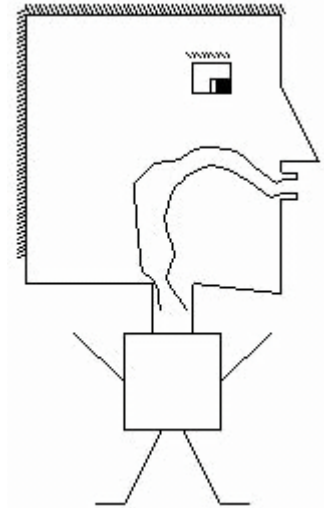


CELEST NUGGET FOR NSF

LEARNING TO SPEAK

The DIVA model of speech production is being developed by Frank Guenther and his CELEST colleagues. DIVA provides an answer to the question, “How does our brain learn to move the tongue, jaw, and lips in order to mimic speech sounds we hear with our ears?” Specifically, the model describes the computations performed by the auditory, somatosensory, and motor cortical areas that allow our brains to transform a desired auditory signal, such as a spoken word or phrase, into precisely timed activations of the more than 40 muscles that move the speech articulators in order to produce that phrase.

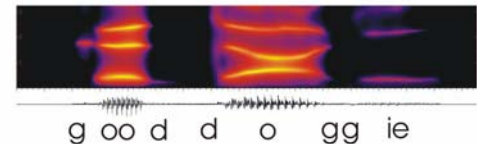
Like an infant, the model uses a babbling phase to learn about the relationship between articulator movements and the sounds they produce. The model also learns the inverse of this relationship; that is, the muscle activation patterns needed to achieve desired sounds. After babbling, the model can be presented with a new word (in the form of a speech sound signal derived from a human speaker pronouncing the word), and within a few practice attempts the model learns the muscle commands needed to produce that word intelligibly.



The figure at right illustrates the model’s first attempts to speak a new phrase, “good doggie”. Each panel illustrates the acoustic energy pattern as a function of time. The top panel illustrates the acoustic energy pattern when an adult human says “good doggie”. The lower panels show the model’s first attempt, which is a poor match to the human speaker, through the ninth attempt, which now closely mimics the human speaker. The improvements with each practice repetition occur because the model updates its commands to the speech muscles based on errors generated on the previous attempt.

By illuminating the neural computations responsible for speech and the ways in which they can break down, the DIVA model can help guide the design of therapies for individuals with disorders such as stuttering, dysarthria, and apraxia of speech. The model’s focus on speech learning also provides insights into the best ways to teach speech to children and learners of a second language.

Sound token presented to model:



Model productions:

