation session with a pedestrian within the social interaction distance. Then the user can use natural language to ask for the location of an item in the item list and the pedestrian will respond appropriately. After that the user may appraise the help and then leave the pedestrian. The vendors do not know the locations of items. Thus, if the user asks the vendors about the locations of the items, the vendors will reply that they do not know. The vendors are modeled for engaging in a simulated dialogue if the user asks for an item. In case the vendor sells that item, it will respond affirmatively and the item will be displayed on the table. After that the user can pick up the object and put it into a digital bag. In case the vendor does not have the item, it will answer *I don't have it*.

3.5 Pilot study on evaluation of VHs

We conducted a pilot study for evaluating the emotional behaviors of the VHs. In the pilot study, users watched animation clips in which a VH performed behaviors of a specific emotion, including walking styles, facial expressions, and body postures accordingly. Then the users answered a questionnaire to rate the VH's emotional valence [38] that is the extent to which the associated emotion is positive or negative. We used a questionnaire with a Five-point Likert scale (very negative, negative, neutral, positive, very positive). A total of 22 participants (Male=18, Female=4, Age Mean=23) rated 67 video clips in a random order. The results showed that the participants' classification accuracy of VH's emotional valence was 80% or higher.

4 Research Hypotheses and Expected Outcomes

Based on the existing literature we considered that answering the following research question can be of a great contribution to the literature:

To what extent does the native vs. foreign language mediates the impact on the users' emotions when interacting with a crowd of emotional virtual humans? Two hypotheses are as follows.

1. H1: *Language familiarity will elicit stronger emotion contagion effects on users.*
2. H2: *Users in the Positive condition and Negative condition will report stronger positive and negative emotionally responses, respectively.*

Research question 1 is based on the existing literature that declares that emotional contagion should be stronger when users are exposed to stimulus presented in the users' native language vs. a foreign second language [9]. Our contribution relies that this has not been examined with VHs in a crowd simulation in a context that involves a task. Regarding H1, we based this hypothesis from the literature that states that emotion contagion effect is stronger when users are exposed to stimulus in their native language [13,39]. Regarding H2, this hypothesis is based on the emotion contagion theory that states that emotion can be transferred between persons [24].

5 Experiment Methodology

5.1 Study Design

This study presented a 3x4 between subjects design. The 3 language familiarity conditions (English Native, English Foreign, Mandarin Native) and 4 emotional virtual humans crowds (Positive, Negative, Neutral and Mixed). Figure 2 shows the experiment design. The three language familiarity groups adopted the language that the users' and the virtual agents shared during the simulated dialogue in the experiment. These language familiarity groups were USA in English (UEN), Taiwan in English (TEN) and Taiwan in Mandarin (TMA). For the condition UEN, the data were collected in the US and participants possessed proficient familiarity with English. The condition TEN was conducted in Taiwan with participants whose English was their second language. And lastly, the condition TMA was also conducted in Taiwan but Mandarin was the participants' native language.

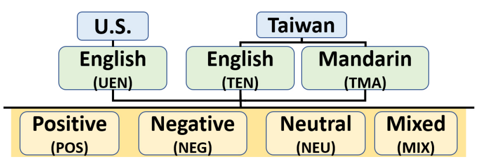

Figure 2: Graph showing the experiment design.

The four crowd conditions were Positive, Negative, Neutral, and Mixed. These presented different emotional valences of the VHs' verbal and non-verbal expressions, including body postures, arm and hand gestures, eye-gaze behavior, and voice emotional prosody (Table 1). For modeling the positive and negative facial expressions, we used the Facial Action Coding System manual (FACS) [16] while for creating the body posture emotional expressions, we relied on literature that classified scientifically positive and negative body posture expressions [59,72]. The virtual agents in the Positive group (POS) displayed positive valenced emotional verbal and non-verbal behaviors, and expressed enthusiasm and interest to engage in simulated dialogue with the users. On the contrary, the Negative crowd (NEG) included sad and unhappy non-verbal communicative behaviors, and showed little interest in conversing with the users by acting apathetically and uninterested. Additionally, VHs in the Neutral condition (NEU) expressed no emotional disposition. Finally, the Mixed (MIX) condition included VHs which are from the positive, negative, and neutral groups equally distributed.

Table 1: The verbal and non-verbal behaviors included in the different emotion conditions. The mixed group consists of positive, negative, and neutral emotional VHs.

	Face Expression	Gaze/Body Expression	Voice Emotional Prosody
POS	Smile, smirk	Relaxed, energetic	Fast paced
		High amplitude in movements	Energetic
		Joint gaze	High frequency variability
NEU	Neutral	Natural movements	Medium paced
		Natural gaze and head motion	Normal
		Natural gaze	Natural variabiltiy
NEG	Sad	Slow movements	Slow paced
		Gaze and head down	Monotonous
		Gaze avertion	Low frequency variabiltiy

Also, it is important to point out that the conversational routines were somewhat focused and task oriented to minimize speech recognition errors. These conversational routines included saying "Hello", "Where can I find {Item}", "Thank you." Moreover, only when users conversed with the agents, the dialogue interface displayed the proper words and word suggestions supported.

Finally, the behaviors of the virtual agents were carefully designed following guidelines from the interpersonal communications literature. Moreover, a specialist in communication with an emphasis in interpersonal communication reviewed and approved the facial expressions, body postures, voice emotional intonation, and overall behaviors of all the emotional conditions of the virtual human crowd. The English and Mandarin voices for the VHs were recorded from real humans. An expert in Human Interpersonal Communications gave suggestions for creating positive and negative emotional speech. The behaviors presented in the current submission were validated in a previous peer-reviewed manuscript. For further details please refer to [65]. Finally, there is evidence that the emotional expressions portrayed by the virtual agents are recognized universally across cultures [17,46,58,67].

5.2 Procedure

First, the research proctor explained the experiment and answered any questions from the participants. Then the experiment began after the participants' consent was obtained. Initially, the participants filled out a demographics questionnaire about their age, gender, and familiarity with virtual reality. Next, the participants started the "Acclimation phase" for learning how to perform the operations needed in the main experiment. This phase was conducted in a virtual environment scenario different from the one used in the main experiment. This phase ended when the participants completed the task and expressed that they were ready to start the main experiment. For the main experiment, the participants were immersed through a head-mounted display in a virtual environment market and were tasked to collect 10 items. To collect all items, the participants needed to engage in simulated dialogue with virtual humans through natural speech. Only during simulated dialogue routines, a graphical user interface (GUI) appeared showing the list of collected and remaining objects, and hints on how to communicate with the virtual humans (Figure. 1). The simulated dialogue routine was carefully and thoroughly created to avoid any misunderstanding on how to communicate properly with the virtual humans. After collecting all the objects, the participants filled out the Differential Emotional Survey (DES) survey and Social Presence questionnaire, and were thanked for their time. The participants attended the study on a volunteer basis, and the study was approved by the respective institution's review board (IRB). Finally, the proctors in the USA and Taiwan followed the same strict and consistent protocol for conducting the experiment. This guide consisted of a script describing step by step description on how the scientists should welcome participants, guide them through the different experiment phases, and dismiss them.

5.3 Measures

5.3.1 Differential Emotions Survey (DES)

We measured the participants' emotional disposition after the interaction with the Differential Emotions Survey (DES). This questionnaire is meant to be applied to measure the emotional state of participants only after some form of treatment, as the statements in the questionnaire were directly tied to the experience the participants received. Therefore, as part of the guideline of the DES questionnaire, we applied this survey to measure the effects of the independent variable medium of communication to emotion contagion as per the intent of the inventory. For further information on the use of the DES inventory, please refer to [32]. This survey contained 30 items scored from 1 to 7 to express how strongly they felt each item (0 being Never and 7 being Extreme) [31]. The results were grouped into 10 categories namely, *interest*, *enjoyment*, *surprise*, *sadness*, *anger*, *contempt*, *fear*, *guilt*, *shame*, and *shyness*.

5.3.2 Co-Presence

We measured Co-presence by the *Internal Consistency and Reliability of the Networked Minds scale* [22]. Co-presence is defined as the degree of initial awareness, allocated attention, the capacity for both content and affective comprehension, and the capacity for both affective and behavioral interdependence with an entity during an interaction. We presented only the "Co-presence" dimension due to space limitations. Finally, we included this measure as it can be used to examine to what extent users felt higher sense of personal and common ground with the virtual agents in relation to the language familiarity condition [10, 11].

5.4 Participants

We determined the sample size needed for a total of 12 independent groups which are 4 Crowd groups (Positive, Negative, Neutral, and Mixed) X 3 Regional Languages (UEN, TEN, and TMA) for obtaining a medium to high effect size in G*power [18] based on Cohen's Effect size convention [12]. We calculated a medium to high effect size of 0.32; and set $(1-\beta) = 0.08$ and alpha = 0.05, and found that a total of 180 participants (15 per condition) would be needed.

We recruited 181 participants who possessed the necessary level of language proficiency required for participating in this study. The following is the experiment sample description per condition: UEN (POS=15, NEG=15, NEU=15, MIX=15; Male=46, Female=14, Age Mean=25.7), TEN (POS=15, NEG=15, NEU=15, MIX=16; Male=37, Female=24, Age Mean=23.5), and TMA (POS=15, NEG=15, NEU=15, MIX=15; Male=30, Female=30, Age Mean= 23.2). We tested the participants in the TEN condition to insure that they were familiar, comfortable, and fluent in using English as a foreign language. No participant reported any cybersickness sensation.

6 Results

6.1 Analysis procedure

We conducted non-parametric statistical analysis on the data for the DES and Social Presence questionnaires. We conducted a Kruskal-Wallis H test for comparing the multiple independent variables (e.g., language groups and emotional groups). In our case, the data values were ordinal. In case of significant differences between groups, we conducted a Mann-Whitney U test analysis for post-hoc pairwise comparisons between levels of the independent variables.

6.2 Co-Presence

We measured and analyzed the participants' co-presence scores in the experiment (See Section 5.3.2 for information about this survey) in a similar manner to DES using non-parametric statistical analysis.

The Kruskall-Wallis H test showed significant differences in multiple categories of this survey. However, we selected only the variable co-presence due to space limitations in the manuscript. Table 2 shows the result and Figure 3 shows the significant post-hoc differences between language groups.

Table 2: Co-Presence significant differences in Emotion by Language conditions interaction.

	Co-Presence		
	Positive	Negative	Neutral
Chi-Square	H(2)=6.25	H(2)=10.66	H(2)=6.08
p-Value	0.04	0.005	0.048
Mean/(SD)	UEN=4.06 (0.68) TEN=4.09 (0.57) TMA=3.59 (0.57)	UEN=4.04 (0.54) TEN=3.30 (0.58) TMA=3.47 (0.59)	UEN=3.99 (0.53) TEN=3.46 (0.58) TMA=3.61 (0.71)

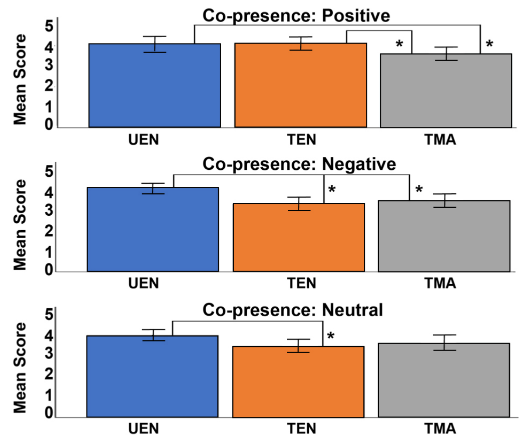

Figure 3: Co-presence differences - Language per emotion interaction differences. Denote * as $p < 0.05$.

In the Positive crowd condition, the participants of both UEN and TEN groups experienced higher sense of co-presence than those in the TMA group. However, the participants in the Negative crowd condition of the UEN group experienced higher sense of co-presence than those in both TEN and TMA groups. Lastly, the emotionally Neutral crowd elicited significantly higher sense of co-presence in the UEN group over the TEN group.

6.3 Differential Emotional Survey (DES)

To examine the emotional impact of the participants after they interacted with the virtual humans, we administered the Differential Emotions Survey (DES) at the end of the experiment (See Section 5.3.1 about the DES survey). To thoroughly inspect the emotional impression after interplay with the crowds of affective virtual humans, we processed the data in two ways: 1) We compared the different emotional crowds groups (POS, NEG, NEU, and MIX) across different conditions of Language Regions (UEN, TEN, and TMA); 2) We compared the emotional impact within each condition of Language Regions.

6.3.1 Analysis of Emotion within Language Regions

TEN - Emotion. First, we analyzed the different groups of emotional crowds, namely Positive, Negative, Neutral and Mixed within the TEN condition on the 10 different DES categories. The Kruskal-Wallis H test reported significant difference only in the *Interest/Excitement*, *Enjoy/Joy*, *Surprise/Startle*, *Distress/Anguish*, *Anger*, and *Fear/Terror* emotions. The participants in the Positive group reported significantly higher *Interest/Excitement* scores than those in the Neutral and Mixed groups. Also, the participants in the

Positive group reported significantly higher *Enjoy/Joy* scores than those in the other three emotional groups. However, the participants in the Mixed group reported significantly higher *Distress/Anguish* scores than those in the Positive and Neutral groups. Furthermore, the participants in the Negative group reported significantly higher *Surprise/Startle* and *Fear/Terror* scores than those in the Negative and Neutral groups. Finally, the participants in the Negative group reported significantly higher *Anger* scores than those in the other three groups. Figure 4 shows Mann-Withney post-hoc significant differences; and Table 3 presents the statistics.

UEN - Emotion. The UEN condition presented significant differences between groups for the *Distress/Anguish* emotion. The participants interacting with the Negative crowd reported significantly higher scores than those in the Neutral and Mixed groups. Table 4 presents the Chi-square, p values, mean, and standard deviations values; and Figure 6 shows Mann-Withney post-hoc significant comparisons.

TMA - Emotion. In the TMA condition, the analysis revealed significant differences between groups in the *Anger* and *Fear/Terror* emotions. The participants in the Mixed crowd reported higher *Anger* and *Fear/Terror* than those in the (Positive, Neutral) and (Negative, Neutral) crowds, respectively. Table 4 presents the statistics and Figure 6 shows Mann-Withney post-hoc significant comparisons.

6.3.2 Analysis of Emotion between Language Regions

We computed a Kruskal-Wallis H test for comparing how the participants in the UEN, TEN, and TMA were affected emotionally when interacting with the virtual human crowds. Table 5 shows the Chi-square, p-values, mean, and standard deviations.

In the Positive condition, the analysis showed that in the *Interest/Excitement* and *Enjoy/Joy* emotions the participants in the TEN (Taiwan English) condition scored significantly higher than those in the UEN and TMA language conditions. Figure 4-A shows the Mann-Withney post-hoc significant differences for the Positive crowd condition.

In the Negative condition, the analysis results showed that in the *Distress/Anguish* emotion the participants in the TEN condition scored significantly higher than those in the TMA language condition. Furthermore, in the *Fear/Terror* the participants in the TEN condition scored significantly higher than those in the UEN and TMA language conditions. Figure 4-B shows the Mann-Withney post-hoc significant differences in *Distress/Anguish* and *Fear/Terror* for the Negative crowd condition.

In the Neutral condition, the analysis results showed that in the *Distress/Anguish* emotion the TMA participants scored significantly higher than those in the UEN and TEN language conditions. However, in the *Anger* emotion the TEN participants scored significantly higher than those in UEN and TMA conditions. Figure 4-C shows the Mann-Withney post-hoc significant differences in *Distress/Anguish* and *Anger* for the Neutral crowd condition.

In the Mixed condition, the analysis results showed that in the *Distress/Anguish* the UEN participants scored significantly lower than those in the TEN and TMA conditions. In *Anger*, and *Fear/Terror* emotions the TMA participants scored significantly higher than those in the UEN and TEN language conditions. Figure 4-D shows the Mann-Withney post-hoc significant differences in *Distress/Anguish*, *Anger*, and *Fear/Terror* for the Mixed crowd condition.

7 Discussion and Limitations

Significant findings were found on Co-Presence between the different language regions (Figure 3). Also, significant differences were found on the analysis of emotion between UEN, TEN and TMA language Regions (Figures 4), and within groups of the emotional crowds (Figures 5 and 6).

In response to Question 1: "*To what extent does the native vs. foreign language mediates the impact on the users' emotions when interacting with a crowd of emotional virtual humans?*."

Results show that emotion was strongly mediated by the users' language familiarity. The examinations of the emotional crowds within the each group show evidence that language was a strong emotional contagion mediator (Figures 5 and 6). The language familiar groups scored UEN (USA in English) and TMA (Taiwan in Mandarin) were less affected by the emotional crowds than the participants of the less familiar language group, TEN (Taiwan in English).

Comparing the results of each crowd condition (POS, NEG, NEU, and MIX) on each language condition (UEN, TEN, and TMA), the users in the TEN group (interaction using an unfamiliar language) showed stronger emotional effects in both positive and negative emotions as compared to those interacting with familiar languages (UEN and TMA).

In response to H2: "*Users in the Positive condition and Negative condition will report stronger positive and negative emotional responses, respectively.*"

The results revealed that the participants in the positive condition did report significant strong positive emotions. But users in the negative group did not showed high negative reactions. However, a similar effect can be appreciated in the emotional analysis namely Neutral and Mixed groups. Thus, it is not possible to claim that emotional contagion happened.

The DES analysis in the UEN condition (USA in the English language) revealed minimal significant differences between groups; only the Distress-Anguish category was different between the groups (see Figure 6). Likewise, the DES analysis result on the four different emotional crowds in the TMA (Taiwan in Mandarin) revealed only a significant difference in the Fear and Terror category (Figure 6). These two conditions presented virtual agents that talked in the users' native language namely English (UEN) and Mandarin (TMA).

However, results in TEN condition were interesting and surprising. The analysis showed that when the participants interacted with the virtual agents in a foreign language (Taiwanese participants using English language), they felt high positive emotions and low negative affect during interaction with the virtual agents in all the emotional groups. This effect provides evidence that conversing in a foreign language seems to have exacerbated the users' emotions.

Co-presence. The participants that interacted with the agents in the English language exhibited higher co-presence scores in the Positive, Negative, and Neutral conditions. The Mixed condition did not show any significant differences (Figure 3). This dimension measures the extent that users believe they are not alone, are aware of the others, and others are aware of them in the virtual environment. We expected that users interacting in their natural language exhibit higher co-presence scores since this could elicit a stronger sense of commonality with the virtual agents [10]. However, this effect only is observed in the UEN group and not in the TMA language condition. These results are interesting and further investigations are deserved to examine how cultural differences could potentially play a role in this result. Furthermore, the eastern physical appearance of the characters could have affected the perceived co-presence to the western TEN and TMA participants since reports state that physical appearance affects the user experience [44].

A limitation of this study was that the virtual humans produced short utterances during simulated dialogue. Even though this was a careful decision made to constrain the dialogue only related to the task, users over-expected the virtual humans speech's capabilities. This could have detrimentally influenced the co-presence effect, perhaps also explaining the results found with regards to H2. For future examinations, we will explore how high fidelity conversational virtual humans with small-talk capabilities affect mediate and moderate the emotion contagion effect. Another limitation is that we

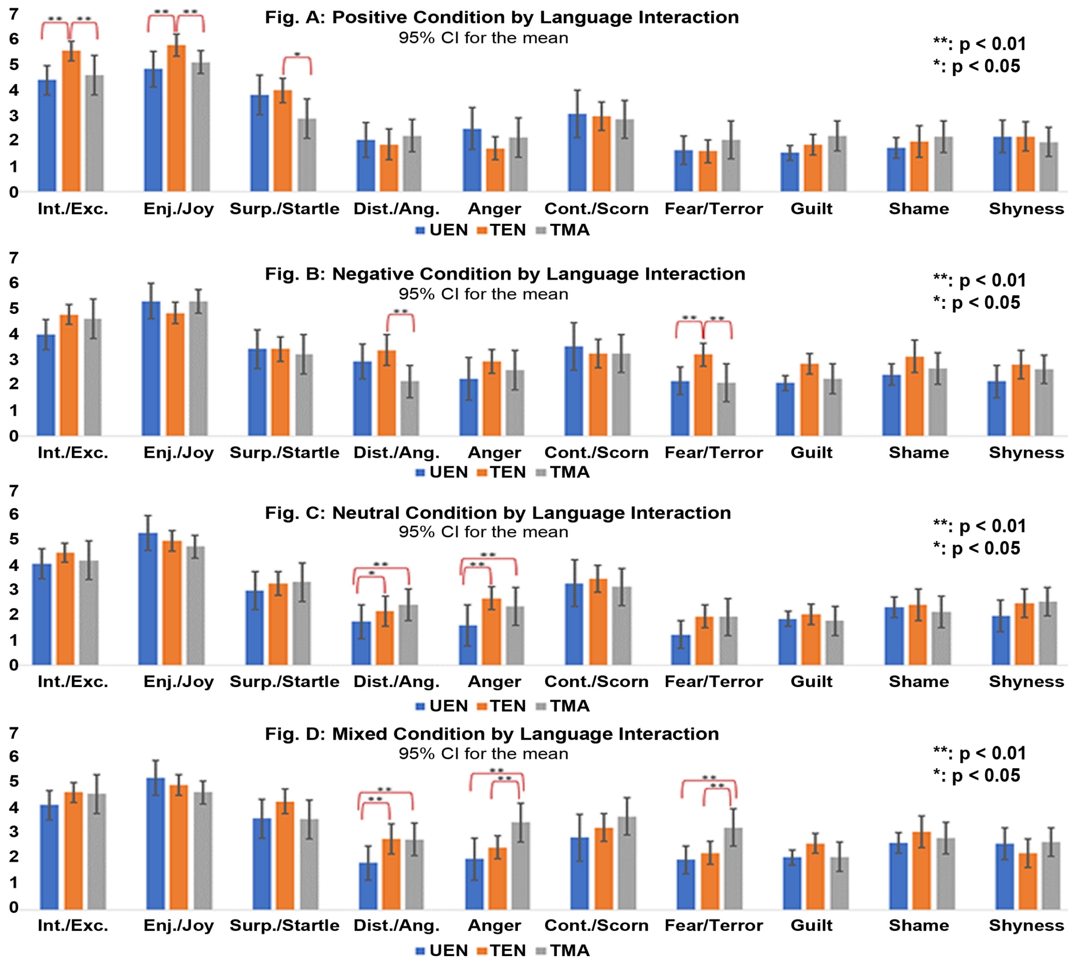

Figure 4: DES results for virtual crowd conditions across all the language conditions.

Table 3: TEN (Taiwan in English) Condition - DES. Significant variables, Chi-Square, p-Values, Mean, Standard Deviations.

	TEN - Emotions					
	Int./Exc.	Enj./Joy	Surp./Startle	Dist./Ang.	Anger	Fear/terror
Chi-Square	H(3)=8.11	H(3)=8.35	H(3)=8.68	H(3)=14.06	H(3)=8.15	H(3)=13
p Value	0.044	0.039	0.034	0.003	0.043	0.005
Mean/(SD)	POS=5.53 (0.76)	POS=5.75 (0.83)	POS=3.97 (0.93)	POS=1.86 (1.18)	POS=1.71 (0.89)	POS=1.6 (0.89)
	NEG=4.75 (1.25)	NEG=4.81 (1.01)	NEG=3.39 (1.3)	NEG=3.35 (1.29)	NEG=2.91 (1.2)	NEG=3.18 (1.4)
	NEU=4.51 (1.16)	NEU=4.97 (0.85)	NEU=3.28 (1.07)	NEU=2.15 (0.95)	NEU=2.68 (1.45)	NEU=1.95 (1.25)
	MIX=4.66 (1.21)	MIX=4.95 (1.03)	MIX=4.31 (1.03)	MIX=2.82 (1.09)	MIX=2.48 (0.94)	MIX=2.26 (1.0)

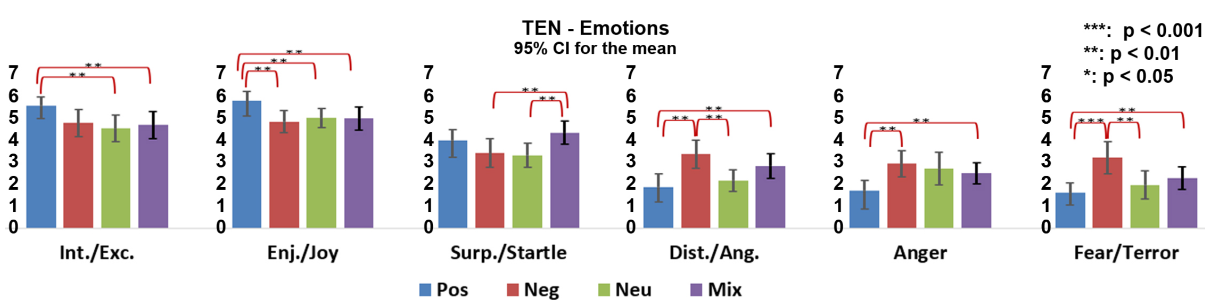

Figure 5: The post-hoc significant differences between the emotional crowds in the TEN condition (Taiwan in English).

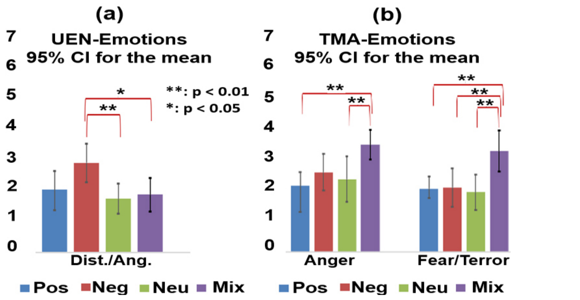

Figure 6: The post-hoc significant differences in the UEN and TMA language conditions. Left: UEN Distress/Anguish post-hoc significant difference. Right: TMA Anger and Fear/Terror post-hoc significant differences. In TMA, under the MIX condition, there were Positive and Neutral virtual humans. Participants could compare them with the Negative virtual humans and would feel that the Negative virtual humans were unfriendly. Thus, the overall participant impression was affected in the MIX condition.

Table 4: UEN (USA in Enlgish) and TMA (Taiwan in Mandarin) Conditions - DES.

	UEN	TMA	
	Dist./Ang.	Anger	Fear/Terror
Chi-Square	H(3) = 8.16	H(3) = 8.34	H(3) = 12.02
p-Value	0.043	0.039	0.007
Mean/(SD)	POS=2.04 (1.33)	POS=2.13 (1.51)	POS=2.04 (1.46)
	NEG=2.91 (1.61)	NEG=2.57 (1.8)	NEG=2.08 (0.96)
	NEU=1.74 (1.25)	NEU=2.35 (0.96)	NEU=1.93 (1.26)
	MIX=1.87 (1.03)	MIX=3.46 (1.41)	MIX=3.26 (1.2)

Table 5: Emotion by Language Region Results - Kruskal Wallis Results

	Positive		Negative	
	Int./Exc.	Enj./Joy	Dist./Ang.	Fear/Terror
Chi-Square	H(2)=9.51	H(2)=6.05	H(2)=6.29	H(2)=6.38
p-Value	0.009	0.049	0.043	0.041
Mean/(SD)	UEN=4.37 (1.16)	UEN=4.82 (1.37)	UEN=2.91 (1.61)	UEN=2.15 (1.35)
	TEN=5.53 (0.76)	TEN=5.75 (0.83)	TEN=3.35 (1.29)	TEN=3.18 (1.46)
	TMA=4.57 (1.51)	TMA=5.08 (0.90)	TMA=2.13 (0.68)	TMA=2.08 (0.96)

	Neutral		Mixed		
	Dist./Ang.	Anger	Dist./Ang.	Anger	Fear/Terror
Chi-Square	H(2)=6.1	H(2)=8.91	H(2)=9.4	H(2)=9.36	H(2)=9.66
p-Value	0.047	0.012	0.009	0.009	0.008
Mean/(SD)	UEN=1.74 (1.25)	UEN=1.6 (1.5)	UEN=1.87 (1.03)	UEN=2.02 (1.24)	UEN=2.0 (1.29)
	TEN=2.15 (0.95)	TEN=2.68 (1.45)	TEN=2.82 (1.09)	TEN=2.48 (0.94)	TEN=2.26 (1.0)
	TMA=2.42 (1.11)	TMA=2.35 (0.96)	TMA=2.8 (0.8)	TMA=3.46 (1.41)	TMA=3.26 (1.2)

did not consider personality traits of participants. Personality traits and states of a participant affect the willingness of the participant in social interaction [20].

8 Conclusion

We evaluated the users' emotional reactions after interaction with a crowd of virtual humans that exhibited distinct emotional expressions in either the participants' native language in two regions (USA and Taiwan) or in a foreign language that participants were generally familiar with (Taiwan in English).

The DES results from the TEN (Taiwan in English), UEN (USA in English) and TMA (Taiwan in Mandarin) conditions across all emotional groups (Positive, Negative, Neutral, and Mixed) provide evidence that language familiarity did affect the users' emotional contagion. In the conditions where users were familiar speaking and listening the language (UEN and TMA), or using a familiar medium of communication with virtual humans, users were weakly emotionally impacted by their virtual interlocutors (see Figure 6). However, when participants in Taiwan interacted in English (TEN), a somewhat unfamiliar due to lack of frequent use but a known medium of communication, they exhibited significantly stronger emotions to their virtual interlocutors (see Figure 5.) There is evidence that speaking a second language can elicit an effect named Foreign language anxiety, also known as Xenoglossophobia. This is the feeling of unease, worry, nervousness and apprehension experienced in learning or using a second or foreign language. The feelings may stem from any second language context whether it is associated with the productive skills of speaking and writing or the receptive skills of reading and listening [29, 43]. Language anxiety could explain the emotions that participants in the TEN group showed. Even though the interactions with the virtual humans are basic, intrinsic aspects such as pronunciation perhaps could have had elicited embarrassment to the participants causing language anxiety. Overall, the stronger effects were found in the condition with the lowest language familiarity, namely TEN group. Independent of the emotional valence (positive, negative or neutral), These results are concurrent with the evidence that users when speaking a language that are not fully familiarized, they could experience a higher level of emotional impact. Furthermore, speaking in a second language could elicit higher visual attention and focus in order to understand better the message from the other person. Thus, the participants elicited higher emotional reactions. A related study showed that during simulated conversations with affective agents, users' visual attention is strongly related to emotional outcome [64]. Finally, it is important to highlight that the verbal and non-verbal behaviors that the virtual humans expressed in our crowd simulation are universally recognized. These behaviors were carefully controlled for eliminating confounding factors. However, a variable that we could not control to full extent and deserve further examination is the interpretation of these behaviors cross-culturally.

The current data analysis results show that language familiarity potentiated strongly the emotional contagion effect. This can be appreciated in the TEN group, where users in a language that is not their primary language (Taiwanese users interacting with the digital agents in English). On the other hand, users in USA that conversed in English and users in Taiwan that interacted in their native Mandarin showed weaker emotional effects in most DES dimensions. These results seem to suggest that language, rather than the agent's appearance or region culture, played a stronger role as an emotional contagion mediator.

The implications of this study are that designers of virtual agent simulations may need to consider the medium of communication or language familiar employed in natural dialogue, as it may potentially adversely affect the participants' emotional reactions in an unexpected manner, and may consider including some techniques to bridge the social distance and elevate the emotional state of the participants in those situations.

Our future work will focus on the influence of social conversations (such as small talk) that can improve the participants' positive emotional reactions in interactions with virtual agents in a foreign language. There is evidence that emotional engagement is important in learning [36, 53]. It would be compelling to design emotional agents which can elicit different levels of emotional arousal in learning systems that support native and foreign languages. Furthermore, personality traits of participants play an important role in our results. Recent studies show that the personality state [20] of a person depends on characteristics of the person and situational factors [28]. Future studies should classify users into groups of different traits for comparing the user responses.

Acknowledgments

This work was partly supported by grant no. MOST 109-2221-E-009-121-MY3 (Taiwan) and the Higher Education Sprout Project of the National Chiao Tung University and Ministry of Education (MOE), Taiwan. The authors would like to thank all the participants.

REFERENCES

[1] Social communication with virtual agents: The effects of body and gaze direction on attention and emotional responding in human observers. *International Journal of Psychophysiology*, 97(2):85 – 92, 2015.

[2] R. B. Adams Jr and R. E. Kleck. Perceived gaze direction and the processing of facial displays of emotion. *Psychological science*, 14(6):644–647, 2003.

[3] J. M. Allbeck and N. I. Badler. Creating crowd variation with the ocean personality model. 2008.

[4] G. Aschwanden, J. Halatsch, and G. Schmitt. Crowd simulation for urban planning. In *Proceedings of eCAADe*, pp. 493–500, 2008.

[5] N. Bee, S. Franke, and E. Andreé. Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–7. IEEE, 2009.

[6] M. H. Bond and T.-M. Lai. Embarrassment and code-switching into a second language. *Journal of Social Psychology*, 126(2):179–186, 1986.

[7] A. Borges Fortes Neto, C. Pelachaud, and S. R. Musse. Giving emotional contagion ability to virtual agents in crowds. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, and S. Kopp, eds., *Intelligent Virtual Agents*, pp. 63–72. Springer International Publishing, 2017.

[8] M. Bulut, S. Lee, and S. S. Narayanan. Analysis of emotional speech prosody in terms of part of speech tags. In *Eighth Annual Conference of the International Speech Communication Association*, pp. 1–4, 2007.

[9] C. L. Caldwell-Harris. Emotionality differences between a native and foreign language: theoretical implications. *Frontiers in psychology*, 5:1055, 2014.

[10] H. H. Clark. Communities, commonalities. *Rethinking linguistic relativity*, (17):324, 1996.

[11] H. H. Clark. *Using language*. Cambridge university press, 1996.

[12] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[13] K. L. Colbeck and J. S. Bowers. Blinded by taboo words in l1 but not l2. *Emotion*, 12(2):217, 2012.

[14] L. Coviello, Y. Sohn, A. D. Kramer, C. Marlow, M. Franceschetti, N. A. Christakis, and J. H. Fowler. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315, 2014.

[15] A. de Lima Bicho, R. A. Rodrigues, S. R. Musse, C. R. Jung, M. Paravisi, and L. P. Magalhães. Simulating crowds based on a space colonization algorithm. *Computers & Graphics*, 36(2):70 – 79, 2012.

[16] P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system: The manual on cd rom. *A Human Face, Salt Lake City*, pp. 77–254, 2002.

[17] P. Ekman and D. Keltner. Universal facial expressions of emotion. *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, pp. 27–46, 1997.

[18] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[19] V. N. Giri. Nonverbal communication theories. *Encyclopedia of Communication Theory*, pp. 690–694, 2009.

[20] D. Gundogdu, A. N. Finnerty, J. Staiano, S. Teso, A. Passerini, F. Pianesi, and B. Lepri. Investigating the association between social interactions and personality states dynamics. *Royal Society open science*, 4(9):170194, 2017.

[21] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold, M. Durbin, M. S. Edmonson, J. L. Fischer, D. Hymes, S. T. Kimball, W. La Barre, , J. E. McClellan, D. S. Marshall, G. B. Milner, H. B. Sarles, G. L. Trager, and A. P. Vayda. Proxemics. *Current Anthropology*, 9(2/3):83–108, 1968. doi: 10.1086/200975

[22] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004.

[23] C. L. Harris, A. Aycicegi, and J. B. Gleason. Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24(4):561, 2003.

[24] E. Hatfield, J. T. Cacioppo, and R. L. Rapson. Emotional contagion. *Current directions in psychological science*, 2(3):96–100, 1993.

[25] D. Helbing, I. J. Farkas, P. Molnar, and T. Vicsek. Simulation of pedestrian crowds in normal and evacuation situations. *Pedestrian and evacuation dynamics*, 21(2):21–58, 2002.

[26] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Physical review E*, 51:4282–4286, 1995.

[27] R. S. Hessels. How does gaze to faces support face-to-face interaction? a review and perspective. *Psychonomic Bulletin & Review*, pp. 1–26, 2020.

[28] K. T. Horstmann and M. Ziegler. Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality*, 34(6):1037–1059, 2020. doi: 10.1002/per.2266

[29] E. K. Horwitz, M. B. Horwitz, and J. Cope. Foreign language classroom anxiety. *The Modern language journal*, 70(2):125–132, 1986.

[30] P.-H. Huang and S.-K. Wong. Emotional virtual crowd on task completion in virtual markets. *Computer Animation and Virtual Worlds*, 29(3-4):e1818, 2018.

[31] C. E. Izard. *The Differential Emotions Scale: DES IV-A;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions]*. University of Delaware, 1993.

[32] C. E. Izard and P. B. Read. *Measuring Emotions in Infants and Children: Volume 1*. Number 3. Cambridge University Press, 1982.

[33] T. W. Jensen. Emotion in languaging: languaging as affective, adaptive, and flexible behavior in social interaction. *Frontiers in Psychology*, 5:720, 2014.

[34] H. Jiang, W. Xu, T. Mao, C. Li, S. Xia, and Z. Wang. Continuum crowd simulation in complex environments. *Computers & Graphics*, 34(5):537–544, 2010.

[35] S. K. Johnson. I second that emotion: Effects of emotional contagion and affect at work on leader and follower outcomes. *The Leadership Quarterly*, 19(1):1–19, 2008.

[36] W. L. Johnson, J. W. Rickel, J. C. Lester, et al. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education*, 11(1):47–78, 2000.

[37] P. Kanyuk. Virtual crowds in film and narrative media. *Simulating Heterogeneous Crowds with Interactive Behaviors*, p. 217, 2016.

[38] C. Kauschke, D. Bahn, M. Vesker, and G. Schwarzer. The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in Psychology*, 10:1654, 2019. doi: 10.3389/fpsyg.2019.01654

[39] B. Keysar, S. L. Hayakawa, and S. G. An. The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, 23(6):661–668, 2012.

[40] S. Kim, S. J. Guy, D. Manocha, and M. C. Lin. Interactive simulation of dynamic crowd behaviors using general adaptation syndrome theory. In *Proceedings of the ACM SIGGRAPH symposium on interactive 3D graphics and games*, pp. 55–62, 2012.

[41] C. Li, P. Lv, D. Manocha, H. Wang, Y. Li, B. Zhou, and M. Xu. Acsee: Antagonistic crowd simulation model with emotional contagion and evolutionary game theory. *IEEE Transactions on Affective Computing*, pp. 1–1, 2019. doi: 10.1109/TAFFC.2019.2954394

[42] F. Li and S. Wong. Animating agents based on radial view in crowd simulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 101–109, 2016.

[43] P. D. MacIntyre and R. C. Gardner. The subtle effects of language anxiety on cognitive processing in the second language. *Language learning*, 44(2):283–305, 1994.

[44] C. S. Oh, J. N. Bailenson, and G. F. Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:114, 2018.

[45] B. L. Omdahl and C. O'Donnell. Emotional contagion, empathic concern and communicative responsiveness as variables affecting nurses' stress and occupational commitment. *Journal of Advanced Nursing*, 29(6):1351–1359, 1999.

[46] C. Parkinson, T. T. Walker, S. Memmi, and T. Wheatley. Emotions are understood from biological motion across remote cultures. *Emotion*, 17(3):459, 2017.

[47] A. Pavlenko. Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6):405–428, 2012.

[48] T. Pejsa, S. Andrist, M. Gleicher, and B. Mutlu. Gaze and attention management for embodied conversational agents. *ACM Trans. Interact. Intell. Syst.*, 5(1):3:1–3:34, Mar. 2015.

[49] N. Pelechano. *Modeling realistic high density autonomous agent crowd movement: social forces, communication, roles and psychological influences*. University of Pennsylvania, 2006.

[50] M. A. Projects. Expressive speech, (accessed 31 August, 2020).

[51] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx. Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture. *Virtual Reality*, 17(4):307–321, 2013.

[52] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx. Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior*, 34:58–68, 2014.

[53] P. Redmond, L.-A. Abawi, A. Brown, R. Henderson, and A. Heffernan. An online engagement framework for higher education. *Online Learning*, 22(1):183–204, 2018.

[54] A. Rıos, D. Mateu, and N. Pelechano. Follower behavior in a virtual environment. *Virtual Humans and Crowds for Immersive Environments (VHCIE)*, 2018.

[55] A. Robb, R. Kopper, R. Ambani, F. Qayyum, D. Lind, L.-M. Su, and B. Lok. Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences. *IEEE transactions on visualization and computer graphics*, 19(4):662–670, 2013.

[56] L. Saifi, A. Boubetra, and F. Nouioua. An approach for emotions and behavior modeling in a crowd in the presence of rare events. *Adaptive Behavior*, 24(6):428–445, 2016.

[57] D. Sauter, F. Eisner, P. Ekman, and S. K. Scott. Universal vocal signals of emotion. In *31st Annual Meeting of the Cognitive Science Society (CogSci 2009)*, pp. 2251–2255. Cognitive Science Society, 2009.

[58] D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6):2408–2412, 2010.

[59] K. Schindler, L. Van Gool, and B. De Gelder. Recognizing emotions expressed by body pose: A biologically inspired neural model. *Neural networks*, 21(9):1238–1246, 2008.

[60] G. Schoenewolf. Emotional contagion: Behavioral induction in individuals and groups. *Modern Psychoanalysis*, 15(1):49–61, 1990.

[61] B. Ulicny and D. Thalmann. Crowd simulation for interactive virtual environments and vr training systems. In *Computer animation and simulation 2001*, pp. 163–170. Springer, 2001.

[62] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics research*, pp. 3–19. 2011.

[63] M. Volante, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics*, 22(4):1326–1335, 2016.

[64] M. Volonte, R. G. Anaraky, B. Knijnenburg, A. T. Duchowski, and S. V. Babu. Empirical evaluation of the interplay of emotion and visual attention in human-virtual human interaction. In *ACM Symposium on Applied Perception 2019*, pp. 1–9, 2019.

[65] M. Volonte, Y. Hsu, K. Liu, J. Mazer, S. Wong, and S. Babu. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 293–302, 2020.

[66] M. Volonte, A. Robb, A. T. Duchowski, and S. V. Babu. Empirical evaluation of virtual human conversational and affective animations on visual attention in inter-personal simulations. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 25–32. IEEE, 2018.

[67] S. Volynets, E. Glerean, J. K. Hietanen, R. Hari, and L. Nummenmaa. Bodily maps of emotions are culturally universal. *Emotion*, 2019.

[68] Q. Wang, M. Wedel, L. Huang, and X. Liu. Effects of model eye gaze direction on consumer visual processing: Evidence from china and america. *Information & Management*, 55(5):588–597, 2018.

[69] F. N. Willis. Initial speaking distance as a function of the speakers' relationship. *Psychonomic Science*, 5(6):221–222, 1966.

[70] D. Wolinski, M. Lin, and J. Pettré. Warpdriver: context-aware probabilistic motion prediction for crowd simulation. *ACM Trans. on Graph.*, 35(6):164, 2016.

[71] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino. Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics*, 20(4):626–635, 2014.

[72] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx. Robot mood is contagious: effects of robot body language in the imitation game. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 973–980. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[73] M. Xu, X. Xie, P. Lv, J. Niu, H. Wang, C. Li, R. Zhu, Z. Deng, and B. Zhou. Crowd behavior simulation with emotional contagion in unexpected multihazard situations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[74] M.-L. Xu, H. Jiang, X.-G. Jin, and Z. Deng. Crowd simulation and its applications: Recent advances. *Journal of Computer Science and Technology*, 29(5):799–811, 2014.