

# Potential of Augmented Reality Platforms to Improve Individual Hearing Aids and to Support More Ecologically Valid Research

Ravish Mehra, Owen Brimijoin, Philip Robinson, and Thomas Lunner

An augmented reality (AR) platform combines several technologies in a system that can render individual “digital objects” that can be manipulated for a given purpose. In the audio domain, these may, for example, be generated by speaker separation, noise suppression, and signal enhancement. Access to the “digital objects” could be used to augment auditory objects that the user wants to hear better. Such AR platforms in conjunction with traditional hearing aids may contribute to closing the gap for people with hearing loss through multimodal sensor integration, leveraging extensive current artificial intelligence research, and machine-learning frameworks. This could take the form of an attention-driven signal enhancement and noise suppression platform, together with context awareness, which would improve the interpersonal communication experience in complex real-life situations. In that sense, an AR platform could serve as a frontend to current and future hearing solutions. The AR device would enhance the signals to be attended, but the hearing amplification would still be handled by hearing aids. In this article, suggestions are made about why AR platforms may offer ideal affordances to compensate for hearing loss, and how research-focused AR platforms could help toward better understanding of the role of hearing in everyday life.

**Key words:** Augmented reality, Ecologically valid, Hearing aids.

**Abbreviations:** AR, augmented reality; EEG, encephalography; HL, hearing loss; HRTF, head-related transfer function; NH, normal hearing; SNR, signal-to-noise ratio; WHO, World Health Organization

(*Ear & Hearing* 2020;41:140S–146S)

## INTRODUCTION

Cochlear damage and detrimental changes in central auditory system processing have consequences that reach far beyond a poor speech reception threshold in noise. The effects of cochlear damage are multifaceted, including impairments in absolute sensitivity, frequency selectivity, loudness perception and intensity discrimination, temporal resolution, temporal integration, pitch perception and frequency discrimination, as well as sound localization and other aspects of binaural and spatial hearing (Moore 1996). People with hearing loss have a higher susceptibility to noise or competing source interference, often requiring 5 up to 10 dB better signal-to-noise ratios (SNRs) compared to the normal hearing (NH) listener with the same speech-in-noise performance (e.g., Killion & Niquette 2000), even if the loss of sensitivity/audibility is rectified (Humes 2007).

Inability to hear well has behavioral, social, and cognitive consequences that reach far beyond a poor speech reception in

noise. People with hearing loss carry more cognitive load to cope with complex acoustic environments like noisy restaurants. Even when speech is fully understood, the listener must spend more effort than NH listeners (Pichora-Fuller *et al.* 2016; Ohlenforst *et al.* 2017). Given this increased mental load, it is no surprise that overcoming hearing loss (HL) consumes more working memory resources, and consequently reduces memory (Rönnerberg *et al.* 2013). Additionally, the impaired auditory system cannot resolve auditory objects in the same way as NH listeners (Shinn-Cunningham & Best 2008). Choi *et al.* (2014) have shown that for NH listeners, attended auditory objects can obtain 10 dB higher neural gain than unattended sources. This neural gain is compromised with HL (Petersen *et al.* 2017). The high demands on the person with hearing loss’ brain will, in many instances, create social withdrawal (Rutherford *et al.* 2018).

The most common treatment for hearing loss is fitting with hearing aids. In particular, multichannel wide dynamic range compression is the tool of choice to solve the audibility problem in modern hearing aids. This approach enhances the perception of soft sounds while keeping louder sounds within a comfortable range; however, sufficiently increasing the intelligibility of speech in noisy environments remains a challenge. Today’s digital hearing aids have solved acoustic feedback and own voice perception problems and are able to present the dynamic changes of various environments at comfortable loudness levels (Kollmeier & Kiessling 2018).

However, current hearing aid technologies cannot match the user’s needs in complex everyday situations such as conversation with several persons at a cocktail party, in a restaurant or pub, or in a vehicle. Even with additional features like speech processors, directional microphones, frequency transpositions, etc., the most advanced devices provide only modest additional benefits (Humes *et al.* 1999; Larson *et al.* 2000; Magnusson *et al.* 2013; Brons *et al.* 2014; Cox *et al.* 2014; Picou *et al.* 2015). Current ear-centered, multimicrophone hearing aid solutions have limited spacing between microphones, and current state-of-the-art beamforming and machine-learning technologies do not allow for the required source separation (Kollmeier & Kiessling 2018) and sound enhancement (Denk *et al.* 2019). This ear-centric form factor also puts tight constraints on the computation and memory resources available due to limited battery capacity and power budget.

There is thus a genuine need for new technologies that help the people with hearing loss to give additional benefit in a cocktail party, in a restaurant or pub, or in a vehicle. From here on, we will simply refer to such situations as *problematic listening situations*.

Here, we suggest that an augmented reality (AR) platform may give such additional benefits. An AR platform is an interdependent hardware, software, and algorithmic system that consists of a collection of constituent technologies (optics and

---

Facebook Reality Labs Research, Redmond, Washington, USA.

Copyright © 2020 The Authors. *Ear & Hearing* is published on behalf of the American Auditory Society, by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

displays, graphics, audio, eye-tracking, and computer vision). An AR platform can either be a single device or collection of interlinked wearable devices working together. Typical form factor manifestations of an AR device could be glasses with additional accessories like headsets or hearables. Various configurations may support a wide range of sensing, inference, computation, and display capabilities.

We first present a section where AR is introduced and how AR could support compensation for hearing loss, followed by a section where the main AR technologies are outlined. Last, we present a section with perspectives about how an AR platform striving for more ecological validity could be used in hearing research, and some of the challenges and unresolved issues for AR platforms.

Our focus is on technical solutions for mitigating the negative consequences of hearing loss and ignore other potential means such as professional counseling and communication training (e.g., Hickson *et al.* 2007; Oberg *et al.* 2014).

### AUGMENTED REALITY: COMPENSATING FOR HEARING LOSS

AR is a class of technologies, which enables us to create virtual stimuli that can be merged with our real world. This contrasts with the accompanying term *Virtual Reality*, where the virtual stimuli completely replace the real-world stimuli, see Hohmann *et al.* (this supplement, pp. 31S-38S) for a discussion of this concept.

These virtual stimuli can be in the form of *digital objects* placed in our real-world surroundings (e.g., virtual television on the living room wall) or could be a digital representation of a person (also known as a *virtual avatar*) at a distance, communicating with us via a telepresence application (*virtual telepresence*). Done well, these virtual avatars could be so realistic that our brains believe the person is in our real-world space; a much better way to communicate than over the phone or even video calling. Assistive features may enable us to see or hear with higher fidelity by overlaying enhancements to natural signals, or just enhance real auditory objects in the scene. With the help of assistive hints, we may be able to process information faster and remember more information longer. In the case of impaired sensory modalities, this would enable us to improve sensory abilities (*perceptual superpowers*).

#### Solving the Cocktail-Party Problem

Figure 1 describes the principal issues that need to be considered to solve the cocktail-party problem technically: (1) a system that detects the *listener's intent* (which sound sources could be of interest and which are of interest at the moment); (2) a *speaker separation* system that isolates the speakers (“digital objects”) with sufficient signal-to-noise improvement\*; (3) a system that exploits *noise suppression* which could be a pair of headphones or hearables that attenuates external sounds; and (4) a *signal enhancement* system that recombines the “digital

\*A sufficient SNR improvement would, in this context, probably be an improvement in the range of 10 to 12 dB, to raise the listener from low speech intelligibility scores to close to 100% speech intelligibility in problematic listening situations, assuming a psychometric function with a typical slope of about 10%/dB. Further SNR improvements may be counterproductive because these may break the “connected discourse between parties in the conversation (Archer-Boyd *et al.* 2018).”

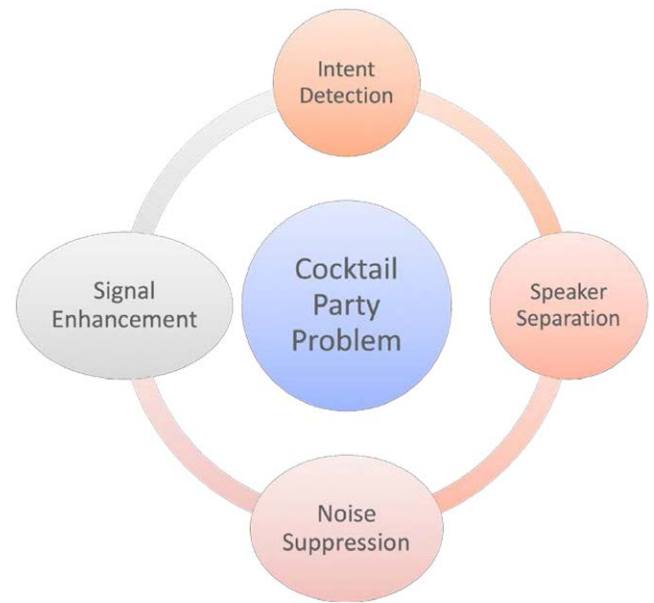


Fig. 1. The cocktail-party problem lies at the intersection of multiple research problems, such as intent detection, speaker separation, noise suppression and signal enhancement.

objects” based on the listener’s intent, with an enhancement of X dB of the currently attended “digital object.” The extraction of digital objects is a key feature of AR and is distinct from what is possible in noise-reduction hearing aids or remote microphones that merely enhance one object at a time, or the object in front of the listener.

#### Multimodal, Ego-Centric Sensing

This is where the idea of an AR platform in support of hearing aids really begins to take shape. New AR glasses could support a larger number of microphones. Additionally, an AR platform could include multimodal sensors, including video, depth, and infrared cameras; inertial measurement units, magnetometers and other motion tracking systems; and many other sensors, which could be used to tackle the hard problems of intent detection, speaker separation, and noise suppression. The section “AR: Hearing-Enhancing Devices” below describes in more detail how these sensors could work together, see especially Figure 2 for an overview.

If successfully implemented, the system could also be used to gather more ecologically valid data in research projects that aim to better understand the role of hearing in real life, that is supporting Purpose A (Understanding) in the current workshop (Keidser *et al.* this supplement, pp. 5S-19S).

An AR platform could serve as a frontend to current and future hearing solutions. The path for the proposed AR platform would follow years of development and evaluation of research platforms. During this time, the AR platform could serve as a technological enabler for improved hearing-related interventions, that is, supporting Purpose B (Development) of the current workshop (Keidser *et al.* this supplement, pp. 5S-19S).

#### Machine-Learning Backbone

Strong artificial intelligence and machine learning frameworks unleash the potential to present completely new solutions

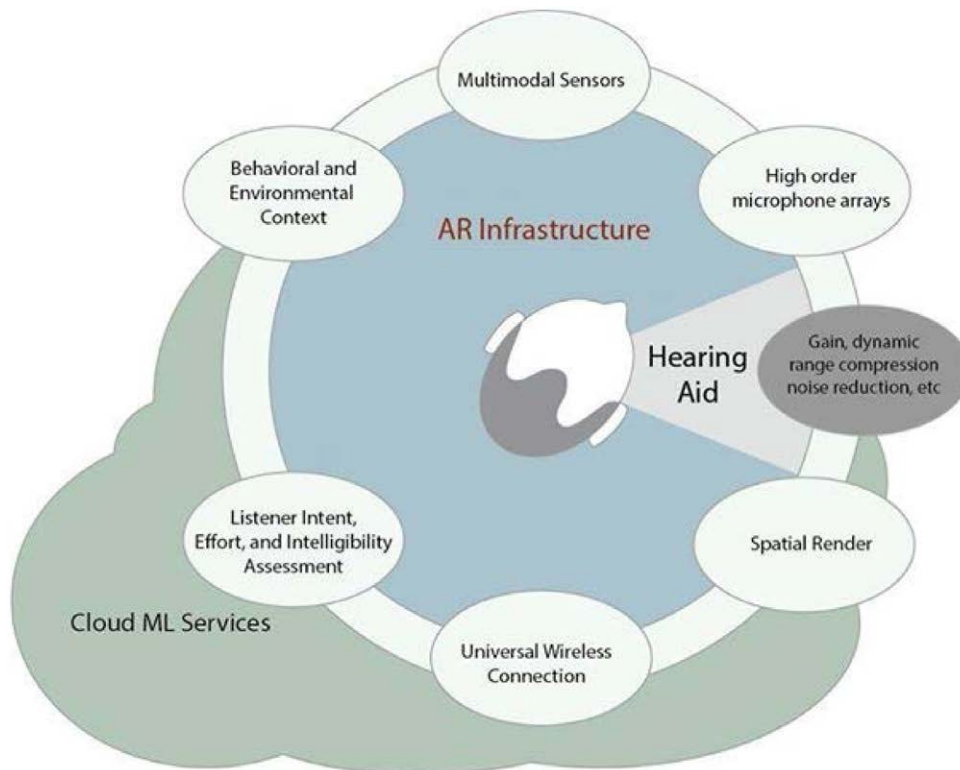


Fig. 2. A proposed AR hearing-enhancing device framework for solving the cocktail-party problem

for *problematic listening situations*. For example, in the Looking to Listen at the Cocktail-Party project, Ephrat *et al.* (2018) presented a deep network-based model that incorporates both visual and auditory signals to solve the problems presented in a cocktail-party situation. The method demonstrated a clear advantage over state-of-the-art audio-only speech separation in cases of mixed speech. Furthermore, other recent developments in deep learning single-channel source separation are promising for HL compensation applications (Chen *et al.* 2016; Chen *et al.* 2017; Wang & Chen 2018), especially if combined with an AR platform.

Another way to support people with hearing loss in problematic listening situations would be to give real-time speech-to-text captioning displayed in the AR glasses display system (e.g., Dufraux *et al.* 2019). In Live Transcribe [a mobile accessibility app designed for the deaf and people with hearing loss (Slaney *et al.* this supplement, pp. 131S-139S)], the researchers demonstrated real-time transcription of speech and sound to text on the screen. Even when acoustic transmission to the listener fails, AR glasses coupled with hearing aids could still allow the listener to participate, if not directly hear.

### Socially Acceptable Form Factor

Self-stigmatization reduces the uptake and use of devices that are perceived as making one look aged, or handicapped. These perceptions potentially influence the 75% of those who could benefit from hearing aids, but do not use them (e.g., Kochkin 2000; Meister *et al.* 2008).

Here, we assume that a pair of AR glasses connected to the cloud (the AR platform) and connected to a pair of hearing aids is used as a communication platform, which would offer

a socially accepted platform that has widespread use. Though glasses and hearing aids both serve to assist the senses, eye-glasses carry much less stigma (Dos Santos *et al.* 2020), and are often a fashion statement. Piggybacking on a fashionable form factor could reduce social stigma and encourage the use of AR glasses with hearing aids.

### Summary of Arguments in Favor of the use of AR Glasses to Support Compensation for Hearing Loss

A combination of multimodal ego-centric sensing, a machine-learning (ML) backbone, and a socially acceptable form factor point toward a future where an AR platform could become the ideal choice to help overcome challenges in compensating for hearing loss; at least, there seems to be high potential for the proposed framework. In the remainder of this paper we will elaborate on the above factors and their integration as a system to offer solutions to the cocktail-party problem.

### AR: HEARING-ENHANCING DEVICES

An AR platform could provide an ideal framework to support hearing aids. The ideal configuration of such a framework, however, is an open question, because what we *do* with its additional capabilities will determine the utility of the resulting framework. Here we detail one potential configuration of the AR platform, see Figure 2, that is a pair of AR glasses connected to the cloud and some input device (e.g., a smartphone), as well as a connection to a pair of hearing aids. This version of the AR platform (AR glasses, cloud, hearing aids, and input device) is in this section called *AR hearing-enhancing device* to not confuse with other AR platforms intended for other



purposes. This section is divided into eight subsections, each discussing a separate set of capabilities.

### Intelligent Initial Fitting and Ongoing Parameter Adjustment

To match the AR hearing-enhancing device well to the listeners' needs, it should be able to import settings from a qualified audiologist, or have self-adjustment properties (e.g., Sabin *et al.* 2020). An AR hearing-enhancing device would also be able to conduct its own assessment of a listener's needs and adjust its settings to approach the optimal values for the current situation. Here, we describe two categories of interactive parameter manipulations, user-driven input and automatic inference.

For the first category there is growing evidence that users can reach a setting that they are seemingly satisfied with and that differs from a prescribed setting (Kuk & Pape 1992; Moore *et al.* 2005; Dreschler *et al.* 2008; Abrams 2017; Boymans & Dreschler 2012; Boothroyd & Mackersie 2017; Mackersie *et al.* 2019; Sabin *et al.* 2020).

The second category of interactive parameter manipulations should be ongoing assessment of settings; the AR device should be capable of automatic inference from the user's hearing performance and make adjustments without requiring any explicit interaction from the user. One example could be where the hearing-enhancing device itself discovers the hearing thresholds. Christensen *et al.* (2018a,b) have shown that using ear-EEG is a feasible method for hearing threshold-level estimation in subjects with sensorineural hearing loss. Another way to assess hearing performance would be to make direct EEG measurements of speech intelligibility from an AR hearing-enhancing device. Several research reports indicate the possibility of attaining reliable correlation between physiological EEG signals and behavioral speech intelligibility (Vanthornhout *et al.* 2018; Das *et al.* 2018).

### High-Order Microphone Arrays

Multichannel enhancement via multimicrophone beamforming (e.g., Aroudi *et al.* 2018; Moore *et al.* 2018, 2019) and deep learning (e.g., Chen *et al.* 2016, 2017; Wang & Chen 2018) have been suggested to capture and enhance the signals. The glasses form factor allows for multichannel speech enhancement, where improvements in the SNR can be on the order of 10 to 20 dB under certain circumstances (see Doclo, 2003). Accuracy is paramount, and it should go without saying that the greater a device's capacity for increasing the SNR, the more catastrophic the consequences of a misidentification of the signal of interest. To solve this problem, which is a restated version of the cocktail-party problem, the device must determine what signal its user is attempting to attend to. Leveraging information from many microphones, both locally and remotely located, to determine the conversational state, what sources are available in the environment, and which ones the user is interacting with most, is just one of a host of other tools at the disposal of a full AR hearing-enhancing device. Multimodal sensing is central, but its integration remains a significant challenge to achieve a highly reliable prediction of listener attention. Sophisticated statistical models must be constructed to accept all these data and output a trustworthy estimate of the currently attended sound or sounds. Such a model must also be able to take into account new noises that suddenly appear, or signals that emerge, such as someone

new calling the user's name, or a waiter approaching the table with a menu.

To enable explicit control, the AR hearing-enhancing device must also allow user-interface-driven source selection, providing a means for the user to actively select desired sources. This could take many forms, from a tap to a gaze-based interface, but is necessary for scenarios where the device, however sophisticated, is unable to establish what the user wishes to hear.

### Context Awareness

A listener will have different needs based on whether they are at home watching TV, driving a car, or sitting in a lively restaurant with many friends and family. A successful AR-enhanced hearing device must be able to adaptively adjust and adapt its settings based on knowledge of its surroundings, and real-time noninvasive evaluation of listener performance. This will require awareness of the device's physical surroundings, such as the user's location (home, supermarket, restaurant, bus, etc.); its own position, orientation, and velocity within the local environment; the position and orientation of other sound sources in space; and the characteristics of the reverberation and noise properties of the space it is occupying. Scene classification in hearing aids is currently based on traditional parameter estimations (Büchler *et al.* 2005) or small feature sets (Townend *et al.* 2018). With large feature sets, deep neural network models outperform traditional parametric estimation methods and achieve the best performance (Li *et al.* 2017).

Knowing the user's location is not very useful without also knowing what they are trying to do in that place. Is the user in a car straining to concentrate on driving in the rain, or casually talking to a fellow passenger? This second class of context awareness is behavioral state. The AR hearing-enhancing device must be able to determine whether the user is engaged in conversation with one or more people, either locally or remote. This must be updated in real-time to cope with changes in conversation partner locations, as well as new partner additions or subtractions. Such sophisticated systems are not implausible: Fridman *et al.* (2018) showed that using 3D convolutional neural networks achieves 86.1% accuracy for predicting task-induced cognitive load in a sample of 92 subjects from video alone.

### Listener Intent

Listener intent, or what a listener wants to hear at any moment, is an elusive signal; we need a technical solution that can learn its markers. Untangling this knot is no trivial task, but an AR platform offers capabilities that may help. Several studies have suggested utilizing eye-tracking (Hart *et al.* 2009; Hládek *et al.* 2016; Kidd 2017; Favre-Felix *et al.* 2017, 2018, Reference Note 1; Hládek *et al.* 2018; Roverud *et al.* 2018) or wearable electroencephalography (EEG) solutions (O'Sullivan *et al.* 2015; Van Eyndhoven *et al.* 2017; Fuglsang *et al.* 2017; Fiedler *et al.* 2017; Han *et al.* 2019) to determine which sound source in a complex scene a listener would like to attend. Not only is this a tricky task, we need to do it quickly enough to follow turn-taking actions and task switches (Monsell 2003) in a conversation. If the AR-enhanced device can make this determination accurately, all manners of digital signal processing, noise reduction, and machine learning-based speech enhancement techniques could be more effectively leveraged for the hearing aid (e.g., Chen *et al.* 2016;

Chen *et al.* 2017; Aroudi *et al.* 2018; Wang & Chen 2018). If the process is too slow, an unsatisfactory new version of the awkward turn taking that happens on laggy video conferences will result. While the means to track listener intention quickly and accurately enough to keep up with a dynamic communication situation is an as-yet unsolved research problem, speech-in-speech performance improvements by enhancing the “digital objects” steered by eye-tracking have been demonstrated (e.g., Favre-Felix, Reference Note 1). Although promising, EEG solutions are still in their infancy due to robustness issues (Alickovic *et al.* 2019). Nonetheless, eye-tracking cameras and/or electrodes may be part of the technical solution to solve the cocktail-party problem.

### High-Output High-Fidelity Spatial Render

A good AR hearing-enhancing device would require acoustic drivers that are efficient and low distortion, even at high sound pressure levels.

Excellent spatial rendering, with full environmental context awareness is also required. Wang *et al.* (2020) showed that beamforming with full-bandwidth spatialization supported speech localization and produced better speech reception thresholds than conditions without spatial rendering or with rendering only in the high-frequency region. Spatial rendering includes the ability to spatialize arbitrary signals to world-fixed and sound source-fixed locations.

The auditory system has been shown to adapt to altered spectral cues of sound location, which presumably provides the basis for recalibration to changes in the shape of the ear over a lifetime (Carlile 2014). Thus, such spatialization would be best performed with individualized head-related transfer functions (HRTFs) (Middlebrooks 1999) and perceptually correct estimation of room acoustics. This information can be preprocessed in the AR hearing-enhancing device and transmitted to the hearing aid.

### Universal “No-Latency” Encrypted Wireless Connectivity

The device should be able to connect to as many audio sources as possible. Device pairing should be intuitive and secure, and the connections established must be bidirectional, transmitting and receiving with no latency. In this case, “no-latency” is ideally less than 1 ms, which would remove the practical constraints that are imposed by the transmission of audio, leaving more time to perform sophisticated digital signal processing and machine learning-based signal enhancement. Giordani and Polese (2020) reviewed the state-of-the-art latencies and found that while 5G is currently above 1 ms, 6G will be significantly below 1 ms. Critically, all connections must be encrypted to ensure security and privacy. Examples of required connections include smartphones, public address and information systems, emergency broadcasts, remote microphones, and other consumer electronics. Special connections to other devices such as power aids and cochlear implants must be enabled for cases when the user’s hearing damage is too extensive to be remediated acoustically. For ideal operation, the system would be paired with a next-generation T-loop system. The most likely candidate to replace it is WiFi due to its ubiquity, but there are connection, interfacing, and transmission latency issues that would have to be solved.

### Extended AR Capability

An ideal AR hearing-enhancing device would be capable of leveraging both multisensory input and *output* to increase

intelligibility. Speech understanding is not a purely acoustic phenomenon, and many other sensory modalities can contribute to or detract from intelligibility. Being able to see lip movements (e.g., MacLeod & Summerfield 1987; Grant 2001) or related head movements (Hadley *et al.* 2019) significantly aids in speech comprehension. AR is inherently multisensory, so the device should make full use of all the systems, such as cameras for scene understanding and motion tracking systems, for multimodal integration to improve intelligibility.

### Challenges to Be Met

Given the above framework for AR hearing-enhancing devices, there are many aspects that need research and maturation of technologies; some technologies are more mature than others. For example, beamforming has already been implemented in teleconferencing systems, while individualized HRTF spatialization and machine learning-based multichannel microphone processing for speech enhancement are both still active research fields. Deep learning for context awareness is similarly only at the research stage. Cloud connectivity with 5G systems is being implemented worldwide, but as discussed above, processor-heavy speech processing algorithms need cloud connections with less than 1 ms latency, which likely means waiting for 6G cloud connectivity.

## AR: ECOLOGICAL VALIDITY IN HEARING RESEARCH

AR hearing-enhancing devices as described above needs a lot of research before being ready for everyday use. AR platforms could be used in research contexts, and with further evaluation, the platforms could provide data of progressively more ecological validity.

For example, in laboratory experiments with eye-trackers and motion trackers in realistic multiperson situations, Hadley *et al.* (2019) found that increased background noise led to increased gaze to the speaker’s mouth. To strive for even more ecologically valid findings, a research AR platform with eye-tracking and motion tracking, as sketched above, could be used to collect comparable everyday life data. Everyday life representing the highest possible ecological validity across sources of stimuli, environment, context of participation, task, and individual variables has been defined by Keidser *et al.* in this supplement, pp. 5S-19S.

Hohmann *et al.* (this supplement, pp. 31S-38S) describe how virtual reality could be used to obtain more ecologically valid findings in the laboratory by introducing more realistic test environments. Creating avatars in a research AR platform, one could strive for even more ecological validity in hearing research, because such studies could be performed in everyday life settings.

Grimm *et al.* (this supplement, pp. 48S-55S) showed that body motion captured by sensors can be used in the laboratory to better understand the role of hearing. A research AR platform could in principle capture the same kind of motion data in everyday life and thus strive for even more ecologically valid outcomes.

Ecological momentary assessment (Holube *et al.* this supplement, pp. 79S-90S; Smeds *et al.* this supplement, pp. 20S-30S) has been proposed as a highly desirable development in hearing research to obtain more ecologically valid findings. However, the ecological momentary assessments may temporarily take the listener out of (e.g., social) context when making

the assessments, and valuable everyday life factors could be lost. Using a research AR platform would make it possible to study hearing behaviors in real life without having to interfere with the listener's natural behavior.

## CONCLUSION

If an AR framework as proposed in this paper becomes a reality in the future, it could impact the 30 million people with hearing loss in the United States, and the 466 million people in the world with disabling hearing loss (6.1% of the world's population, WHO, 2020), affording many advantages to the experience of traditional hearing aids alone. The advantages include significantly improved speech intelligibility in problematic listening environments where the device understands the listener's intent. The combination of AR glasses, cloud computing, and traditional hearing aids to an AR hearing-enhancing device has the potential help people with hearing loss beyond what is possible with current hearing aids.

As a research tool, AR platforms in the form of AR hearing-enhancing devices could help the field of hearing science strive toward greater ecological validity with the goals of better understanding hearing in everyday life and of improved hearing interventions.

To achieve the potential benefits outlined in this article, there are major challenges still to be solved in the development of AR hearing-assistance platforms. That said, progress is being made, and we believe that AR devices will remove the serious constraints posed by the form factor of current hearing aids, while adding leaps in functionality that will constitute a step-change in terms of a listener's ability to follow speech in noisy reverberant backgrounds.

## ACKNOWLEDGMENTS

All authors were employees of Facebook at the time of manuscript preparation.

Address for correspondence: Thomas Lunner, Facebook Reality Labs Research, Redmond, WA 98052, USA. E-mail: thlu@fb.com

Received May 29, 2020; accepted August 20, 2020.

## REFERENCES

- Abrams, H. (2017). Hearing loss and associated comorbidities: What do we know? *Hearing Review*, 2017;24:32–35.
- Alickovic, E., Lunner, T., Gustafsson, F., Ljung, L. (2019). A tutorial on auditory attention identification methods. *Front Neurosci*, 13, 153.
- Archer-Boyd, A. W., Holman, J. A., Brimijoin, W. O. (2018). The minimum monitoring signal-to-noise ratio for off-axis signals and its implications for directional hearing aids. *Hear Res*, 357, 64–72.
- Aroudi, A., Marquardt, D., Doclo, S. (2018). EEG-based auditory attention decoding using steerable binaural super directive beamformer. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 851–855.
- Boothroyd, A., & Mackersie, C. (2017). A “Goldilocks” approach to hearing-aid self-fitting: user interactions. *Am J Audiol*, 26(3S), 430–435.
- Boymans, M., & Dreschler, W. (2012). Audiologist-driven versus patient-driven fine tuning of hearing instruments. *Trends Amplif*, 16, 49–58.
- Brons, I., Houben, R., Dreschler, W.A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends Hear*, 18.2331216514553924
- Büchler, M., Allegro, S., Launer, S., Dillier, N. (2005). Sound classification in hearing aids inspired by auditory scene analysis. *EURASIP J Appl Signal Processing*, 18, 2991–3002.
- Carlile, S. (2014). The plastic ear and perceptual relearning in auditory spatial perception. *Front Neurosci*, 8, 237.
- Carlile, S., & Keidser, G. (2020). Conversational interaction is the brain in action: Implications for the evaluation of hearing and hearing interventions. *Ear Hear*, 41(Suppl 1), 140S–146S.
- Chen, J., Wang, Y., Yoho, S. E., Wang, D., Healy, E. W. (2016). Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises. *J Acoust Soc Am*, 139, 2604–2612.h
- Chen, Y., Luo, N., Mesgarani, N. (2017). Deep attractor network for single-microphone speaker separation. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 246–250.
- Choi, I., Wang, L., Bharadwaj, H., Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hear Res*, 314, 10–19.
- Christensen, C. B., Harte, J. M., Lunner, T., Kidmose, P. (2018a). Ear-EEG-based objective hearing threshold estimation evaluated on normal hearing subjects. *IEEE Trans Biomed Eng*, 65, 1026–1034.
- Christensen, C. B., Hietkamp, R. K., Harte, J. M., Lunner, T., Kidmose, P. (2018b). Toward EEG-assisted hearing aids: Objective threshold estimation based on ear-EEG in subjects with sensorineural hearing loss. *Trends Hear*; 2 22, 2331216518816203.2
- Cox, R. M., Johnson, J. A., Xu, J. (2014). Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. *Gerontology*, 60, 557–568.
- Das, N., Bertrand, A., & Francart, T. (2018). EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *J Neural Eng*, 6, 066017.
- Denk, F., Ewert, S. D., Kollmeier, B. (2019). On the limitations of sound localization with hearing devices. *J Acoust Soc Am*, 146, 1732.
- Doclo, S. (2003). Multi-microphone Noise Reduction and Dereverberation Techniques for Speech Applications. <ftp://ftp.esat.kuleuven.be/pub/SISTA/doclo/phd/phd.pdf>
- Dos Santos, A. D. P., Ferrari, A. L. M., Medola, F. O., Sandnes, F. E. (2020). Aesthetics and the perceived stigma of assistive technology for visual impairment. *Disabil Rehabil Assist Technol*, DOI: 10.1080/17483107.2020.1768308.
- Dreschler, W. A., Keidser, G., Convery, E., Dillon, H. (2008). Client-based adjustments of hearing aid gain: The effect of different control configurations. *Ear Hear*, 29, 214–227.
- Dufraux, A., Vincent, E., Hannun, A., Brun, A., Douze, M. (2019). Lead2Gold: Towards exploiting the full potential of noisy transcriptions for speech recognition. *IEEE Auto Speech Recog Understanding Workshop*. SG, Singapore, 2019, pp. 78–85, doi: 10.1109/ASRU46091.2019.9003981.
- Ephrat, A., Mosseri, I., Lang, O., Dekel, T., Wilson, K., Hassidim, A., Freeman, W. T., Rubinstein, M. (2018). Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation. *ACM Trans Graph*, 37, 112:1–112:11
- Favre-Félix, A., Graversen, C., Dau, T., Lunner, T. (2017). Real-time estimation of eye gaze by in-ear electrodes. *PConf Proc IEEE Eng Med Biol Soc*, 6447324, 4086–4089.1
- Favre-Félix, A., Graversen, C., Hietkamp, R. K., Dau, T., Lunner, T. (2018). Improving speech intelligibility by hearing aid eye-gaze steering: conditions with head fixated in a multitalker environment., *Trends Hear*, 22, 1–11. 1
- Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *J Neural Eng*, 14, 036020.
- Fridman, L., Reimer, B., Mehler, B., Freeman, W. (2018). Cognitive load estimation in the wild. *The 2018 CHI Conference*. Available at: <https://doi.org/10.1145/3173574.3174226>. doi:10.1145/3173574.3174226
- Fuglsang, S. A., Dau, T., Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage*, 156, 435–444.
- Giordani, M., & Polese, M. (2020). Towards 6G networks: Use cases and technologies. *IEEE Commun Mag*, 58, 55–61.
- Grant, K. W. (2001). The effect of speechreading on masked detection thresholds for filtered speech. *J Acoust Soc Am*, 109, 2272–2275.
- Hadley, L. V., Brimijoin, W. O., Whitmer, W. M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Sci Rep*, 9, 10451.



- Han, C., O'Sullivan, J., Luo, Y., Herrero, J., Mehta, A.D., Mesgarani, N. (2019). Speaker-independent auditory attention decoding without access to clean speech sources. *Sci Adv*, 5, eaav6134.
- Hart, J., Onceanu, D., Sohn, C., Wightman, D., Vertegaal, R. (2009). The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, M. Winckler (Eds), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (PART 1)*, Vol 5726, pp. 19–35.
- Hickson, L., Worrall, L., Scarinci, N. (2007). A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. *Ear Hear*, 28, 212–230.
- Hládek, L., Brimijoin, W. O., Porr, B. (2016). Horizontal eye tracking using EOG: A step towards attention driven hearing aids. Basic Auditory Science, University of Cambridge. p. 2007.
- Hládek, L., Porr, B., Brimijoin, W. O. (2018). Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography. *PLoS One*, 13, e0190420.
- Hohmann, V., Paluch, R., Krüger, M., Meis, M., Grimm, G. (2020). The Virtual Lab: Realization and application of virtual sound environments. *Ear Hear*, 41(Suppl 1), 31S–38S.
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol*, 18, 590–603.
- Humes, L. E., Christensen, L., Thomas, T., Bess, F. H., Hedley-Williams, A., Bentler, R. (1999). A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. *J Speech Lang Hear Res*, 42, 65–79.
- Keidser, G., Naylor, G., Brungart, D., Caduff, A., Campos, J., Carlile, S., Carpenter, M., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rappoport, F., Slaney, M., Smeds K (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear Hear*, 41(Suppl 1), 5S–19S.
- Killion, M. C., & Niquette, P. A. (2000). What can the pure-tone audiogram tell us about a patient's SNR loss? *Hear J*, 53, 46–53.
- Kidd, G. Jr. (2017). Enhancing auditory selective attention using a visually guided hearing aid. *J Speech Lang Hear Res*, 60, 3027–3038.
- Kochkin, S. (2000). MarkeTrak V: "Why my hearing aids are in the drawer": The consumer's perspective. *Hear J*, 53, 34, 36, 39–41.
- Kollmeier, B., & Kiessling, J. (2018). Functionality of hearing aids: state-of-the-art and future model-based solutions. *Int J Audiol*, 57 (sup3), S3–S28.
- Kuk, F., & Pape, N. (1992). The reliability of a modified simplex procedure in hearing aid frequency-response selection. *J Speech Lang Hear Res*, 35, 418–429.
- Li, J., Dai, W., Metzger, F., Qu, S., Das, S. (2017). A comparison of deep learning methods for environmental sound detection. doi:10.1109/ICASSP.2017.7952131
- Larson, V. D., Williams, D. W., Henderson, W. G., Luethke, L. E., Beck, L. B., Noffsinger, D., Wilson, R. H., Dobie, R. A., Haskell, G. B., Bratt, G. W., Shanks, J. E., Stelmachowicz, P., Studebaker, G. A., Boysen, A. E., Donahue, A., Canalis, R., Fausti, S. A., Rappaport, B. Z. (2000). Efficacy of 3 commonly used hearing aid circuits: A crossover trial. NIDCD/VA Hearing Aid Clinical Trial Group. *JAMA*, 284, 1806–1813.
- Mackersie, C., Boothroyd, A., Lithgow, A. (2019). A "Goldilocks" approach to hearing aid self-fitting: Ear-canal output and speech intelligibility index. *Ear Hear*, 40, 107–115.
- MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*, 21, 131–141.
- Magnusson, L., Claesson, A., Persson, M., Tengstrand, T. (2013). Speech recognition in noise using bilateral open-fit hearing aids: the limited benefit of directional microphones and noise reduction. *Int J Audiol*, 52, 29–36.
- Meister, H., Walger, M., Brehmer, D., von Wedel, U. C., von Wedel, H. (2008). The relationship between pre-fitting expectations and willingness to use hearing aids. *Int J Audiol*, 47, 153–159.
- Middlebrooks, J. C. (1999). Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J Acoust Soc Am*, 106, 1493–1510.
- Monsell, S. (2003). Task switching. *Trends Cogn Sci*, 7, 134–140.
- Moore, B. C. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear*, 17, 133–161.
- Moore, A., Haan, J., Pedersen, M., Naylor, P., Brookes, M., Jensen, J. (2019). Personalized signal-independent beamforming for binaural hearing aids. *J Acoust Soc Am*, 145, 2971.
- Moore, A., Xue, W., Naylor, P., Brookes, M. (2018). Noise covariance matrix estimation for rotating microphone arrays. *IEEE/ACM Trans Audio Speech Lang Proc*, 27, 519–530.
- Moore, B., Marriage, J., Alcantara, J., Glasberg, B. (2005). Comparison of two adaptive procedures for fitting a multichannel compression hearing aid. *Int J Audiol*, 44, 345–357.
- O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C. (2015). Attentional Selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, 25, 1697–1706.
- Oberg, M., Bohn, T., Larsson, U. (2014). Short- and long-term effects of the modified Swedish version of the Active Communication Education (ACE) program for adults with hearing loss. *J Am Acad Audiol*, 25, 848–858.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear Res*, 351, 68–79.
- Petersen, E. B., Wöstmann, M., Obleser, J., Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *J Neurophysiol*, 117, 18–27.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear*, 37, 5S–27S.
- Picou, E. M., Marcrum, S. C., Ricketts, T. A. (2015). Evaluation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss. *Int J Audiol*, 54, 162–169.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci*, 7, 31.
- Roverud, E., Best, V., Mason, C. R., Streeter, T., Kidd, G. Jr. (2018). Evaluating the performance of a visually guided hearing aid using a dynamic auditory-visual word congruence task. *Ear Hear*, 39, 756–769.
- Rutherford, B. R., Brewster, K., Golub, J. S., Kim, A. H., Roose, S. P. (2018). Sensation and psychiatry: Linking age-related hearing loss to late-life depression and cognitive decline. *Am J Psychiatry*, 175, 3.
- Sabin, A. T., Van Tasell, D. J., Rabinowitz, B., Dhar, S. (2020). Validation of a self-fitting method for over-the-counter hearing aids. *Trends Hear*, 24, 2331216519900589.
- Slaney, M., Lyon, R. F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., Kanevsky, D., Savla, S., Cerf, V. (2020). Ecological auditory measures for the next billion users. *Ear Hear*, 41(Suppl 1), 131S–139S.
- Smeds, K., Gotowiec, S., Wolters, F., Herrlin, P., Larsson, J., Dahlquist, M. (2020). Selecting scenarios for hearing-related laboratory testing. *Ear Hear*, 41(Suppl 1), 20S–30S.
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif*, 12, 283–299.
- Townend, O., Nielsen, J.B., Ramsgaard J. (2018). Real-life applications of machine learning in hearing aids. *Hear Rev*, 25, 34–37.
- Van Eyndhoven, S., Francart, T., Bertrand, A. (2017). EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Trans Biomed Eng*, 64, 1045–1056.
- Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *J Assoc Res Otolaryngol*, 19, 181–191.
- Wang, D., & Chen, J. (2018). Supervised speech separation based on deep learning: An overview. *IEEE/ACM Trans Audio Speech Lang Process*, 26, 1702–1726.
- Wang, L., Best, V., Shinn-Cunningham, B. G. (2020). Benefits of beamforming with local spatial-cue preservation for speech localization and segregation. *Trends Hear*, 24, 2331216519896908.

## REFERENCE NOTE

1. Favre-Félix, A., Harrison, M., Bhuiyan, T., Graversen, C., Dau, T., Lunner, T. Assessing speech intelligibility and behavior when steering audio via head and eye movements in a multi-talker scenario. Submitted to *Trends in Hearing*.