# Harnessing the power of artificial intelligence to transform hearing healthcare and research

Nicholas A. Lesica[1], Nishchay Mehta[2,3], Joseph G. Manjaly[2,3], Li Deng[4], Blake S Wilson[5], Fan-Gang Zeng[6]

[1] Ear Institute, University College London, UK

[2] Royal National ENT Hospital & University College London Hospitals, UK

[3] NIHR UCLH BRC Deafness and Hearing Problems Theme, Ear Institute, University College London, UK

[4] Department of Electrical and Computer Engineering, University of Washington, USA

[5] Departments of Surgery, Biomedical Engineering, and Electrical and Computer Engineering, Duke University, USA

[6] Center for Hearing Research, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, Otolaryngology - Head and Neck Surgery, University of California Irvine, USA

**Abstract**

The advances in artificial intelligence (AI) that are transforming many fields have yet to make an impact in hearing. Hearing healthcare continues to rely on a labor-intensive service model that fails to provide access to the majority of those in need, while hearing research suffers from a lack of computational tools with the capacity to match the complexities of auditory processing. This Perspective is a call for the AI and hearing communities to come together to bring about a technological revolution in hearing. We describe opportunities for rapid clinical impact through the application of existing technologies and propose directions for the development of new technologies to create true artificial auditory systems. There is an urgent need to push hearing forward toward a future in which AI provides critical support for the testing of hypotheses, the development of therapies, and the effective delivery of care worldwide.

**Introduction**

Hearing was once at the forefront of technological innovation. The ***cochlear implant (CI)***, which restores hearing through direct electrical stimulation of the auditory nerve, was a revolutionary advance and remains the most successful neural prosthetic in terms of both performance and penetration[1,2]. Even hearing aids, now considered staid, once led the way in the miniaturization of digital electronics[3]. But innovation has stalled and hearing healthcare is struggling to meet a growing global burden; the vast majority of those with hearing loss do not receive treatment, and those who do often receive only limited benefit.

Recent advances in artificial intelligence (AI) have the potential to transform hearing. Machines have already achieved human-like performance in hearing-related tasks such as automatic speech recognition (ASR) and natural language processing (NLP). AI is also starting to have an impact in medicine; for example, eye screening technologies based on deep neural networks (DNNs) are already in worldwide use. But there are few applications related to hearing per se and AI remains absent from hearing healthcare. In this Perspective, we describe opportunities to use existing technologies to create clinical applications with widespread impact, as well as the potential for new technologies that faithfully model the auditory system to enable fundamental advances in hearing research.

The disconnect between AI and hearing has deep roots. In contrast to modern machine vision, which began with the explicit goal of mimicking the visual cortex[4] and continues to draw inspiration from the visual system[5], work in modern machine hearing has never prioritized biological links. The earliest attempts at ASR were, in fact, modeled on human speech processing, but this approach was largely unsuccessful. The first viable ASR systems arose only after the field made a deliberate turn away from biology (with rationale neatly summarized by IBM's Frederick Jelinek: "Airplanes don't flap their wings"[6]) to focus on modelling the statistical structure of the temporal sequences in speech and language via hidden Markov models.

The recent incorporation of deep neural networks into machine hearing systems has further improved their performance in specific tasks, but it has not brought machine hearing any closer to the auditory system in a mechanistic sense. Biological replication is not necessarily a requirement: many of the important clinical challenges in hearing can be addressed using models with no relation to the auditory system[7] (e.g. DNNs for image classification), or models that mimic only certain aspects of its function[8,9] (e.g. DNNs for sound source separation). But for the full potential of AI in hearing to be realized, new machine hearing systems that match both the function of the auditory system and key elements of its structure are needed.

We envision a future in which the natural links between machine and biological hearing are leveraged to provide effective hearing healthcare across the world and enable progress on hearing's most complex research challenges. To motivate this future, we first provide a brief overview of the auditory system and its disorders and describe the potential of AI to address urgent and important needs in hearing healthcare. We then outline the steps that must be taken to bridge the present disconnect between AI and hearing and suggest directions for future work to unite the two fields in working toward the development of true artificial auditory systems.

## The auditory system and its disorders

The auditory system is a marvel of signal processing. Its combination of microsecond temporal precision, sensitivity over more than five orders of sound magnitude, and flexibility to support tasks ranging from sound localization to music appreciation is still without parallel in other natural or artificial systems. This remarkable performance is achieved through a complex interplay of biomechanical, chemical, and neural components that implement operations such as signal conditioning, filtering, feature extraction, and classification in interconnected stages across the ear and brain to create the experience of auditory perception (Fig. 1a).

The complexity of the auditory system is reflected in its disorders. The system is susceptible to disruption at any of its stages, resulting in a variety of perceptual impairments such as ***deafness*** (a loss of sensitivity to sounds), ***hyperacusis*** (an increase in sensitivity that causes sounds to become uncomfortable or painful) or ***tinnitus*** (the constant perception of a phantom sound, often a ringing or whistling). In order to help identify the underlying causes of a perceptual impairment, hearing assessments are designed to provide clinicians with a wide range of data reflecting the status of the different processing stages, including: mechanical and acoustic measurements of the ear; electrophysiological and imaging measurements of the ear and brain; and psychoacoustic and cognitive measurements of perception (Fig. 1b-d).

Despite this wealth of data, the diagnosis and treatment of hearing disorders are often problematic. The primary difficulties arise from the multifactorial nature of the disorders and our limited understanding of their mechanistic underpinnings. A particular perceptual impairment can be associated with many different pathologies, and a particular pathology can be associated with many different perceptual impairments. AI can help to disentangle the links between pathologies and perceptual impairments to improve diagnosis and treatment, as well as to advance our understanding of the fundamentals of hearing and provide insight into the causes of complex disorders.

In Table 1, we provide an overview of opportunities for AI to address a range of challenges in hearing and specify the scale of the problem underlying each challenge, the nature of the technology needed to solve the problem, and the current state of the art. We address each of these challenges in detail in the sections below.

## Applying existing technologies to meet pressing needs in hearing healthcare

The need for improved hearing healthcare is urgent: hearing disorders are a leading cause of disability, affecting approximately 500 million people worldwide and costing nearly $750 billion annually[10]. The current care model, which is heavily reliant on specialized equipment and labor-intensive clinician services, is failing to cope: approximately 80% of those who need treatment are not receiving it[10]. Fortunately, many of the most pressing problems in hearing healthcare can be framed as classification or regression problems that can be solved by training existing AI technologies on the appropriate clinical datasets. In this section, we give examples of how AI could make an impact in two areas of hearing healthcare: clinical inference and automated service.

*Clinical inference*

The use of information about a patient and their symptoms to identify a condition, predict its course, and determine the optimal treatment is fundamental to all healthcare. Existing technologies such as convolutional neural networks (CNNs) are well suited to such problems and have already achieved excellent performance in many diagnostic tasks. The application of these technologies to hearing could bring immediate improvements in the diagnosis and treatment of some of the most common conditions.

One example is ***middle ear infection*** (otitis media), which is the most frequent reason for children to visit the doctor, take antibiotics, and have surgery[11]. Despite its prevalence, the diagnosis of different middle ear conditions by clinicians remains problematic: accuracy has been estimated at 50% for non-specialists and 75% for specialists[12]. Worse still, the great majority (>80%) of those with middle ear conditions live in low- and middle-income countries (LMICs) with little or no access to care at all. Thus, the application of AI to the diagnosis of middle ear conditions could bring dramatic improvements in both efficacy and accessibility.

Proof of concept has already been established. For example, one recent effort used transfer learning to train publicly-available CNNs (e.g. Inception-V3) on a database of ear drum images (Fig. 1d) to identify six different middle ear conditions with 90% accuracy[13]. Commercial products based on similar technology have recently become available[14]. If such products can be used reliably during routine health checks without the need for specialist resources, their impact will be profound.

Beyond diagnosis, there is also uncertainty regarding the appropriate course of treatment for many hearing conditions that AI could help to resolve. For example, if there is a persistent buildup of fluid in the middle ear, ***grommets*** (tubes) can be inserted into the ear drum to ventilate the middle ear, allowing the fluid to drain out and improving hearing. But performing this procedure in children is resource intensive and carries risk. Since many cases resolve spontaneously, surgery is not usually performed until after several months of “watchful waiting” to identify persistent cases. The development of applications with the ability to consider ear drum images together with other information about patient history, genetics, etc. to predict time to resolution could help to avoid either unnecessary waiting or unnecessary surgery.

Assembling the comprehensive datasets required to make the best use of AI for clinical inference in hearing healthcare will be a challenge. In high-income countries where care is available, patients are often served by specialists across multiple sectors, with each holding vital pieces of information. Efforts are underway to join existing hearing datasets[15] and create new disease or treatment registries for analysis[16]. But technologies developed based on data from high-income countries may not be appropriate for use in LMICs with different populations, so it is critical to ensure that resources are allocated to building datasets that faithfully reflect the global burden of hearing loss[10].

*Automated service*

At present, nearly all hearing healthcare services -- from initial screening and consultation through to follow-up and rehabilitation -- are provided in-person by highly-trained staff using specialized equipment. This “high-touch” model restricts care to places where the required resources are readily available, thus excluding many LMICs, as well as remote locations in high-income countries[17]. COVID-19 has exacerbated the problem: even in places with the required resources, vulnerable patients may be unwilling or unable to seek in-person care and staff may be unable to provide it safely[18]. Fortunately, many of the most common services in hearing healthcare can be readily automated or controlled remotely through telemedicine.

One such service is the measurement of an ***audiogram***, the standard clinical test for hearing loss. An audiogram is obtained by presenting tones at different frequencies and intensities to determine a listener's sensitivity threshold for each frequency. The automation of this process in standard clinical conditions (i.e. with medical-grade earphones in a sound-proof chamber) is straightforward, and recent studies demonstrated that approaches based on active learning and Gaussian process regression can provide more comprehensive measurements in less time than the standard manual approach[19,20].

The challenge in designing automated audiogram measurement applications is that neither the specifics of the equipment nor the environment can be guaranteed in a non-clinical setting[21]. AI can potentially help by framing the problem as audiogram inference rather than audiogram measurement. Given a sufficient training dataset of paired audiograms measured under ideal and non-ideal conditions (perhaps supplemented by data augmentation) along with calibration routines to determine background noise levels, earphone properties, etc., it should be possible to infer the true audiogram from non-ideal measurements.

Another example of a basic service that could be readily automated is the ***fitting or mapping*** of a CI, a procedure in which a clinician establishes the dynamic range of electrical stimulation by adjusting the current emitted while asking the listener to report the magnitude of their sensation. This procedure is performed when the implant is first activated as well as periodically thereafter to compensate for ongoing changes in the device, the stimulation interface, and the brain. Proof-of-concept studies have established that an automated fitting using Bayesian networks can achieve results that are comparable to a standard fitting[22] and that the process can be done by the patient themselves without the need for a clinician[23].

Automated service for CIs could significantly improve both access and outcomes. Most LMICs have few, if any, CI surgeons, so patients are often forced to travel a great distance to receive their implant. But without follow-up services such as device adjustment and speech training, the full potential benefit of the CI will not be realized[24]. Thus, technologies that allow follow-up services to be provided at home or in a local clinic could have a dramatic impact. In the long-term, access to CIs may be increased even further through AI-assisted surgery. While fully-automated implantation is unlikely in the near future, supporting technologies for surgery planning and real-time image enhancement[25,26] could enable surgeons with limited experience.

**Mimicking auditory function to improve the performance of hearing devices**

There are not yet any biological treatments for most forms of hearing loss, so treatments are generally limited to the provision of assistive devices (Fig. 2). For profound deafness, the only available option is to provide direct electrical stimulation of the auditory nerve through a CI. For mild or moderate loss of hearing, a ***hearing aid (HA)*** may be able to help the ear process sound by providing suitable amplification. The signal processing in hearing devices improved rapidly during their early development but in recent years progress has been stagnant[27–29]. This is not due to lack of effort: the number of research papers and patents related to hearing devices continues to grow exponentially[29,30]. The real problem is the complexity of the challenges involved in improving real-world device performance and the inability of traditional engineering approaches to meet them.

Some commercial devices are already using AI in a limited capacity. For example, there are devices that automatically adjust their settings based on the user's current environment (e.g. indoors or outdoors) using either pre-trained DNNs[31] or active learning with Gaussian processes to track each individual user's preferences over time[32]. Work is ongoing to allow future devices to combine the capacity of DNNs with adaptive personalization by collecting continuous data from each user (e.g. through ASR or sensor-based measures of listening effort).

But the most promising use of AI in hearing devices is in replicating or enhancing functions that are normally performed by the auditory system[33]. By using DNNs to transform incoming sounds, AI could dramatically improve the signal processing in hearing devices. This approach is particularly well suited to address the most common problem reported by device users: difficulty understanding speech in a setting with multiple talkers or substantial background noise (the so-called “cocktail party” problem). Recent work has already demonstrated that DNNs can improve the understanding of speech-in-noise for device users. In just a few years, this “deep denoising” has progressed rapidly from separating the voice of a known talker from steady-state noise to separating multiple unknown talkers in reverberant environments[34].

With denoising DNNs, hearing devices can parse complex acoustic environments just as the brain normally would, using source separation and selective attention to turn speech-in-noise into speech-in-quiet. Commercial products including deep denoising are already available[35,36]. While the real-world performance of these products has not been rigorously tested, lab studies using deep denoising have demonstrated that the performance of HA users in recognition tasks can match or even exceed normal levels[37]. Similar approaches being developed for CIs[38,39] and hybrid electro-acoustic devices[40] have also produced promising initial results.

Separating different sound sources is a critical first step toward helping listeners overcome difficulties understanding speech in noise in the real world. But the real challenge is determining which sound source to amplify. In some situations, e.g. a single talker in a background of continuous fan noise, the source is of interest may be obvious. But in others, e.g. a room full of multiple talkers, a source that is of primary interest one minute may become a distraction the next. To address this problem, efforts are underway to bring hearing devices under “cognitive control” by monitoring the brain’s selective attention. When a listener is attending to a particular sound source, the fluctuations in their brain’s neural activity track the fluctuations in the amplitude of the attended source. Thus, the attended source can be inferred from correlations between recorded neural activity and possible sources of interest. Initial studies suggest that recordings that are sufficient to identify the attended source can be obtained from a single electrode within the ear canal, which could easily be integrated with a hearing device[41–43].

Analysis of neural activity may also enable the development of entirely new processing strategies to improve hearing by correcting the distortions in neural activity caused by hearing loss[44]. In the case of CIs, for example, the ideal sound-to-current mapping would be that which evokes the same patterns of neural activity as those that would be present in the auditory system of a listener with normal hearing. Such strategies have always been a goal of hearing device designers, but the inability of traditional engineering approaches to account for the highly nonlinear nature of auditory processing has limited progress. Deep learning provides a powerful new set of tools that can overcome this limitation and identify the sound-to-current mapping which comes closest to the ideal. Given a large dataset of neural activity from normal hearing listeners to serve as a target, a neural network could be trained to find the sound-to-current mapping that minimizes the difference between CI-evoked neural activity and the target.

Using AI to steer the neural activity in the impaired system back toward normal can be viewed as an attempt to enhance homeostatic plasticity. The auditory system has its own mechanisms that appear to act with similar purpose; for example, the brain responds to hearing loss by increasing the gain that it applies to weakened inputs from the ear, presumably to restore overall activity levels to normal[45] (see further discussion of plasticity below). Brain plasticity is not sufficient on its own to correct all of the distortions in neural activity that are caused by hearing loss, but AI-derived sound transformations may be able to correct some of the distortions that the brain cannot. Approaches aimed at correcting distortions in neural activity could be used to develop novel processing strategies for any device -- traditional acoustic HAs, bone-anchored mechanical devices, or transcranial stimulators -- and optimized for any type of sound -- music[46], non-tonal or tonal language[47], or environmental sounds.

Another promising approach is to move beyond hearing devices per se toward a more comprehensive augmented reality (AR) system that can enhance the brain’s own multi-modal capacities[48]. Systems of integrated wearable and associated devices with a variety of multi-modal sensors will eventually become common and have the potential to provide powerful platforms to support deaf people (see Box 1). For example, to enable better speech understanding, AR glasses could implement eye tracking to aid inference of the current sound source of interest along with real-time speech-to-text captioning for instances when auditory perception fails.

Integrating the various technologies for sound or multi-modal processing to provide a seamless user experience will be a challenge[49]. For sound processing during an in-person conversation, the maximum tolerable latency is around 10 ms[50]; any transformation of the sound, e.g. denoising, must be performed on this timescale. This latency requirement presents a dilemma: the capacity for running complex DNNs in an on-ear device, even for inference only, is limited, but offloading to a coprocessor on a paired device introduces an additional delay. One possible solution is a hybrid system in which a sound transformation runs continuously with low latency in an on-ear device while a paired device adjusts the parameters of the sound transformation on slightly slower timescale[35]. Other operations, such as personalization or adjustments based on changes in the listener's environment, can be performed on a much slower timescale, either on a paired device or in the cloud.

**Developing new technologies for machine hearing to empower hearing research**

There is little doubt that the application of current AI technologies to hearing could improve care for many common conditions by making basic services more accessible and enabling devices to restore or enhance auditory function. But there are also many complex disorders for which current technologies may prove insufficient to overcome our lack of understanding. One important example is tinnitus, which affects 15% of people worldwide and is often debilitating[51]. While the phenomenology of tinnitus is simple, developing effective treatments for it is difficult because the underlying mechanisms remain poorly understood[52]. For other conditions such as ***auditory processing disorders*** (e.g. difficulty understanding speech-in-noise despite audiometrically "normal" hearing), providing effective care is even more difficult, as there is little agreement on diagnosis, let alone on treatment[53,54].

The difficulties associated with complex hearing disorders stem from the fact that they are emergent properties of aberrant network states (as opposed to consequences of identifiable molecular or cellular pathologies). Current technologies for regression and classification may be able to improve care for these disorders by identifying reliable biomarkers or other objective measures within complex data to allow for more accurate diagnosis and treatment[55,56]. But a more ambitious approach is for AI and hearing researchers to work together to create new artificial networks for hearing that share key mechanistic features with the auditory system.

If an artificial system is to serve as a surrogate for testing manipulations that cannot be performed on the auditory system itself (either at all, or at the required scale), biological replication will help to ensure that any conclusions drawn from observations made in silico will also hold true in vivo. Artificial auditory systems could provide a powerful framework for the generation and testing of new hypotheses and serve as a platform for developing potential treatments for network-level disorders[57]. In the following sections, we highlight three critical aspects of hearing that artificial auditory systems will need to incorporate: temporal processing, multi-modal processing, and plasticity.

*Temporal processing*

Natural sounds evolve over many different timescales and some, such as speech and music, are defined by the complex patterns that they exhibit across timescales. The brain tracks and groups the amplitude fluctuations across the different frequencies emitted by individual sound sources in order to create distinct perceptual objects. Disruption of this temporal processing is thought to underlie auditory processing disorders[58], as well as the hearing difficulties that are associated with other complex conditions such as dyslexia[59] or schizophrenia[60].

Individual neurons in the auditory system exhibit various forms of selectivity for different time intervals. In some cases, such as the extraction of the microsecond interaural time differences that indicate the location of a sound, there is clear evidence suggesting the presence of a dedicated neural circuit[61]. But the processing of timescales from hundreds of milliseconds to seconds appears to rely on a complex interplay between distributed networks in different brain areas[62]. For example, the judgement of sound intervals of several seconds appears to rely not only on the auditory system but also on the network dynamics in the striatum[63]. Thus, understanding the aspects of hearing that rely on temporal processing requires understanding how sensitivity to intervals and patterns emerges in networks from the intrinsic properties of neurons and the synapses that connect them.

There have recently been several new network architectures developed for multi-timescale processing of speech and language, such WaveNet[64] and the Transformer[65]. These networks achieve impressive performance in many tasks, but bear little resemblance to the auditory system. To be useful as models of hearing per se, artificial networks must not only process temporal information as effectively as the brain, but also do so through comparable mechanisms, such as recurrency. One recent study in which recurrent neural networks were trained to perform a variety of tasks that relied on the analysis of temporal intervals found that they exhibited a number of phenomena that have been observed in the brain[66]. For example, the representations of temporal and non-temporal information occupied orthogonal subspaces of neural activity, as has been observed in prefrontal cortex[67] and the network followed stereotypical dynamical trajectories that were scaled to match the timescale of a task, as has been observed in medial frontal cortex[68]. Further work along these lines is needed to go beyond the analysis of time intervals to tasks involving the processing of complex temporal patterns that are typical of natural sounds.

*Multi-modal processing*

To accurately model the auditory system, artificial networks must ultimately integrate other sensorimotor modalities with the flexibility to perform a wide range of different tasks just as the brain does[69]. The auditory system did not develop in isolation and it does not function in isolation; thus, it cannot be accurately modeled in isolation. The ears are just one of many sources that provide information to the brain and the integration of information from different sources is evident even at early stages of processing[70]. Explicit attempts to model multi-modal properties in isolation are unlikely to be useful (beyond providing a compact description of the phenomena). But if networks with appropriate features are trained on a wide variety of tasks, multi-modal flexibility will emerge just as it has in the brain.

In one recent study, recurrent neural networks trained to perform 20 different cognitive tasks exhibited clustering and compositionality, i.e. they developed distinct groups of units specialized for simple computations that appeared to serve as building blocks for more complex tasks[71]. These properties persisted across changes in some network hyperparameters but not others: the formation of clusters depended strongly on the choice of activation function and occurred only when all tasks were trained in parallel. When tasks were trained sequentially using continual learning techniques (mimicking human learning in adulthood), specialized clusters were replaced by mixed selectivity. These results highlight the need to accurately model both the internal properties of a system and its developmental environment. For the auditory system, it may be appropriate to use parallel training for early stages of processing to model brainstem circuits that evolved to carry out general encoding or elementary computations (or, alternatively, unsupervised learning with generative frameworks, as has proven effective for pre-training ASR and NLP systems[72,73]). For the late stages of processing, sequential training may be more appropriate to model cortical networks with the flexibility to perform a range of multi-modal tasks.

*Plasticity*

The auditory system never stops changing. This plasticity is what allows the brain to learn new tasks and to match the allocation of its limited resources to the task at hand. But it is also the root of several complex hearing problems. For example, tinnitus, often described as a ringing in the ear, is actually a ringing in the brain. A prevailing theory is that following a prolonged loss of input from the ear, the brain responds with increased central gain that amplifies spontaneous neural activity to a level that is perceptible. But this simple idea is difficult to reconcile with experimental data. While increased spontaneous activity with tinnitus has been widely observed at the earliest stages of the auditory system, it does not necessarily propagate to later stages[52]. Furthermore, tinnitus does not actually impair auditory perception[74]. Other network-level theories have been proposed, such as increased central variance[75], disrupted multi-modal plasticity[76], or dysrhythmia of thalamocortical oscillations[55], but definitive evidence is lacking. Accurate network models of the auditory system that include realistic forms of plasticity might be a way to differentiate among the various hypotheses.

Such models could also help to improve prognosis, rehabilitation and training following the restoration of hearing. With CIs, for example, there is currently a large variation in benefit across patients that is difficult to explain[77]. One hypothesis is that the benefit provided by a CI ultimately depends on the degree to which plasticity allows the brain to adapt to the new information that it is receiving from the ear. Many different forms of training to encourage this plasticity have been explored but none has proven widely effective[78]. Artificial networks that accurately model auditory plasticity after hearing restoration would allow for a systematic exploration of different training strategies to determine the conditions under which each is optimal. Given the limited number and heterogeneity of people receiving CIs, it is unlikely that such optimization could ever be achieved through studies of human users. Of course, there is no guarantee that training strategies that are optimal for the artificial system will prove useful for human users. But the likelihood of successful translation will be increased if the key features of the artificial and biological systems are closely matched.

**Toward artificial auditory systems**

Faithful replication of the auditory system will require the design of new networks that are well matched to the structure of the system and the perceptions that it creates. Attempts to model hearing using CNNs have had some success[79,80]. One recent study trained an encoder-decoder network to reproduce complex cochlear mechanics with high accuracy[81]. Such demonstrations that artificial networks can capture the required input-output transformations are a critical first step toward developing artificial auditory systems. But on a mechanistic level, the architecture of CNNs is a poor match for the auditory system[82]. The tiling of space by neurons with similar receptive fields in the visual system that inspired CNNs has no analog in the ear or central auditory system, nor does the translational invariance achieved in CNNs through weight sharing and subsequent pooling. Auditory objects are not translationally invariant with respect to their primary representational dimension, frequency; in fact, a translation in frequency can be a key distinction between, for example, different speech phonemes.

It may be possible to make CNNs more like the auditory system by introducing new features. One example is the introduction of heterogeneous pooling (i.e. pooling across different subsets of convolutional units) to provide some invariance to small changes in frequency (such as those related to voice pitch) while maintaining sensitivity to the large frequency shifts that distinguish phonemes[83]. But, ultimately, new architectures will be required. The inclusion of recurrent features is likely to be critical, since feedback connections are present at all levels of the auditory system and contribute to temporal and multi-modal processing and plasticity[84]. Including such features in networks may also improve their efficiency as well as their fidelity as models of the brain; while many recurrent networks have feedforward equivalents, the recurrent version typically has fewer parameters[5].

An example of the power of new designs is the inclusion of recurrent features in capsule networks for vision [85], which were inspired by the columnar nature of cortical microcircuitry. These features allow the network to capture local invariances (to, for example, skew or rotation) that are not easily captured by traditional CNNs, which improves the robustness of object recognition performance. Capsule networks also accurately reproduce aspects of visual perception that CNNs cannot, such as those related to crowding (the masking of an object by its neighbors)[86]. Networks with similar features may also be useful for hearing; visual crowding is analogous to auditory informational masking[87], and the transformations between "place coding" and "rate coding" in capsule networks are a hallmark of auditory processing[85]. New versions of these networks with the flexibility to share computations across different representations could provide a starting point for developing models with the multi-timescale and multi-modal capabilities of the auditory system[88].

**Outlook**

The current model of hearing healthcare improves the lives of millions of people every year. But it is far from optimal: children with middle ear conditions are triaged to "watchful waiting" while their development is disrupted; people with tinnitus are subject to treatment by trial-and-error, often with little or no benefit; and the deaf are provided with devices that don't allow them to understand speech in noise or enjoy music. And those are the lucky ones: most people with hearing conditions live in LMICs with little or no access to treatment or support of any kind.

Despite the potential for AI to produce dramatic improvements, it has yet to make a significant impact. We have described opportunities for AI to reshape hearing healthcare with the potential for immediate benefit on the diagnosis and treatment of many common conditions. For this potential to be realized, coordinated effort is required with AI developers working to turn current technologies into robust applications and hearing scientists and clinicians ensuring both the availability of appropriate data for training and responsive clinical infrastructure to support rapid adoption.

Transforming hearing healthcare will not be easy. Firstly, there are important ethical considerations regarding appropriate use of technologies, data privacy, and liability that have not yet been resolved[7]. Secondly, the inertia associated with the current service model is strong. The market for devices is highly concentrated, and excessive regulation and restricted distribution have protected incumbents and stifled innovation[89,90]. These problems have recently been recognized and action is being taken to reduce barriers and promote market disruption[91]. But additional efforts will be required to incentivize device manufacturers and service providers to enter underdeveloped markets in LMICs where the need is most urgent.

We have also outlined ways in which AI could be applied beyond healthcare to play a critical part in future hearing research. Artificial networks that provide accurate models of auditory processing, with parallel computations across multiple timescales, integration of inputs from multiple modalities, and plasticity to adapt to internal and external changes have the potential to revolutionize the study of hearing. But to realize this potential, AI and hearing researchers must work together to coordinate experiments on artificial networks and the auditory system with the goal of identifying the aspects of structure and function that are most important.

Ongoing collaboration between AI and hearing researchers would create a win-win situation for both communities and also help to ensure that new technologies are well matched to the needs of users[92,93]. The computational strategies implemented by the ear and brain evolved over many millennia under strong pressure to be highly effective and efficient. Thus, new AI tools modeled after the auditory system have the potential to be transformative not only for hearing, but also for other domains in which efficient and adaptive multi-scale, multi-modality, and multi-task capabilities are critical. This is not the first call for the AI and hearing communities to come together[94], but, given the immense opportunities created by recent developments, we are hopeful that it will be the last.

**Acknowledgements**

We are grateful to Shievanie Sabesan, Dani Sive, and Andreas Fragner for their help with this work, Krystal Cachola and JoAnne Gu for the artwork in Figure 2, and members of the Lancet Commission on Hearing Loss for helpful discussions.

## References


1. Wilson, B. S. & Dorman, M. F. Cochlear implants: A remarkable past and a brilliant future. *Hear. Res.* **242**, 3–21 (2008).
2. Zeng, F.-G., Rebscher, S., Harrison, W. V., Sun, X. & Feng, H. Cochlear Implants: System Design, Integration and Evaluation. *IEEE Rev. Biomed. Eng.* **1**, 115–142 (2008).
3. Levitt, H. A Historical Perspective on Digital Hearing Aids: How Digital Technology Has Changed Modern Hearing Aids. *Trends Amplif.* **11**, 7–24 (2007).
4. Lindsay, G. W. Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *J. Cogn. Neurosci.* 1–15 (2020) doi:10.1162/jocn_a_01544.
5. van Bergen, R. S. & Kriegeskorte, N. Going in circles is the way forward: the role of recurrence in visual inference. *Curr. Opin. Neurobiol.* **65**, 176–193 (2020).
6. Lohr, S. Frederick Jelinek, Who Gave Machines the Key to Human Speech, Dies at 77. *The New York Times* (2010).
7. Wasmann, J.-W. A. *et al.* Computational Audiology: New Approaches to Advance Hearing Health Care in the Digital Age. *Ear Hear.* (2021) doi:10.1097/AUD.0000000000001041.
8. Wei, Y. *et al.* A Review of Algorithm Hardware Design for AI-Based Biomedical Applications. *IEEE Trans. Biomed. Circuits Syst.* **14**, 145–163 (2020).
9. Purwins, H. *et al.* Deep Learning for Audio Signal Processing. *IEEE J. Sel. Top. Signal Process.* **13**, 206–219 (2019).
10. World Health Organization. *World report on hearing*. (2021).
11. Rovers, M. M., Schilder, A. G., Zielhuis, G. A. & Rosenfeld, R. M. Otitis media. *The Lancet* **363**, 465–473 (2004).
12. Pichichero, M. E. & Poole, M. D. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Arch. Pediatr. Adolesc. Med.* **155**, 1137–1142 (2001).
13. Cha, D., Pae, C., Seong, S.-B., Choi, J. Y. & Park, H.-J. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine* **45**, 606–614 (2019).
14. World's first otoscope with artificial intelligence (AI) image classification of ear diseases. https://hearxgroup.com/blog/world-first-otoscope-ai-image-classification-of-ear-diseases.html (2020).
15. NIHR. Hearing health – Health Informatics Collaborative. https://hic.nihr.ac.uk/hearing+health (2020).
16. Sing Registry. The Genetic Sensorineural Hearing Loss Registry. http://singregistry.com (2020).
17. Swanepoel, D. W. *et al.* Telehealth in audiology: the need and potential to reach underserved communities. *Int. J. Audiol.* **49**, 195–202 (2010).
18. Swanepoel, D. W. & Hall, J. W. Making Audiology Work During COVID-19 and Beyond. *Hear. J.* **73**, 20 (2020).
19. Barbour, D. L. *et al.* Online Machine Learning Audiometry. *Ear Hear.* **40**, 918–926 (2019).
20. Schlittenlacher, J., Turner, R. E. & Moore, B. C. J. Audiogram estimation using Bayesian active learning. *J. Acoust. Soc. Am.* **144**, 421–430 (2018).

21. Sandström, J., Swanepoel, D., Laurent, C., Umefjord, G. & Lundberg, T. Accuracy and Reliability of Smartphone Self-Test Audiometry in Community Clinics in Low Income Settings: A Comparative Study. *Ann. Otol. Rhinol. Laryngol.* **129**, 578–584 (2020).

22. Meeuws, M. *et al.* Computer-assisted CI fitting: Is the learning capacity of the intelligent agent FOX beneficial for speech understanding? *Cochlear Implants Int.* **18**, 198–206 (2017).

23. Meeuws, M., Pascoal, D., Janssens de Varebeke, S., De Ceulaer, G. & Govaerts, P. J. Cochlear implant telemedicine: Remote fitting based on psychoacoustic self-tests and artificial intelligence. *Cochlear Implants Int.* **21**, 260–268 (2020).

24. Friedner, M., Nagarajan, R., Murthy, A. & Frankfurter, R. Embracing Multiple Normals — A 12-Year-Old Boy in India with a Cochlear Implant. *N. Engl. J. Med.* **381**, 2381–2384 (2019).

25. Weber, S. *et al.* Instrument flight to the inner ear. *Sci. Robot.* **2**, (2017).

26. Kratchman, L. B. *et al.* Design of a Bone-Attached Parallel Robot for Percutaneous Cochlear Implantation. *IEEE Trans. Biomed. Eng.* **58**, 2904–2910 (2011).

27. Lesica, N. A. Why Do Hearing Aids Fail to Restore Normal Auditory Perception? *Trends Neurosci.* **41**, 174–185 (2018).

28. Wilson, B. S. Getting a decent (but sparse) signal to the brain for users of cochlear implants. *Hear. Res.* **322**, 24–38 (2015).

29. Zeng, F.-G. Challenges in Improving Cochlear Implant Performance and Accessibility. *IEEE Trans. Biomed. Eng.* **64**, 1662–1664 (2017).

30. Zeng, F.-G. Do or Die for Hearing Aid Industry. *Hear. J.* **68**, 6 (2015).

31. Oticon More™ Technology | Polaris™ l For Professionals. https://www.oticon.com/professionals/brainhearing-technology/more-technology.

32. Artificial intelligence in hearing aids | Widex Professionals. https://uk.widex.pro/en-gb/evidence-technology/technological-excellence/artificial-intelligence-in-hearing-aids.

33. Slaney, M. *et al.* Auditory Measures for the Next Billion Users. *Ear Hear.* **41**, 131S (2020).

34. Wang, D. & Chen, J. Supervised Speech Separation Based on Deep Learning: An Overview. *IEEEACM Trans. Audio Speech Lang. Process.* **26**, 1702–1726 (2018).

35. Whisper I The New Hearing Aid That Gets Better Over Time. https://whisper.ai/.

36. Krisp | HD Voice with Echo & Noise Cancellation. https://www.krisp.ai/.

37. Healy, E. W., Johnson, E. M., Delfarah, M. & Wang, D. A talker-independent deep learning algorithm to increase intelligibility for hearing-impaired listeners in reverberant competing talker conditions. *J. Acoust. Soc. Am.* **147**, 4106 (2020).

38. Goehring, T., Keshavarzi, M., Carlyon, R. P. & Moore, B. C. J. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *J. Acoust. Soc. Am.* **146**, 705 (2019).

39. Lai, Y.-H. *et al.* Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients. *Ear Hear.* **39**, 795–809 (2018).

40. Wang, N. Y.-H. *et al.* Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **29**, 184–195 (2021).

41. An, W. W., Pei, A., Noyce, A. L. & Shinn-Cunningham, B. Decoding auditory attention from single-trial EEG for a high-efficiency brain-computer interface. in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* 3456–3459 (2020). doi:10.1109/EMBC44109.2020.9175753.

42. Fiedler, L. *et al.* Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *J. Neural Eng.* **14**, 036020 (2017).

43. O'Sullivan, J. A. *et al.* Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex N. Y. N 1991* **25**, 1697–1706 (2015).

44. Shi, L.-F., Carney, L. H. & Doherty, K. A. Correction of the peripheral spatiotemporal response pattern: a potential new signal-processing strategy. *J. Speech Lang. Hear. Res. JSLHR* **49**, 848–855 (2006).

45. Cai, S., Ma, W.-L. D. & Young, E. D. Encoding intensity in ventral cochlear nucleus following acoustic trauma: implications for loudness recruitment. *J. Assoc. Res. Otolaryngol. JARO* **10**, 5–22 (2009).

46. Nogueira, W., Nagathil, A. & Martin, R. Making Music More Accessible for Cochlear Implant Listeners: Recent Developments. *IEEE Signal Process. Mag.* **36**, 115–127 (2019).

47. Liu, H., Peng, X., Zhao, Y. & Ni, X. The effectiveness of sound-processing strategies on tonal language cochlear implant users: A systematic review. *Pediatr. Investig.* **1**, 32–39 (2017).

48. Mehra, R., Brimijoin, O., Robinson, P. & Lunner, T. Potential of Augmented Reality Platforms to Improve Individual Hearing Aids and to Support More Ecologically Valid Research. *Ear Hear.* **41 Suppl 1**, 140S-146S (2020).

49. Tseng, R.-Y. *et al.* A Study of Joint Effect on Denoising Techniques and Visual Cues to Improve Speech Intelligibility in Cochlear Implant Simulation. *IEEE Trans. Cogn. Dev. Syst.* 1–1 (2020) doi:10.1109/TCDS.2020.3017042.

50. Goehring, T., Chapman, J. L., Bleeck, S. & Monaghan*, J. J. M. Tolerable delay for speech production and perception: effects of hearing ability and experience with hearing aids. *Int. J. Audiol.* **57**, 61–68 (2018).

51. Baguley, D., McFerran, D. & Hall, D. Tinnitus. *The Lancet* **382**, 1600–1607 (2013).

52. Shore, S. E. & Wu, C. Mechanisms of Noise-Induced Tinnitus: Insights from Cellular Studies. *Neuron* **103**, 8–20 (2019).

53. Iliadou, V. & Kiese-Himmel, C. Common Misconceptions Regarding Pediatric Auditory Processing Disorder. *Front. Neurol.* **8**, (2018).

54. Neijenhuis, K. *et al.* An Evidence-Based Perspective on "Misconceptions" Regarding Pediatric Auditory Processing Disorder. *Front. Neurol.* **10**, (2019).

55. Vanneste, S., Song, J.-J. & De Ridder, D. Thalamocortical dysrhythmia detected by machine learning. *Nat. Commun.* **9**, 1103 (2018).

56. Palacios, G., Noreña, A. & Londero, A. Assessing the Heterogeneity of Complaints Related to Tinnitus and Hyperacusis from an Unsupervised Machine Learning Approach: An Exploratory Study. *Audiol. Neurootol.* **25**, 174–189 (2020).

57. Verhulst, S., Altoè, A. & Vasilkov, V. Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hear. Res.* **360**, 55–75 (2018).

58. Kopp-Scheinpflug, C. & Tempel, B. L. Decreased temporal precision of neuronal signaling as a candidate mechanism of auditory processing disorder. *Hear. Res.* **330**, 213–220 (2015).

59. Farmer, M. E. & Klein, R. M. The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychon. Bull. Rev.* **2**, 460–493 (1995).

60. Carroll, C. A., Boggs, J., O'Donnell, B. F., Shekhar, A. & Hetrick, W. P. Temporal Processing Dysfunction in Schizophrenia. *Brain Cogn.* **67**, 150–161 (2008).

61. Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of sound localization in mammals. *Physiol. Rev.* **90**, 983–1012 (2010).

62. Paton, J. J. & Buonomano, D. V. The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions. *Neuron* **98**, 687–705 (2018).

63. Gouvêa, T. S. *et al.* Striatal dynamics explain duration judgments. *eLife* **4**, e11386 (2015).

64. Oord, A. van den *et al.* WaveNet: A Generative Model for Raw Audio. *ArXiv160903499 Cs* (2016).

65. Vaswani, A. *et al.* Attention Is All You Need. *ArXiv170603762 Cs* (2017).

66. Bi, Z. & Zhou, C. Understanding the computation of time using neural network models. *Proc. Natl. Acad. Sci.* **117**, 10530–10540 (2020).

67. Murray, J. D. *et al.* Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci.* **114**, 394–399 (2017).

68. Wang, J., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102–110 (2018).

69. Zhang, C., Yang, Z., He, X. & Deng, L. Multimodal Intelligence: Representation Learning, Information Fusion, and Applications. *IEEE J. Sel. Top. Signal Process.* **14**, 478–493 (2020).

70. Bizley, J. K. & Dai, Y. Non-auditory processing in the central auditory pathway. *Curr. Opin. Physiol.* **18**, 100–105 (2020).

71. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297–306 (2019).

72. Brown, T. B. *et al.* Language Models are Few-Shot Learners. *ArXiv200514165 Cs* (2020).

73. Deng, L., Hinton, G. & Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: an overview. in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 8599–8603 (2013). doi:10.1109/ICASSP.2013.6639344.

74. Zeng, F.-G., Richardson, M. & Turner, K. Tinnitus Does Not Interfere with Auditory and Speech Perception. *J. Neurosci. Off. J. Soc. Neurosci.* **40**, 6007–6017 (2020).

75. Zeng, F.-G. Tinnitus and hyperacusis: Central noise, gain and variance. *Curr. Opin. Physiol.* **18**, 123–129 (2020).

76. Shore, S. E., Roberts, L. E. & Langguth, B. Maladaptive plasticity in tinnitus--triggers, mechanisms and treatment. *Nat. Rev. Neurol.* **12**, 150–160 (2016).

77. Zhao, E. E. *et al.* Association of Patient-Related Factors With Adult Cochlear Implant Speech Recognition Outcomes: A Meta-analysis. *JAMA Otolaryngol Head Neck Surg* **146**, 613–620 (2020).

78. Drouin, J. R. & Theodore, R. M. Leveraging interdisciplinary perspectives to optimize auditory training for cochlear implant users. *Lang. Linguist. Compass* **14**, e12394 (2020).

79. Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V. & McDermott, J. H. A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron* **98**, 630-644.e16 (2018).

80. Keshishian, M. *et al.* Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models. *eLife* **9**, e53445 (2020).

81. Baby, D., Van Den Broucke, A. & Verhulst, S. A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications. *Nat. Mach. Intell.* 1–10 (2021) doi:10.1038/s42256-020-00286-8.

82. Deng, L. & Li, X. Machine Learning Paradigms for Speech Recognition: An Overview. *IEEE Trans. Audio Speech Lang. Process.* **21**, 1060–1089 (2013).

83. Deng, L., Abdel-Hamid, O. & Yu, D. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 6669–6673 (2013). doi:10.1109/ICASSP.2013.6638952.

84. Schofield, B. R. Central Descending Auditory Pathways. in *Auditory and Vestibular Efferents* (eds. Ryugo, D. K. & Fay, R. R.) 261–290 (Springer, 2011). doi:10.1007/978-1-4419-7070-1_9.

85. Sabour, S., Frosst, N. & Hinton, G. E. Dynamic routing between capsules. in *Advances in neural information processing systems* 3856–3866 (2017).

86. Doerig, A., Bornet, A., Choung, O. H. & Herzog, M. H. Crowding reveals fundamental differences in local vs. global processing in humans and machines. *Vision Res.* **167**, 39–45 (2020).

87. Zhang, M., Denison, R. N., Pelli, D. G., Le, T. T. C. & Ihlefeld, A. Informational masking vs. crowding — A mid-level trade-off between auditory and visual processing. *bioRxiv* 2021.04.21.440826 (2021) doi:10.1101/2021.04.21.440826.

88. Hinton, G. How to represent part-whole hierarchies in a neural network. *ArXiv210212627 Cs* (2021).

89. Committee on Accessible and Affordable Hearing Health Care for Adults, Board on Health Sciences Policy, Health and Medicine Division, & National Academies of Sciences, Engineering, and Medicine. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. (National Academies Press, 2016).

90. President's Council of Advisors on Science and Technology, (PCAST). Aging America & Hearing Loss: Imperative of Improved Hearing Technologies. (2016).

91. Warren, E. & Grassley, C. Over-the-Counter Hearing Aids: The Path Forward. *JAMA Intern. Med.* **177**, 609–610 (2017).

92. Davies-Venn, E. & Glista, D. Connected hearing healthcare: the realisation of benefit relies on successful clinical implementation. *ENT & Audiology News* https://www.entandaudiologynews.com/features/audiology-features/post/connected-hearing-healthcare-the-realisation-of-benefit-relies-on-successful-clinical-implementation (2019).

93. Lindsell, C. J., Stead, W. W. & Johnson, K. B. Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem. *JAMA* **323**, 2141–2142 (2020).

94. Lyon, R. F. Machine Hearing: An Emerging Field. *IEEE Signal Process. Mag.* **27**, 131–139 (2010).

95. National Association of the Deaf. Community and Culture – Frequently Asked Questions. *Community and Culture – Frequently Asked Questions* https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/ (2020).

96. Feng, G. *et al.* Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1022–E1031 (2018).

97. Zhou, Z. *et al.* Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat. Electron.* 1–8 (2020) doi:10.1038/s41928-020-0428-6.

98. Bance, M. Hearing and aging. *CMAJ* **176**, 925–927 (2007).

99. Denys, S., Latzel, M., Francart, T. & Wouters, J. A preliminary investigation into hearing aid fitting based on automated real-ear measurements integrated in the fitting software: test–retest reliability, matching accuracy and perceptual outcomes. *Int. J. Audiol.* **58**, 132–140 (2019).

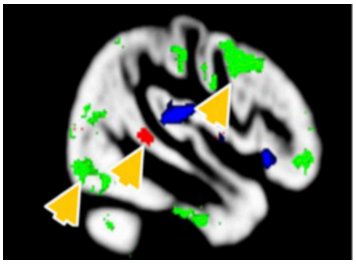
100. Saremi, A. *et al.* A comparative study of seven human cochlear filter models. *J. Acoust. Soc. Am.* **140**, 1618 (2016).

## Box 1 | Artificial intelligence to supporting multiple normals

Hearing healthcare is focused on treating deafness, but this outcome is not always feasible or even desirable. Not all people with hearing loss view it as a problem to be fixed[95]. While AI can certainly transform restorative treatments for deafness, it's impact could be even larger for those who remain deaf. Much of the disability associated with deafness arises from the fact that hearing is currently required for engagement in society. AI has the potential to bring about a new societal model with support for "multiple normals," in which alternative modes of engagement are readily available[24].

### Supporting informed decision making

The benefit that an individual receives from a CI can vary widely. Given that a CI also has downsides -- significant upfront and ongoing costs, risks and complications associated with surgery, continued dependence on associated support and services, etc. -- decisions about whether to undergo implantation can be difficult. Accurate predictions of benefit would be a great help; unfortunately, such predictions are not currently available. Attempts to explain variation in CI outcomes through traditional approaches have been largely unsuccessful[77]. But efforts to apply AI to the problem have produced promising initial results.



A brain image indicating areas (red and green) where pre-implantation morphology was predictive of CI benefit, such as occipital and pre-frontal cortices, and areas (blue) that were impacted by deafness but were not predictive of benefit, such as primary auditory cortex. Image from (Feng et al., 2018) (permission requested).

In one recent study, a support vector machine classifier was used to predict improvements in speech perception in children after implantation[96]. The inputs to the classifier were morphological measures of neural preservation from MRI images in higher-level auditory and cognitive regions. Based on these image data alone, the correlation between the classifier prediction and the actual benefit observed 6 months after implantation approached 0.5. With further development to build predictive models that fuse image data with other measures of auditory structure and function (see Fig. 1) and other patient data, much more accurate predictions may be possible.

### Supporting hearing-optional communication

It is becoming increasingly easy to imagine a world in which deafness is not a disability, as AI is already making many settings more inclusive. In higher education, for example, much of the content is delivered as structured communication from teacher to students through technology platforms on which accessibility features are now readily available; standard software, such as Powerpoint, has the capacity to provide captions in multiple languages in real-time during ongoing presentations. The recent switch to remote learning because of COVID-19, which requires all communication between teachers and students to be routed through technology platforms, provides an opportunity to make accessibility features part of standard leaning models by default.

A translation device with stretchable sensor arrays on each finger attached to a wireless circuit board on the wrist. Image from (Zhou et al., 2020) (permission requested).

Supporting alternative modes of unstructured social communication is more challenging, as many deaf people communicate through signed rather than spoken language. But technologies for real-time automated translation can potentially bridge this gap. One recent study demonstrated the potential for a glove-like device that tracks finger movements to enable translation from American Sign Language to English[97]. This technology required the coordinated development of hardware that is comfortable, durable, and flexible and associated software to classify signals from the device using support vector machines. Though the overall accuracy of the system in this initial study was 98%, the vocabulary was limited to only 11 gestures, so much more work is needed to enable use of the full complement of gestures as well as integration with facial and other movements. Applications based on such technology have the potential to support natural communication not only between deaf and hearing people but also between deaf people from different countries, each of which has its own unique signed language.

## Figure legends

### Figure 1 | The auditory system and its disorders

**(a)** The major processing stages of the auditory system. Sound that enters the ear canal causes vibrations of the ear drum. These vibrations are transmitted by the ossicle bones in the middle ear to the fluid-filled cochlea in the inner ear. Hair cells in the inner ear amplify and transduce motion of the cochlear fluid into electrical signals that are sent to the brain. These signals are processed by several specialized pathways in the brainstem and the resulting information is integrated in the cortex to produce a coherent auditory experience. Some of the key functions performed at each processing stage are indicated in the boxes. Image modified from[98] (permission requested). **(b)** Examples of objective measures used in hearing assessment. Each panel describes one measure and provides a schematic illustration of the associated results from a patient with (dark blue) and without (light blue) a hearing condition. Key differences are indicated by the arrows. **(c)** Examples of subjective measures used in hearing assessment. **(d)** Example of imaging used in hearing assessment.

### Figure 2 | Artificial intelligence for the hearing devices of the future

**(a)** The key elements of future hearing devices. Current hearing devices use a microphone to pick up sound, which is amplified and filtered before being digitized for signal processing; the processing parameters are fixed after fitting by an audiologist; the processed digital signals are converted to either an analog signal delivered to a speaker in hearing aids (HAs) or an electrical signal delivered to electrodes in cochlear implants (CIs) (bottom-left insert). **(b)** Examples of how AI could transform the experience of a deaf person throughout their entire life. The boxes indicate the current state-of-the-art (Now) and the potential for improvement (With AI) in screening and diagnosis (left), devices and implantation (middle), and fitting and therapy (right).

**Table 1 | Top challenges for artificial intelligence in hearing**

Each row provides summary information about a particular challenge, including its scale (the number of people in need, in millions), whether or not the technology must replicate aspects of the auditory system, the current Technology Readiness Level of potential AI-based solutions, and the key next steps to be taken.

(* Assuming only high-end devices; # Commercial AI-based technologies are available but efficacy is unknown)

| | Challenge | Summary | Scale (M) | Auditory | TRL | Next steps | See |
|---|---|---|---|---|---|---|---|
| **1** | Treatment of middle ear conditions | Identification of condition from ear drum images; prediction of disease/treatment course from disparate data | >100 | No | 3-4[13,#] | Build products; collect big data | Fig 1 |
| **2** | Automated audiogram measurement | Inference of hearing thresholds from subjective measures | > 100 | No | 3[19,20] (AI-led) 9[21] (self-led) | Build products; increase robustness | Fig 1 |
| **3** | Automated fitting of hearing devices | Inference of optimal device settings from objective and subjective measures | 10-100 | No | 3-4[23,99,#] | Build products; gain users | Fig 2 |
| **4** | Speech denoising for hearing devices | Amplification of sound of interest and suppression of background noise | 10-100 | Function only | 3-4[37,38,#] | Build products; increase flexibility | Fig 2 |
| **5** | Cognitive control for hearing devices | Inference of sound of interest from measurements of brain activity | 1-10* | Function only | 3-4[41] | Build practical systems | Fig 2 |
| **6** | Multi-modal integration (AR) for devices | Fusion of information from different modalities to enhance perception | 1-10* | Function only | 2-3[48] | Build practical systems | Fig 2 |
| **7** | Treatment of profound deafness | Identification of condition from disparate data; prediction of disease/treatment course from disparate data | 1-10 | No | 2-3[96] | More research; collect big data | Box 1 |
| **8** | Automated translation of signed language | Inference of intended meaning from motion and image data | 1-10 | Function only | 2-3[97,#] | Build practical systems | Box 1 |
| **9** | Treatment of tinnitus | Identification of reliable biomarkers or other objective measures / understanding of fundamental problem | 10-100 | No; Structure & function | 1-2[55,56] | More research; collect big data | |
| **10** | Artificial auditory systems | Development of models that match both the structure and function of the auditory system | ? | Structure & function | 8-9[100] (ear) 1-2 (brain) | More research | |