Adaptive Robot Discourse for Language Acquisition in Adulthood

Ronald Cumbal
Division of Speech, Music and Hearing
KTH - Royal Institute of Technology
Stockholm, Sweden
ronaldcg@kth.se

***Abstract*—Acquiring a second language in adulthood differs considerably from the approach taken at younger ages. Learning rates tend to decrease during adolescence, and socio-emotional characteristics, like motivation and expectations, take a different perspective for adults. In particular, acquiring *communicative competence* is a stronger objective for older learners, as an appropriate use of language in social contexts ensures a better community immersion and well-being. This skill is best attained through interactions with proficient speakers, but if this option is not available, social robots present a good alternative for this purpose. However, to obtain optimal results, a robot companion should adapt to the learner's proficiency level and motivation continuously to encourage speech production and increase fluency. Our work attempts to achieve this goal by developing an adaptive robot that modifies its spoken dialogue strategy, and visual feedback, to reflect a student's knowledge, proficiency and engagement levels in situated interactions for long-term learning.**

***Index Terms*—Dialogue system, second language learning, social robot, communicative competence, situated interaction**

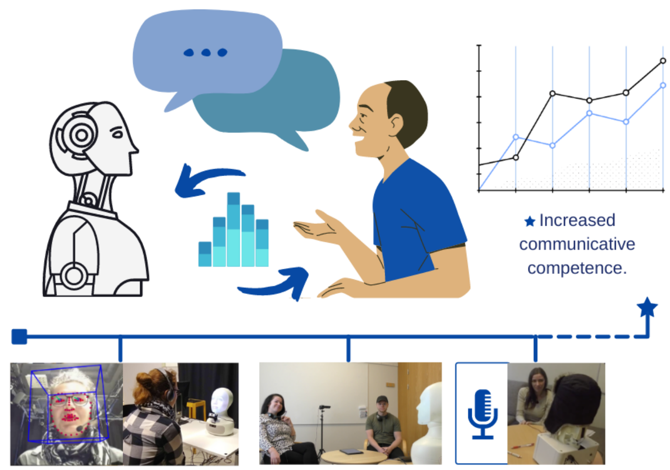

Fig. 1. An adaptive strategy to encourage communication in adult L2 learners and previous studies developed on interactive L2 practice with a social robot.

I. Introduction

There are varied reasons for someone to start learning a new language in adulthood. In the case of older adults, second language (L2) learning is suggested as a beneficial process for healthier cognition, as well as contributing to emotional and social well-fare [1], [2]. For younger adults, the process is most often linked to international migration, where the ability to learn the language of destination shapes the odds of successfully integrating to the new society [3], [4]. Learning a second language in adulthood is a challenging endeavor as progress can be hindered by a drastic reduction in the rate of learning that occurs during adolescence [5], [6]. Hence, providing alternatives that can support adult learners is important for their progress and integration. In particular, achieving *communicative competence*, i.e., being able to adequately communicate in social contexts is the greater goal of L2 acquisition. A common way to attain this goal is to engage in encouraging practice interactions with proficient speakers. However, the opportunities to talk with native speakers may be scarce in some environments. Social robots, then, present an alternative for practice activities in situated interactions. Previous works have shown promising results for infant and child language acquisition, but there are fewer studies with adult learners [7]. Notably, the process of acquiring a language in adulthood differs from the approach taken at early age, as cognition (i.e., performance) and affective states (e.g., motivation, attitudes and expectations) differ. Consequently, the adaption of a robot's discourse (i.e., dialogue strategy) and spoken output characteristics, e.g., intonation, will have to be different to establish rapport with adult learners and support communicative competence. Given this foundation, our primary research question is defined as:

How can a robot companion adapt its discourse to support adult second language (L2) learners in achieving communicative competence?

II. Method and Research Question

Recent approaches on educational robots have shown the importance of adaptation and personalization to improve learning performance [8], [9]. These works, as well as most previous studies, focus on vocabulary and grammar practice to design pedagogical activities for learners [7]. Our work attempts to build interactions above this basic knowledge layer, where utterance formulation, speech fluency and linguistic content, in a context-appropriate framework, are the learning objectives. For this purpose, adaptive dialogue systems have been proposed based on Reinforcement Learning frameworks [10] or Partially Observable Markov Decision Process (POMDP) [11].

Our approach resembles these alternatives where the dialogue policy optimizes the content and formulation of the robot's utterances based on the (context-aware) knowledge model of the learner. This cognitive model is a statistical representation of the learner's utterance composition and (accumulative) use of phrases, shared by the robot's knowledge domain, which are further mapped for social/situational appropriateness through semantic features generated by Language Models (e.g., BERT [12]) and semantic networks (e.g., ConceptNet [13]).

As the quality of the interaction does not depend only on the dialogue content and strategy, it is critical to assess the affective state of the learner. Various works have highlighted the importance of tracking a learner's task and emotional engagement [14]–[16]. This distinction attempts to differentiate disinterested participants from a user that displays reduced participation from knowledge limitations. The engagement level of the student is therefore used to modify the prosody characteristics of the robot (e.g., intonation) as well as gazing behaviour to influence the learner into participating in a more active and engaged manner [17]. Our setting is finally structured on interactions occurring with pairs of learners, as this configuration can generate learner collaboration in a natural manner. Following these considerations, we further segment the research question into a set of sub-questions:

- How can we determine and accurately evaluate a learner's language knowledge and communicative competence?
- What dialogue strategies could a robot companion use to encourage a learner's progression in communicative competence?

III. Past, present and future work

The initial steps we took toward developing robot-led conversation practice was oriented toward understanding events that might decrease the learners' motivation for this activity [18]. We manipulated the robot's spoken output to elicit events of learner uncertainty (or confusion). A thorough analysis of the interactions determined that uncertainty was better characterized in levels of intensity, than in a simple binary classification. Using this knowledge we annotated a corpus of dyad and multi-party practice conversations to build an automatic detector of confusion using multimodal features of the learner. While model performance was lower than we had hoped for, we found that pauses and facial markers contributed with a majority of the information in this classification task. These findings will be useful to select multimodal features to build our future model for engagement detection.

In a second study we switched the focus towards creating a robot behavior that could balance participation in a spoken game between native (L1) and L2 speakers [17]. In this setting, differences in proficiency are expected to generate unequal participation. The robot, then, modified its gazing behavior to stimulate the participant that spoke the least amount of time. Our results indicated that this adaptive gaze behavior could successfully balance speaking participation, but it did not increase the total speaking time. Notably, this finding indicates that social cues, specifically gazing, will enhance a robot's communicative characteristics and, thus, can be used in further steps of our development to shape learners' behaviours.

Furthermore, as the interaction between the robot and learners depends, for the most part, on the speech technology used in the dialogue system, we evaluated available Automatic Speech Recognition systems (ASR) to compare their accuracy with L1 and L2 speakers of Swedish [19]. Our findings corroborated that ASR performance for non-native speakers is worse than for L1 speakers, increasing up to a double of the Word Error Rate (WER) of native speakers in social conversations. Our recent work further checked how these results may alter the performance of the robot's autonomous dialogue system, finding that proposed automatic utterance selection methods performed adequately for different dialogue strategies [20]. We further plan to adapt the language model of the best-performed ASR to improve speech recognition.

In preparation for a long-term study, we are currently carrying out focus interviews with teachers of L2 courses to co-design an interactive activity that will become part of the curriculum of a Swedish course during an entire academic term. Finally, as previously presented in the first research sub-question, we plan to use the collected datasets from previous studies, to model a student's proficiency based on linguistic and prosody features, first using the self-reported proficiency levels and later extending the labels to include intermediate levels. We will test this model in a pilot study using the game setup of our previous second study.

IV. Contribution

This research project is motivated by expanding our understanding of situated dialogues and embodied conversational agents. While our current focus is language acquisition, there are various scenarios where participants talking with a robot companion could benefit from an adaptive communicative strategy that could influence the participants' behavior toward more positive behaviors. Although there are still challenges in fully autonomous conversational robots, the social construct of speech presents a wide range of opportunities for human-robot interaction (HRI) [21]. We see that the perspective of linking cognitive models (i.e., language knowledge) and a social concept (i.e. communicative competence), in context-aware dialogues, is a novel approach in social robot research that may contribute to the larger HRI community.

Finally, we expect that the next step in our future work, that of modelling language proficiency, could present a better and more accurate solution to track the progression of learners in any language course. Ideally, this model could be deployed as a stand-alone system that generates proficiency levels with greater granularity.

Our global objective is that robots that adapt their communicative interaction to learner proficiency levels not only free human teacher time, but provide realistic practice for communicative competence.

Acknowledgment

This work is supported by the Marcus and Amalia Wallenberg foundation under grant MAW 2020.005

REFERENCES

[1] K. Cheng, Y. Deng, M. Li, and H. M. Yan, "The impact of l2 learning on cognitive aging," *ADMET and DMPK*, vol. 3, no. 3, pp. 260–273, 2015.

[2] E. Eguz, "Learning a second language in late adulthood: benefits and challenges," *Educational Gerontology*, vol. 45, no. 12, pp. 701–707, 2019. [Online]. Available: https://doi.org/10.1080/03601277.2019.1690273

[3] A. Adsera and M. Pytlikova, "The role of language in shaping international migration," Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London, CReAM Discussion Paper Series 1206, Feb. 2012. [Online]. Available: https://ideas.repec.org/p/crm/wpaper/1206.html

[4] IOM, "World migration report 2020," 2020.

[5] S. Steber and S. Rossi, "The challenge of learning a new language in adulthood: Evidence from a multi-methodological neuroscientific approach," *PLOS ONE*, vol. 16, no. 2, pp. 1–23, 02 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0246421

[6] J. K. Hartshorne, J. B. Tenenbaum, and S. Pinker, "A critical period for second language acquisition: Evidence from 2/3 million English speakers," *Cognition*, vol. 177, no. April, pp. 263–277, 2018. [Online]. Available: https://doi.org/10.1016/j.cognition.2018.04.007

[7] N. Randall, "A survey of robot-assisted language learning (rall)," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 1, pp. 1–36, 2019.

[8] H. W. Park, I. Grover, S. Spaulding, L. Gomez, and C. Breazeal, "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 687–694, 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3846

[9] T. Schodde, L. Hoffmann, S. Stange, and S. Kopp, "Adapt, explain, engage—a study on how social robots can scaffold second-language learning of children," *ACM Trans. Human-Robot Interact.*, vol. 9, no. 1, pp. 1–27, 2020.

[10] J. Magyar, M. Kobayashi, S. Nishio, P. Sincak, and H. Ishiguro, "Autonomous robotic dialogue system with reinforcement learning for elderlies with dementia," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2019-Octob, pp. 3416–3421, 2019.

[11] F. Yuan, A. Sadovnik, R. Zhang, D. Casenhiser, E. J. Paek, S. O. Yoon, and X. Zhao, "A Simulated Experiment to Explore Robotic Dialogue Strategies for People with Dementia," 2021. [Online]. Available: http://arxiv.org/abs/2104.08940

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[13] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," 2018.

[14] O. Engwall, R. Cumbal, J. D. Àguas Lopes, M. Ljung, and L. Månsson, "Identification of low-engaged learners in robot-led second language conversations with adults," *ACM Transactions on Human-Robot Interaction*, 2022. [Online]. Available: http://dx.doi.org/10.1145/3503799

[15] P. Dao and M. Sato, "Exploring fluctuations in the relationship between learners' positive emotional engagement and their interactional behaviours," *Language Teaching Research*, vol. 25, no. 6, pp. 972–994, 2021.

[16] S. Nakamura, L. Phung, and H. Reinders, "The effect of learner choice on l2 task engagement," *Studies in Second Language Acquisition*, vol. 43, no. 2, pp. 428–441, 2021.

[17] S. Gillet, R. Cumbal, A. Pereira, J. Lopes, O. Engwall, and I. Leite, "Robot gaze can mediate participation imbalance in groups with different skill levels," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 303–311.

[18] R. Cumbal, J. Lopes, and O. Engwall, "Detection of listener uncertainty in robot-led second language conversation practice," in *Proceedings of the 2020 International Conference on Multimodal Interaction*, 2020, pp. 625–629.

[19] R. Cumbal, B. Moell, J. Lopes, and O. Engwall, "You don't understand me!: Comparing asr results for l1 and l2 speakers of swedish," in *Proceedings Interspeech 2021*, 2021, pp. 4463–4467.

[20] O. Engwall, J. Lopes, and R. Cumbal, "Is a wizard-of-oz required for robot-led conversation practice in a second language?" *International Journal of Social Robotics*, 2022. [Online]. Available: https://doi.org/10.1007/s12369-021-00849-8

[21] R. K. Moore, "Talking with robots: Opportunities and challenges," *arXiv preprint arXiv:1912.00369*, 2019.