Human Interaction and Emerging Technologies (IHIET 2023), Vol. 111, 2023, 175–184
https://doi.org/10.54941/ahfe1004022

Assessing Engagement of the Elderly in Active Listening From Body Movement

Yuji Tanabe and Hiromitsu Shimakawa

Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

ABSTRACT

In recent years, the problem of apathy among the elderly has run into a serious problem as the population ages. Active listening, a type of counseling technique, is useful to address the problem. This study proposes a method for estimating the conversational state in active listening using voice and body movement data to facilitate for participants to exchange their words. In the study, the body movement of the elderly person is recorded as well as the voice of both participants. A hidden Markov model, to which those data are fed, estimates the latent conversational state during active listening. The body movement data with a small variance estimates three explicit states of the listener's speech, the speaker's speech, and silence, as well as two implicit states of the speaker's thinking and the speaker's laughter. On the other hand, the persona of large body movement variation indicates the same three explicit states, as well as the implicit state of the speaker's giving responses and an uninterpretable state.

Keywords: Active listening, Elderly, Apathy, Conversation, Hidden Markov model

INTRODUCTION

As the number of elderly people increases due to the aging of the population, it has got serious problems that many elderly people suffer from apathy syndromes (Brodaty et al., 2010; Ligthart et al., 2012) in elderly care facilities. Active listening, a counseling technique, is effective to solve the problem. Generally, the more people talk, the more they improve their willingness to live. Though elderly people are poor at engaging in healthy exercises, they have little difficulty talking about various experiences in their life. Aiming at improving their mood, listeners provide them with opportunities to talk a lot in active listening. Listeners are dispatched to various elderly people in active listening. Especially, at the beginning of the listening session, listers are strangers to elderly people. Since elderly people are not motivated to talk to strangers, listeners often fail to create atmospheres to get the elderly people to talk actively. To make matter worse, listeners are very few compared with elderly people. Many volunteers have to play the roles of listeners in active listening. Most listeners are not familiar with talks with elderly people, which makes it difficult for the listeners to lead a successful session in a short period. If we have a method to estimate the state of conversation to determine whether the elderly people and listeners are actively engaging in the conversation, we would be able to find a clue to bring successful active listening. The method would facilitate active listening so that many elderly people can improve their mental health. It is also expected to reduce the burden on the staff working in elderly care facilities. In the existing active listening, Cialdini et al. (1996) argue that persuasiveness in dialogue depends on the closeness of the participants. Using the result, Schulman et al. (2009) attempt to model active listening to reduce human effort. However, they have concluded that it is difficult to improve the closeness of participants in short active listening sessions. We should note that the method by Schulman's group only assumes two types of non-human aspects: text-only and conversational agents. They neglect the listening skills that humans use on the elderly. Such skills offered by listeners to the elderly are seen as a useful means of encouraging the elderly to talk. Active listening modeling should take the skills of the listener side into account. To achieve a model that facilitates active listening, it is necessary to estimate the conversational state of people participating in it. In addition, we should avoid any invasive methods giving physical loads to participants. Estimating a conversational state, human listeners would present conversational clues to people they want to lead to active talk, considering the characteristics of the people. The paper proposes a method to estimate the conversational state of participants of active listening, under what humans use to prompt conversation. It also pays attention to body movement and voice acquired during active listening using sensors, because those data can be collected without interfering with the participants of active listening. The paper presents experimental results where the human-centric method promotes conversation, considering the characteristics of persons who talk about their experiences to listeners.

Facilitation of Active Listening Taking Conversation State Into Account

Even if the speaker is actively talking while active listening, the conversation will not continue without a response from the listener. To facilitate conversation, listeners should have skills that prompt speakers to talk. Rogers et al. (1957) discuss the importance of the following three types of useful active listening skills. Self-agreement; listeners should always grasp the speakers' true meaning by asking back the speakers what they do not understand during active listening. Empathic understanding; Listeners should take speakers' viewpoints. Unconditional positive regards; Listers should always accept what speakers present in a positive manner, without evaluating it good or bad, showing their preference or avoidance. Note that, among the above three skills, empathic understanding and unconditional positive regard require listeners to speak spontaneously. To realize smooth listening, it is not realistic to expect these skills from listeners when there is no prior familiarity between speakers and listeners. It is advisable to always keep self-agreement in their mind, making effective use of the empathic understanding and the unconditional positive regard when speakers' speech is spontaneous.

Estimation of Intimacy Through Conversational State

The study refers to how the conversation proceeds well in active listening as the degree of intimacy. During active listening, the speaker and the listener transit states that differ in the liveliness of the conversation. The study assumes it comes from a variation of their intimacy. When the conversation turns out inactive, the listener can promote the speaker to talk actively because the listener can use skills to increase intimacy on that timing. It is essential to know the current state of the conversational on the spot. This study proposes an estimation model of the conversational state, to reveal features in each state. The methodological overview of the study is shown in Figure 1. The study assumes listeners who use the three types of skills useful for active listening intentionally reproduce a situation of high intimacy. During active listening, the speaker and the listener transit states, each of which differs in voice and body movement due to excitement variation. To estimate each state, the body movement of the speaker and the voice of the speaker and the listener are acquired during active listening using sensors (Mitsui Chemicals, n.d.). As the intimacy increases, the listener gets interested in the speaker. the listener wants to know more about the speaker. The speaker trusts the listener in the state. The speaker tries to let the listener know better himself or herself. As the intimacy between the participants increases, not only their topics would become more convincing but also their conversations should get more lively. It is expected that a high degree of intimacy facilitates the active listening process.

In actual active listening, it is not unusual for the speaker and the listener to engage in active listening while they meet each other for the first time. In such situations, the listener does not know what topics should be picked up to initiate conversation as well as the speaker does not know what to talk about, which often fails active listening. An awkward atmosphere at the beginning often warns persons who are not used to active listening, especially the elderly. They may be reluctant to converse from then on. To avoid such problems, it is considered practical to prepare topics that many people are likely to easily talk about. In order to make the initial states as equivalent as possible for every pair, the study gets all pairs to watch the same short

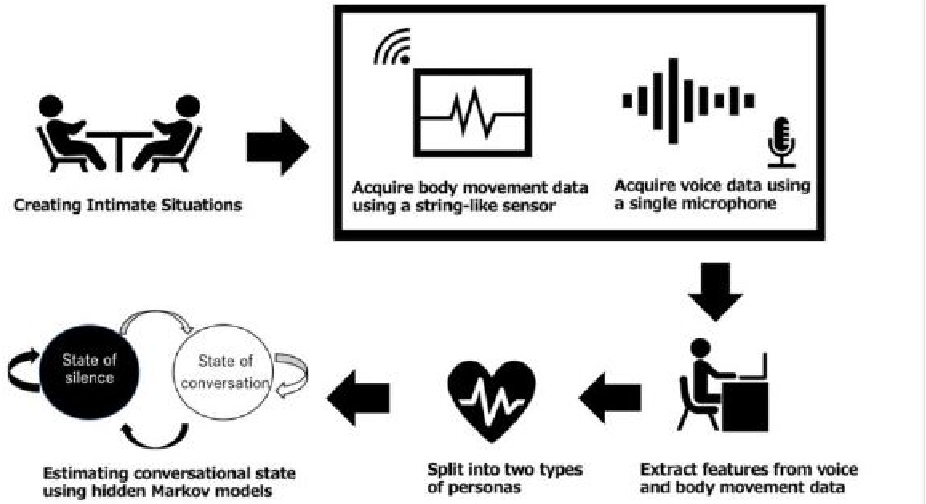

Figure 1: The methodological overview of the study.

commercial at the beginning. During the active listening session, the listener asks questions related to the situation that is the motif in the commercial to unify the topics.

The study identifies the conversation state from the viewpoint of whether the speaker actively talk. Since states are latent, they must be identified with observable data during a conversation. As those data, the research uses voice and body movements during conversation. Body movements during conversation vary with each person. Persons are divided into two types of personas. The first is a persona whose body movements are small during no conversation, while large for active conversation. In the other persona, body movements are always large, regardless of the activeness of speaking. In the first persona, body movements are large when the conversation is active. It facilitates the estimation of the conversational state. It is expected that the estimation of the state is likely to be successful. On the other hand, for the second persona, the body movements are large regardless of the conversational state, which indicates that the body movements are not important in the estimation of the conversational state. It is afraid that the state estimation may not be performed well. Due to these differences, the study proposes to estimate the conversation state for each of the two personas.

The more words are exchanged between the speaker and the listener, the more the conversation will go well. This indicates that the state of each time point in the conversation is influenced by the previous state. In other words, conversational data should be treated as time series data. In addition, the conversational state at each time point during the conversation is a hidden state that cannot be observed. To estimate the hidden state during active listening, the study applies a hidden Markov model, which is a kind of machine learning model to estimate unobserved states from the observed time series data. The features calculated from the body movement data and the conversation volume collected from the speech data are used as observable data to classify all states into three categories.

- The first is a state of active utterance from the listeners.
- The second is the state in which the speaker is active in utterance.
- The third is a silent state.

Experiments to Obtain Biometric Information During Active Listening

An experiment is conducted to estimate the state of conversation using voice and body movement data in active listening. The purpose of this study is to confirm the effectiveness of the skills by examining body movements and speech sounds during active listening. Elderly people generally present smaller movements and talk in weaker voices. To distinguish differences in body movement and voice depending on states of active listening, college students, not the elderly, play the roles of the speakers in the experiment. We can expect university students to provide far more vivid differences in body movements and speech sounds than elderly people.

The experiment verifies the following points:

- precise estimation of conversation state using hidden Markov models, and
- The usefulness of body movement data in estimating conversation state.

Table 1. The actual labeled items.

Silent state	Neither the listener nor the speaker issues utterances.
Active state of speech of the listener	Only the listener issues utterances.
Active state of speech of the speaker	The speaker issues utterance. (It may include utterances of the listener.)

To avoid variation of active listening skills in each experiment, the listener in the experiment is fixed to one specific person. The experiment is conducted as follows.

1. Speakers watch a commercial on a PC to unifiy the topic.
2. The listener asks questions on the content of the CM to the speaker to get his or her episodes related to it for about 5 minutes.During the period, the body movement of the speaker and voice of the both persons are collected using a pressure sensor attached to the seat of a chair and a microphone placed in front of the two persons. The pressure sensor consists of a metal wire, which detects changes in capacitance on the wire due to pressure.
3. A hidden Markov model is trained with the time series data obtained in Step 2, to map each time point in the time series data to the obtained hidden states.
4. The meaning of each hidden state is inferred from actual contexts at that time.

In this experiment, the spectrogram for each frequency is extracted from the time-series data of body movement and speech using the Fourie transform. In addition, the maximum, the minimum, the mean, and the variance are calculated for a time window of a fixed length. The conversation states are estimated, feeding these features as explanatory variables. The actual labeled items are shown in Table 1.

However, some of the features might be unnecessary for the estimation of conversational states. Those unnecessary features might prevent the hidden states assumed in this study from detection. The study uses RandomForest (Breiman, 2001), a machine learning algorithm, to calculate the importance of features. It selects features that are important for estimating conversational states. Since RandomForest is a supervised learning method, manual labeling is necessary for every second, actually listening to the speech content. The label is used as the objective variable. To select important features, the data for manual labeling should include all states to be labeled. A preliminary experiment has shown that all states are included if 80 seconds of data were obtained for each subject. The interval to be labeled for Random Forest is set to 80 seconds in this experiment.

Consideration

The study classifies the speakers into personas according to body movements. The features of the personas are listed below.

1. The first person takes large changes in body movement during an active conversation, while small changes in body movement when an conversation is not active.
2. The other persona always takes large changes in body movements regardless of conversations.

The study refers to the first as a weak vibration persona, while the second as a strong vibration persona. The accuracy of the prediction is calculated using RandomForest. For the initial prediction using only voice data, the accuracy is 0.56 and 0.61 for the weak vibration persona and the strong vibration persona, respectively. It is not high for either persona. After the addion of body movements, the prediction achieves highly accurate results of 0.96 and 0.99 for the weak vibration persona and the strong vibration persona, respectively. It indicates that body movements are useful for estimating the conversational state. For the weak vibration persona, the mean of the body movement data is the most important variable, followed by the mean of the voice, the variance of the voice, and the maximum value of the voice. It indicates that body movements are more important than speech data for estimating conversational states of the weak vibration persona. In contrast, in the results for the strong vibration persona, the top three variables are the maximum of voice, the variance of voice, and the mean of voice, followed by the variance of body movement data and the mean of body movement data. It turns out speech data are more important than body movement data for the estimation of conversational states of the strong vibration persona. It is noteworthy that, for both personas, the power spectrum of the frequency is less important for the estimation of the conversational state.

In this study, a hidden Markov model is used to estimate the conversational state. The identification of the state at each time point is attempted, assuming the three states shown in Table 1, which are assumed in the preliminary experiment. However, it is unable to estimate the state successfully. The cause is thought to be the existence of a hidden state other than the above three states. Another trial of estimation assuming 5 states fits better. In addition to the three assumed states, two other states are considered to be hidden. For the two hidden states, differences arose for each persona. Let us subdivide the conversation state for each persona. Figures 2 and 3 show the results of state estimation for the weak vibration persona.

Figure 2 is a graph showing the output of the color-coded states and body movement data. Figure 3 shows a graph of the color-coded states and the output of voice data. First, let us discuss the three states that are estimated in the experimental phase. In the light blue states, the voice is strong and the body movements are small. Iit can be estimated that the listeners are in an active state of speech. The actual speech in the light blue colored area contains the speech of the listeners. In the green state, the values of body movement data are larger than those in the light blue state, but are small compared with those in the overall state. The green state is assumed to be a silent state because of the small value of the voice data. The green state in the actual speech reveals the utterance of the speakers. One of the possible

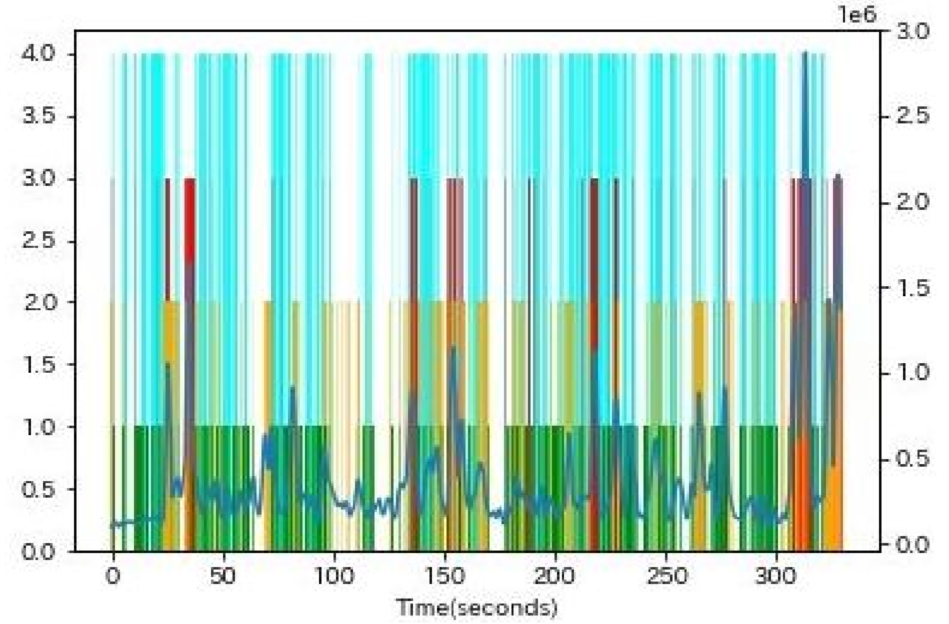

Figure 2: The output of the color-coded states and body movement data (the weak vibration persona).

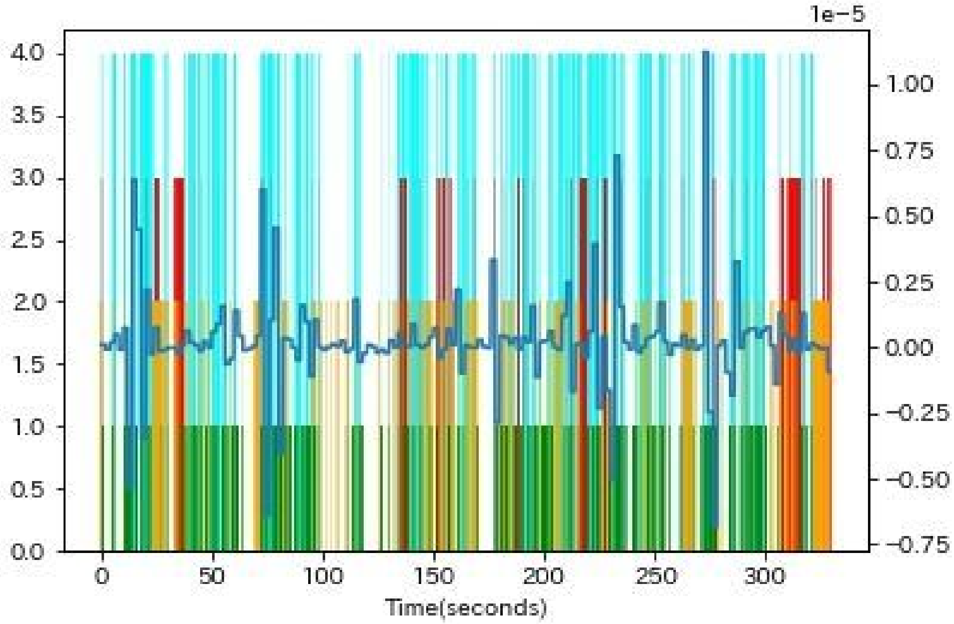

Figure 3: The color-coded states and the output of voice data (the weak vibration persona).

causes is that the microphone used for voice acquisition in the experiment may be placed in front of the listener during the experiment. Due to it, the voices of the listener may be acquired louder while those of the speakers in weaker amplitude. It may causes discrepancy between the estimated green status and the actual status. In the white state, the body movements and the voices are small, which leads to assuming that the subjects stay in a silent state. When the white state is investigated with contents recorded in actual speech data, no speech is confirmed by either the listeners or the speakers. The three states assumed in the preliminary experiment are successfully identified using the hidden Markov model. Secondly, we should give a discussion on the two states that have not been deduced in the experimental phase. In the red state, the body movement is larger while the voices are smaller. The actual voice shows the speakers' laughter, which means the speaker is laughing in the red state. In the yellow state, the body movement are smaller than in the red state, but larger than in all other states. Furthermore, we can confirm that the values of the voice data are small. The sound of the speakers breath being inhaled is confirmed in the yellow state in the actual speech. Before the yellow state, actual speech data shows there are many acts of the listeners asking questions to the speakers. Since the body movement data are large, it is estimated that the speakers are in the state of thinking while taking a deep breath in the yellow state. Figures 4 and 5 show the results of state estimation for the strong vibration persona.

In the red state, the body movement and the voice are large, which indicates that the listeners are in an active state of speech. It is actually confirmed the actual speech contains the utterance of the speakers in the red state. In the green state, both large and small values are observed for both body movement data and voice data. The estimation of the state might fail. However, checking the green state in the actual speech data finds only the utterance of the listener. It leads the green state to a state in which the listener's speech is active. In the white state, both the body movement and voices are

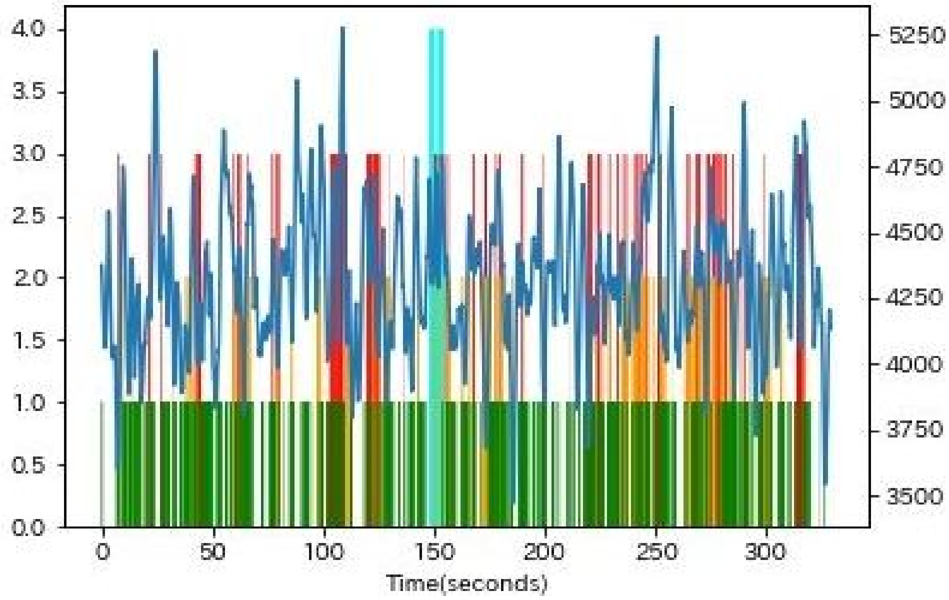

Figure 4: The output of the color-coded states and body movement data (the strong vibration persona).

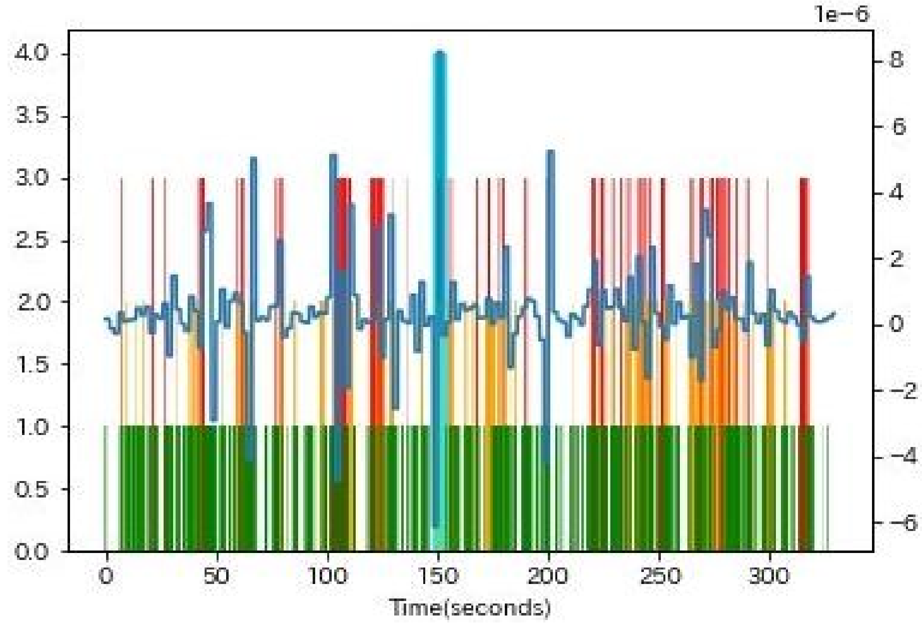

Figure 5: The color-coded states and the output of voice data (the strong vibration persona).

small, which indicates the silent state. In the white states in the actual speech, neither the listener nor the speaker presents utterance. The white state is considered to be a silent one as well as the weak vibration persona. The three states estimated in the experimental phase are successfully estimated using the hidden Markov model. To discuss two states that have not been deduced at the experimental stage. In the yellow state, both small and large values of body movement data are observed. The value of voice data is small. From the actual voice data, the yellow state contains a scene in which the speakers gives a "sop" to the speakers. The yellow state is considered to be a state in which the speakers are listening attentively to the listener. In the light blue state, the body movement data and the voice data are larger. It can be inferred that the speakers are in an active state of speech. The light blue state in the actual speech contains utterances of the speakers. It is considered that the light blue state indicates that the speakers' speech is active. However, no difference is confirmed between the light blue state and the red state. Furthermore, since the time spent in the light blue state is significantly less than in the red state, but no special behavior is observed, it leads to an assumption that the light blue and red states are identical, or that the light blue state has a special meaning which goes over ideas of the analysts. Considering the former case, the estimation for the four states is attempted, but it is impossible to separate the speech of the speakers from that of the listeners. It implies the light blue state is likely to be some special state of the speakers.

CONCLUSION

The paper proposes a method to estimate the state of conversation for each persona classified with body movement to facilitate active listening to elderly people. The weak vibration persona turns out to have 5 states. For the strong vibration persona, four states are identified, but it still has an uninterpretable state. The results suggest body movements can inform the listener of the hidden conversational state of the speakers. The information lets the listener know the timing to use skills to increase intimacy, which can keep active listening smooth.

In the future, it will be necessary to acquire video recordings during the experiment to interpret the state that cannot be interpreted in the present study.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to Professor Hiromitsu Shimakawa of Ritsumeikan University for his invaluable guidance in conducting this research and writing this paper.

REFERENCES

Breiman. Random forests. Machine Learning, Vol. 45, No. 1, pp. 5–32, 2001.

Brodaty H, Altendorf A, Withall A, et al. Do people become more apathetic as they grow older? A longitudinal study in healthy individuals. *Int Psychogeriatr.* 2010;22(3): 426–436.

Cialdini, R. B. and Goldstein, N. J. Social influence: compliance and conformity. Annu Rev Psychol, 55, (2004) 591–621 Vol. 39, No. 2, pp. 86–98, 1996.

Daniel Schulman, Timothy Bickmore. Persuading users through counseling dialogue with a conversational agent. Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology. April 2009. Article No. 25, Pages 1–8.

Ligthart SA, Richard E, Fransen NL, et al. Association of vascular factors with apathy in community-dwelling elderly individuals. *Arch Gen Psychiatry.* 2012;69(6): 636–642.

Mitsui Chemicals, "Tension sensor Piezola", https://jp.mitsuichemicals.com/jp/service/product/piezoelctric- line.htm

Rogers, C. R. "The necessary and sufficient conditions of therapeutic personality change." Journal of Consulting Psychology, 21(2), 95–103, (1957).