Does the Voice Reveal More Emotion than the Face? a Study with Animated Agents

Joshua E. Heller, Nicoletta Adamo(✉), Nandhini Giri, and Derek Larson

Purdue University, West Lafayette, IN 47907, USA
{amehrez,nadamovi,girin,dglarson}@purdue.edu

Abstract. In general, people tend to identify the emotions of others from their facial expressions, however recent findings suggest that we may be more accurate when we hear someone's voice than when we look only at their facial expression. The study reported in the paper examined whether these findings hold true for animated agents. A total of 37 subjects participated in the study: 19 males, 14 females, and 4 of non-specified gender. Subjects were asked to view 18 video stimuli; 9 clips featured a male agent and 9 clips a female agent. Each agent showed 3 different facial expressions (happy, angry, neutral), each one paired with 3 different voice lines spoken in three different tones (happy, angry, neutral). Hence, in some clips the agent's tone of voice and facial expression were congruent, while in some videos they were not. Subjects answered questions regarding the emotion they believed the agent was feeling and rated the emotion intensity, typicality, and sincerity. Findings showed that emotion recognition rate and ratings of emotion intensity, typicality and sincerity were highest when the agent's face and voice were congruent. However, when the channels were incongruent, subjects identified the emotion more accurately from the agent's facial expression than the tone of voice.

Keywords: Facial Expressions · Animated Agents · Perception · Emotions

1 Introduction

Research has shown that animated pedagogical agents (APA) can effectively pro- mote learning (Schroeder et al. 2013; Adamo et al. 2021). However, many questions remain unanswered, particularly concerning their emotional design. With a growing understanding of the complex interplay between emotions and cognition, there is a need to develop life-like agents that provide both effective expert guidance and convincing emotional interactions with the learner (Kim and Baylor 2007). One goal of our research is to develop APAs that can convey clearly perceivable emotions through speech, facial expressions, and body gestures. The study conducted for this paper was a step in this direction, as it focused on how emotions are expressed through voice and face.

Human emotions can be expressed using several modalities: vocal and facial expressions, arm and hand gestures, trunk rotation, head rotation, and leg movements. The face is cited as being the most used resource in identification of emotions (Noroozi et al. 2018).

X. Fang (Ed.): HCII 2023, LNCS 14047, pp. 229–242, 2023.
https://doi.org/10.1007/978-3-031-35979-8_18

However, research by Kraus (2017), suggests that people may be more accurate when hearing another's voice than when solely considering their facial expression. The study reported in the paper examined whether Kraus' findings hold true for animated agents. More specifically, it investigated whether animated agents' facial expressions are a more effective channel for conveying emotions that the tone of voice. The study examined perception of animated agents' emotions in the context of multi-sense communication, e.g., voice + facial expression.

2 Related Work

In animated agents, emotions can be expressed through facial expressions, body movements, and speech. Facial expressions and speech are the modalities that have been studied the most in HCI, computer science and psychology.

The face is cited as being the most used resource in identification of emotional states (Noroozi et al. 2018). The ability of the face to convey emotions and the facial deformations associated with different emotions have been documented in Ekman and Friesen's Facial Action Coding System (FACS) (Ekman and Friesen 1978). The FACS allows animators to draw upon methods outside their own practice to create facial expressions that communicate emotions effectively to the audience (Buchanan 2009). For example, the use of the FACS in the creation of the character Gollum in Lord of the Rings – The Two Towers (2002), resulted in a character which was widely regarded by critics as emotionally believable (Kerlow 2014).

Several approaches for representing facial expressions in animated agents exist. Some computational frameworks are based on discrete representation of emotion; others on dimensional models; and others on appraisal theories (Pelachaud 2009). Approaches that are based on the expression of standard emotions (Ekman and Friesen 1975; Ekman 2003) compute new expressions as a mathematical combination of the parameters of predefined facial expressions (Becker and Wachsmuth 2006; Pandzic and Forcheimer 2002). Approaches based on dimensional models use a 2 dimensional--valence and arousal (Garcia-Rojas et al. 2006) or 3 dimensional--valence, arousal, and power (Albrecht et al., 2005) representation of facial emotions. A new expression is created by mixing the facial parameters of the expressions of the closest standard facial emotions in the representation space. A few approaches use fuzzy logic to compute the combination of expressions of the six standard emotions (Duy Bui et al. 2004), or the combination of facial regions of several emotions (Pelachaud, 2009). Some approaches are based on Scherer's appraisal theory (Scherer, 2001) and model a facial expression as a sequence of the facial articulations that are displayed consecutively as a result of cognitive estimates (Paleari & Lisetti, 2006).

Body movements are particularly important for expressing emotions when the agent is framed in a medium to long shot (Anasingaraju and Adamo-Villani 2020; Meyer et al. 2021), or to convey emotions that are less susceptible to social editing (Ekman and Friesen, 1974). Bodily cues have been shown to be very effective for discriminating between intense positive and intense negative affective states (Avezier et al. 2012). Emotional states can be conveyed through body movement modulation (e.g., the manipulation of motion parameters such as speed or amplitude) or movement type (e.g., a specific body gesture) (Cheng et al. 2020) or a blend of both (Karg et al. 2013). A combination of modalities that include both facial and bodily gestures improve recognition rate of emotion compared to facial cues alone by 35%. The best rate of recognition uses a combination of facial and body gestures with the inclusion of voice (Gunes et al. 2015).

Studies in marketing and psychology, suggest that voice, including both speech content and the linguistic and paralinguistic vocal cues (e.g., pitch, cadence, speed, and volume), is a particularly powerful channel for perceiving the emotions of other people (Kraus 2017). Simon-Thomas et al. (2009) examined how well brief vocal bursts could communicate 22 different emotions. Results showed that vocal bursts can communicate emotions like anger, fear, and sadness, and also less-studied states and highlighted the voice as a rich modality for emotion expression/perception. Kraus (2017) conducted five experiments to test the hypothesis that voice-only communication elicits higher rates of emotion recognition accuracy than vison-only and multi-sense communication in the context of social interactions. Findings support the hypothesis and challenge the primary role of facial expressions in emotion recognition. In a study by Zaki et al. (2009) social targets were filmed while discussing emotional autobiographical events. A group of subjects watched the videos and inferred the targets' emotional states while having access to only visual or auditory information, or both. Findings suggest that auditory, and especially verbal information, is critical to accurate detection of emotion. In a study by Gesn and Ickes (1999) participants viewed video segments of simulated psychotherapy sessions and attempted to identify each client's emotional state. Results showed that, in this particular context, emotion recognition accuracy was primarily dependent upon verbal, rather than nonverbal, cues.

In conclusion, several studies have investigated the role of voice for recognizing people's emotions, however, to our knowledge, no studies have attempted to examine the role of voice versus facial cues for identifying the affective states of animated agents. The work reported in the paper aims to fill this gap.

3 Methodology

The study aimed to answer the following research question: do facial expressions override tone of voice when perceiving the emotion of an animated agent? Drawing from prior research on animated agents and best practices in character animation (Williams 2012), the first hypothesis was that participants' recognition of the emotions displayed by the animated agents would be more accurate when discerned from the agents' facial expressions than from their tones of voice. The second hypothesis was that when comparing the ratings of typicality, sincerity, and intensity of emotions, stimuli with congruent voice/facial expression channels would be rated higher than stimuli with incongruent channels. The third hypothesis was that participants' gender, age and animation experience would have a significant effect on emotion recognition and on typicality, sincerity, and intensity ratings.

The study used a within subject design and collected both quantitative and qualitative data. The independent variables included facial expression-tone of voice congruence (yes/no), participants' gender (male, female, other), age, and animation experience, and emotion type (happy, angry, neutral). The dependent variables were the participants' emotion recognition rate and the participants' ratings of the emotion's intensity, typicality, and sincerity on a 5-point Likert scale (1 = low; 5 = high).

3.1 Subjects

37 people participated in the study. 19 were males, 14 females, and 4 of non-disclosed gender. The age range was 18–56 years old; the animation experience of the subjects ranged from no experience to high experience.

3.2 Materials

Stimuli. The stimuli were 18 animation clips: 9 featured a male agent and the remaining 9 a female agent. The agents exhibited three different facial emotions (happy, angry, neutral), each one paired with three different voice lines spoken in three separate tones (happy, angry, neutral). For example, for the happy emotion, one animation featured an agent with a happy face/happy voice, another an agent with a happy face/angry voice, and a third an agent with a happy face/neutral voice. The facial emotions were based on Ekman and Friesen Facial Action Coding System (Ekman and Friesen 1978). The different facial expressions and tones of the voice lines were validated through a pilot study with eight subjects. Figure 1 illustrates four frames extracted from the animation clips (two per agent).

The agents' models and rigs were downloaded from the internet and the facial expressions were animated by an experienced character animator using keyframe animation in Autodesk Maya software; the characters were manually lip synced to the voice lines. The two characters were rigged with identical facial skeletal deformation systems and the animations were created based on the Facial Action Coding System (FACS) (Ekman and Friesen 1978). Action Units (AU) 6 + 12 were used for the happy expressions, and Action Units 4 + 5 + 7 + 23 for the angry expressions; the neutral facial expressions did not feature any facial deformations. Characters occasionally blinked and slightly changed their gaze direction. The characters were framed from the neck up at a three-quarters angle; no background elements were included in the videos to keep the viewers focused on the characters' faces. The animation frame rate was 24 fps; the length of the videos ranged between 11 and 16 s and varied depending on the voice line being spoken, but each expression and voice line had equal duration.

Evaluation Instrument. The evaluation instrument was an online questionnaire consisting of 54 questions. The first set of questions gathered demographic data, namely participant gender, age range, and animation experience. Then, the 18 stimuli videos were presented to the participants in random order and 4 questions were provided after each video. One question was a multiple-choice question that asked the subjects to identify the emotion they believed the character in the video was conveying (e.g., happy, angry, or other). If a subject selected the "other" option, they were asked to specify what emotion specifically came to mind when viewing the video. The other three questions asked the participants to rate the typicality (how usual the emotion was), sincerity (how convincing/genuine the emotion was) and intensity (how strong the emotion was) of the

Fig. 1. Frames showing the animated agents used in the study. Happy male agent (top left), angry male agent (top right), happy female agent (bottom left), angry female agent (bottom right)

emotion displayed by the character on a 5-point Likert scale ($1 = \text{low}$; $5 = \text{high}$). Finally, subjects could offer additional comments about the emotion displayed by the character.

3.3 Procedure

The online questionnaire was sent to the participants using several methods including email, online chat threads, and text messages. The questionnaire could be taken at any time and at any location and subjects were only permitted one attempt at the survey. The researcher did not assist respondents in answering the questions beyond providing the instructions included in the survey.

4 Data Analysis

4.1 Cross-Modal Identification of Emotion

Rstudio was used to run a power analysis to determine the needed sample size. The results of the power analysis, which used data from a pilot study with 7 subjects, showed that at least 32 samples were needed for a power of 0.8 at a significance level of 0.05. The analysis of cross-modal identification of emotion was conducted in Python; the subjects' responses were analyzed using a mixed-effect ordinal regression and compared to the fixed responses provided by the experiment. Figures 2–4 show bar graphs that summarize the data. The x-axis represents the facial emotions expressed by the character, and the y-axis represents the number of participants who selected that emotion; each figure relates to a different tone of voice (happy in Fig. 2; neutral in Fig. 3, angry in Fig. 4). The data presented in Figs. 2–4 show a high recognition rate of the correct emotion when the channels were congruent with one another. However, even when the channels were incongruent, subjects identified the emotion based more heavily on the facial expression than the tone of voice. An angry expression was the easiest to identify given any condition, while a happy expression was the most difficult to identify, being perceived more as a neutral expression. Despite this confusion, the facial expression prevailed more often in identifying an emotion than the tone of voice. To analyze the data related to specific emotions, each emotion was given a value that was inputted into an equation to identify its probability. The equation below was employed to calculate the probability of an event as a function of the predictor values for the model and Table 1. The random effect is denoted as τid for each participant:

$$logit(P(Y \leq j)) = \log \frac{P(Y \leq j)}{P(Y > j)} = intercept_j + \beta_{j2} x_{j2} + \ldots + \tau_{id}; j = 1,$$

The j value is the representation of the emotional value, 1 is for angry, 2 is for happy, and 3 is for neutral. For example, the equation measures the probability of an angry expression when j equals 1. Since Y has to be less than or equal to 1, the only option that can be placed within that field is 1 (angry). This was compared to the estimates of the other remaining emotions of happy or neutral, as the second parameter calls for Y to be greater than 1, meaning only 2 (happy) or 3 (neutral) can fit in that field. This is further summarized when compared to the alternative formula meant to calculate probabilities:

$$logit(P(Y \leq 2)) = \log \frac{P(Y \leq 2)}{P(Y > 2)} = intercept_2 + \beta_{2i} x_{2i} + \ldots + \tau_{id},$$

This functions exactly as the previous one did. However, it is meant to iden- tify the probability changes of both happy and angry compared to the baseline neutral expression. Inversely, the function can be written in the form below to identify the probability of a happy value:

$$P(Y = 2) = P(Y \leq 2) - P(Y \leq 1) = \frac{e^{a_2 + \beta x}}{1 + e^{a_2 + \beta x}} - \frac{e^{a_1 + \beta x}}{1 + e^{a_1 + \beta x}}$$

After discovering the meaning behind the exploratory studies presented in Figs. 2–4, the meaning behind the patterns was analyzed further utilizing a GLIMMIX procedure in SAS. This enabled the analysis of random variables to be accounted for when inferring relationships between fixed effect predictors and the outcome. Table 1 further elaborates the analysis; a Type III test was conducted to assess whether the significant values offer sufficient statistical data for each factor. Table 1 indicates that the facial variable, tone variable, and their interactions are significant in this model.

To further confirm the validity of the mixed-effect ordinal regression, we cre- ated a confusion matrix to summarize the model performance, as shown in Fig. 5. The confusion matrix shows the software's selections of emotions based on estimates gathered from the data on the x-axis in comparison to the true selections of emotions from participants on the y-axis. The color of the regions is in relation to the frequency of the selections based on these comparisons. For example, the middle square represents a total of 246 participant responses that match the predicted label of happy while the bottom middle square represents 96 responses interpreting the predicted happy emotion as other, as well as the top middle square representing 74 responses interpreting the predicted happy emotion as angry. Refer to Table 2 for the response frequency from the GLIMMIX procedure.

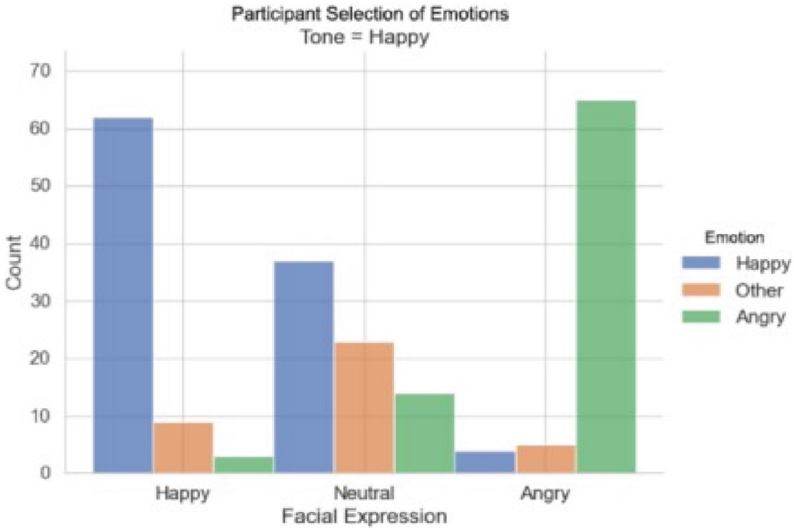

Fig. 2. Participant selection of emotions based on perceived expression and happy tone of voice.

Table 3 presents a series of estimates provided from fixed effects based around the participant's input and the computer's predictions of accuracy values. Estimates were pulled from the data using a cumulative logit parameter to illustrate how the subject responses change from one set of expressions to another. These estimates also showed how significant the contrast of factors was between both expressions and tones of voice. Neutral emotions were the baseline measurement for the estimates, so they always remained 0. The p-value in the last column on the right indicates the significance of each effect. From the results, we can see that the "angry" and "happy" levels of the facial factor both play statistically significant roles in the model. Moreover, combined with previous probability calculations, we can infer that changing the facial level from neutral to angry is estimated to increase the probability of an angry emotion.

4.2 Analysis of Ratings of Emotion Typicality, Sincerity and Intensity

Descriptive statistics show that for the happy emotion, the ratings of typicality, sincerity and intensity were highest when the emotional channels, e.g., facial expression and voice, were congruent (Table 4). These findings support our hypothesis. Surprisingly, for the

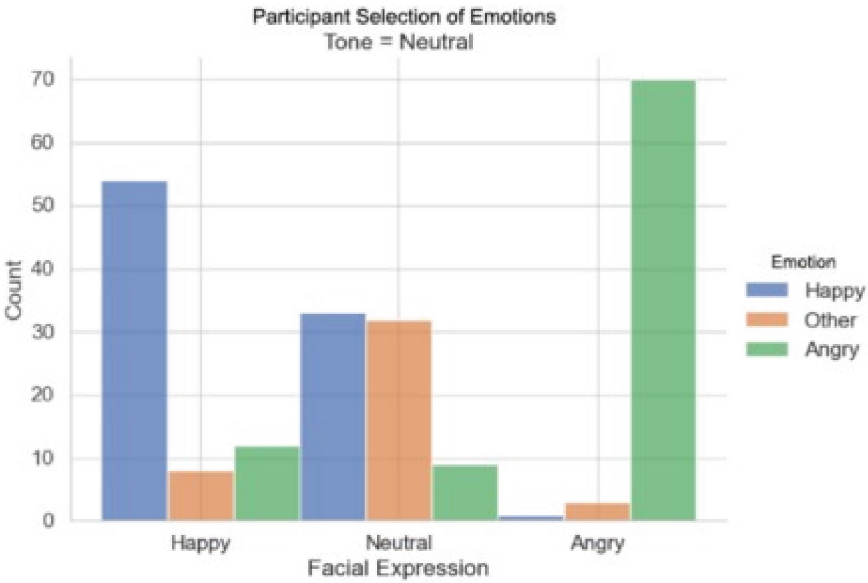

Fig. 3. Participant selection of emotions based on perceived expression and neutral tone of voice.

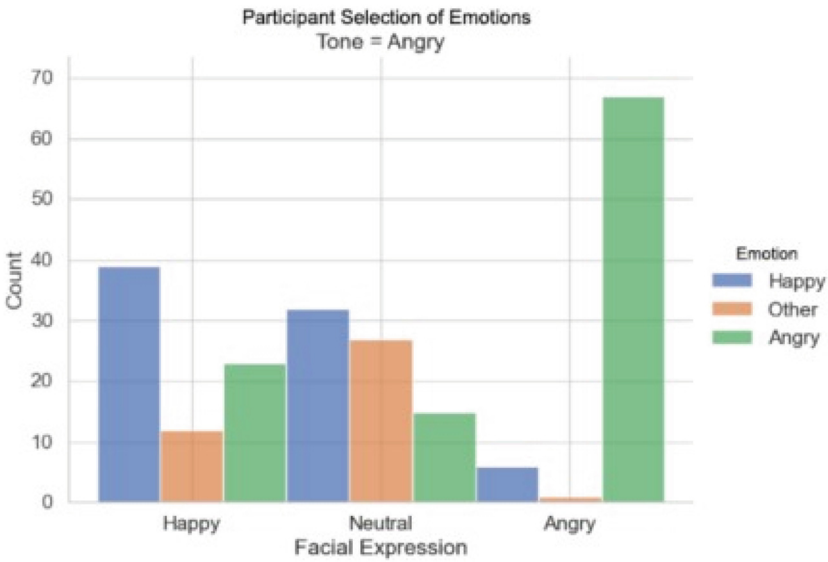

Fig. 4. Participant selection of emotions based on perceived expression and angry tone of voice.

angry emotion, the ratings of typicality, sincerity and intensity were highest when the agent showed an angry face and spoke using a neutral tone of voice rather than an angry tone of voice (Table 5). For the neutral emotion, the congruent condition had the highest ratings, however the differences in ratings between conditions were small (Table 6).

Table 1. Tests of the statistical significance of a variable's ability to reject the null hypothesis based on the factors' two-way interactions.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Facial	2	650	105.00	<.0001
Tone	2	650	5.30	0.0052
Facial*Tone	4	650	4.25	0.0021
gender	2	26.11	0.09	0.9185
age	2	26.95	0.43	0.6549
experience	2	26.72	0.09	0.9110

Table 2. Respective values for each emotion for formulaic input alongside the sum of each selection from the survey.

Response Profile		
Ordered Value	value	Total Frequency
1	Angry	193
2	Happy	210
3	Neutral	263

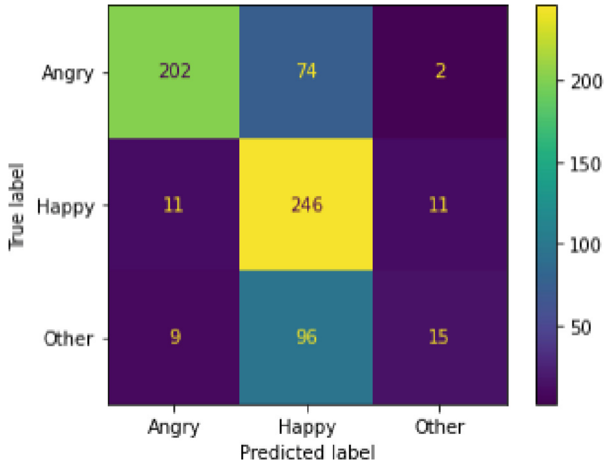

Fig. 5. Confusion Matrix of participant selections compared to computer-predicted selections. The accuracy of these two data sets is represented by the scale to the left of the matrix.

Differences Based on Participants' Age, Animation Experience and Gender. The data was analyzed using a Type III regression method placed on a 5-point Likert scale on

Table 3. Fixed-effect estimates and their significance levels.

Solutions for Fixed Effects											
Effect	value	Facial	Tone	gender	age	experience	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	Angry						-2.1967	1.1624	29.13	-1.89	0.0688
Intercept	Happy						-0.1258	1.1581	29.13	-0.11	0.9142
Facial		Angry					4.0679	0.4126	650	9.86	<.0001
Facial		Happy					0.7457	0.3356	650	2.22	0.0266
Facial		Neutral					0	.	.	.	.
Tone			Angry				0.09007	0.3422	650	0.26	0.7925
Tone			Happy				-0.3969	0.3529	650	-1.12	0.2611
Tone			Neutral				0	.	.	.	.
Facial*Tone		Angry	Angry				-1.8287	0.5176	650	-3.53	0.0004
Facial*Tone		Angry	Happy				-0.4315	0.5349	650	-0.81	0.4202
Facial*Tone		Angry	Neutral				0	.	.	.	.
Facial*Tone		Happy	Angry				-0.4359	0.4746	650	-0.92	0.3588
Facial*Tone		Happy	Happy				0.5450	0.4792	650	1.14	0.2558
Facial*Tone		Happy	Neutral				0	.	.	.	.
Facial*Tone		Neutral	Angry				0	.	.	.	.
Facial*Tone		Neutral	Happy				0	.	.	.	.
Facial*Tone		Neutral	Neutral				0	.	.	.	.
gender				1			-0.2927	0.7343	25.92	-0.40	0.6934
gender				2			-0.2973	0.7877	25.86	-0.38	0.7089
gender				3			0	.	.	.	.
age					2		0.07560	0.8832	26.82	0.09	0.9324
age					4		-0.5555	1.0488	26.92	-0.53	0.6007
age					5		0	.	.	.	.

Table 4. Comparison of typicality, sincerity, and intensity rates for congruent and incongruent happy stimuli - happy facial expression

Happy/Happy	Mean	Median	SD
Typicality	3.57	3	0.89
Sincerity	3.62	3	1.13
Intensity	3.7	3	1.15
Happy/Angry	**Mean**	**Median**	**SD**
Typicality	3.2	3	0.86
Sincerity	3.3	3	1.13
Intensity	3.01	3	0.93
Happy/Neutral	**Mean**	**Median**	**SD**
Typicality	3.2	3	0.86
Sincerity	3.09	3	1.13
Intensity	3.27	3	1.12

Table 5. Comparison of typicality, sincerity, and intensity rates for congruent and incongruent channels - angry facial expression

Angry/Happy	Mean	Median	SD
Typicality	3.05	3	0.98
Sincerity	3.28	3	1.14
Intensity	3.43	3	1.21
Angry/Angry	**Mean**	**Median**	**SD**
Typicality	3.43	3	0.68
Sincerity	3.62	4	0.87
Intensity	3.35	3	1.01
Angry/Neutral	**Mean**	**Median**	**SD**
Typicality	3.5	3	0.74
Sincerity	3.92	4	0.93
Intensity	3.7	4	0.92

Table 6. Comparison of typicality, sincerity, and intensity rates for congruent and incongruent neutral stimuli (neutral facial expression)

Neutral/Happy	Mean	Median	SD
Typicality	2.89	3	0.9
Sincerity	2.64	3	1.19
Intensity	2.31	3	1.21
Neutral/Angry	**Mean**	**Median**	**SD**
Typicality	2.85	3	0.95
Sincerity	2.66	3	1.14
Intensity	2.34	3	1.25
Neutral/Neutral	**Mean**	**Median**	**SD**
Typicality	2.97	3	0.92
Sincerity	2.78	3	1.11
Intensity	2.6	3	1.2

each of the demographic values. The regression was created in SAS and conducted using a MIXED procedure, which allows the random variables to be compared holistically with all fixed variables. The equation used to calculate the data for the mixed effect regression model was as follows:

$$typicality = \alpha + \beta_{angry} * I(Angry) + \beta_{happy} * I(Happy) + \in + \tau_{id},$$
$$\in \sim N(0, \sigma^2), \quad \tau_{id} \sim N(0, \sigma_{id}^2), \quad I(\cdot) \; indicator\ function$$

The typicality value can be replaced with either intensity or sincerity de- pending on the response being calculated within the function. We treat the participants within the study as a random effect to account for the correlation for each individual subject.

Results of the analysis showed that participants' age and animation experience had no effect on intensity, typicality and sincerity ratings across all emotions and conditions (p-values > 0.05). Participants' gender did not have significance for intensity and typicality ratings but had a significant effect on sincerity ratings. Female participants gave significantly higher sincerity ratings than the other participants across all emotions and across all conditions (p-values < 0.05).

5 Discussion and Conclusion

The study reported in the paper examined the roles of facial expressions and tone of voice in the perception of animated agents' emotions. Findings from the study support the first and second hypotheses: emotion recognition rate and ratings of emotion intensity, typicality and sincerity are highest when the agent's face and voice are congruent. However, when the channels are incongruent, subjects tend to identify the emotion more accurately from the agent's facial expression than the tone of voice. Results also show an effect of participants' gender on emotion sincerity ratings, thus supporting in part the third hypothesis. Female participants perceived the emotions displayed by the agents as significantly more sincere than the other participants across all emotions and conditions.

The study had several limitations that could be overcome in future research. The sample size was small (37 subjects) and participants' genders were not represented equally. Future studies should use a larger number of subjects with a more even distribution of genders.

The study comprised only two stylized characters and three voice lines spoken in three different tones. Part of the results could be due to the intrinsic design characteristics of the characters and to the intrinsic linguistic and paralinguistic cues of the voice lines. Future experiments should include a higher variety of voice lines and characters with different degrees of stylization. For instance, it would be interesting to examine whether participants' perception of the agents' emotions is different for stylized versus realistic characters. It could be possible that the face is the strongest emotion expression modality for cartoon characters while the voice is the strongest emotion expression channel for realistic gents.

The study focused on only two of Ekman's six basic emotions. Future experiments should consider the other four emotions. The short duration of the stimuli animations may have affected the participants ability to perceive the characters' emotions in all conditions. Future research should use animation sequences of longer duration and with two or more characters interacting with each other.

A substantial body of research has examined the role of voice for recognizing people's emotions, however, to our knowledge, no studies have attempted to examine the role of vocal versus facial cues for identifying the affective states of animated agents. To advance knowledge in this area, there is a need for experimental research studies that systematically investigate the extent to which specific affective channels contribute to enhance the perception of animated characters' emotions. The study reported in the paper is a small step in this direction. Studies like this not only have important implications for research, but also for practice. They can provide useful guidelines for animators and instructional designers for enhancing the appeal and emotional impact of animated characters and, possibly, the educational effectiveness of animated pedagogical agents.

Acknowledgements. This work is supported in part by NSF- RETTL Collaborative Research: Using Artificial Intelligence to Transform Online Video Lectures into Effective and Inclusive Agent-Based Presentations; Award # 2201019. The authors would like to thank Purdue Statistical Consulting Service for their help with the data analysis.

References

Schroeder, N., Adesope, O.O., Barouch Gilbert, R.: How effective are pedagogical agents for learning? a meta-analytic review. J. Educ. Comput. Res. **49**(1) (2013). https://doi.org/10.2190/EC.49

Adamo, N., et al.: Multimodal affective pedagogical agents for different types of learners. In: Russo, D., Ahram, T., Karwowski, W., DiBucchianico, G., Taiar, R. (eds.) Intelligent Human Systems Integration 2021. AISC, vol. 1322, pp. 218–224. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68017-6_33

Kim, Y., Baylor, A.L.: Pedagogical agents as social models to influence learner attitudes. Educ. Technol. **47**(01), 23–28 (2007)

Noroozi, F., Kaminska, D., Corneanu, C., Sapinski, T., Escalera, S., Anbarjafari, G.: Survey on EMOTIONAL BODY GESTURE RECOGNITION. IEEE Trans. Affect. Comput. 1–20 (2018)

Buchanan, A.: Facial expressions for empathic communication of emotion in animated characters. Animation studies (2009). https://journal.animationstudies.org/andrew-buchanan-facial-expressions-for-empathic-communication-of-emotion-in-animated-characters/

Kraus, M.W.: Voice-only communication enhances empathic accuracy. Am. Psychol. **72**(7), 644–654 (2017)

Ekman, P., Friesen, E.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)

Kerlow, I.: Creative Human Character Animation: The Incredibles vs. The Polar Express. vfxworld.com (2014)

Pelachaud, C.: Modelling multimodal expression of emotion in a virtual agent. Phil. Trans. R. Soc. B **2009**(364), 3539–3548 (2009). https://doi.org/10.1098/rstb.2009.0186

Ekman, P., Friesen, W.: Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues. Prentice-Hall Inc., Eaglewood Cliffs (1975)

Ekman, P.: Emotions revealed. Times Books (US)/Weidenfeld & Nicolson, New York/London (2003)

Becker, C., Wachsmuth, I.: Modeling primary and secondary emotions for a believable communication agent. In: Proceedings of International Workshop on Emotion and Computing, in conj. with the 29th Annual German Conference on Artificial Intelligence (KI2006), Bremen, Germany, pp. 31–34 (2006)

Pandzic, I.S., Forcheimer, R. (eds) MPEG4 Facial Animation – The Standard, Implementations and Applications. John Wiley and Sons, Chichester, UK (2002)

Garcia-Rojas, A., et al.: Emotional face expression profiles supported by virtual human ontology. Comp. Anim. Virtual Worlds **17**, 259–269 (2006). https://doi.org/10.1002/cav.130

Meyer, Z., Adamo, N., Benes, B.: Bodily expression of emotions in animated agents. In: Bebis, G., et al. (eds.) Advances in Visual Computing, vol. 13018, pp. 475–487. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-90436-4_38

Bui, D., Heylen, D., Poel, M., Nijholt, A.: Combination of facial movements on a 3D talking head. In: Computer Graphics International, pp. 284–290 (2004)

Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking. In: Scherer, K., Schorr, A., Johnstone, T. (eds.) Appraisal Processes in Emotion: Theory, Methods, Research, pp. 92–119. Oxford University Press, New York (2001)

Paleari, M., Lisetti, C.L.: Psychologically grounded avatars expressions. In: Proceedings of First Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence, Bremen, Germany (2006)

Anasingaraju, S., Adamo-Villani, N., Dib, H.N.: The contribution of different body channels to the expression of emotion in animated pedagogical agents. Int. J. Technol. Human Interact. (IJTHI), **16**(4) (2020). https://doi.org/10.4018/IJTHI.202010010

Ekman, P., Friesen, W.: Detecting deception from the body or face. J. Pers. Soc. Psychol. **29**, 288–298 (1974)

Avezier, H., Trope, Y., Todorov, A.: Body cues, not facial expressions, discriminate between intense positive and negative emotions. Science **338**(6111), 1225–1229 (2012)

Cheng, J., Zhou, W., Lei, X., Adamo, N., Benes, B.: The effects of body gestures and gender on viewer's perception of animated pedagogical agent's emotions. In: Kurosu, M. (ed.) Human-Computer Interaction. Multimodal and Natural Interaction. LNCS, vol. 12182, pp. 169–186. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49062-1_11

Karg, M., Samadani, A. -A., Gorbet, R., Kühnlenz, K., Hoey Kulić, D.: Body movements for affective expression: a survey of automatic recognition and generation. IEEE Trans. Affect. Comput. **4**(4), 341–359 (2013). https://doi.org/10.1109/T-AFFC.2013.29

Gunes, H., Shan, C., Chen, S., Tian, Y.: Bodily expression for automatic affectrecognition. In: Emotion Recognition, pp. 343–377. Wiley (2015)

SimonThomas, E.R., Keltner, D.J., Sauter, D., SinicropiYao, L., Abramson, A.: The voice conveys specific emotions: evidence from vocal burst displays. Emotion **9**, 838–846 (2009). https://doi.org/10.1037/a0017810

Zaki, J., Bolger, N., Ochsner, K.: Unpacking the informational bases of empathic accuracy. Emotion **9**, 478–487 (2009). https://doi.org/10.1037/a0016551

Gesn, P.R., Ickes, W.: The development of meaning contexts for empathic accuracy: channel and sequence effects. J. Pers. Soc. Psychol. **77**, 746–761 (1999). https://doi.org/10.1037/0022-3514.77.4.746

Williams, R.: The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators. Farrar, Straus and Giroux. 4th Edn. (2012)