

Listening Skills Assessment through Computer Agents

Hiroki Tanaka

Graduate School of Information Science, Nara Institute of
Science and Technology, Japan
hiroki-tan@is.naist.jp

Hidemi Iwasaka

Developmental Center for Child and Adult, Shigisan
Hospital, Japan
iwasaka@heartland.or.jp

Hideki Negoro

Center for Special Needs Education, Nara University of
Education, Japan
gorosan@nara-edu.ac.jp

Satoshi Nakamura

Data Science Center, Nara Institute of Science and
Technology, Japan
s-nakamura@is.naist.jp

ABSTRACT

Social skills training, performed by human trainers, is a well-established method for obtaining appropriate skills in social interaction. Previous work automated the process of social skills training by developing a dialogue system that teaches social skills through interaction with a computer agent. Even though previous work that simulated social skills training considered speaking skills, human social skills trainers take into account other skills such as listening. In this paper, we propose assessment of user listening skills during conversation with computer agents toward automated social skills training. We recorded data of 27 Japanese graduate students interacting with a female agent. The agent spoke to the participants about a recent memorable story and how to make a telephone call, and the participants listened. Two expert external raters assessed the participants' listening skills. We manually extracted features relating to eye fixation and behavioral cues of the participants, and confirmed that a simple linear regression with selected features can correctly predict a user's listening skills with above 0.45 correlation coefficient.

CCS CONCEPTS

• **Human-centered computing** → *Interaction design*;

KEYWORDS

Listening; computer agents; social skills training; prediction

ACM Reference Format:

Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2018. Listening Skills Assessment through Computer Agents. In *2018 International Conference on Multimodal Interaction (ICMI '18)*, October 16–20, 2018, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3242969.3242970>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMI '18, October 16–20, 2018, Boulder, CO, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5692-3/18/10...\$15.00

<https://doi.org/10.1145/3242969.3242970>

1 INTRODUCTION

Social communication skills are critical factors that influence human life. Persistent social skill deficits impede those with such afflictions from forming relationships or succeeding in social situations. Social skills training (SST), a general psychosocial treatment through which people with social difficulties can obtain appropriate social skills, is widely used by teachers, therapists, and trainers. Automating the SST process would simplify the acquisition of such social skills by those who require them.

It may also be easier for those with social difficulties to use computer agents than to directly interact with a human. Using computer agents in SST is motivated by the fact that even though people with social difficulties have difficulty during social communication, they also show good or sometimes even superior systemizing skills [3, 9]. Systemizing is the drive to analyze or build systems and understand and predict behavior in terms of underlying rules and regularities. The use of systematic computer-based training for people who need to improve their social skills can exploit the following facts: 1) such people favor computerized environments because they are predictable, consistent, and free from social demands; 2) they can work at their own speed and level of understanding; 3) training can be repeated over and over until the goal is achieved; and 4) interest and motivation can be maintained through computerized rewards.

Previous works conducted social skills training using computer agents [6], for instance, in the contexts of narrative, public speaking, and emotional regulation [14, 26, 27, 31].

The Bellack method (named step-by-step SST) [5] is a well established and widely used approach. It defines the framework of SST, and defines four basic skills in SST: speaking skills, listening skills, asking skills, and expressing feeling of discomfort. Here, important aspects of listening skills are: 1) looking a partner in the eye, 2) nodding, and 3) repeating keywords of the conversation partner [5]. Listening skills considered in this study are not for hearing and understanding of speech [4, 7]. Listening skills are to explicitly express that one is listening to the partner's speech [5, 29]. One study showed that people with social difficulties don't tend to look conversation partners in the eye [15].

In spite of the importance of listening skills, most automated SSTs focused on users' speaking skills. Okada et al. proposed assessing interaction skills considering listening attitudes [23]. There have been differently motivated works designed to generate (model) human-like head tilting and nodding on humanoid robots or virtual agents based on analyzing human behaviors [12, 13, 16, 19, 21]. Ward et al. proposed listening-skills training



Figure 1: Interaction with computer agents.

that produces immediate feedback, although they did not use computer agents and focused only on back-channeling behavior [30]. Also, one previous work showed that personality [10] was related to empathic listening skills [25].

In this paper, we hypothesized that the SST process in listening skills can be automated for interaction between humans and computer agents. As a first step, we analyzed a part of the automatic assessment of listening skills. We collected listening data between interaction of graduate students and computer agents, and investigated the possibility of automatically assessing user’s listening skills.

2 COMPUTER AGENTS

MMDAgent [18] was used as the computer agent. We used default parameters for the agent’s speech such as speaking rate and voice pitch.

Four Japanese people (two males and two females) created the agent’s spoken sentences. Here, one person is a license psychiatrist, who had more than three years of experience with SST, and one is a licensed speech therapist. We created three types of tasks: Speaking, Listening 1, Listening 2. We explain them in detail as follows.

- (1) Speaking: The user tells a recent memorable story to the computer agent. This module follows the same procedure as a previous work [27].
- (2) Listening 1: The user listens to the agent’s recent memorable story. Table 1 shows sentences we created (translated into English). This supposes casual social small talk.
- (3) Listening 2: The user listens to a procedure of how to make a telephone call. These sentences are shown in Table 1 (translated into English). They are designed for a more serious situation such as job training.

Regarding Listening 1 and Listening 2, the agent spoke for about one minute. There were several three- or five-second pauses between the sentences. During the pause, the agent nodded her head if the user said something, and waited three more seconds after the final user utterance.

3 DATA COLLECTION

3.1 Participants

We recruited 27 participants (6 females and 21 males, with a mean age of 25.1, SD: 2.13) from the ***. We confirmed that participants had no hearing difficulties by directly asking them. The first author explained the experiment to the participants to obtain informed consent. The participants completed continuous Speaking (60 sec), Listening 1 (60-90 sec), Listening 2 (60-90 sec) sessions. The completion time changed according to how many times the participants spoke.

3.2 Procedure

The first author explained how to use the system by playing an example video showing head nodding and backchannel feedback.

Data was collected in a soundproof room, using a laptop PC (IBM ThinkPad). A WebCam (ELECOM UCAM-DLY300TA) was placed on top of the laptop, and an eye-tracker (Tobii X2-30) was placed on the bottom of the laptop screen. We turned off the light in the room to minimize external distractions (Figure 1).

After collecting data, we conducted two questionnaires explained in a later section from all participants. Total amount of time for all procedures was approximately 20 minutes. From the collected data, we calculated the following eye fixation, video, and audio features. These features were selected based on previous studies [13, 16, 21], specifically important aspects of listening skills from the Bellack method of SST [5].

3.3 Eye Fixation

An IV-filter was applied to the raw eye-gaze data. We calculated the following features: 1) standard deviation of horizontal fixation axis, 2) standard deviation of vertical fixation axis. We also manually categorized areas of interest as follows: 3) eyes, 4) mouth, 5) face, 6) other.

3.4 Video Annotation

We coded head nodding (video) and speech (audio) using the ELAN tool. The following information was coded by one male annotator: backchannel feedback (e.g. “un”, “hai” in Japanese), repetition of agent’s utterance (paraphrase), question, miscellaneous utterance, head nod (once), head nod (twice), and head nod (three or more times). Here, we defined more than one second as separate coding.

As an output of our coding, we extracted the following features: 1) the number of backchannel feedback instances, 2) the number of repetitions, 3) the number of questions, 4) the number of miscellaneous utterances, and 5) the number of nods. Here, we simplified that the number of nods is counted as the total amount of head nods of once, twice, and three or more times based on [21].

3.5 Social Responsiveness Scale and Big Five Personality Test

We conducted two questionnaires: the Social Responsiveness Scale (SRS) [8], which is related to autistic traits based on the DSM-V [1], and the Big Five Personality Test [11]. The Big Five Personality Test consists of the following subareas: extraversion, agreeableness, conscientiousness, neuroticism, and openness. The relationship between these two questionnaires was investigated in [28].

Table 1: Sentences spoken by computer agents. <pause> denotes three seconds of silence, and <long pause> denotes five seconds of silence.

<p><Listening 1> The other day, my friends and I went to a fashionable cafe in Kyoto that was advertised in a magazine. <pause> I drank a caffe latte. The foam on top of the cafe latte had pictures of animals, and it was so cute. So I took a picture. I also uploaded it to Instagram. <pause> After that, because my friend came by car, we went to Kiyomizu Temple by car. I found a souvenir shop on a side street on the way to Kiyomizu Temple. After thinking a long time, I finally bought a very delicious roll cake. <pause> We reached Kiyomizu Temple. At the temple, the autumn leaves were very beautiful. I think that cherry blossoms are also beautiful, so I would like to go there in the spring as well. That’s all. Is there anything you would like to ask? <long pause> Thank you.</p>
<p><Listening 2> First, you dial number and connect; then tell the person your name and affiliation. <pause> After that, please say the name and affiliation of the person who you would like to talk to, and ask to be connected. <pause> When you are connected to the person in charge, state your message briefly. <pause> If the person in charge is not present, say you will call back, and then hang up. <pause> It is important to make a phone call at a proper time. To make a phone call at midnight or early in the morning will annoy people, so avoid doing so. <pause> This is the end of the explanation. Do you have any questions? <long pause> Thank you.</p>

Table 2: Top five features. Brackets denote correlation coefficient (: $p < 0.01$, *: $p < 0.05$).**

Rank.	Listening 1	Listening 2
1	# of nods (0.51**)	# of backchannel (0.55**)
2	# of questions (0.42*)	# of nods (0.42*)
3	# of repetitions (0.25)	SD of horizontal (-0.25)
4	# of miscellaneous (0.22)	# repetitions (0.23)
5	eyes (0.21)	face (-0.14)

3.6 Ground Truth of Listening Skills

Two licensed clinical psychologists (females), who had more than three years of experience with SST, rated listening skills as well as speaking skills by watching videos exported from the Tobii video recorder. Our instruction to the raters was that you should pay attention to the participants’ impression in addition to behaviors such as eye movement, head nodding, facial expression and speech as with the usual SST. We also directed that after watching multiple videos, they should evaluate our participants’ overall listening skills with Likert scores on a scale of 1 (not good) to 7 (good) [27].

Kappa statistics of the two raters were calculated using a weighted Kappa set to be 1 (on the diagonal) and decreased weights off the diagonal [17]. The following are weighted Kappa correlation coefficients: 0.37 (Speaking), 0.47 (Listening 1), and 0.59 (Listening 2). This is a fair to moderate agreement in accordance with [17]. Also, we calculated Pearson’s correlation coefficients of two raters as follows: 0.44 (Speaking), 0.46 (Listening 1), and 0.66 (Listening 2) (all, $p < 0.05$). Finally, we averaged the two raters’ score for further analysis.

4 EXPERIMENTAL EVALUATION

This section represents our experimental evaluation of the collected data. After analyzing the relationship between each feature,

we finally evaluated our prediction model toward automatic listening skills assessment. In most of this study, Spearman’s correlation coefficient was used to observe correlation. We selected good model persons who scored above five in both listening skills according to [2, 27] to be used as good examples in SST.

First, we analyzed the relationship between each question and listening skills. Then, we normalized extracted features using the z-score normalization. Regarding automatic assessment, we used multiple linear regression, which is a very simple linear approach to predicting listening skills. Leave-one-person-out cross validation was performed. We automatically selected features based on Akaike’s information criterion (AIC) in a stepwise algorithm on the training set. Moreover, the random forest regression was used to assess listening skills as a non-linear model. We set the number of variables tried at each split as four.

Finally, after confirming normality by the Kolmogorov-Smirnov test, we calculated the Pearson’s correlation coefficient between actual values and predicted values. We also calculated the root mean square error (RMSE).

5 RESULTS

5.1 Correlation Analysis

The correlation coefficient of Listening 1 and Speaking was 0.31 ($p=0.10$), that of Listening 2 and Speaking was 0.41 ($p=0.03$), and that of Listening 1 and Listening 2 was 0.54 ($p=0.003$). Five persons were selected as good model persons. Although we found no significant differences between other metrics (SRS and Big Five Personality) and listening, we observed SRS mean values of 47.2 (SD: 20.5) for good model persons and 68.27 (SD: 19.8) for other persons (Wilcoxon rank sum test (one-tailed): $p=0.03$).

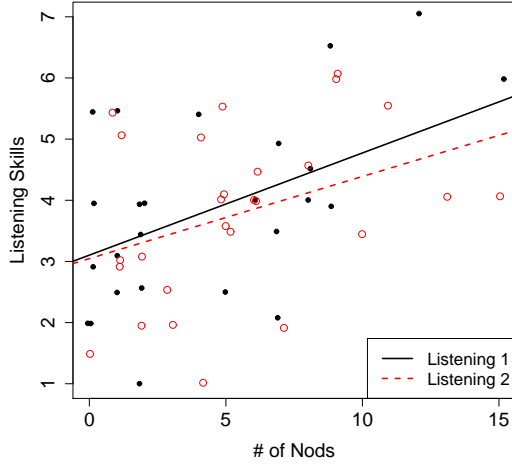


Figure 2: Relationship between listening skills and # of nods. Jitters were added to avoid overlapping same data points.

Regarding eye fixation, we found mean values of 40% for eyes, 7% for mouth, 88% for face, and 5% for other within all fixation points in all participants. Table 2 indicates top five correlated features to listening skills. We can see that # of nods was significantly related to the listening skills. Figure 2 shows a relationship between these two attributes. Figure 3 shows two examples from persons who scored 6 (P12) and 2 (P9) in terms of head nodding and backchannel feedback in Listening 2. We confirmed that persons with a low listening score tended to nod only during the agent’s pauses. In contrast, persons with a high listening score also nodded and uttered at other times. For example, they tended to respond at positions of specific keywords, commas and periods of the agent’s transcripts within the sentence according to pitch of agent’s speech.

5.2 Listening Skills Assessment

For the linear regression, we found a correlation coefficient between the predicted value and actual values as follows: Listening 1 was 0.45 ($p=0.01$), and Listening 2 was 0.47 ($p=0.01$). The RMSE was 1.58 (Listening 1) and 1.19 (Listening 2). In contrast, the random forest regression obtained the following correlation coefficients: 0.26 (Listening 1) and 0.34 (Listening 2), which were lower performance than the linear regression. This indicates the feature set might be linearly related to listening skills.

Here, we tried to find important features identified from the linear regression. We counted remaining times of each feature in each fold. We confirmed that mouth, face, eyes, the number of nods, and the number of repetitions, were important for Listening 1, and SD of vertical fixation axis, face, the number of backchannel feedback instances, the number of miscellaneous utterances were important for Listening 2.

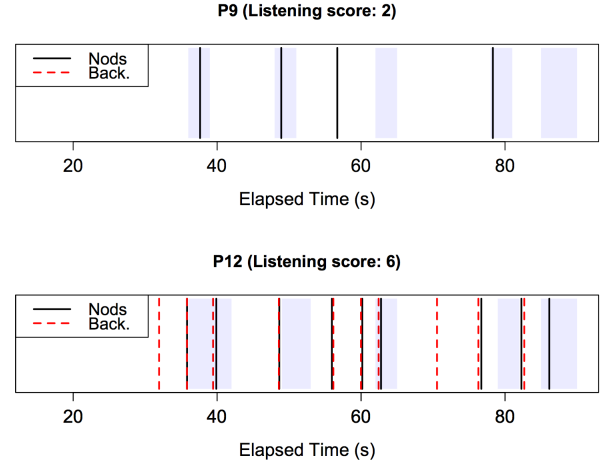


Figure 3: Timing of head nodding and backchannel feedback. Colored areas denote agent’s pauses.

6 DISCUSSION

We collected data from three types of settings in human-computer interaction. Several behavioral features were coded and extracted based on [5, 13, 16, 21], and the linear regression model achieved prediction of 0.45 in Listening 1 and 0.47 in Listening 2 of correlation coefficients. We confirmed that the correlation coefficient of two raters was 0.46 (Listening 1) and 0.66 (Listening 2), and that our prediction model achieved similar prediction in Listening 1 and a previous work on speaking skills [27]. In the case of Listening 2, the human raters were more agreed than our prediction model. In Listening 2, we found that P5 (female) was predicted as 3.7 in our model; in contrast, the two raters rated 5 and 7. Although P5 responded with not very much backchannel feedback and nodding, she had appropriate timing, speech was loud and clear, and she smiled. We need to consider and extract such additional information to improve our model [16]. In the future, we should compare persons who obtained a high SRS score and others in order to know the normal range of human behavior in this regard. We also found that the amount of backchannel feedback was more important in Listening 2 than in Listening 1 because Listening 2 was a more serious type of interaction, requiring explicit cues to show one was listening.

This study did not investigate the effects of human-computer interaction and human-human interaction. A previous work suggested that people treat computers as real people, showing that people are polite to computers [24]. In contrast, a recent work found that the dynamics of facial expressions differ for users interacting with a human or a virtual agent [22]. We need to consider these effects in the future. We will integrate our listening-skills assessment into the automation framework [20] and test it on people with autism spectrum disorders.

ACKNOWLEDGMENTS

The part of this work was supported by JSPS KAKENHI Grant Numbers JP17H06101, and JP18K11437.

REFERENCES

- [1] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: Dsm-5*. Amer Psychiatric Pub Incorporated. <https://books.google.co.jp/books?id=ElbMlwEACAAJ>
- [2] Albert Bandura. 1978. Social learning theory of aggression. *Journal of communication* 28, 3 (1978), 12–29.
- [3] S. Baron-Cohen, J. Richler, D. Bisarya, N. Gurnathan, and S. Wheelwright. 2003. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 1430 (Feb 2003), 361–374.
- [4] J. G. Barry, D. Tomlin, D. R. Moore, and H. Dillon. 2015. Use of Questionnaire-Based Measures in the Assessment of Listening Difficulties in School-Aged Children. *Ear Hear* 36, 6 (2015), 300–313.
- [5] A.S. Bellack, K.T. Mueser, S. Gingerich, and J. Agresta. 2013. *Social Skills Training for Schizophrenia, Second Edition: A Step-by-Step Guide*. Guilford Publications. <https://books.google.co.jp/books?id=TSMxAAAAQBAJ>
- [6] Justine Cassell. 2001. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Mag.* 22, 4 (Oct. 2001), 67–83. <http://dl.acm.org/citation.cfm?id=567363.567368>
- [7] Fatih Cigerci and M Gultekin. 2017. Use of digital stories to develop listening comprehension skills. 27 (2017), 252–268.
- [8] John N Constantino, Sandra A Davis, Richard D Todd, Matthew K Schindler, Maggie M Gross, Susan L Brophy, Lisa M Metzger, Christiana S Shoushtari, Reagan Splinter, and Wendy Reich. 2003. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of autism and developmental disorders* 33, 4 (2003), 427–433.
- [9] O. Golan and S. Baron-Cohen. 2006. Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Dev. Psychopathol.* 18, 2 (2006), 591–617.
- [10] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37 (2003), 504–528.
- [11] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.
- [12] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating Rapport with Virtual Agents. In *Lecture Notes in Artificial Intelligence; Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA)*, Vol. 4722. Paris, France, 125–128. <http://ict.usc.edu/pubs/Creating%20Rapport%20with%20Virtual%20Agents.pdf>
- [13] Dirk Heylen. 2008. Listening Heads. In *Proceedings of the Embodied Communication in Humans and Machines, 2Nd Zif Research Group International Conference on Modeling Communication with Robots and Virtual Humans (Zif'06)*. Springer-Verlag, Berlin, Heidelberg, 241–259. <http://dl.acm.org/citation.cfm?id=1794517.1794530>
- [14] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My Automated Conversation Coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 697–706. <https://doi.org/10.1145/2493432.2493502>
- [15] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. 2002. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 9 (2002), 809–816.
- [16] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, 127–136. <http://aclweb.org/anthology/W17-5516>
- [17] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977).
- [18] Akinobu Lee, Keiichiro Oura, and Keiichi Tokuda. 2013. Mmdagent - A fully open-source toolkit for voice interaction systems. In *ICASSP*.
- [19] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita. 2012. Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 285–292.
- [20] Fang Liu, Dinoj Surendran, and Yi Xu. 2006. Classification of statement and question intonations in Mandarin. *Proc. 3rd Speech Prosody* (2006), 603–606.
- [21] Fumie Nori, Afia Akhter Lipi, and Yukiko Nakano. 2011. Cultural Difference in Nonverbal Behaviors in Negotiation Conversations: Towards a Model for Culture-adapted Conversational Agents. In *Proceedings of the 6th International Conference on Universal Access in Human-computer Interaction: Design for All and Inclusion - Volume Part I (UAHCI'11)*. Springer-Verlag, Berlin, Heidelberg, 410–419. <http://dl.acm.org/citation.cfm?id=2022591.2022639>
- [22] Magalie Ochs, Nathan Libermann, Axel Boidin, and Thierry Chaminade. 2017. Do You Speak to a Human or a Virtual Agent? Automatic Analysis of User's Social Cues During Mediated Communication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. ACM, New York, NY, USA, 197–205. <https://doi.org/10.1145/3136755.3136807>
- [23] Shogo Okada, Yoshihiko Ohtake, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsuan Huang, Yutaka Takase, and Katsumi Nitta. 2016. Estimating Communication Skills Using Dialogue Acts and Nonverbal Features in Multiple Discussion Datasets. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 169–176. <https://doi.org/10.1145/2993148.2993154>
- [24] Byron Reeves and Clifford Ivar Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.
- [25] Ceri M. Sims. 2017. Do the Big-Five Personality Traits Predict Empathic Listening and Assertive Communication? *International Journal of Listening* 31, 3 (2017), 163–188. <https://doi.org/10.1080/10904018.2016.1202770> arXiv:<https://doi.org/10.1080/10904018.2016.1202770>
- [26] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2017. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS ONE* 12, 8 (08 2017), 1–15. <https://doi.org/10.1371/journal.pone.0182151>
- [27] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2015. Automated Social Skills Trainer. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. ACM, New York, NY, USA, 17–27. <https://doi.org/10.1145/2678025.2701368>
- [28] M. N. Tsai, C. L. Wu, L. P. Tseng, C. P. An, and H. C. Chen. 2018. Extraversion Is a Mediator of Gelotophobia: A Study of Autism Spectrum Disorder and the Big Five. *Front Psychol* 9 (2018), 150.
- [29] Babita Tyagi. 2013. Listening: An important skill and its various aspects. *The Criterion An International Journal in English* 12 (2013), 1–8.
- [30] Nigel G. Ward, Rafael Escalante, Yaffa Al Bayyari, and Thamar Solorio. 2007. Learning to show you're listening. *Computer Assisted Language Learning* 20, 4 (2007), 385–407. <https://doi.org/10.1080/09588220701745825> arXiv:<https://doi.org/10.1080/09588220701745825>
- [31] Ru Zhao, Vivian Li, Hugo Barbosa, Gourab Ghoshal, and Mohammed Ehsan Hoque. 2017. Semi-Automated 8 Collaborative Online Training Module for Improving Communication Skills. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 32 (2017), 32:1–32:20 pages. <https://doi.org/10.1145/3090097>