Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations

Pedro Guillermo Feijóo-García
Mohan Zalake
University of Florida
Gainesville, FL, U.S.A.
pfeijoogarcia@ufl.edu
mohanzalake@ufl.edu

Alexandre Gomes de Siqueira
Benjamin Lok
University of Florida
Gainesville, FL, U.S.A.
agomesdesiqueira@ufl.edu
lok@cise.ufl.edu

Felix Hamza-Lup
Georgia Southern University
Savannah, GA, U.S.A.
fhamzalup@georgiasouthern.edu

ABSTRACT

In the context of mental wellness support, trust and intimacy between a counselor and a patient are necessary to converge healing processes positively. However, convincing students to trust a virtual human for topics regarding mental wellness is a complex problem that requires understanding students' experiences. Based on research that discusses mental health as a concerning topic regarding Computer Science (CS) students, this paper investigates how undergraduate computing-related students perceive virtual humans' expertise on mental wellness support based on demographic resemblance on spoken accent and gender. Four virtual human counselors were developed to conduct the study, as 58 undergraduate computing-related students from two North American universities were recruited and assessed. Our findings suggest that students were less inclined to interact with a male virtual human than a female one. Also, that spoken accents can impact students' perceptions of expertise under students' multilingualism.

CCS CONCEPTS

• Computing methodologies ~ Virtual Reality

KEYWORDS

Virtual reality, culturally relevant computing, virtual humans, conversational agent

ACM Reference format:

Pedro Guillermo Feijoo-Garcia, Mohan Zalake, Alexandre Gomes de Siqueira, Benjamin Lok and Felix Hamza-Lup. 2021. Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations. In *Proceedings of the 21st International Conference on Intelligent Virtual Agents (IVA '21). October 14-17, 2021. Virtual Event, Kyoto, Japan. ACM, New York, NY, USA, 8 pages.* https://doi.org/10.1145/3472306.3478367

IVA '21, September 14–17, 2021, Virtual Event, Japan

ACM ISBN 978-1-4503-8619-7/21/09…$15.00
https://doi.org/10.1145/3472306.3478367

1 INTRODUCTION

Human-to-human relationships are complex and full of factors that affect how individuals trust each other. Factors may vary from demographic characteristics (e.g., same place of origin) to shared backgrounds and stories (e.g., a trip to Europe), and impact how much an individual opens and shares: how much intimacy can exist in a relationship. In the context of mental wellness support, trust and intimacy between a counselor and a patient are necessary to converge healing processes positively [41, 43, 21, 6]: a counselor needs to know the patient good enough to help, as the patient needs to trust enough to share their experiences and traumas. That trade-off of experiences and advice requires professional training in guidance and is generally piloted by counselors during multiple listening sessions to understand their patients. However, the lack of alliance between patients and counselors can lead to early processes' withdrawals [6, 41], leaving patients without the healing they need or that they were looking for.

The study reported in this paper contributes to the broader project RM^3, which stands for *"The right message, from the right messenger, at the right moment."* The project enables online learning students to interact with a virtual human inside their Learning Management System (LMS) to discuss mental health topics. The project's virtual humans (VHs) are proposed and designed as culturally relevant solutions that engage students in informing and connecting them to the right mental health resources when needed. Nevertheless, convincing students to trust in a virtual human for topics regarding mental wellness is a complex problem that requires understanding students' situational factors and the identification of barriers that arise when students' expectations do not match virtual humans' characteristics and behavior.

To contribute to the state of the art on the design of virtual humans for mental wellness support, we conducted a study to understand how individuals perceived virtual human's expertise, regarding two of the virtual human's characteristics: gender and spoken accent. We conducted our study with computing-related undergraduate students who interacted with virtual humans developed to introduce and explain "Gratitude Journaling" as a mental wellness technique [15]. We considered a computing-related context based on existing research on mental health in Computer Science Education (CS Ed) that discusses anxiety, depression, and the Impostor Phenomenon (IP) among CS students [35, 38]. Our study addressed the following research questions:

- RQ1: Can the spoken accent of a virtual human designed for mental wellness conversations impact how undergraduate computing-related students perceive the virtual human's expertise?
- RQ2: Can the gender of a virtual human designed for mental wellness conversations impact how undergraduate computing-related students perceive the virtual human's expertise?

For this study, we considered two hypotheses based on our research questions:

- H1: The spoken accent of a virtual human designed for mental wellness conversations can impact how undergraduate computing-related students perceive the virtual human's expertise.
- H2: The gender of a virtual human designed for mental wellness conversations can impact how undergraduate computing-related students perceive the virtual human's expertise.

2 BACKGROUND

2.1 Computing Education and Mental Wellness

Recent CS Ed studies have addressed the importance of mental wellness and counseling for CS students [35, 38]. Rosenstein et al., [35] studied the prevalence of the Impostor Phenomenon (IP) [7] among CS students in America by using the Clance IP Scale [14] with over 200 junior and senior undergraduate students. The authors found that 57.4% of participants met the diagnostic criteria for IP, also finding that IP was exposed statistically more prevalent in CS than in previously explored STEM disciplines. Similarly, Soares Passos et al., [38] studied the prevalence of anxiety and depression symptoms among 131 CS undergraduate students in Brazil, using two quantitative scales: The Beck Anxiety Inventory (BAI) [4] and the Beck Depression Inventory (BDI) [5]. The authors found that the prevalence of anxiety and depression symptoms was higher among Brazilian CS students than the general Brazilian population and more prevalent in female CS students. Most of the students with high BDA and BDI scores were not in treatment for anxiety or depression [38].

Our study contributes to the design and evaluation of virtual humans to serve counseling and mental wellness for computing-related students. We report how the match and mismatch of human-centered factors such as gender and spoken accent influence computing-related students' judgment with virtual humans designed for mental health support.

2.2 Gender and Virtual Humans

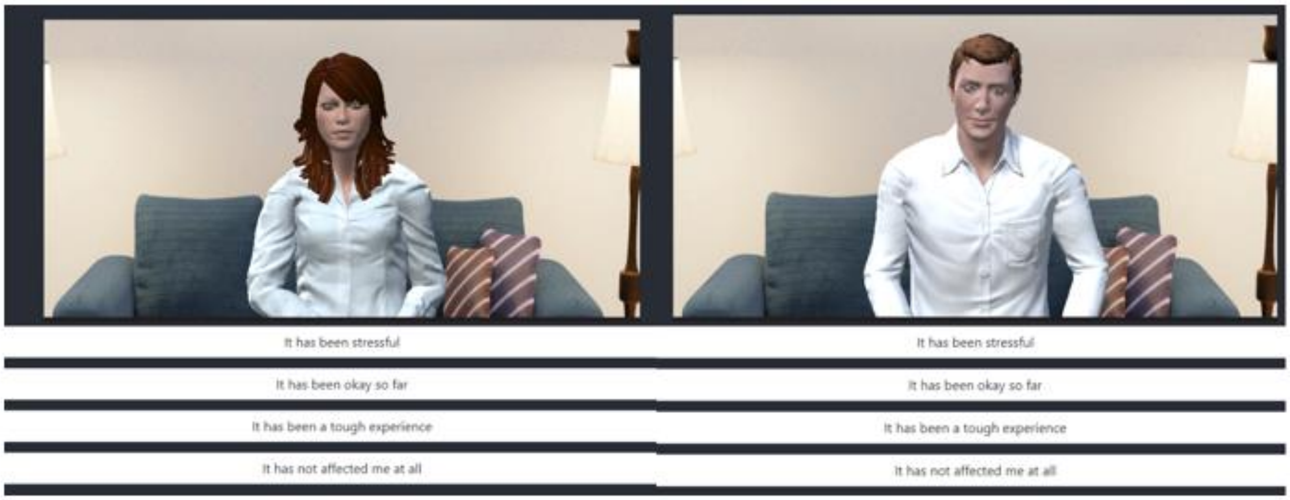
The role of gender in mental wellness scenarios and its impact on mental disorders treatments has been explored in different studies [41, 43, 21, 6]. Some of them have found gender-matching (i.e., same gender between the patient and the therapist) not to influence the outcomes of treatment [41, 21], as some others have found otherwise [43, 6]. For example, Wintersteen et al., [41] conducted a study with over 600 adolescent substance abusers. They observed that gender-matching led to higher earlier alliances than gender-mismatch: male adolescents tended to stay longer in treatment when matched with a male therapist. Contrarily, Zlotnick et al., [43] observed that among patients with depression, gender-matching did not significantly impact the level of depression at the termination of treatment, nor attrition rates, or patients' perceptions about their therapists. In the HCI community, several studies have evaluated the effects virtual humans' gender has on virtual humans' credibility [26, 29, 34, 24, 40]. Similarly, some of them have found virtual humans' gender to impact how they are perceived [26, 29, 34, 24], as some others have reported no effects [40, 37]. In general, as it can be seen in Richards et al., [34] work, assessing virtual humans' characteristics is a complex task that not necessarily outcomes a "silver bullet." Hence, it is essential to understand users, characterize them, and conceive human-centered designed solutions.

Regarding our study's scope, our literature review found no previous work that assessed virtual humans concerning gender in the context of counseling or mental wellness. This literature gap means that our study contributes to the HCI community in the assessment of virtual humans designed especially for counseling, considering their gender and speech cues such as their accent.

2.3 Spoken Accent and Virtual Humans

Multiple studies have reported on the impact accent has on individuals' credibility and judgment in different spoken dialogue scenarios, exploring positive and negative biases in regard to accent's foreignness. Regarding native English listeners in America, some studies have found them presenting negative biases towards foreign accents, with native English listeners finding non-native English speakers' statements less credible [31, 37, 27, 28]. For example, Lev-Ari and Keysar [31] came to that conclusion after assessing adult undergraduate American native English listeners towards three types of accents: American English (i.e., native), mild non-native (e.g., German), and heavy non-native (e.g., Korean). Kinzler et al., [27] came to similar findings with American preschool-aged children, who preferred native-accented speakers to foreign-accented ones for tasks involving linguistic and object recognition. Their work relates to Kinzler et al., previous work [28], which reports on how months-old infants and young children preferred looking at people who had previously spoken to them with a native accent.

Other research has referred to international populations and has found that accent impacts how listeners believe in speakers [20, 17, 23, 16, 22]. For example, Hanzlíková and Skarnitzl [20] revisited Lev-Ari and Keysar [31] with anglophone Czech participants. They found that non-native English listeners were sensitive to non-native English speakers, generally favoring statements coming from native English speakers. In another context, Frances et al., [17] assessed Spaniards native to Barcelona, Spain, towards different Spanish regional accents from Spain and Latin-America.



(a) **(b)**

Figure 1. VIP System – Counseling Virtual Humans

Their results did not find speakers' accents to impact memory or credibility. Nevertheless, the authors explain that their findings could be due to intelligibility. When intelligibility is high and cannot be attributed to linguistic competence, listeners may not perceive a different accent as an indicator of less intellectual ability. Jiang et al., [23] and Foucart et al., [16] went one step further and assessed non-American participants psychophysically with event-related potentials (ERPs). Foucart et al., [16] observed that foreign accents negatively impacted how speakers were evaluated on social variables such as affect, status, and solidarity. Jiang et al. [23] observed that listeners presented increased neural activity by accents familiar to them and discussed how the lack of experience with foreign accents could increase processing needs due to reduced intelligibility. Jiang et al., [23] based their work on Hatzidaki et al., [22], who argued on how listeners discontinued processing social information when facing foreign-accented speech.

Human-Computer Interaction (HCI) researchers have also explored the role that speech has on listeners' perceptions of spoken interfaces' credibility [11, 9, 8, 36], and some of them have done so concerning virtual humans [25, 33, 10]. Dahlbäck et al., [11], for example, explored the similarity-attraction effect between individuals' region and spoken web interfaces' accents to find that users preferred accents similar to their own. That same similarity-attraction effect was observed by Cowan et al. [9] regarding navigation systems' credibility, as their Irish users rated as more trustworthy an Irish accented system than an American accented one. On the other hand, Khooshabeh et al., [25] studied the effect accents posed on virtual humans' credibility. The authors observed that users were more likely to make decisions congruent with their cultural customs after interacting with a culturally-related virtual human—they used accent as the primary paralinguistic cue.

In the context of our study, we contribute by evaluating how accent impacts the way users perceive virtual humans' expertise when the agents are designed to assist in mental wellness counseling. Our literature review found no previous work that assessed virtual humans concerning language nor accent, especially for counseling or mental health assistance. With our study, we pose a novelty contribution to the state-of-the-art to explore strategies towards human-centered design of virtual humans for counseling and mental wellness support.

3 SYSTEM DESIGN

We designed and developed four virtual humans (VHs) using the Virtual Interviewer Platform (VIP) for the study. As described by Zalake et al., the VIP is an internet-based system designed for rapid prototyping of real-time VHs [42]. The VIP system offers a repository of 3D VHs and features a Conversational Script Editor that integrates Ink [3] to allow designers to create and edit VHs' dialogues. The system's architecture includes WebGL to enable cross-platform compatibility [30].

For this study, four VHs were designed to foster mental wellness conversations. They varied regarding their biological sex—female (figure 1-a) or male (figure 1-b), and their spoken English accent. Four accents were used from the Google Cloud Text-to-Speech API [2]: 1) American Female—en-US-Standard-G (FEMALE), 2) International Female—de-DE-Standard-A (FEMALE), 3) American Male—en-US-Standard-D (MALE), and 4) International Male—de-DE-Wavenet-B (MALE). We selected German accents for the international VHs due to the high resemblance between English and German in terms of intonation [19]. Participants were not informed about the accents they were exposed to. We compared VHs with similar spoken intonations (German and English) to avoid potential biases regarding participants' assumptions on VHs' demographics.

All four VHs where light-skinned and addressed the same topic with a constant script on "Gratitude Journaling" [15]. The color of the VHs' skin was decided based on counseling services' demographics in the U.S.A [1], which indicate that most counselors in the U.S.A are white non-Hispanic professionals. The system provided participants with an interface that featured one virtual human at a time to interview with. The interview structure consisted of close-ended multiple-option questions, which the participants responded to using on-screen buttons (see Figure 1). The VHs were in a sitting pose during the interview and had an idle breathing animation.

4 METHOD

We conducted an online-asynchronous study in Spring-2021 with undergraduate sophomore and senior computing-related students at two universities in North America. We collected data via four questionnaires distributed between four groups of students, each one featuring one virtual human. Students' participation was no longer than 60 min for the study, and all communication with them was conducted via email.

4.1 Participants

We report on data that corresponds to a sample population consisting of undergraduate sophomore and senior computing-related-major students (n=58) from two universities in North America. The recruitment process was open to any participant from these colleges regardless of their age, gender, ethnicity, language, or place of origin. Participants' demographics were collected after interacting with the study's virtual humans: there was no prior filter to balance groups regarding demographic characteristics.

Participants' reported age was: 18-20 years of age (n=30), 21-25 years of age (n=21), 26-30 years of age (n=4), and >30 years of age (n=3). We asked participants to report their place of origin (i.e., city and country) and how many languages they were familiar to. Participant's reported place of origin distributed the sample population between US national students (n=44), international students (n=10), and a group of students who did not report their place of origin (n=4). Similarly, students' reported languages distributed the sample population between two groups: monolingual speakers (n=30)—English as the main language, and multilingual speakers (n=28).

We asked participants to report their gender (Female, Male, and non-Binary), as also their ethnicity. Participant's reported gender distributed the sample population between Female (n=24) and Male (n=34). The ethnical groups we obtained were majorly White, Non-Hispanic/Latin American (n=28), Hispanic/Latin American (n=14), and Asian/Pacific Islander (n=11). Five participants reported to belong to other ethnical groups (n=5).

Female participants' reported ethnicity was: White, Non-Hispanic/Latin American (n=10), Hispanic/Latin American (n=7), Asian/Pacific Islander (n=6), and one participant reported other ethnical groups (n=1). On the other hand, male participants' reported ethnicity was: White, Non-Hispanic/Latin American (n=18), Hispanic/Latin American (n=7), Asian/Pacific Islander (n=5), and four participants reported other ethnical groups (n=4).

4.2 Study Design

The study was an online, asynchronous, and two-factorial between-subjects. Participants were recruited from two universities in North America. First, students were given an informed consent document with the details of the study. They were asked to read it and respond to an online survey if they agreed to participate. Participants who agreed to take part in the study were distributed between four groups, each one introducing one sole virtual human (section 3):

- **Group 1 (G1)** introduced a light-skinned female virtual human with an American English accent.
- **Group 2 (G2)** introduced a light-skinned male virtual human with an American English accent.
- **Group 3 (G3)** introduced a light-skinned female virtual human with a non-native English accent.
- **Group 4 (G4)** introduced a light-skinned male virtual human with a non-native English accent.

Table 1 describes the four groups and their participants based on gender and ethnicity. Table 2 describes the four groups based on participants' gender and multilingualism.

Table 1. Participants per Group—Gender and Ethnicity

Gender	*Ethnicity*	*G1 (n=17)*	*G2 (n=11)*	*G3 (n=13)*	*G4 (n=17)*
Female	*Asian / Pacific Islander*	*0*	*4*	*1*	*1*
	White, Non-Hispanic/ Latin American	*2*	*1*	*4*	*3*
	Hispanic / Latin American	*2*	*1*	*1*	*3*
	Other	*1*	*0*	*0*	*0*
Male	*Asian / Pacific Islander*	*1*	*2*	*1*	*1*
	White, Non-Hispanic/ Latin American	*7*	*1*	*4*	*6*
	Hispanic / Latin American	*2*	*1*	*2*	*2*
	Other	*2*	*1*	*0*	*1*

Table 2. Participants per Group—Gender and Multilingualism

Participants per Group - Gender and Ethnicity					
Gender	***Ethnicity***	***Group 1 (n=17)***	***Group 2 (n=11)***	***Group 3 (n=13)***	***Group 4 (n=17)***
Female	***Monolingual***	***2***	***2***	***2***	***3***
	Multilingual	***3***	***4***	***4***	***4***
Male	***Monolingual***	***7***	***2***	***5***	***7***
	Multilingual	***5***	***3***	***2***	***3***

We then asked participants to respond to an online questionnaire of 21 questions that presented three sections in sequence:

1. The **first section** asked seven demographic questions about participants' age, gender, race, ethnicity, place of origin, and spoken languages.
2. The **second section** included a link to the virtual human participants were asked to interact with. The interaction with the virtual human was no longer than 10 min and it addressed "Gratitude Journaling" [15] as the main topic. After interacting with the virtual human, participants were asked to respond to three questions:
    a. Multiple-option close-ended question: What is the topic the virtual human introduced?
    b. 7-point Likert Scale: "The virtual human was an expert on the topic s/he introduced"
    c. Open-ended question: "Please explain your response"
3. The **third section** had five close-ended and six open-ended follow-up questions on participants' perceptions about virtual human's characteristics (e.g., gender, race, ethnicity), as also on users' perceptions about expertise, empathy, and bonding (i.e., create and maintain emotional relations). For the extend of this paper—centered on expertise regarding gender and spoken accent, only two close-ended questions are analyzed in addition to their corresponding follow-up open-ended questions (i.e., "please explain your response"). Responses to the additional questions are intended for future publications. The two questions considered in this paper are:
    a. Close-ended multiple-choice question: In the case of a mental wellness conversation, which kind of a virtual human would you prefer to talk to?
    b. Close-ended 7-scale question: How important do you consider is expertise (i.e., knowledge in a particular field) as a factor to trust in a virtual human?

4.3 Data Analysis

To respond to the questions posed in Section 1, we broke down our analysis into two parts. The first part corresponds to students' responses to the 7-point Likert scales that assessed their perceptions on the VHs' expertise. We began our analysis on students' responses by comparing the four VHs based on their characteristics, initially ignoring students' demographics. Then, we followed up by exploring how demographic subgroups of students behaved between one and another towards the VHs' characteristics. We compared female and male students who interacted with female and male VHs, as also monolingual and multilingual students regarding the VHs' spoken accent. Due to Likert scales' ordinal structure, we conducted our analysis using the non-parametric independent samples Mann-Whitney U test [12] to compare groups and subgroups of participants, two at a time. The Inferential statistics considered the whole set of responses from 1-score to 7-score. Additionally, we considered the most frequent score (i.e., the mode) per Likert scale, as also the frequency of the three highest (5, 6, & 7) and three lowest (1, 2, & 3) options. We did that additional analysis to observe participants' perceptions out of the Likert scale's middle-point (4-score): middle-point responses could mean ambivalence [18] or lack of motivation [13, 32].

The second part refers to students' perceptions on expertise and gender regarding VHs characteristics. Our analysis was percentual and based on participants' gender, and multilingualism. The responses analyzed in this part responded to the third section of the questionnaire (section 4.2)—for the second question we used the non-parametric independent samples Mann-Whitney U test [12] to compare demographic subgroups.

5 FINDINGS AND RESULTS

This section presents findings and results according to the structure described in section 4.3. The four VHs were introduced in section 3 and described in section 4.2.

5.1 Perceptions on Accent and Expertise

Table 3 shows responses' distribution on participants' perceptions of expertise regarding the four VHs that were evaluated. Male virtual humans—G2 and G4, presented higher frequencies of high responses and lower frequencies of low responses than their female counterparts—G1 and G3. Virtual humans who posed an American English accent—G1 and G2, exposed higher frequencies of high responses than the two VHs who posed a non-native English accent—G3 and G4. The female non-native English speaking virtual human—G3 presented the lowest frequency of high responses (69.23%) and the highest frequency of low responses (23.08%). On the other hand, the male American English-speaking virtual human—G2, scored the highest frequency of high responses (81.82%) and the lowest frequency of low responses (9.09%). G2 was also the only virtual human to have seven (36.36%) as one of the two most frequent responses, next to five (36.36%). Nevertheless, we compared the best-rated virtual human—G2, with the worst-rated one—G3, and found no significant difference between them (Mann-Whitney U Test, $p>0.05$).

Table 3. Participants' Perceptions—Groups' Comparison

Group	*High Res. Freq. [%] (5, 6, & 7)*	*Low Res. Freq. [%] (1, 2, & 3)*	*(Mode, Frequency [%])*
G1 (n=17)	*70.59%*	*11.76%*	*(6, 35.29%)*
G2 (n=11)	*81.82%*	*9.09%*	*(7, 36.36%), (5, 36.36%)*
G3 (n=13)	*69.23%*	*23.08%*	*(5, 38.46%)*
G4 (n=17)	*70.59%*	*11.76%*	*(6, 41.17%)*

Table 4 presents responses' frequencies on participants' perceptions of expertise based on participants' skills with multiple languages—monolingual or multilingual. Monolingual participants' rates varied from American English-speaking VHs and non-native English-speaking VHs, being apparently high towards American English-speaking VHs—G1 and G2. However, we found no significant difference between VHs for monolingual subgroups (Mann-Whitney U Test, p> 0.05). We found that for the first three groups—G1, G2, and G3, no significant difference was found between monolingual or multilingual participants—(Mann-Whitney U Test, p> 0.05). Nevertheless, towards the male non-native English-speaking virtual human—G4, we found a significant difference between monolingual and multilingual participants—(Mann-Whitney U Test, p<0.05).

Table 4. Perceptions—Monolinguals and Multilinguals

Group	*Subgroup*	*High Res. Freq. (5, 6, & 7) [%]*	*Low Res. Freq. (1, 2, & 3) [%]*	*(Mode, Frequency [%])*
G1 (n=17)	*Mono. (n=9)*	*66.67%*	*11.11%*	*(5, 33.33%)*
	Mult. (n=8)	*75.00%*	*12.50%*	*(6, 50.00%)*
G2 (n=11)	*Mono. (n=4)*	*100.00%*	*0.00%*	*(7, 75.00%)*
	Mult. (n=7)	*71.43%*	*14.29%*	*(5, 57.14%)*
G3 (n=13)	*Mono. (n=7)*	*71.43%*	*14.29%*	*(5, 42.86%)*
	Mult. (n=6)	*66.67%*	*33.33%*	*(6, 33.33%), (5, 33.33%), (3, 33.33%)*
G4 (n=17)	*Mono. (n=10)*	*50.00%*	*20.00%*	*(6, 33.33%), (4, 33.33%)*
	Mult. (n=7)	*100.00%*	*0.00%*	*(6, 57.14%)*

Table 5. Participants' Perceptions—Females vs Males

Group	*Subgroup*	*High Res. Freq. (5, 6, & 7) [%]*	*Low Res. Freq. (1, 2, & 3) [%]*	*(Mode, Frequency [%])*
G1 (n=17)	*Female (n=5)*	*80.00%*	*20.00%*	*(6, 60.00%)*
	Male (n=12)	*75.00%*	*8.33%*	*(5, 41.67%)*
G2 (n=11)	*Female (n=6)*	*100.00%*	*0.00%*	*(5,50.00%)*
	Male (n=5)	*60.00%*	*20.00%*	*(7, 40.00%)*
G3 (n=13)	*Female (n=6)*	*83.33%*	*16.67%*	*(6, 33.33%), (5, 33.33%)*
	Male (n=7)	*66.67%*	*28.57%*	*(5, 42.86%)*
G4 (n=17)	*Female (n=7)*	*71.43%*	*14.29%*	*(6, 28.57%), (5, 28.57%)*
	Male (n=10)	*70.00%*	*10.00%*	*(6, 50.00%)*

Table 5 reports responses' frequencies on participants' perceptions of expertise based on participants' gender—female or male. In general, female participants were more into providing higher rates compared to the male participants. None of the groups (G1 to G4) presented a significant difference between female or male participants— (Mann-Whitney U Test, p> 0.05).

5.2 Perceptions on Gender and Expertise

Figure 2 presents participants' preferences about VHs' gender—female or male. As it can be seen, 50.00% (n=29) of participants indicated that they were okay with any of both genders, while 37.93% (n=22) preferred a female virtual human. Only 6.90% (n=4) of participants preferred a male virtual human—one female and three males, and 5.17% (n=3) indicated preferring none of them—two males and one female.

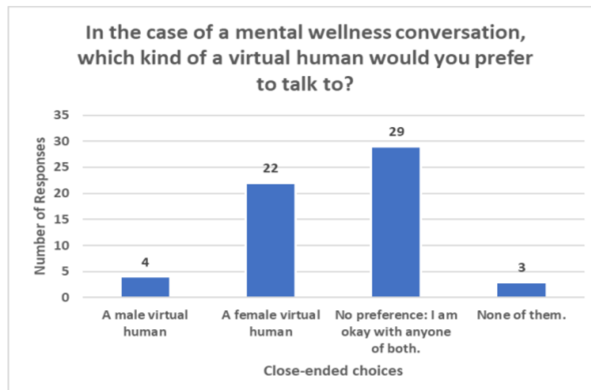


Figure 2. Virtual Humans and Gender Preferences

Table 6. Perceptions about Expertise—Demographic Distribution

	Demographic Distribution [%]			
7-point Likert Scale score	***Multilingual (n=28)***	***Monolingual (n=34)***	***Female (n=24)***	***Male (n=34)***
1	***3.57%***	***0.00%***	***0.00%***	***2.94%***
2	***0.00%***	***3.33%***	***0.00%***	***2.94%***
3	***10.71%***	***10.00%***	***12.50%***	***12.50%***
4	***7.14%***	***20.00%***	***8.33%***	***17.65%***
5	***32.14%***	***30.00%***	***30.00%***	***32.35%***
6	***39.29%***	***20.00%***	***33.33%***	***26.47%***
7	***7.14%***	***16.67%***	***16.67%***	***8.82%***

Table 6 presents participants' perceptions on how important they considered expertise to be as a factor to trust in a virtual human. Responses from multilingual participants (n=28) were not significantly different (Mann-Whitney U Test, p>0.05) than those from monolingual participants (n=30). However, multilingual participants had a higher frequency of responses (46.43%) that rated between six (6) and seven (7) on the Likert scale, compared to monolingual participants (36.67%). Female participants (n=24) were not significantly different (*Mann-Whitney U Test,* p>0.05) from male participants (n=34). However, female participants reported a higher frequency of responses (50.00%) that rated between six (6) and seven (7) on the Likert scale, compared to male participants (35.29%).

6 CONCLUSIONS AND FUTURE WORK

Based on the findings and results presented in this paper, we obtained a positive answer to RQ1—*Can the spoken accent of a virtual human designed for mental wellness conversations impact how undergraduate computing-related students perceive the virtual human's expertise?* Hence, **H1 is accepted:** The spoken accent of a virtual human can impact how computing-related students perceive the virtual human's expertise. Our conclusion is based on the analysis presented in section 5.1 in regard to the male non-native English-speaking virtual human—G4: multilingual students rated higher the non-native English-speaking virtual human. Although that outcome was also expected for the female non-native English-speaking virtual human—G3, we believe that the lack of significant difference for G3 may respond to the students' low preference towards male virtual humans (see Figure 2). Further research may explore the previous rationale to also understand why could accents impact users' perceptions on expertise.

On the other hand, we obtained a negative answer to RQ2—*Can the gender of a virtual human designed for mental wellness conversations impact how undergraduate computing-related students perceive the virtual human's expertise?* Therefore, **H2 is rejected:** Our study did not find any significant difference in students' expertise perception, between groups and demographic subgroups, regarding participants' gender and virtual humans' gender resemblance. However, these findings respect what students responded to in Figure 2: 50% of participants explicitly indicated no preference for gender. For this study, we assessed the question in Figure 2 in the third and last section of the core questionnaire (see Section 4.2): participants were randomly assigned between groups. Further research will attempt to characterize participants ahead to their interactions, to assess gender exclusively as a core virtual human's characteristic.

Currently, the VIP system (Section 3) features text-based and multiple-choice close-ended interaction with our VHs. Next steps will consider spoken open-ended interactions with the VHs: how the interlocution between students and VHs impact perceptions of expertise. Moreover, we will evaluate students' interactions with a more varied catalog of spoken accents to find similarities and differences that can be translated into design frameworks of VHs.

ACKNOWLEDGMENTS

The authors extend their gratitude to Dr. Sanethia Thomas, Ph.D., for her help in recruiting students for this study. Additionally, the authors thank the students who actively participated in this study. Their participation was essential to our research and findings.

REFERENCES

[1] [n.d.]. Counselors. https://datausa.io/profile/soc/counselors
[2] [n.d.]. Speech-to-Text: Automatic Speech Recognition | Google Cloud. https://cloud.google.com/speech-to-text
[3] [n.d.]. What's the difference between inklewriter and ink? https://www.inklestudios.com/ink
[4] Aaron T Beck, Norman Epstein, Gary Brown, and Robert A Steer. 1988. An inventory for measuring clinical anxiety: psychometric properties. Journal of consulting and clinical psychology 56, 6 (1988), 893.
[5] Aaron T Beck, Robert A Steer, and Margery G Carbin. 1988. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical psychology review 8, 1 (1988), 77–100.
[6] Adrian J Blow, Tina M Timm, and Ronald Cox. 2008. The role of the therapist in therapeutic change: Does therapist gender matter? Journal of Feminist Family Therapy 20, 1 (2008), 66–86.
[7] Pauline Rose Clance and Maureen Ann OToole. 1987. The imposter phenomenon: An internal barrier to empowerment and achievement. Women & Therapy 6, 3 (1987), 51–64.
[8] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. Interacting with Computers 31, 4 (2019), 349–371.
[9] Benjamin R Cowan, Derek Gannon, Jenny Walsh, Justin Kinneen, Eanna O'Keefe, and Linxin Xie. 2016. Towards Understanding How Speech Output Affects Navigation System Credibility. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2805–2812.
[10] Scotty D Craig and Noah L Schroeder. 2017. Reconsidering the voice effect when learning from a virtual human. Computers & Education 114 (2017), 193–205.
[11] Nils Dahlbäck, QianYing Wang, Clifford Nass, and Jenny Alwin. 2007. Similarity is more important than expertise: Accent effects in speech interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems. 1553–1556.
[12] JFC De Winter and Dimitra Dodou. 2010. Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). Practical Assessment, Research, and Evaluation 15, 1 (2010), 11.
[13] Bernard DuBois and John A Burns. 1975. An analysis of the meaning of the question mark response category in attitude scales. Educational and Psychological Measurement 35, 4 (1975), 869–884.
[14] Patrick W Edwards, Amos Zeichner, Norma Lawler, and Rachel Kowalski. 1987. A validation study of the Harvey Impostor Phenomenon Scale. Psychotherapy: Theory, Research, Practice, Training 24, 2 (1987), 256.
[15] Carol L Flinchbaugh, E Whitney G Moore, Young K Chang, and Douglas R May. 2012. Student well-being interventions: The effects of stress management techniques and gratitude journaling in the management education classroom. Journal of Management Education 36, 2 (2012), 191–219.
[16] Alice Foucart, Albert Costa, Luis Morís-Fernández, and Robert J Hartsuiker. 2020. Foreignness or processing fluency? On understanding the negative bias toward foreign-accented speakers. Language Learning 70, 4 (2020), 974–1016.
[17] Candice Frances, Albert Costa, and Cristina Baus. 2018. On the effects of regional accents on memory and credibility. Acta psychologica 186 (2018), 63–70.
[18] Lewis R Goldberg. 1981. Unconfounding situational attributions from uncertain, neutral, and ambiguous ones: A psychometric analysis of descriptions of oneself and various types of others. Journal of Personality and Social Psychology 41, 3 (1981), 517.
[19] Esther Grabe. 1997. Comparative intonational phonology: English and German. In Intonation: Theory, models and applications.
[20] Dagmar Hanzlíková and Radek Skarnitzl. 2017. Credibility of native and nonnative speakers of English revisited: Do non-native listeners feel the same? Research in Language 15, 3 (2017), 285–298.
[21] Sherry L Hatcher, Todd K Favorite, Elizabeth A Hardy, Robert L Goode, Linda A DeShetler, and Rosa M Thomas. 2005. An Analogue Study of Therapist Empathic Process: Working With Difference. Psychotherapy: Theory, Research, Practice, Training 42, 2 (2005), 198.
[22] Anna Hatzidaki, Cristina Baus, and Albert Costa. 2015. The way you say it, the way I feel it: emotional word processing in accented speech. Frontiers in psychology 6 (2015), 351.
[23] Xiaoming Jiang, Kira Gossack-Keenan, and Marc D Pell. 2020. To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. Quarterly Journal of Experimental Psychology 73, 1 (2020), 55–79.
[24] Bilge Karacora, Morteza Dehghani, Nicole Kramer-Mertens, and Jonathan Gratch. 2012. The influence of virtual agents' gender and rapport on enhancing

math performance. In Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 34.
[25] Peter Khooshabeh, Morteza Dehghani, Angela Nazarian, and Jonathan Gratch. 2017. The cultural influence model: When accented natural language spoken by virtual characters matters. AI & society 32, 1 (2017), 9–16.
[26] Yanghee Kim, Amy L Baylor, and Entong Shen. 2007. Pedagogical agents as learning companions: the impact of agent emotion and gender. Journal of Computer Assisted Learning 23, 3 (2007), 220–234.
[27] Katherine D Kinzler, Kathleen H Corriveau, and Paul L Harris. 2011. Children's selective trust in native-accented speakers. Developmental science 14, 1 (2011), 106–111.
[28] Katherine D Kinzler, Emmanuel Dupoux, and Elizabeth S Spelke. 2007. The native language of social cognition. Proceedings of the National Academy of Sciences 104, 30 (2007), 12577–12580.
[29] Nicole C Krämer, Bilge Karacora, Gale Lucas, Morteza Dehghani, Gina Rüther, and Jonathan Gratch. 2016. Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. Computers & Education 99 (2016), 1–13.
[30] Catherine Leung and Andor Salga. 2010. Enabling webgl. In Proceedings of the 19th international conference on World wide web. 1369–1370.
[31] Shiri Lev-Ari and Boaz Keysar. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. Journal of experimental social psychology 46, 6 (2010), 1093–1096.
[32] Stephen M Nowlis, Barbara E Kahn, and Ravi Dhar. 2002. Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. Journal of Consumer research 29, 3 (2002), 319–334.
[33] David Obremski, Jean-Luc Lugrin, Philipp Schaper, and Birgit Lugrin. 2021. Nonnative speaker perception of Intelligent Virtual Agents in two languages: the impact of amount and type of grammatical mistakes. Journal on Multimodal User Interfaces (2021), 1–10.
[34] Deborah Richards, Bayan Alsharbi, and Amal Abdulrahman. 2020. Can I help you? Preferences of young adults for the age, gender and ethnicity of a Virtual Support Person based on individual differences including personality and psychological state. In Proceedings of the Australasian Computer Science Week Multiconference. 1–10.
[35] Adam Rosenstein, Aishma Raghu, and Leo Porter. 2020. Identifying the prevalence of the impostor phenomenon among computer science students. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education. 30–36.
[36] Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob Van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, et al. 2015. A survey on perceived speaker traits: Personality, likability, pathology, and the first challenge. Computer speech & language 29, 1 (2015), 100–131.
[37] Youssef Shiban, Iris Schelhorn, Verena Jobst, Alexander Hörnlein, Frank Puppe, Paul Pauli, and Andreas Mühlberger. 2015. The appearance effect: Influences of virtual agent features on performance and motivation. Computers in Human Behavior 49 (2015), 5–11.
[38] Lígia Maria Soares Passos, Christian Murphy, Rita Zhen Chen, Marcos Gonçalves de Santana, and Giselle Soares Passos. 2020. The Prevalence of Anxiety and Depression Symptoms among Brazilian Computer Science Students. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education. 316–322.
[39] Ladina Stocker. 2017. The impact of foreign accent on credibility: An analysis of cognitive statement ratings in a Swiss context. Journal of Psycholinguistic Research 46, 3 (2017), 617–628.
[40] Justyna Swidrak and Grzegorz Pochwatko. 2019. Being Touched by a Virtual Human. Relationships Between Heart Rate, Gender, Social Status, and Compliance.. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 49–55.
[41] Matthew B Wintersteen, Janell L Mensinger, and Guy S Diamond. 2005. Do gender and racial differences between patient and therapist affect therapeutic alliance and treatment retention in adolescents? Professional Psychology: Research and Practice 36, 4 (2005), 400.
[42] Mohan Zalake, Alexandre Gomes de Siqueira, Krishna Vaddiparti, Pavlo Antonenko, Felix Hamza-Lup, and Benjamin Lok. 2020. Towards Rapid Development of Conversational Virtual Humans Using Web3D Technologies. In The 25th International Conference on 3D Web Technology (Virtual Event, Republic of Korea) (Web3D '20). Association for Computing Machinery, New York, NY, USA, Article 34, 2 pages. https://doi.org/10.1145/3424616.3424727.
[43] Caron Zlotnick, Irene Elkin, and M Tracie Shea. 1998. Does the gender of a patient or the gender of a therapist affect the treatment of patients with major depression? Journal of Consulting and Clinical Psychology 66, 4 (1998), 655.