Modeling User Intents as Context in Smartphone-connected Hearing Aids

Maciej Jan Korzepa
Technical University of Denmark
Kongens Lyngby, Denmark
mjko@dtu.dk

Benjamin Johansen
Technical University of Denmark
Kongens Lyngby, Denmark
benjoh@dtu.dk

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

Jan Larsen
Technical University of Denmark
Kongens Lyngby, Denmark
janla@dtu.dk

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

Niels Henrik Pontoppidan*
Eriksholm Research Center
Snekkersten, Denmark
npon@eriksholm.com

ABSTRACT

Despite the technological advancement of modern hearing aids, many users leave their devices unused due to little perceived benefit. This problem arises from the limitations of the current fitting procedure that rarely takes into account 1) the perceptual differences between users not explained by measurable hearing loss characteristics and 2) the variation in context-specific preferences within individuals. However, the recent emergence of smartphone-connected hearings aids opens the door to a new level of context awareness that can facilitate dynamic adaptation of settings to users' changing needs. In this position paper, we discuss how user auditory intents could be modeled as context collected via mobile devices and suggest what kinds of contextual information are relevant when learning situation-specific intents and the corresponding preferences of hearing impaired users. Finally, we illustrate our ideas with several examples of real-life situations experienced by subjects from our study.

KEYWORDS

context awareness; user adaptation; augmented hearing

ACM Reference Format:
Maciej Jan Korzepa, Benjamin Johansen, Michael Kai Petersen, Jan Larsen, Jakob Eg Larsen, and Niels Henrik Pontoppidan. 2018. Modeling User Intents as Context in Smartphone-connected Hearing Aids. In *UMAP'18 Adjunct: 26th Conference on User Modeling, Adaptation and Personalization Adjunct, July 8–11, 2018, Singapore, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3213586.3226211

1 INTRODUCTION

In recent years, hearings aids (HA) have undergone great technological advancements transforming these once bulky, analog devices into powerful, yet discrete wearables. However, despite this substantial change, a considerable fraction of the hearing impaired population fitted with HA does not wear them [20]. One of the most commonly reported reasons for non-use of HA is that they bring users little benefit. However, as at the same time, numerous studies prove the effectiveness of modern HA [8], we rather seek the source of the problem in the limitations of the current fitting procedure.

The current audiological approach bases on prescriptive formulas that determine the frequency-gain curve for a user with specific audiogram i.e. hearing loss characteristics measured as audible hearing thresholds at different frequencies. There are, however, several issues with this approach which lead to suboptimal settings that often do not provide satisfactory level of help to users. First of all, it is well established that hearing loss is not just weakening of neural activity, but also its serious distortion [18] and thus hearing impairment cannot be fully characterized by an audiogram. Killion et al. [15] demonstrated that the ability to understand speech may differ by up to 15-20dB difference in signal-to-noise ratio for subjects with nearly identical audiogram. Similarly, the perceived loudness of soft sounds can vary greatly as shown by LeGoff et al. [17]. Even though modern HA are equipped with advanced signal processing algorithms that go beyond simple amplification, they are rarely taken into account and, without proper control, may even work against wearers. For instance, noise reduction can introduce distortions of spatial cues that might be crucial for some users to distinguish between different auditory streams [18]. All these variations in users' cognitive processing capabilities help to explain why the standard 'one size fits all' approach fails, and indicate that more personalization is needed when fitting HA.

Yet, it has been also established that there are large variations in setting preferences not only between different users, but also within individuals. Keidser et al. [14] demonstrated that the preferred frequency gain characteristics are highly dependent on the auditory environment the user is in. Likewise, Johansen et al. [13] showed that individual users, when given a set of settings varying in terms of omnidirectionality, brightness and noise reduction and freedom to change between them in real, non-clinical environments, exhibit consistent usage patterns of multiple, often very contrasting, settings. These results indicate that user preferences are dependent not only on users' cognition but also on the environments and situations they experience every day. As a result, it is not even 'one size

*Niels Pontoppidan is partly funded by European Union's Horizon 2020 research and innovation program under Grant Agreement 727521 EVOTION.

UMAP'18 Adjunct, July 8–11, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5784-5/18/07…$15.00
https://doi.org/10.1145/3213586.3226211

fits one' approach that should be aimed for, but rather continuous adaptation to individual users' perception in dynamically changing contexts.

Contextual personalization of HA has been researched for many years. Over a decade ago, Dillon et al. [6] presented a concept of HA with trainable frequency-gain curve based on acoustic measurements of environment and ecological momentary assessment, and discussed the potential benefits of such solution. In the following years, the introduction of body-worn gateway devices made it possible to connect the HA with smartphones and prototype intelligent HA systems. Aldaz et al. [4] developed a prototype that used reinforcement learning to discover user preferences based on auditory and geospatial context by prompting users to perform momentary A/B listening tests. However, only with the emergence of the current state-of-the art HA such as Oticon Opn [21], it has become possible to go beyond ecological momentary assessment by continuously tracking both users' interactions with their HA together with auditory context perceived by these devices. For more than a month, Korzepa et al. [16] continuously observed how users switch between different HA settings in different auditory contexts, discussed possibilities and challenges of learning contextual preferences directly from continuous data and suggested the application of conversational AI interfaces and collaborative filtering in this process.

The advancements in the connectivity of hearings aids, accessibility to various sources of contextual data as well as rapidly increasing adoption of smartphones among the elderly open up new possibilities for building context-aware and user-adapted solutions for hearing healthcare. To that end, one of the key elements is the ability to distinguish between different situations hearing-impaired users experience in their daily lives and understand the difficulties they face and the intentions they have in these situations. The purpose of this position paper is to present our views on how to facilitate such context awareness in HA. In Section 2, we explain how user auditory intents can be modeled as context and propose different types of contextual information that are relevant to hearing impaired users. In Section 3, we demonstrate how context can represent the underlying intents based on a few examples of real-life situations experienced by the subjects of our study.

2 USER AUDITORY INTENTS AND CONTEXT

User intent is a common term used in web search domain and refers to the information a user is looking for with a specific goal such as learning/doing something or going somewhere. Web browser and its search engine constitute an interface that attempts to identify what a specific user intent is and provides the user with the most relevant information. Analogically, we define user auditory intent as what a hearing impaired user expects with respect to a specific listening situation. Some examples of auditory intents can be understanding a specific person in a noisy environment, enjoying quiet sounds of nature or zoning out from distracting noises. We will refer to them also simply as user intents in this paper. Likewise, HA constitute an interface that is capable of filtering and processing the content, in this case, different auditory streams.

Even though modern HA have highly advanced signal processing capabilities, they make no attempt whatsoever to identify user intents. In the light of the evidence that users exhibit very contrasting preferences in different situations [13, 16], it is clear that the lack of adaptation leads to wasting the great potential offered by modern HA. Without understanding user intents, these devices will not be able to offer the optimal settings at the right time. However, learning the actual user intents is challenging as it would require getting users' explicit feedback and giving it an actionable form. To address this problem, we assume that we could instead use context, i.e. the state of the user and the situation the user is in, as a representation of user intents. This can be greatly facilitated by the emergence of smartphone-connected hearings aids which bring completely new opportunities for collection and processing of contextual information.

Nonetheless, context awareness is certainly not enough to offer users the optimal settings. HA also need to know user's contextual preferences, or in other words, what settings user prefers or benefits from most in a given situation. User preferences can be inferred by continuously observing user's adjustments of settings and the corresponding context in a non-invasive manner [16] or by asking user to perform ecological momentary assessment, e.g. in the form of A/B tests [4, 23]. Given enough data, multiple streams of contextual information and the corresponding preferences can be also potentially modeled through recurrent neural networks similar to how Rajkomar et al. [22] used multiple layers of healthcare data such as medications or test results as sequential events to predict patient hospitalization outcomes. However, inferring contextual user preferences itself is beyond the scope of this paper as we focus here on discussing the potential of different context sources that might facilitate accurate modelling of user intents. We present them in the following sections.

2.1 Auditory context

Acoustic scene might be the richest source of information that can help to determine user intents. Numerous auditory streams mix together and form so-called soundscape - sound understood as environment that is perceived by humans. Soundscape can consist of e.g. nature sounds, human speech, music or appliance noise. They all might provide useful insights into what users do and what their auditory intent is, and can be represented in a number of ways. Basic information about the soundscape can be extracted from HA as they measure various sound characteristics in different frequency bands and use them as control parameters for signal processing algorithms. [16] represented the auditory context by clustering records based on sound pressure level, noise level, signal-to-noise ratio and modulation characteristics. The authors also used HA' in-built sound classification that labeled soundscapes as quiet, noise, speech in quiet or speech in noise. These characteristics allowed the authors to identify primary patterns related to soundscape.

Another information that can be learned from soundscape is its higher level representation that is connected to a physical location or specific sources of sound. This can be achieved by means of acoustic scene classification (ASC) which has been widely researched for the past two decades. The methods of ASC primarily use probabilistic models (e.g. Hidden Markov Model [7]) or neural networks (e.g. convolutional neural networks [24]) usually with the input in the form of Mel-frequency cepstral coefficients (MFCC)

calculated from short audio frames. ASC has the potential to distinguish between various environments users spend their time at every day such as supermarket, street, bus, restaurant, party, forest or seashore. Such acoustic environment awareness allows to infer much more about user intent. For example, it is very likely that in a restaurant or at a party, the user wants to understand speech despite the surrounding noise while in a forest or at a seashore, the user might rather want to focus on the sounds of nature.

Yet another highly informative component of soundscape is connected with what HA are primarily optimized for - enhancing human voices. The characteristics of human speech such as pitch, timbre or pace vary greatly between speakers dependent on their gender, age, language, possible disorders and many other factors. Additionally, signal processing in HA influences speech differently due to its varied characteristics and, as a result, users' perception of different speakers is challenged [9]. Voice- and speaker-awareness in HA would allow to learn user preferences and personalize settings with respect to voice characteristics. Simple distinction between male and female combined with basic hand-crafted features representing pitch and timbre would be already very informative but one could go even further. Speaker embeddings have been recently widely used for tasks such as speaker recognition or diarization with state-of-the-art results achieved by deep learning methods (e.g. [11, 19]). Speaker embeddings are real vector representations of different speakers that encode distinctive characteristics of their voices. Using such embeddings, HA could not only learn which types of voices need what processing to optimize user's perception without being constrained to a set of arbitrary features that might miss some important characteristics, but also they could help to selectively amplify specific, for example familiar, voices in order to solve the cocktail party problem [10].

2.2 Location

Similar to how acoustic scene analysis gives insight into the user's location, geolocation data can provide information about the acoustic environment and the corresponding user intents. In this way, these two context sources can complement each other by providing information if one is lacking. When both are available, they can be used as labeled data to adapt and optimize ASC model to user-specific environments. Location plays also another important role - it might be mapped to an activity which in turn could be interpreted as specific intents.

The type of location may often be obtained via a public API such as Google Places API [3] or Foursquare Places API [1] based on geographic coordinates. We see particularly big potential in the latter one which, in many countries, has very detailed venue maps and supports adding new, private or public, places. Additionally, venue category structure is very fine and hierarchical (e.g. Arts & Entertainment → Performing Arts Venue → Theater) which can facilitate learning user preferences on different levels of granularity.

Most commonly visited locations can be also learned based on clustering of geospatial coordinates (e.g. by HDBSCAN [12]). This approach might prove helpful especially when there is no access to venue type information. However, as locations expressed as a cluster of coordinates do not carry any semantic information, it is not possible to benefit from learned preferences without knowing to what degree the new locations resemble the familiar ones.

2.3 Time and motion

Some user intents might be related to the way users move. For example, when biking, the user might want to keep maximum omnidirectionality faithfully preserving spatial cues to be aware of the location of other traffic participants and potential dangers, while when in a car, the user might prefer to reduce traffic noise to focus on driving. Motion or activity can be easily predicted using one of many public APIs such as Google's Activity Recognition API [2]. Moreover, motion can be used to track location changes more robustly.

Time is another factor that can support modelling user intents as it often carries information about repeating activities that user is involved in. As shown by Johansen et al.[13], user preferences can greatly change throughout the day and week (especially weekdays vs weekends). Naturally, time context carries lots of uncertainty as what a user does at a specific time may vary greatly, but it might often prove very informative when coupled with other context sources. Time can be also potentially considered as a measure of mental fatigue. Listening in challenging environments requires higher listening effort and increases mental fatigue, especially for hearing-impaired people [5]. Tracking time of day and time spent in challenging listening conditions could conceivably allow to adjust HA settings (e.g. increase noise reduction) according to the estimated level of mental fatigue of a user.

3 DISCUSSION

Relating user intents to a single source of context is naturally prone to errors that might arise not only from limited accuracy of context classification but also from excessive generality. For instance, in a noisy environment with speech, the user might want to understand a specific speaker or zone out from the noisy surroundings. To model user intents, different context sources need to be used in a complementary way. An example from a not yet published study, where we collected data over 2 months capturing user's interactions with the HA and the corresponding context from 10 users, may illustrate both the different components defining the context as well as how they relate to the actual user intents which we learned through interviews with the test persons.

Figure 1 shows how a subject changed his HA program preference and the corresponding context classification for acoustic environment, location, type of location and activity over a period of 12 hours. Between 1pm and 5pm, the subject attended a bridge card game competition. In order to focus on the game, he aimed to attenuate the ambience and chose the program reducing noise ('P4'). Interestingly, the HA does not detect much noise, but mainly speech. In this case, the environment classification alone would not be sufficient to capture user intents and the corresponding program preference. Nor would it explain the preference for a brighter, more intense sound later in the evening, caused by a wish to enhance a TV soundtrack. However, adding the location and time might here generate recognizable patterns.

Another subject reported the benefit of a program offering maximum brightness and amplification of soft sounds during walking

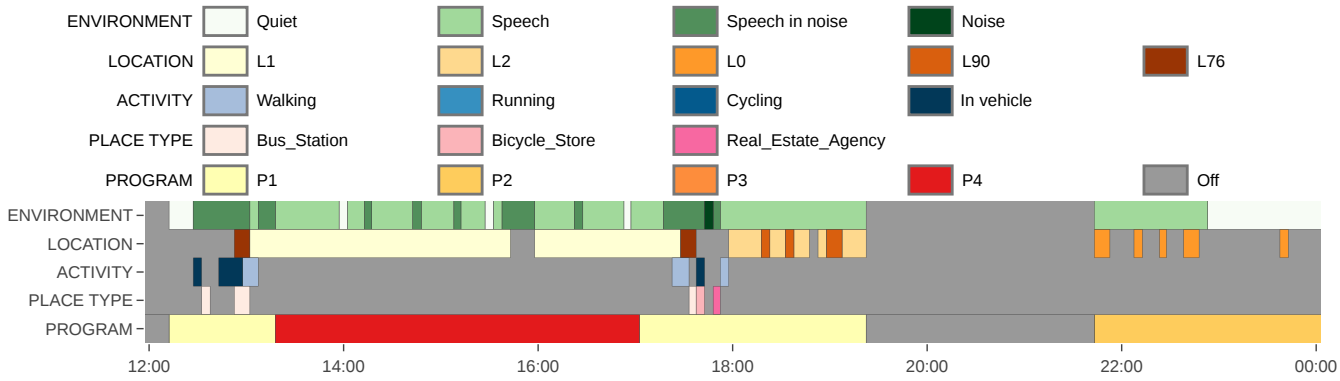

Figure 1: Example of user setting preferences (four programs - P1-P4) juxtaposed with different types of context captured in a continuous manner for a period of 12 hours. Environment is obtained through HA' in-built environment classification, location is represented as cluster membership based on HDBSCAN clustering, activity was estimated by Google's Activity Recognition API and place type was queried using Google Places API.

his dog in the evening when his intention is to enjoy the subtle sounds of nature. In this case, motion combined with acoustic scene and possibly time would be needed to capture and act upon this user's preference. Yet another subject indicated the benefits of using a highly omnidirectional program with some added brightness when his intention is to understand other speakers during lunch in a noisy corporate canteen. In this case, location, acoustic scene and time could be combined to define the user intent. The last example is a subject who generally prefers an omnidirectional, bright setting enhancing the gain in mid and high frequencies, but complains that some female voices get too shrill in that program. Tracking the speakers' voice characteristics might facilitate adjusting the brightness to optimize the user's listening comfort.

The quoted examples serve as yet another proof that hearing impaired users have greatly varying intents and setting preferences that go beyond the need for speech understanding. Understanding intents and personalizing settings with respect to them requires redefining the concept of context awareness in hearing aids. Basic distinction between quiet/noisy and speech/non-speech environments is simply not sufficient to discern between many situations in which user auditory intents differ. However, the new generation of smartphone-connected hearing aids opens the door to infer behavioral patterns from multiple kinds of context that can be obtained through ubiquitous mobile sensors, powerful deep learning techniques and widely available cloud APIs. Combining various soundscape characteristics, location, motion, time and potentially other contextual features not considered in this paper would be a major step towards a whole new level of user-adaption that could unlock the full potential of modern hearing aids.

REFERENCES

[1] 2018. Foursquare Places API. (2018). https://developer.foursquare.com/places-api [Online; accessed 2018-04-15].

[2] 2018. Google Activity Recognition API. (2018). https://developers.google.com/location-context/activity-recognition/ [Online; accessed 2018-04-16].

[3] 2018. Google Places API. (2018). https://developers.google.com/places/ [Online; accessed 2018-04-15].

[4] Gabriel Aldaz, Sunil Puria, and Larry J. Leifer. 2016. Smartphone-Based System for Learning and Inferring Hearing Aid Settings. *Journal of the American Academy of Audiology* 27, 9 (2016), "732–749". https://doi.org/doi:10.3766/jaaa.15099

[5] Fred H. Bess and Benjamin W. Y. Hornsby. 2014. Commentary: Listening Can Be Exhausting-Fatigue in Children and Adults With Hearing Loss. *Ear and Hearing* 35, 6 (2014), 592–599. https://doi.org/10.1097/AUD.0000000000000099

[6] Harvey Dillon, Justin A. Zakis, Hugh McDermott, Gitte Keidser, Wouter Dreschler, and Elizabeth Convery. 2006. The trainable hearing aid: What will it do for clients and clinicians? 59 (04 2006), 30.

[7] AJ Eronen, VT Peltonen, JT Tuomi, AP Klapuri, S Fagerlund, T Sorsa, G Lorho, and J Huopaniemi. 2006. Audio-based context recognition. *Ieee Transactions on Audio Speech and Language Processing* 14, 1 (2006), 321–329. https://doi.org/10.1109/TSA.2005.854103

[8] Melanie A. Ferguson, Padraig T. Kitterick, Lee Yee Chong, Mark Edmondson-Jones, Fiona Barker, and Derek J. Hoare. 2017. Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database of Systematic Reviews* 2017, 9 (2017), CD012023. https://doi.org/10.1002/14651858.CD012023.pub2

[9] Huiwen Goy, M. K. Pichora-Fuller, Pascal Van Lieshout, and Kathryn Arehart. 2013. Quality of older voices processed by hearing aids: Acoustic factors explaining inter-talker differences. *Proceedings of Meetings on Acoustics* 19 (2013), 060133, 060133. https://doi.org/10.1121/1.4800468

[10] Yusuf Isik, Jonathan Le Roux, Zhuo Chen, Shinji Watanabe, and John R. Hershey. 2016. Single-Channel Multi-Speaker Separation using Deep Clustering [arXiv]. *Arxiv* (2016), 4 pp., 4 pp.

[11] Arindam Jati and Panayiotis Georgiou. 2018. Neural Predictive Coding using Convolutional Neural Networks towards Unsupervised Learning of Speaker Characteristics. (2018).

[12] Vincent S. Longbing Cao Guandong Xu Motoda Hiroshi Jian Pei, Tseng (Ed.). 2013. Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7819, 2 (2013), 160–172. https://doi.org/{}10.1007/978-3-642-37456-2_14

[13] Benjamin Johansen, Michael Kai Petersen, Maciej Jan Korzepa, Jan Larsen, Niels Henrik Pontoppidan, and Jakob Eg Larsen. 2018. Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data. *Computers* 7, 1 (2018).

[14] Gitte Keidser, Chris Brew, Scott Brewer, Harvey Dillon, Frances Grant, and Lydia Storey. 2005. The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input. *International Journal of Audiology* 44, 11 (2005), 656–670, 656–670.

[15] Mead C Killion. 2002. New thinking on hearing in noise: a generalized articulation index. (2002), 57–76 pages. https://doi.org/10.1055/s-2002-24976

[16] Maciej Jan Korzepa, Benjamin Johansen, Michael Kai Petersen, Jan Larsen, Jakob Eg Larsen, and Niels Henrik Pontoppidan. 2018. Learning preferences and soundscapes for augmented hearing. In *Companion Proceedings of the 23rd International on Intelligent User Interfaces: 2nd Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces (HUMANIZE).*

[17] Nicolas Le Goff. 2015. *Amplifying soft sounds - a personal matter.* Technical Report February.

[18] Nicholas A. Lesica. 2018. Why Do Hearing Aids Fail to Restore Normal Auditory Perception? *Trends in Neurosciences* 41, 4 (2018), 174–185. https://doi.org/10.1016/j.tins.2018.01.008

[19] Chao Li, Xiaokong Ma, Bing Jiang, Xiangang Li, Xuewei Zhang, Xiao Liu, Ying Cao, Ajay Kannan, and Zhenyao Zhu. 2017. Deep Speaker: an End-to-End Neural Speaker Embedding System. (2017).

[20] Abby McCormack and Heather Fortnum. 2013. Why do people fitted with hearing aids not wear them? *International Journal of Audiology* 52, 5 (2013), 360–368. https://doi.org/10.3109/14992027.2013.769066

[21] Oticon. 2017. Oticon Opn product guide. (2017). https://www.oticon.co.za/-/media/oticon/main/pdf/master/opn/pbr/177406uk_pbr_opn_product_guide_17_1.pdf [Online; accessed 2017-12-17].

[22] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael D. Howell, Claire Cui, Greg S. Corrado, and Jeffrey Dean. 2018. Scalable and accurate deep learning with electronic health records. *Npj Digital Medicine* 1, 1 (2018). https://doi.org/10.1038/s41746-018-0029-1

[23] Oliver Townend, Jens Brehm Nielsen, and Ditte Balslev. 2018. Real-life Applications of Machine Learning in Hearing Aids. (2018). http://www.hearingreview.com/2018/05/real-life-applications-machine-learning-hearing-aids-2/ [Online; accessed 2018-05-27].

[24] Chien-Yao Wang, Andri Santoso, and Jia-Ching Wang. 2017. Acoustic scene classification using self-determination convolutional neural network. *2017 Asia-pacific Signal and Information Processing Association Annual Summit and Conference (apsipa Asc). Proceedings* (2017), 019–22, 019–22. https://doi.org/10.1109/APSIPA.2017.8281995