

Listen to this

Hearing and vision are powerful and important senses for interacting with our surroundings. So far, advances in the area of machine vision have been the most prominent, but machine hearing research that closely mimics the complex sound processing in the human ear has exciting opportunities to offer.

Machine hearing technology is familiar to many of us who use Google Assistant, Apple's Siri, Amazon's Alexa or other voice assistants. Speech recognition has been one of many application areas that have been boosted by machine learning advances this past decade. However, current machine hearing doesn't come close in performance to human hearing, which can easily pick up different types of sounds, voices and music from a noisy background and from different directions.

So far, machine hearing has generally not been as biologically inspired by auditory physiology as computer vision has been inspired by the hierarchical flow of information in the brain's visual cortex. To date, state-of-the-art machine hearing is far from capturing the intricate sound processing of the inner ear. Sound waves are produced when the air is mechanically disturbed. These sound waves are measured by two sensors — the ears — which provide signals to the brain. The function of the auditory system is to sense, process and understand sound. For instance, hearing can perform complex tasks, such as recognizing what is said, who said it, and the emotion of the speaker. Hearing is critical for knowing where things are, including events we cannot see, such as noises in the dark. The brain performs complex operations on sounds, such as segregating sounds from multiple people talking at a party, and directing attention to a sound source of interest, as when someone says your name, a phenomenon known as the cocktail party effect. Finally, hearing is the gateway of one of the great art forms, music.

Artificial intelligence approaches that would more closely capture biological audition could lead to big gains for machine hearing — for example, enabling voice assistants or robots that operate in varying and challenging acoustic environments. Moreover, advances in machine hearing could improve medical technology to help the many people who suffer from hearing disorders. In fact, the development of hearing aids has already been a technological success story: small hearing aids that could be worn inside or behind the ear were developed in the 1950s and 1960s, and were one of the first widespread applications of the transistor, which was invented in 1947. Another breakthrough invention, developed mainly in the 1970s, are cochlear implants, which directly stimulate the auditory (cochlear) nerve. However, these devices still rely on the cochlea, which provides the complex processing necessary for hearing. Unfortunately, for many complex hearing disorders, such as tinnitus, no good solution is currently available.

Extensive research has gone into developing auditory models that capture the complicated nonlinear processing of the ear. While the inner ear can be computationally modelled, doing so requires intensive calculations that are too slow for real-time applications or too simple to approximate the richness of hearing. A promising interdisciplinary approach is to combine computational auditory neuroscience with machine-learning-based audio processing. In an [Article](#) in this issue, Baby, Van Den Broucke and Verhulst leverage the ability of artificial neural networks to approximate complex functions. They trained a convolutional neural network model, called

CoNNear, to predict responses of the basilar membrane inside the cochlea to speech sounds. The target membrane responses were simulated with a computational model, developed by the authors, that represents the cochlea as a transmission line. The proposed model replicates a variety of results from the cochlear mechanics literature when tested on synthetic tone and click stimuli, which differ from the training set stimuli. These results could enable substantial advances in computational auditory research, such as developing more human-like automatic speech recognition systems and improved hearing aids.

The CoNNear model of Baby et al. provides an important link between the physiology of the inner ear and machine hearing. Just as machines have learnt to 'see' with human-like computer vision in several applications, they can learn to 'hear' like humans. In an article on machine hearing, written about a decade ago, Richard F. Lyon noted, figuratively, that our computers are mostly deaf, having little idea what the sounds they store and process actually represent¹. He continued: "leveraging our knowledge of the amazing capabilities of the mammalian cochlea and auditory brain is a goal that will keep this field busy for a while and that will provide rewards on many fronts." Interdisciplinary research from neuroscience, psychology, machine learning and engineering is poised to reap such rewards. □

Published online: 17 February 2021
<https://doi.org/10.1038/s42256-021-00313-2>

References

1. Lyon, R. F. *IEEE Signal Process. Mag.* 27, 131–139 (2010).