Original Article

Effects of remote counseling using an embodied agent

Iwasaki Kazuko, Graduate School of Science and Technology, Kyoto Institute of Technology
*Kida Noriyuki, Faculty of Arts and Sciences, Kyoto Institute of Technology

Abstract

Objectives: We developed a remote counseling environment showing the counselors' facial expressions in real-time shown to clients as an embodied agent and quantitatively evaluated the length of utterances and pauses.

Methods: We provided mock counseling using the "dog and rose role-playing technique" to ten healthy adults as clients and one clinical psychologist as a counselor, in the following counselor's facial expression conditions presented on a display placed in front of the clients. The conditions were voice, agent, and real facial expressions, each condition presented for ten minutes.

Results: Pause length during a turn in the agent condition was significantly longer than in the voice condition. Moreover, in-turn pause length (the length of pausing time after clients start uttering) in the agent condition was significantly longer than in the voice or real facial expression conditions.

Conclusion: Clients might be able to stay silent easily and have a dialogue with themselves in the agent condition because they have less interpersonal tension, look at the counselor's nods or facial expressions as an image, and not feel that they must immediately respond to the counselor.

Keywords: Avatar, utterance time, counselor, client

Introduction

People often feel stress in modern society under various settings, such as work, study, interpersonal relationships, and family settings, which has increased the need for psychological counseling (counseling). Psychological counselors (counselors) help clients reduce their worries and stress through conversations, in which not only verbal information using words as clues but also nonverbal information using clues other than words, such as facial expressions, nodding, among others, is indispensable (Ekman, 1982; Katsikitis, 1997; Nakamura, 2009). In counseling settings, counselors anticipate clients' psychological conditions based not only on clients' utterances but also on pauses between utterances, facial expressions, eyes, and gestures. Based on anticipation, counselors express empathy with clients using facial expressions or gestures by repeating clients' words or giving affirmative responses, accepting clients' personalities, and valuing their thoughts. Therefore, counseling has usually adopted a one-to-one, face-to-face style.

Recently, many people have tended not to seek expert support, including counseling, despite the need for such support (Andrews et al., 2001). Therefore, it is necessary to develop a support system for people with psychological problems to obtain expert support. Clients sometimes feel stress related to interpersonal relationships when they have face-to-face counseling for the first time, and counselors need to reduce their tension. Adopting embodied agents might effectively relieve speakers' tension in conversation settings, and studies have been conducted in different situations to

Accepted on 23 November, 2021
*: Corresponding Author

achieve this objective (Kitamura, 2016; Hayashi, 2012). The present study focused on avatars as embodied agents.

An embodied agent is an interface presented on a screen behaving like a person or an animal, which is increasingly used in various situations as a communication tool. Image recognition and computer graphic functions have improved dramatically recently, and Virtual You Tubers, presenting avatars by converting their facial expressions using computer graphics in real-time, have appeared. The possibility of using avatars in counseling has increased through combining remote techniques and embodied agents. Imaizumi et al. (2018) suggested interview techniques using an embodied agent for people with difficulty in face-to-face communication to increase their utterances. Remote counseling using an embodied agent might also be possible.

The effects of embodied agents are usually verified based on participants' evaluation of impressions and behaviors (Nakagawa et al., 2019). This study focused on time-related information during a conversation, including the time of utterances and pauses, to quantitatively evaluate utterance performance in counseling by referring to a preceding study on interview settings (Imaizumi, 2018). Generally, clients are expected to face themselves and dialogue with themselves in counseling. However, it is difficult to evaluate the frequency the depth of their self-dialogue. Therefore, we perceived a pause as an approach to the inner self. Kandabashi (1990, 2016) indicated, "A rich silence is the basis of rich words." Staying silent is a type of expression for clients. During counseling, clients' silences (pauses) are a time for thinking freely, which is the time to face themselves. Silence is the process necessary for relieving clients' worries and stress, and extended silence is desirable.

The present study developed a counseling environment in which a counselors' facial expressions were shown in real-time to clients as an embodied agent, substituted mock counseling, quantitatively assessed the length of time for utterances and pauses, and verified the effects of counseling using an embodied agent. Moreover, we developed two conditions for comparison: counselors' facial expressions invisible condition, and counselors' facial expressions visible condition.

Methods

1. Participants

Healthy adults (N = 10) played the role of clients, and a clinical psychologist played the role of a counselor. We had explained the purpose and methods of the experiment to participants in advance. The present study was conducted after obtaining approval from the ethics committee of the Kyoto Institute of Technology.

2. Environmental settings

We conducted mock counseling using a one-to-one individual style. A counselor and a client stayed in different soundproof rooms. We placed a personal computer connected to a camera and a microphone and a display with a speaker facing the counselor and client. The image and the voice obtained by the camera and microphone were output from the display and speaker in the other room using an HDMI LAN cable.

We converted the counselor's facial expressions using facial-recognition-image transformation software (FaceRig) to reflect facial expressions and directions on a 3D character and presented the character on the display facing the client. We did not convert the client's facial expressions and present them on the display facing the counselor. Moreover, we did not edit the voice and presented it as it was.

3. Mock counseling

We used the dog and rose role-playing technique (Otsuji, 2004), often used in counselors' training, in which clients play the role of a dog or a rose and consult about their worries with a counselor. We provided explanations about the technique to participants in advance. Participants practiced conversation for 20 minutes to form a rapport with the counselor before starting the experiment.

We set up the following three conditions of the counselor's facial expressions presented to clients; voice, agent, and real facial expression conditions. We only output the voice without presenting facial expressions in the voice condition. We transformed the counselor's facial expressions using the software (FaceRig) in the agent condition and displayed them as an older woman's facial expressions. In the real facial expression condition, we did not convert the counselor's facial expressions and presented them as they were. We conducted 10-minutes mock counseling in each condition randomly for each participant, counterbalancing the order of the conditions by using random numbers to prevent order effects.

4. Recording and analysis

We recorded the situation in which clients had mock counseling using a video camera and assessed the feature values related to utterance using voice software (Audacity) by referring to Miyake et al. (2004). We defined the time from the counselor finishing utterance to the client finishing utterance as the length of one turn (turn length). When there was no utterance exceeding one second during one turn, we defined it as a pause and the pause duration as pause length, calculating the pause length in one turn. We perceived the time from removing the pause length from the turn length as the utterance length. Moreover, we classified the pause as pre-turn and in-turn. The former was the pause from the counselor finishing an utterance to the client starting an utterance, and the latter was the pause after the client starts an utterance, calculating pre-turn pause length and in-turn pause length (Figure 1). All the times were measured in seconds. After finishing mock counseling, we conducted an interview survey on the ease of talking.

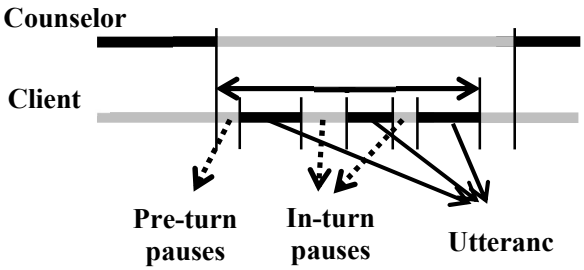

Figure 1: Utterances and pauses in one turn

Results

1. Utterance length and pause length

Table 1 shows each parameter's mean and standard deviation (SD) for turns related to clients' utterances under each condition. We examined the difference among the three conditions using Friedman's test, which indicated no significant main effect of conditions on the number of turns, whole turn length, or one turn length.

Table 2 shows the mean and SD of each parameter related to clients' utterances and pauses under each condition. We examined differences among the three conditions using Friedman's test. Utterance length in one turn did not show a significant difference, whereas pause length in one turn showed a significant difference. The results of Bonferroni's multiple comparisons indicated that the pause length under the agent condition was significantly longer than under the voice condition. Moreover, in-turn pause length showed a significant difference, whereas pre-turn pause length did not. The results of Bonferroni's multiple

Table 1: The number of turns and their length

	Voice condition	Agent condition	Real facial expression condition	χ^2	p
The number of turns	13.1±6.3	11.7±4.3	12.8±5.9	0.222	.895
The length of the whole turn (second)	391±74	423±52	429±70	1.400	.497
The length of one turn (second)	38±21	43±23	39±17	2.600	.273

Table 2. Utterance lengths and pause lengths

	Voice condition	Agent condition	Real facial expression condition	χ^2	p
Utterance length in a turn (seconds)	29±22	31±20	28±18	1.400	.497
Pause length in a turn (seconds)	8.6±3.8	12.4±3.7	11.5±7.0	6.200	.045 *
Pre-turn pause length in a turn (seconds)	4.8±4.4	4.6±2.8	6.3±5.5	0.200	.905
In-turn pause length in a turn (seconds)	3.9±1.7	7.8±3.0	5.2±3.9	6.513	.039 *

*, p<.05

comparisons indicated that the in-turn pause length under the agent condition was significantly longer than under the voice and real facial expression conditions.

2. Interview survey on the easiness of talking

We conducted an interview survey with clients on the easiness of speaking after the experiment. Eight participants in the voice responded, "I could rely just on the counselor's response. I felt I must keep talking and was scared of silence." On the other hand, two participants responded, "I felt easy talking because the counselor was invisible."

In the agent condition, seven participants responded, "I felt less interpersonal tension," five responded, "I did not need to be too considerate of the counselor," and eight said, "I felt easy talking because facial expressions were visible." On the other hand, two participants answered, "I felt anxious about the agent's unnatural eye movements and actions."

Five participants in the real facial expression condition responded, "I felt easy talking because I felt relaxed at seeing the counselor's facial expressions, and smile, during pauses." On the other hand, one participant responded, "I felt uncomfortable about the silence while being watched."

Discussion

1. Comparisons among three conditions on turns and pause lengths:

The number of turns between the counselor and clients did not significantly differ among the three conditions. The counselor was blind to the clients' conditions, which resulted in a well-controlled experimental setting. The turn length did not significantly differ among the conditions; however, the pause length differed significantly. Previous studies on conversations using an embodied agent recreated interview settings, whereas few studies have recreated counseling settings. Moreover, Nakagawa et al.

(2019) examined counseling using virtual reality (VR) and assessed mood changes after sessions without conducting time analyses. This study introduced quantitative time analysis into counseling settings using an embodied agent and obtained a new finding that pause length differed among the conditions, which was not the case with the utterance length.

The pause length was longest under the agent condition, followed by the real facial expression condition, whereas it was significantly short under the voice condition. The interview survey results indicated that many participants in the voice condition responded that they could rely just on the counselor's response, felt they must keep talking, and felt anxious about making pauses during the conversations. In contrast, specific participants responded that they felt easy talking. They could relax because they were not being watched by the counselor, which led to short pause lengths. When clients cannot see the counselor's face, regardless of an actual counselor or an agent, clients cannot obtain nonverbal information such as nods and facial expressions, and might feel difficulty in stopping talking, which might have pressurized clients, and made them anxious, leading to decreased pause lengths. Remaining silent and having a self-dialogue is essential in counseling (Itobayashi, 2009; Gendlin, 1978). Participants might not have had sufficient time for self-dialogue under the voice condition, which might be considered inferior to the agent or real facial expression condition.

In the agent condition, participants responded that they felt less interpersonal tension because they did not need to consider the counselor's face. Since the counselor's nods and facial expressions were presented as an image, the clients did not feel they must reply immediately, felt less anxious about making the counselor wait, which resulted in inadequate time for self-dialogue by staying silent.

Silence might reflect clients' inner conditions, including resistance and indifference. In this study, however, we conducted mock counseling after forming rapport between clients and the counselor by practicing the dialogue in advance. Therefore, silence in this study reflected the clients' self-dialogue rather than resistance or indifference.

2. Pre-turn pause and in-turn pause: comparisons among three conditions

The pre-turn pause did not show a significant difference among the three conditions, whereas the in-turn pause did show a significant difference, which was the longest under the agent condition, compared to the voice or real facial expression conditions. We should note that the in-turn pause under the real face condition was shorter than under the agent condition. In the interview survey, specific participants under the real facial expression condition responded that they felt easy talking because they were relaxed by seeing the counselor's facial expressions and feeling that the counselor was listening to them. The counselor's nonverbal information, such as nods or facial expressions, might provide a feeling of relief to clients, facilitating their utterance. Clients try to keep talking in the real facial expression condition by reading the counselor's subtle facial expressions, resulting in tension. On the other hand, clients might not get nervous in the agent condition because the agent just provides some information about the counselor. As a result, clients might efficiently conduct self-dialogue based on their utterances.

As described above, clients could obtain little nonverbal information, get nervous, and not have time to be silent for self-dialogue in the voice condition. In the real facial expression condition, on the other hand, clients obtained too much nonverbal information, recognized subtle eye movements, became nervous, and did not

facilitate self-dialogue, leading to short pause lengths. Counseling using an embodied agent might effectively decrease clients' interpersonal tension, depending on the type of agent.

Using embodied agents might help protect counselors' privacy in actual counseling settings. Moreover, clients suffering from domestic violence or having gender identity disorders might strongly prefer counselors' appearance. Using embodied agents might be helpful for such clients to develop an environment where they can have counseling at ease. Based on the above, the present study results are considered practically helpful.

3. Limitations and future perspectives

Utterance length did not show a significant difference among the three conditions. We predicted that pauses and utterances would increase in the agent condition, which was not supported by the result. Whether the counselor is real or embodied might not affect the ease of speaking or utterance length. Conversely, a study using an embodied agent in interview settings indicated a decrease in utterance length, which might be caused by the sense of strangeness due to image display delays, among other factors (Imaizumi et al., 2018). The agent's mouth movements and the counselor's utterances were not wholly consistent in the present study, which might have prevented utterance length from increasing. Future studies should examine whether improving image transformation functions, such as reduced delays and improvements in links between the mouth shape and pronunciation, might affect utterance length.

The limitations of the present study include the following We used a gentle and friendly older woman's image as the agent, which might have given a sense of relief to clients. Moreover, using an agent might not have increased pause lengths, i.e., facilitated self-dialogue, but the agent's characteristics might have increased pause lengths. Therefore, caution is advised when generalizing these results. On the other hand, impressions of the agent might differ depending on the person. An image generally arousing empathy might evoke a sense of aversion in some clients. It is suggested that we examine different types of agents that would lead to self-dialogue in different clients.

References

Andrews G, Issakidis C, Carter G, 2001. Shortfall in mental health service utilization. British Journal of Psychiatry, 179, 417-425.

Ekman P, Friesen WV, Ellsworth P, 1982. What are the relative contributions of facial behavior and contextual information to the judgment of emotion? In P.Ekman (Ed.), Emotion in the human face. Cambridge University Press.111-127.

Gendlin ET, 1978. Focusing. New York: Bantam Books.

Hayashi Y, Cooper E, Kryssanov V, Urao A, & Kobayashi H, 2012. Taiwa ejento tono komyunikeshon ni okeru shinritokusei (Psychological characteristics of communication with a conversational agent). International Journal of Affective Engineering, 11, 3, 459-467.

Imaizumi Y, Nakamura R, Kamibayashi N, 2018. Hi-mensetsusha no hatsugen wo unagasu tame no gijinka ejento mensetsu no teian to kouka kenshou (Suggesting and effect verification of interviews using embodied agents to promote interviewees' utterances). the 80th national Conference of Information Processing Society of Japan, 4, 149-150.

Itobayashi T, 2009. Kaunseringu ni okeru chinmoku no imi ni tuite (The meaning of silence in counseling). The Japanese Journal of Humanistic Psychology, 27, 67-68.

Kandabashi J, 1990. Seishinryouhou mensetsu no kotsu (Tips of psychotherapy interviews). IwasakiAcademic Publisher.

Kandabashi J. 2016. Chiryou no tame no seishinbunseki noto (Psychotherapy notebook for treatment). Sogensha.

Katsikitis M, 1997. The classification of facial expressions of emotion: A multi-dimensional-scaling approach. Perception, 26, 613-626.

Kitamura A, Hayashi Y, 2016. Gijinka ejento tono intarakushon wo sokushin suru yoin: deforume-do no sousa to myakuhaku wo mochiita jikkenteki kentou (Factors facilitating interaction with an embodied agent: An experimental investigation using deformed design and pulse rate). Proceedings of the 33rd Japanese Cognitive Science Society. 853-859.

Miyake Y, Tatsumi Y, Sugihara S, 2004. Kougo-hatsuwa ni okeru hatsuwa-chou to Hatsuwa-kankaku no jikanteki kaisousei (Hierarchy between Speech Duration and Speech Interval in Alternate Speech). Transactions of the Society of Instrument and Control Engineers. 40, 6, 670-678.

Nakamura T, 2009. Komyunikeshon ni okeru "Ma" no kansei-jouhou-shinrigaku (Psychological Study of "Ma" (pause) in Communication). Journal of the Phonetic Society of Japan. 13, 1, 40-52.

Nakagawa S, Saegusa H, Endo R, Naruse K, Guan L W, Kuniyoshi Y, 2019. VR tele-kaunseringu ni motozuku serapisuto-kuraianto-kan intarakushon ni kansuru hyouka (Evaluation of Therapist-Client Interaction Based on New VR Tele-counseling Approach). Transactions of the 24th Virtual Reality Society of Japan. 4D-04.

Otsuji T, Hirano K, 2004. Inubara-hou (jiko-shouchou-teki kaunsera kunren gihou) no chiryouteki igi (The Therapeutic Significance of Dog and rose Role-Playing Technique (Self-symbolic Counselor Training Technique)). The Japanese Journal of Medical Health. 19, 1, 49-60.