Designing audiologist bots fusing soundscapes and user feedback

Alessandro Pasta
Technical University of Denmark
Kongens Lyngby, Denmark
alpas@dtu.dk

Michael Kai Petersen
Eriksholm Research Centre
Snekkersten, Denmark
mkpe@eriksholm.com

Kasper Juul Jensen
Oticon A/S
Smørum, Denmark
kjen@oticon.com

Jakob Eg Larsen
Technical University of Denmark
Kongens Lyngby, Denmark
jaeg@dtu.dk

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).
CHI 2020, Workshop on Conversational Agents for Health and Wellbeing, April 26, 2020, Honolulu, HI, USA

Abstract
Despite having different audiological preferences, hearing aid users are usually provided with default settings. This lack of personalization is due to a scarcity of audiological resources and to the difficulty of optimizing hearing aid settings in the clinic. Implementing a conversational agent allows to automatically gather user feedback in real-world environments, while monitoring the soundscape, in order to recommend personalized settings. We outline a conversational agent model that interprets user utterances as audiological intents and fuses user feedback and soundscape features to predict the most likely preferred hearing aid setting. Subsequently, we propose two use cases for a conversational agent, that envisage two different interactions to address distinct user needs: troubleshooting and contextual personalization.

Author Keywords
conversational agents; recommender systems; personalization; hearing healthcare; hearing aids

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI); Natural language interfaces;** *User centered design; Ambient intelligence;*

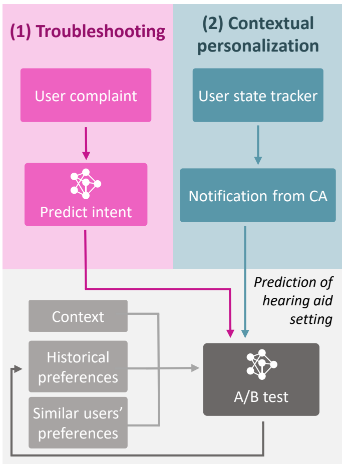

Figure 1: Overview of two different conversational agent use cases with different objectives: Troubleshooting (1) and Contextual personalization (2). If the objective is troubleshooting (1), the interaction is initiated by the user and the complaint is translated into an audiological intent. If the objective is contextual personalization (2), the CA tracks user's state and notifies the user when an alternative setting could enhance the experience. Based on the context, the historical preferences and, potentially, the audiological intent, the CA proposes an A/B setting pair. The user selects the preferred setting and the memory network is updated.

Introduction

Hearing aid users are currently prescribed amplification solely based on a hearing test, which measures hearing thresholds at different frequencies but does not capture individual differences in the ability to understand speech in noise [2] and in the loudness perception of sounds [13]. Despite having different audiological preferences [9], users are usually provided with default settings. Hearing care professionals assume that such default settings suffice [5] and subsequently, during follow-up visits, might modify the gain or noise reduction settings based on user recollections of past listening experiences [10]. The resulting lack of personalization is due to both the scarcity of audiological resources [14] and the difficulty of optimizing hearing aid settings based on user descriptions from recollection in the clinic, rather than dynamically adjusting the settings in the actual listening scenario. Furthermore, the same hearing aid user may have very contrasting preferences depending on the context [8].

Internet-connected hearing aids constitute a paradigm shift in hearing healthcare, as the devices can now potentially be complemented with a smartphone app capable of recommending truly personalized settings [11]. Previous research has shown that gathering user preferences in real-world environments makes it possible to define a device configuration which is preferred over a configuration personalized in a standard clinical workflow [15]. A conversational agent (CA) might autonomously gather user feedback in real-world environments and collect information about the soundscape, to learn the audiological preferences and eventually personalize the device settings. This requires a domain-specific mapping of user utterances to audiological intents [7], as well as the integration of contextual data into the dialogue management.

In this paper, we outline a CA model, which is trained by combining natural language understanding with sequential patterns of soundscape features, to predict the most likely preferred hearing aid setting. Subsequently, as shown in Figure 1, we propose two use cases for a CA:

1. **Troubleshooting**. By eliciting active user participation, the CA understands user complaints and fine-tunes the settings prescribed by the audiologist in the first visit.
2. **Contextual personalization**. By proactively asking the user to compare two alternative settings in specific real-world environments, the CA learns to contextually personalize the listening experience.

Model design

Recent research in CAs has progressed towards reaching human-level understanding within a general conversation framework [1]. Social CA evaluation metrics, reflecting whether responses are logically coherent and sufficiently specific, are equally relevant for the hearing healthcare domain. However, designing an audiologist CA entails domain-specific requirements for customized actions which go beyond dialogue management. Task-oriented dialogue systems often need to keep track of the context across multiple domains and store it as vectors in memory slots [12], while applying attention mechanisms to select the most relevant parameters in order to predict the next action [18].

Due to the requirement for our CA to interpret user utterances referring to audiological intents, we cannot rely on generic pretrained word embeddings. Instead, we build

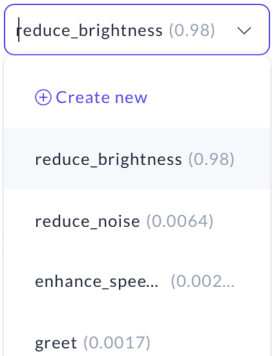

Figure 2: Interactive learning of a CA model. Prediction, based on highest probability, for mapping the word embeddings of the utterance "the sounds are too sharp" to a user intent.

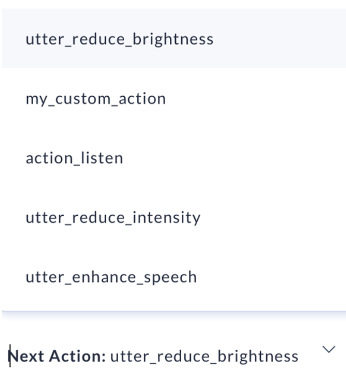

Figure 3: Interactive learning of a CA model. Prediction, based on vector similarity, for selecting the next action in response to the user utterance "the sounds are too sharp" and specific soundscape parameters.

from scratch a natural language understanding (NLU) module trained on audiological terms. In order to learn domain-specific word embeddings, we implement a neural network using frequently occurring user complaints as inputs and potential CA audiological actions as labels [6]. The model is trained by maximizing the similarity between user utterances and associated labels [20]. As shown by Figure 2, this allows to map utterances (e.g. "the sounds are too sharp") into audiological intents (e.g. "reduce high-frequency gain" or, in short, "reduce_brightness"). The most likely audiological intent is predicted based on the highest ranked probability (e.g. 0.98).

Subsequently, in order to predict the most likely next CA action, a dialogue management model is implemented in another neural network, using a transformer architecture similar to BERT [19]. Both the audiological intent embeddings learned by the NLU model, as well as feature vectors capturing the corresponding auditory environment, are forwarded to the second neural network. This dialogue management model compares the cosine similarities of the vectors to predict the most likely next CA action, based on previously learned contextual dialogue patterns. As shown in Figure 3, a user complaint (i.e. "the sounds are too sharp"), merged with the soundscape parameters stored in memory slots, results in a predicted next CA action (i.e. proposing the user an alternative setting defined by lower high-frequency amplification or, in short, "utter_reduce_brightness").

In practical terms, we implement both the NLU and the dialogue management models by training two TensorFlow Neural Networks using the Rasa open source framework [17]. As displayed in Figures 2 and 3, the interactive learning mode allows to accelerate the CA training, by providing feedback and fixing any mistakes.

Discussion

The CA model outlined above might be deployed in different ways, shaping different conversational experiences to address different user needs.

1. Troubleshooting

After the user has been prescribed a default amplification based on the hearing test, she usually tries the hearing aids for some weeks. Given that the prescriptive formula does not guarantee user satisfaction, some fine-tuning based on users' subjective reactions is usually needed [10]. A successful fine-tuning requires the user to be able to communicate her experiences and the professional to be able to interpret and translate them into an adjustment of hearing aid settings [4]. However, this is a time-consuming, yet not optimal procedure, since fine-tuning and additional hearing tests (e.g. speech-in-noise test) performed in the clinic environment do not guarantee a significant advantage over a default prescription [3, 16].

A CA might address users' issues and fine-tune newly acquired hearing aids during the trial phase. Since, in this scenario, the objective is to solve a problem experienced by the user, the interaction is initiated by the user herself. The CA maps the complaint expressed by the user into an audiological intent. In parallel, data describing the context and the current hearing aid settings are inserted into the dialogue. These contextual features, together with the user input, enable the CA to generate a setting adjustment potentially capable of solving the problem experienced by the user. Furthermore, the dialogue allows to gather immediate user feedback on the proposed setting adjustment.

As exemplified in Figure 4, a user might perceive that the sounds are too sharp and initiate a dialogue to report the complaint to the CA. The CA would translate the complaint into an audiological intent and would respond by decreas-

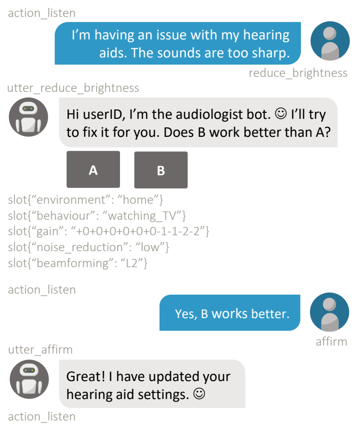

Figure 4: Troubleshooting dialogue initiated by a user complaint. The CA translates the complaint into an audiological intent, responds by decreasing high-frequency amplification and asks the user for her feedback on the new setting.

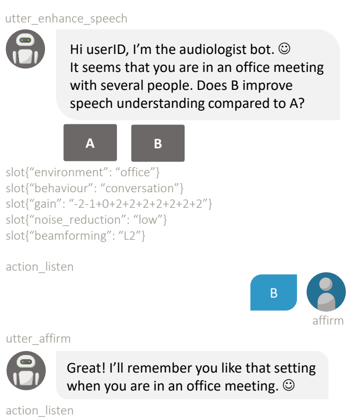

Figure 5: Contextual personalization initiated by the CA, based on tracked user state. The CA monitors the context, asks the user to compare two settings and learns from user feedback.

ing high-frequency amplification. It would ask the user for her feedback on the new setting and remember it, gradually learning the preferred fine-tuning actions in response to users’ complaints. This solution enables solving audiological issues as soon as they arise, by gathering user feedback in real-world environments. It, thereby, potentially provides an effective tool for troubleshooting and reduces the clinical workload.

2. Contextual Personalization
After the hearing aid user has completed the fine-tuning process, there is still potential for improving her listening experience, as user preferences vary based on the context [8]. Although users potentially benefit from a contextually personalized hearing aid setting, collecting user preferences in real-world environments is a time-consuming process.

A CA might autonomously gather user feedback in real-world environments and learn to contextually recommend personalized hearing aid settings based on the historical choices of the user and on the preferences of other users in similar environments. In this scenario, the interaction is initiated by the CA, that tracks user’s state, by monitoring the environment and the current device setting, to suggest an alternative setting. Since the user might benefit from a more effective setting, while not having an explicit complaint in mind, the interaction is simplified.

As shown in Figure 5, the CA might ask the user to compare two alternative settings: the current setting and one that, according to the algorithm, is the most likely to improve user satisfaction in that sound environment. The user would compare the two settings and choose the preferred one. This solution would allow to get an immediate user feedback on alternative settings tried in real-world environments. By iteratively gathering user feedback and learning user preferences in different situations, the CA could, eventually, autonomously personalize hearing aid settings based on the context.

In order for this solution to be successfully implemented, a challenge posed by possible user fatigue needs to be addressed. A hearing aid user might be frequently moving between different contexts and might feel overwhelmed if the CA continuously prompts her to test different settings. In order to avoid an excessive number of notifications, the CA would have to learn when hearing aid personalization is particularly relevant and only prompt the user in those situations. The relevance of personalization is potentially determined by several factors, such as the distance between the current setting and the alternative one, and the perceived usefulness of adjusting an audiological parameter in that situation.

Acknowledgements
We would like to thank Oticon A/S and Eriksholm Research Centre.

REFERENCES
[1] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2020. Towards a Human-like Open-Domain Chatbot. (2020).

[2] Mead C. Killion. 2002. New Thinking on Hearing in Noise: A Generalized Articulation Index. *Seminars in Hearing* 23 (January 2002), 057–076. DOI: http://dx.doi.org/10.1055/s-2002-24976

[3] David R. Cunningham, Kristie J. Williams, and L. Jane Goldsmith. 2001. Effects of Providing and Withholding Postfitting Fine-Tuning Adjustments on Outcome Measures in Novice Hearing Aid Users. *American*

Journal of Audiology 10, 1 (2001), 13–23. DOI: http://dx.doi.org/10.1044/1059-0889(2001/001)

[4] C. Elberling and K.V. Hansen. 1999. *Auditory Models and Non-Linear Hearing Instruments (Rasmussen, Arne Nørby and Osterhammel, Poul Aabo and Ture Andersen and Torben Poulsen)*. Holmens Trykkeri, Copenhagen, Denmark, Chapter Hearing Instruments: Interaction with User Preference, 341–357.

[5] David B. Hawkins and Jodi A Cook. 2003. Hearing aid software predictive gain values: How accurate are they? *The Hearing Journal* 56 (2003), 26,28,32,34.

[6] Lorienne Jenstad, Dianne Tasell, and Chiquita Ewert. 2003. Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology* 14 (10 2003), 347–60.

[7] Benjamin Johansen, Michael K. Petersen, Niels H. Pontoppidan, and Jakob E. Larsen. 2019. Modelling User Utterances as Intents in an Audiological Design Space. In *Workshop on Computational Modeling in Human-Computer Interaction (CHI 2019)*.

[8] Benjamin Johansen, Michael K. Petersen, Maciej J. Korzepa, Jan Larsen, Niels H. Pontoppidan, and Jakob E. Larsen. 2018. Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data. *Computers* 7, 1 (2018). DOI:http://dx.doi.org/10.3390/computers7010001

[9] Gitte Keidser and Harvey Dillon. 2006. What's New in Prescriptive Fitting Down Under?. In *Hearing Care for Adults*. 133–142.

[10] S. Kochkin, D.L. Beck, L.A. Christensen, C. Compton-Conley, P.B. Kricos, B.J. Fligor, J.B. McSpaden, H.G. Mueller, M.J. Nilsson, J.L. Northern, T.A. Powers, R.W. Sweetow, B. Taylor, and R.G. Turner. 2010. MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success. *Hearing Review* 17 (4 2010), 12–34.

[11] Maciej J. Korzepa, Benjamin Johansen, Michael K. Petersen, Jan Larsen, Jakob Eg Larsen, and Niels H. Pontoppidan. 2018. Modeling User Intents as Context in Smartphone-Connected Hearing Aids. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 151–155. DOI: http://dx.doi.org/10.1145/3213586.3226211

[12] Adarsh Kumar, Peter Ku, Anuj Goyal, Angeliki Metallinou, and Dilek Hakkani-Tur. 2019. MA-DST: Multi-Attention-Based Scalable Dialog State Tracking. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*.

[13] Dirk Oetting, Volker Hohmann, Jens Appell, Birger Kollmeier, and Stephan Ewert. 2017. Restoring Perceived Loudness for Listeners With Hearing Loss. *Ear and Hearing* 39 (11 2017), 1. DOI: http://dx.doi.org/10.1097/AUD.0000000000000521

[14] World Health Organization. 2013. *Multi-Country Assessment of National Capacity to Provide Hearing Care*. Geneva, Switzerland.

[15] Alessandro Pasta, Michael K. Petersen, Kasper J. Jensen, and Jakob Eg Larsen. 2019. Rethinking Hearing Aids as Recommender Systems. In *Proceedings of the 4th International Workshop on Health Recommender Systems co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19)*. 11–17. http://ceur-ws.org/Vol-2439/4-paginated.pdf

[16] Lu-Feng Shi, Karen A. Doherty, Tammy M. Kordas, and Joseph Pellegrino. 2007. Short-term and

long-term hearing aid benefit and user satisfaction: a comparison between two fitting protocols. *Journal of the American Academy of Audiology* 18 6 (2007), 482–95.

[17] Rasa Technologies. 2020. Build contextual chatbots and AI assistants with Rasa. (2020). Retrieved April 8, 2020 from `https://rasa.com/docs/rasa/`

[18] Vladimir Vlasov, Akela Drissner-Schmid, and Alan Nichol. 2018. Few-Shot Generalization Across Dialogue Tasks. *CoRR* abs/1811.11707 (2018). `http://arxiv.org/abs/1811.11707`

[19] Vladimir Vlasov, Johannes E. M. Mosig, and Alan Nichol. 2019. Dialogue Transformers. (2019).

[20] Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston. 2017. StarSpace: Embed All The Things! *CoRR* abs/1709.03856 (2017). `http://arxiv.org/abs/1709.03856`