

# Perfecting pitch perception

December 20 2021, by Jennifer Michalowski

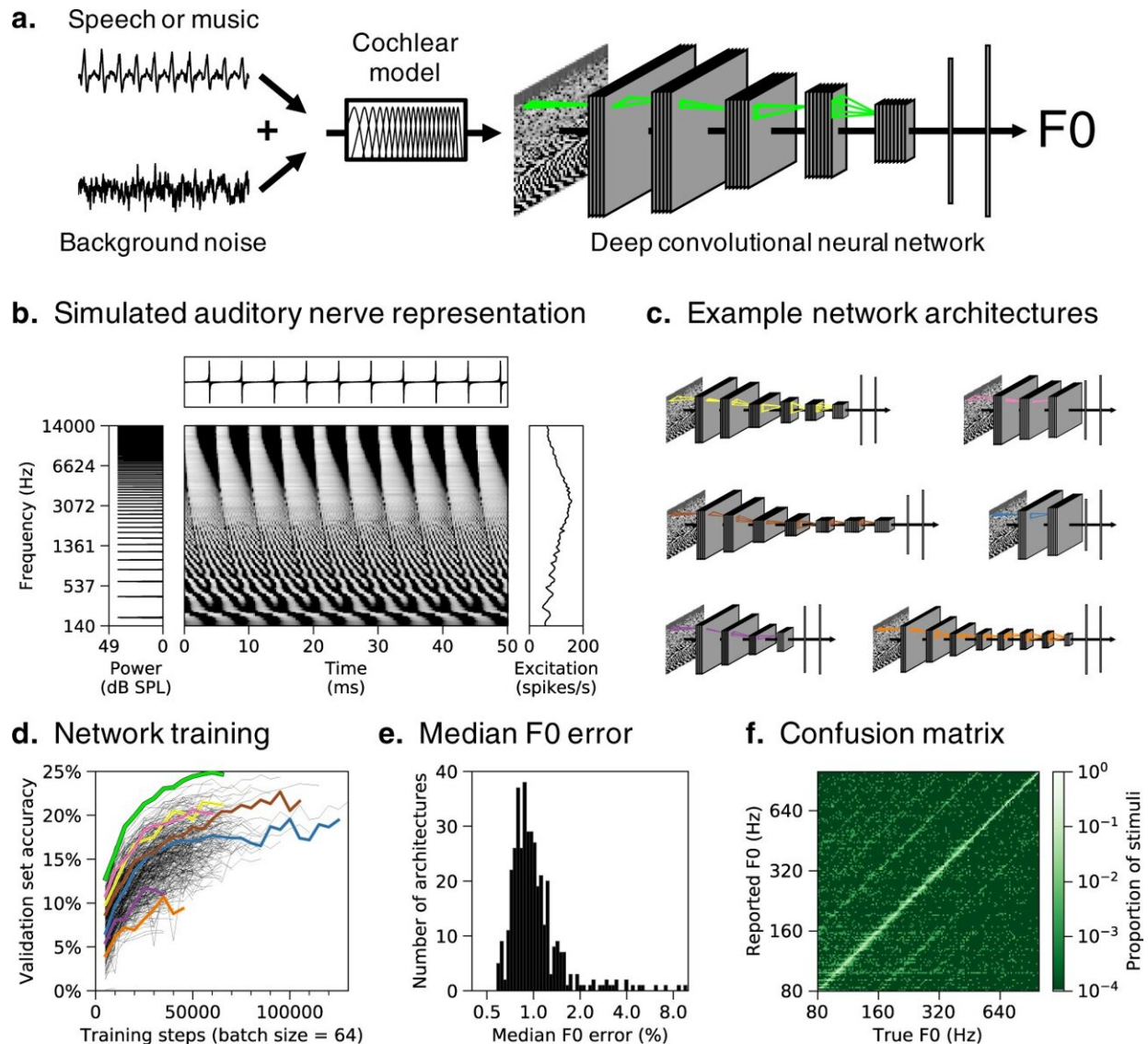


Fig. 1: Pitch model overview. a Schematic of model structure. DNNs were trained to estimate the F0 of speech and music sounds embedded in real-world background noise. Networks received simulated auditory nerve representations

of acoustic stimuli as input. Green outlines depict the extent of example convolutional filter kernels in time and frequency (horizontal and vertical dimensions, respectively). b Simulated auditory nerve representation of a harmonic tone with a fundamental frequency (F0) of 200 Hz. The sound waveform is shown above and its power spectrum is shown to the left. The waveform is periodic in time, with a period of 5ms. The spectrum is harmonic (i.e., containing multiples of the fundamental frequency). Network inputs were arrays of instantaneous auditory nerve firing rates (depicted in greyscale, with lighter hues indicating higher firing rates). Each row plots the firing rate of a frequency-tuned auditory nerve fiber, arranged in order of their place along the cochlea (with low frequencies at the bottom). Individual fibers phase-lock to low-numbered harmonics in the stimulus (lower portion of the nerve representation) or to the combination of high-numbered harmonics (upper portion). Time-averaged responses on the right show the pattern of nerve fiber excitation across the cochlear frequency axis (the “excitation pattern”). Low-numbered harmonics produce distinct peaks in the excitation pattern. c Schematics of six example DNN architectures trained to estimate F0. Network architectures varied in the number of layers, the number of units per layer, the extent of pooling between layers, and the size and shape of convolutional filter kernels d Summary of network architecture search. F0 classification performance on the validation set (noisy speech and instrument stimuli not seen during training) is shown as a function of training steps for all 400 networks trained. The highlighted curves correspond to the architectures depicted in a and c. The relatively low overall accuracy reflects the fine-grained F0 bins we used. e Histogram of accuracy, expressed as the median F0 error on the validation set, for all trained networks (F0 error in percent is more interpretable than the classification accuracy, the absolute value of which is dependent on the width of the F0 bins). f Confusion matrix for the best-performing network (depicted in a) tested on the validation set. Credit: DOI: 10.1038/s41467-021-27366-6

**New research from MIT neuroscientists suggests that natural soundscapes have shaped our sense of hearing, optimizing it for the kinds of sounds we most often encounter.**

In a study reported Dec. 14 in the journal *Nature Communications*, researchers led by McGovern Institute for Brain Research associate investigator Josh McDermott used computational modeling to explore factors that influence how humans hear [pitch](#). Their model's pitch perception closely resembled that of humans—but only when it was trained using music, voices, or other naturalistic sounds.

Humans' ability to recognize pitch—essentially, the rate at which a sound repeats—gives melody to music and nuance to spoken language. Although this is arguably the best-studied aspect of human hearing, researchers are still debating which factors determine the properties of pitch perception, and why it is more acute for some types of sounds than others. McDermott, who is also an associate professor in MIT's Department of Brain and Cognitive Sciences, and an Investigator with the Center for Brains, Minds, and Machines (CBMM) at MIT, is particularly interested in understanding how our nervous system perceives pitch because cochlear implants, which send electrical signals about sound to the brain in people with profound deafness, don't replicate this aspect of human hearing very well.

"Cochlear implants can do a pretty [good job](#) of helping people understand speech, especially if they're in a quiet environment. But they really don't reproduce the percept of pitch very well," says Mark Saddler, a [graduate student](#) and CBMM researcher who co-led the project and an inaugural graduate fellow of the K. Lisa Yang Integrative Computational Neuroscience Center. "One of the reasons it's important to understand the detailed basis of pitch perception in people with normal hearing is to try to get better insights into how we would reproduce that artificially in a prosthesis."

## **Artificial hearing**

Pitch perception begins in the cochlea, the snail-shaped structure in the

inner ear where vibrations from sounds are transformed into [electrical signals](#) and relayed to the brain via the [auditory nerve](#). The cochlea's structure and function help determine how and what we hear. And although it hasn't been possible to test this idea experimentally, McDermott's team suspected our "auditory diet" might shape our hearing as well.

To explore how both our ears and our environment influence pitch perception, McDermott, Saddler, and Research Assistant Ray Gonzalez built a [computer model](#) called a [deep neural network](#). Neural networks are a type of machine learning model widely used in automatic speech recognition and other artificial intelligence applications. Although the structure of an artificial neural network coarsely resembles the connectivity of neurons in the brain, the models used in engineering applications don't actually hear the same way humans do—so the team developed a new model to reproduce human pitch perception. Their approach combined an artificial neural network with an existing model of the mammalian ear, uniting the power of machine learning with insights from biology. "These new machine-learning models are really the first that can be trained to do complex auditory tasks and actually do them well, at human levels of performance," Saddler explains.

The researchers trained the neural network to estimate pitch by asking it to identify the repetition rate of sounds in a training set. This gave them the flexibility to change the parameters under which pitch perception developed. They could manipulate the types of sound they presented to the model, as well as the properties of the ear that processed those sounds before passing them on to the neural network.

When the model was trained using sounds that are important to humans, like speech and music, it learned to estimate pitch much as humans do. "We very nicely replicated many characteristics of human perception ... suggesting that it's using similar cues from the sounds and the cochlear

representation to do the task," Saddler says.

But when the model was trained using more artificial sounds or in the absence of any background noise, its behavior was very different. For example, Saddler says, "If you optimize for this idealized world where there's never any competing sources of noise, you can learn a pitch strategy that seems to be very different from that of humans, which suggests that perhaps the human pitch system was really optimized to deal with cases where sometimes noise is obscuring parts of the sound."

The team also found the timing of nerve signals initiated in the cochlea to be critical to pitch perception. In a healthy cochlea, McDermott explains, nerve cells fire precisely in time with the sound vibrations that reach the inner ear. When the researchers skewed this relationship in their [model](#), so that the timing of nerve signals was less tightly correlated to vibrations produced by incoming sounds, pitch perception deviated from normal human hearing.

McDermott says it will be important to take this into account as researchers work to develop better cochlear implants. "It does very much suggest that for [cochlear implants](#) to produce normal pitch [perception](#), there needs to be a way to reproduce the fine-grained timing information in the auditory nerve," he says. "Right now, they don't do that, and there are technical challenges to making that happen—but the modeling results really pretty clearly suggest that's what you've got to do."

**More information:** Mark R. Saddler et al, Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception, *Nature Communications* (2021). [DOI: 10.1038/s41467-021-27366-6](https://doi.org/10.1038/s41467-021-27366-6)

*This story is republished courtesy of MIT News*

([web.mit.edu/newsoffice/](http://web.mit.edu/newsoffice/)), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Perfecting pitch perception (2021, December 20) retrieved 26 April 2024 from <https://medicalxpress.com/news/2021-12-pitch-perception.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.