

Listening Skills Assessment through Computer Agents

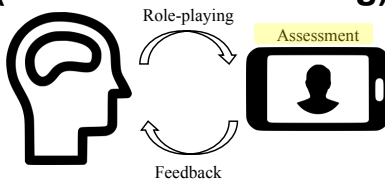
Hiroki Tanaka¹, Hidemi Iwasaka², Hideki Negoro³, and Satoshi Nakamura¹

¹Nara Institute of Science and Technology, ²Shigisan Hospital, ³Nara University of Education, Japan
hiroki-tan@is.naist.jp

Towards automated SST (Social Skills Training)

SST framework [Bellack, A. S. 2004]

Num.	Procedure
1	Defining target skills
2	Modeling
3	Role-play
4	Feedback
5	Reinforcement
6	Homework



We propose assessment of user listening skills during conversation with computer agents

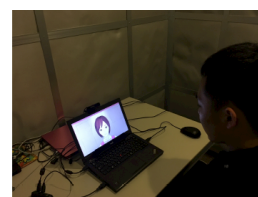
This presentation

- Motivation of listening skills assessment
- Agents and data collection
- Evaluation based on correlation

1. Introduction

- SST is a general psychosocial treatment through which people with social difficulties can obtain appropriate social skills
- Previous works conducted SSTs using computer agents, for instance, in the contexts of interview, public speaking, and emotional regulation [Hoque et al., 2013, Zhao et al., 2017, Tanaka et al., 2017]
- Most automated SSTs focused on users' speaking skills
- We analyzed a part of the assessment of listening skills

- Feature extraction
Eye fixation, manual video annotation (use ELAN)
- Two licensed clinical psychologists rated listening skills by watching videos (ground truth: 1 to 7)



2. Computer agents

- MMDAgent was used as the computer agent
- Four Japanese people (two males and two females) created the agent's spoken sentences
- We created three types of tasks:
 1. Speaking: The user tells a recent memorable story to the computer agent
 2. Listening 1: The user listens to the agent's recent memorable story. *This supposes casual social small talk*
 3. Listening 2: The user listens to a procedure of how to make a telephone call. *They are designed for a more serious situation such as job training*

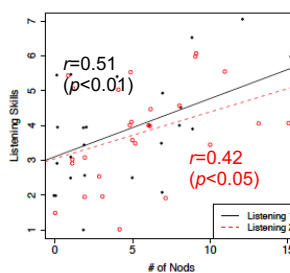
Listening 1-
The other day, my friends and I went to a fashionable cafe in Kyoto that was advertised in a magazine. <pause>
I drank a cafe latte. The foam on top of the cafe latte had pictures of animals, and it was so cute.
So I took a picture. I also uploaded it to Instagram. <pause>
After that, because 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.

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.

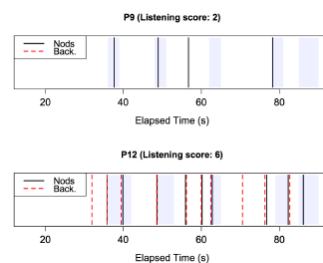
4. Evaluation based on correlation

Top five important features which correlates rated listening skills (**: $p < 0.01$, *: $p < 0.05$)

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)



Relationship between listening skills and # of nods



Timing of head nodding and backchannel feedback. Colored areas denote agent's pauses

Multi linear regression with selected features based on AIC can predict a unseen user's listening skills with **0.45** (Listening 1) and **0.47** (Listening 2) correlation coefficient

3. Data collection

- 27 participants (6 females and 21 males, with a mean age of 25.1, SD: 2.13)
- We conducted the Social Responsiveness Scale and the Big Five Personality Test
- We recorded interaction between computer agents and participants

5. Discussions and future works

- The correlation coefficient of two raters was 0.46 (Listening 1) and 0.66 (Listening 2)
- Our prediction model achieved similar prediction in Listening 1
- Integrate our listening-skills assessment with timing into the automation framework
- Test it on people with autism spectrum disorders