Linguistic Features during Speech Utterances in the Context of Social Skills Training

Hiroki Tanaka[1], Hidemi Iwasaka[2], Hideki Negoro[3] and Satoshi Nakamura[1]

***Abstract*— Previous works automated the process of social skills training by developing an embodied conversational agent. This paper investigates the linguistic features that can be integrated into a training framework for automated social skills training. We prepared two datasets: narratives spoken by adults and children/adolescents with autism spectrum disorders. Data were collected in human and human interaction, and human social skills trainer(s) rated the speakers' overall speaking skills. We analyzed speech content based on such linguistic features as type and token, word vector representation, conjunction usage, and parse tree depth. Finally, we confirmed the important features: the number of tokens and the mean of depth tree.**

Clinical relevance: **This study can be utilized to easily identify language features that must be improved.**

I. Introduction

Social skills training (SST) is a psychosocial treatment through which people with social difficulties can obtain appropriate social skills [1]. Automating the SST process will simplify the acquisition of social skills. Previous works automated parts of SST using embodied conversational agents for improving speaking skills, listening skills [2], and job interviews [3]. SST consists of role-playing and feedback in terms of multimodal behaviors. However, most of the automated SSTs focused on analyzing the correlations with paralinguistic features and nonverbal behaviors (e.g., prosody and facial expression) although speech contents are also highly correlated with social skills [2]. This study attempt to analyze speech contents that are related to social skills. We manually collected two datasets that include the speech of adults and children/adolescents with autism spectrum disorders. We computed some linguistic features and calculated the correlation coefficient with overall speaking skills that were rated by experienced human social skills trainers.

II. Method

Research Ethics Committee of the Nara Institute of Science and Technology approved this experiment. We recruited 18 native Japanese-speaking adults, and nine male children/adolescents with autism spectrum disorders (ages 7-19) without intellectual disability ($IQ > 70$). The participants first told a recent fun story to a person for one minute, used the social training system [2], and repeated it to the same person. A male and female social skills trainers evaluated the participants' overall speaking skills on a 1 to 7 Likert scale respectively. They watched randomly ordered videos and rated the scores. Manual transcription was done on the above two datasets. We separated two sentences when a pause exceeded 0.5 seconds or the raters judged to be different utterances. We extracted following linguistic features: 1) the number of tokens in each speaking instance, 2) the type-token ratio (TTR), which represents the ratio of the total vocabulary to the overall words, 3) the number of conjunctions, 4) the number of filler utterances as "umm" or "eh", 5) the similarity of the word vector representation as we used a pre-trained Word2vec model with 200 dimensions (`http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/`). We calculated the word vector representation of ten words and sequentially shifted the word with a 50% window until the end of the speech. We averaged the cosine similarity of these adjacent frames, 6) the mean of tree depth. We used KNP [4] for the parse tree.

TABLE I

Correlation coefficients to overall speaking skills

	Tokens	TTR	Conj.	Fillers	Similarity	Depth
Adults	0.50*	-0.09	0.03	0	0.10	0.38*
Children	0.57*	0	0.53*	0.29	-0.07	0.43†

* represents $p < 0.05$, and † represents $p < 0.1$ compared to no correlation.

III. Results and Conclusions

Table I represents the Pearson's correlation coefficient for each feature. The number of tokens, the number of conjunctions, and the mean of tree depth are significantly correlated to overall speaking skills. These features can be used for future automated SSTs.

References

[1] A. Bellack, K. Mueser, S. Gingerich, and J. Agresta, *Social Skills Training for Schizophrenia, Second Edition: A Step-by-Step Guide*. Guilford Publications, 2013.

[2] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura, "Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders," *PLoS ONE*, vol. 12, no. 8, 2017.

[3] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated conversation coach," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 697–706.

[4] D. Kawahara and S. Kurohashi, "A fully-lexicalized probabilistic model for japanese syntactic and case structure analysis," in *HLT-NAACL*. USA: Association for Computational Linguistics, 2006, pp. 176–183.

*Funding was provided by the Core Research for Evolutional Science and Technology (Grant No. JPMJCR19A5) and Japan Society for the Promotion of Science (Grant Nos. JP17H06101 and JP18K11437).

[1]Hiroki Tanaka and Satoshi Nakamura are with the Nara Institute of Science and Technology, Ikoma, Nara, Japan `hiroki-tan@is.naist.jp`

[2]Hidemi Iwasaka is with the Department of Psychiatry, Nara Medical University, Kashihara, Nara, Japan

[3]Hideki Negoro is with the Center for Special Needs Education, Nara University of Education, Nara, Japan