

A Computational Model of Human Vocal Imitation

Matthew Caren

MIT, Cambridge, Massachusetts, United States

Kartik Chandra

MIT, Cambridge, Massachusetts, United States

Joshua Tenenbaum

MIT, Cambridge, Massachusetts, United States

Jonathan Ragan-Kelley

MIT, Cambridge, Massachusetts, United States

Karima Ma

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Abstract

Humans have the remarkable intuitive ability to use their voices to imitate any sound they hear. For example, imagine conveying the sound of an engine to a mechanic, a birdcall to a friend, or a strange new synthesizer to a musician. Vocal imitation is fundamental to human communication, and may even form the basis of early language learning. But it is also mysterious: how do we use the limited affordances of a vocal tract to communicate novel sounds far beyond its reach? We propose that the answer lies in recursive social reasoning: we present a computational model of vocal imitation that extends the Rational Speech Acts framework with a simulated vocal tract for the speaker, and a feature-based model of auditory perception for the listener. Without fitting any parameters, our model accurately predicts the types of phones human speakers choose when imitating a variety of real-world sounds ($R^2 = 0.809$).