Probing a Neural Network Model of Sound Change for Perceptual Integration

Cerys Hughes
ceryshughes@linguist.umass.edu
Linguistics Society of America 2023 UMassAmherst

What humans do

CONTRAST SHIFT

Cues to a contrast can change over time

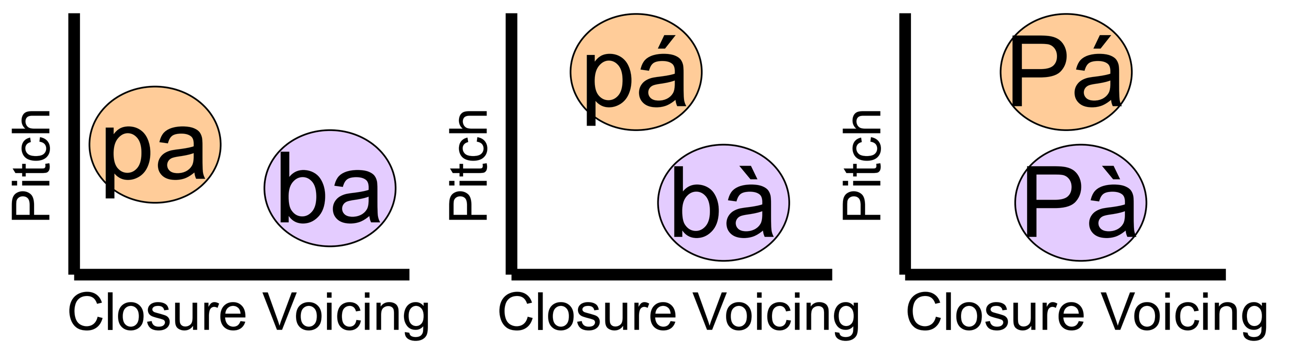

Fig 1. Contrast shift schematic adapted from Kang (2014)

- Yang (2019): more likely to shift cues contribute to the same **Integrated Perceptual Property** (**IPP**).
  - **Remains to be computationally implemented**

INTEGRATED PERCEPTUAL PROPERTIES

- Combined **auditory dimensions**
- Cues not perceived independently
- Example: **spectral continuity** (Kingston et al 2008)
  - Associated with stop voicing
  - Low F1, long voicing

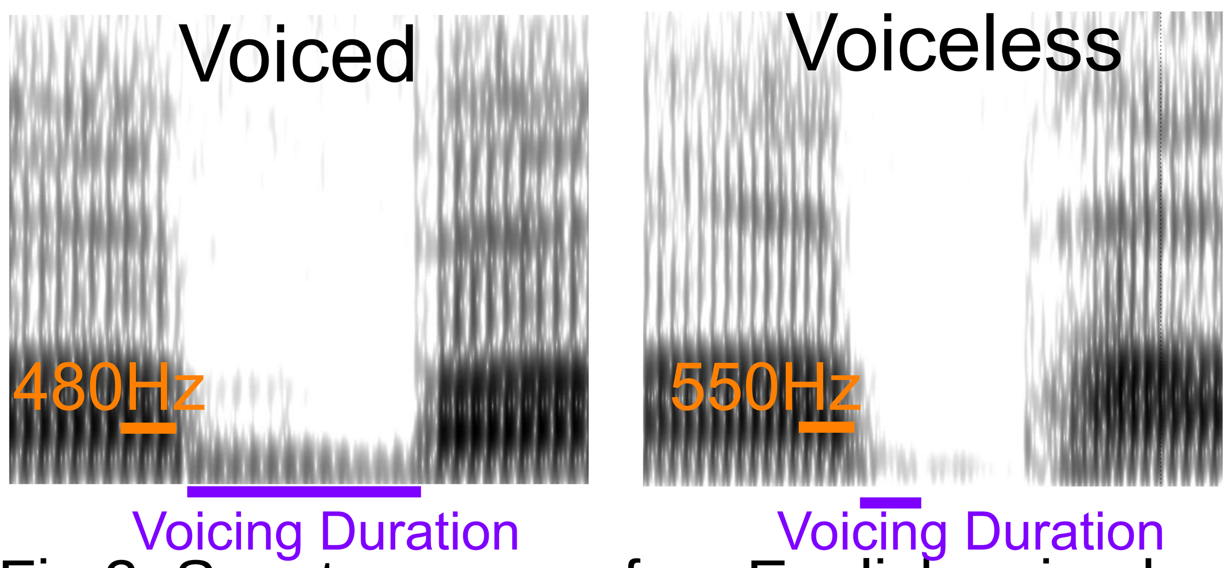

Fig 2. Spectrograms of an English voiced stop (high spectral continuity) and a voiceless stop (low spectral continuity)

- Diagnose using **perceptual distances** and linear regression

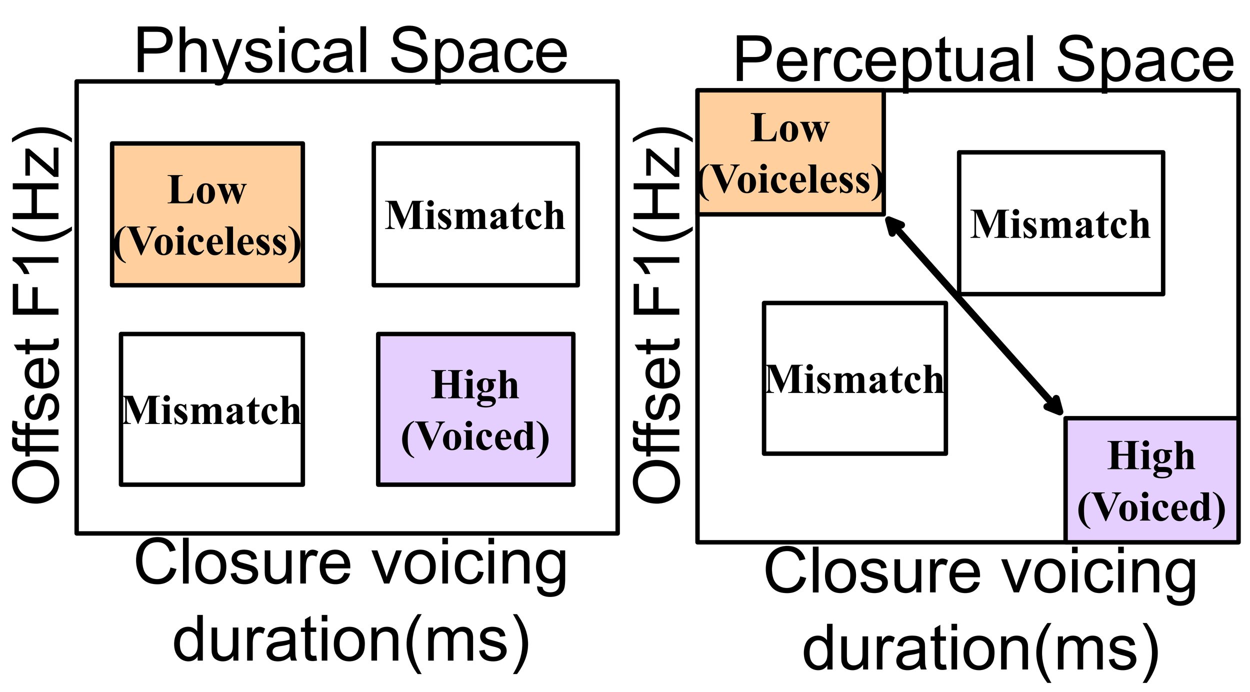

Fig 3. Example Garner paradigm plot showing stretching along the IPP dimension, but not the opposite dimension. (Garner 1974, Kingston et al. 2008)

What machines do

CONVOLUTIONAL NEURAL NETWORKS

- Can a neural network model of sound change (Beguš 2020) implement the IPP account of contrast shift?
  - Raw audio input
  - Speech perception component: **Convolutional Neural Network (CNN)**
- **Are a CNN's acoustic representations like IPPs?**
- Relevant property: pattern-recognizing **filters**

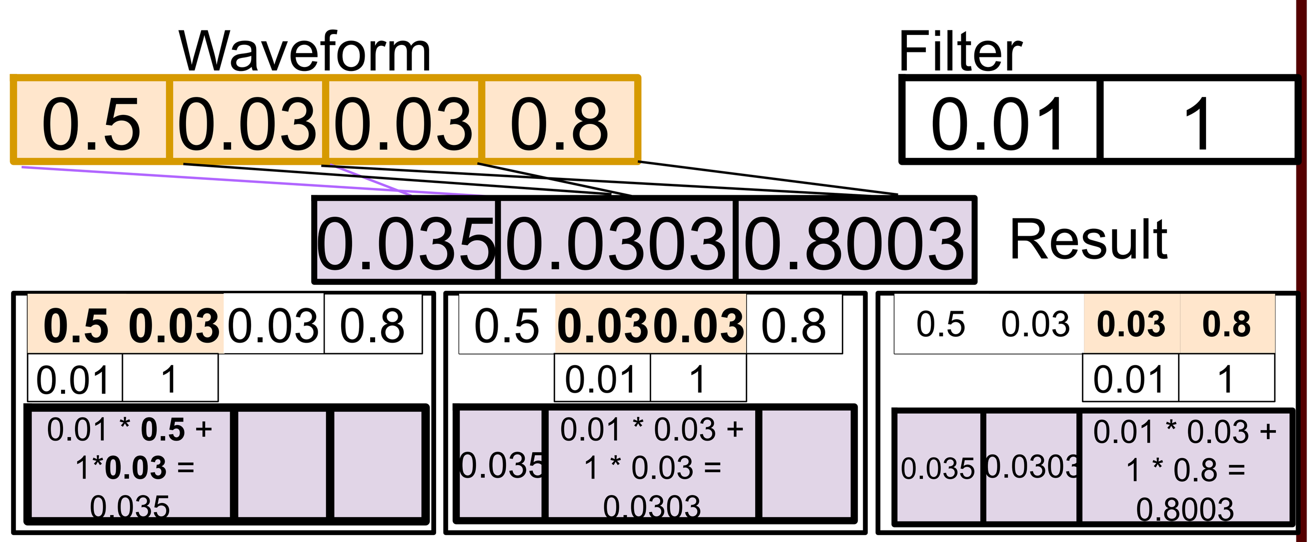

Fig 4. Step-by-step example calculation of how a filter detects patterns in the input signal.

- **Will the network combine spectral continuity cues into the same set of filters, like the human IPP?**

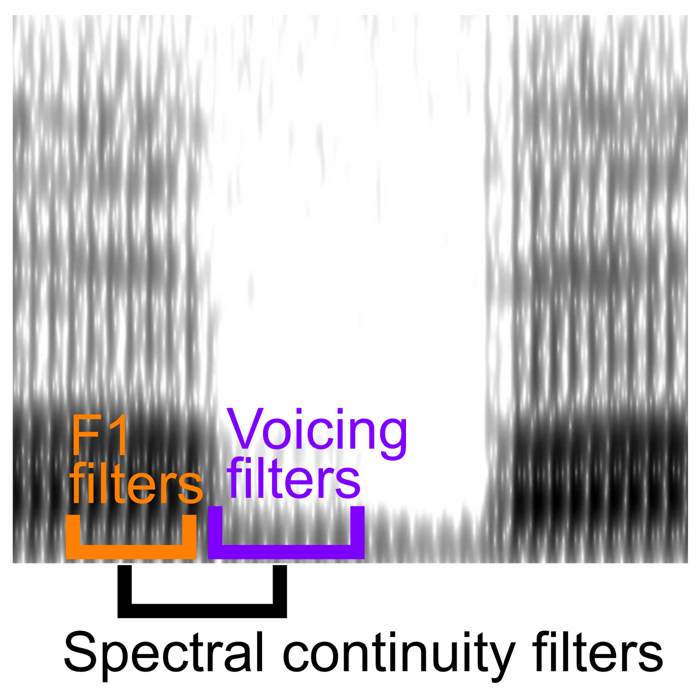

Fig 5. Visual representation of filters that a CNN could learn, independent or IPP

EXPERIMENT

- Isolate CNN component by replicating its architecture and giving it a simple classification task:
  - English voiced vs voiceless stops (MIT SCG 2005)
- Adapt human paradigm and compare with results of Kingston et al. (2008)

Comparing humans and machines

TESTING HUMANS VS CNNs

- Humans: perceptual distances from discrimination task (Kingston et al. 2008)
- CNNs: perceptual distances from internal vector representations (Ward 2019)

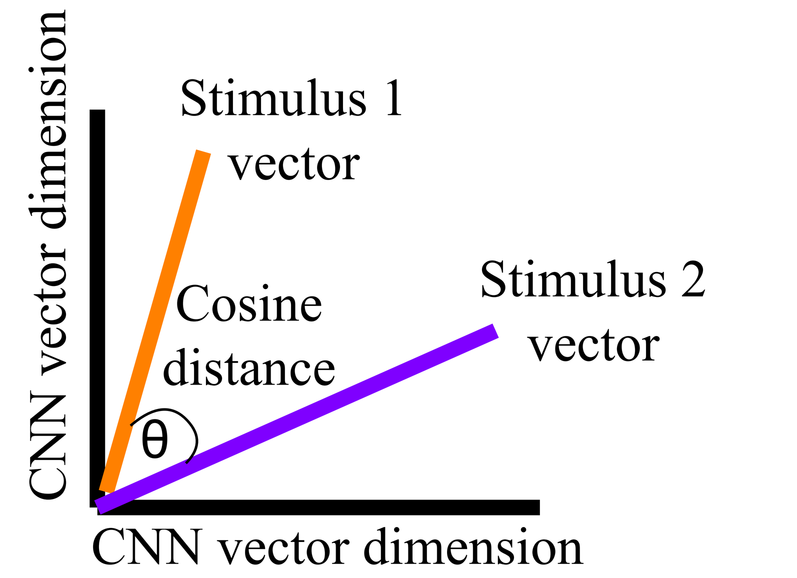

Fig 6. Geometry of cosine distance

RESULTS

- **CNNs show a different integration pattern from the humans' (Kingston et al. 2008) for each pair of cues investigated.**

Cue Pair	Human	CNN
f0, closure voicing	Integrated	Not integrated
F1, closure voicing	Integrated	Weakly integrated on non-IPP dimension ($p=0.0305$)
F0, closure duration	Not integrated	Integrated ($p < 0.001$)
F1, closure duration	Not integrated	Integrated ($p < 0.001$)

Fig 7. Summary comparison of CNN results and human results. Human results from Kingston et al. (2008)

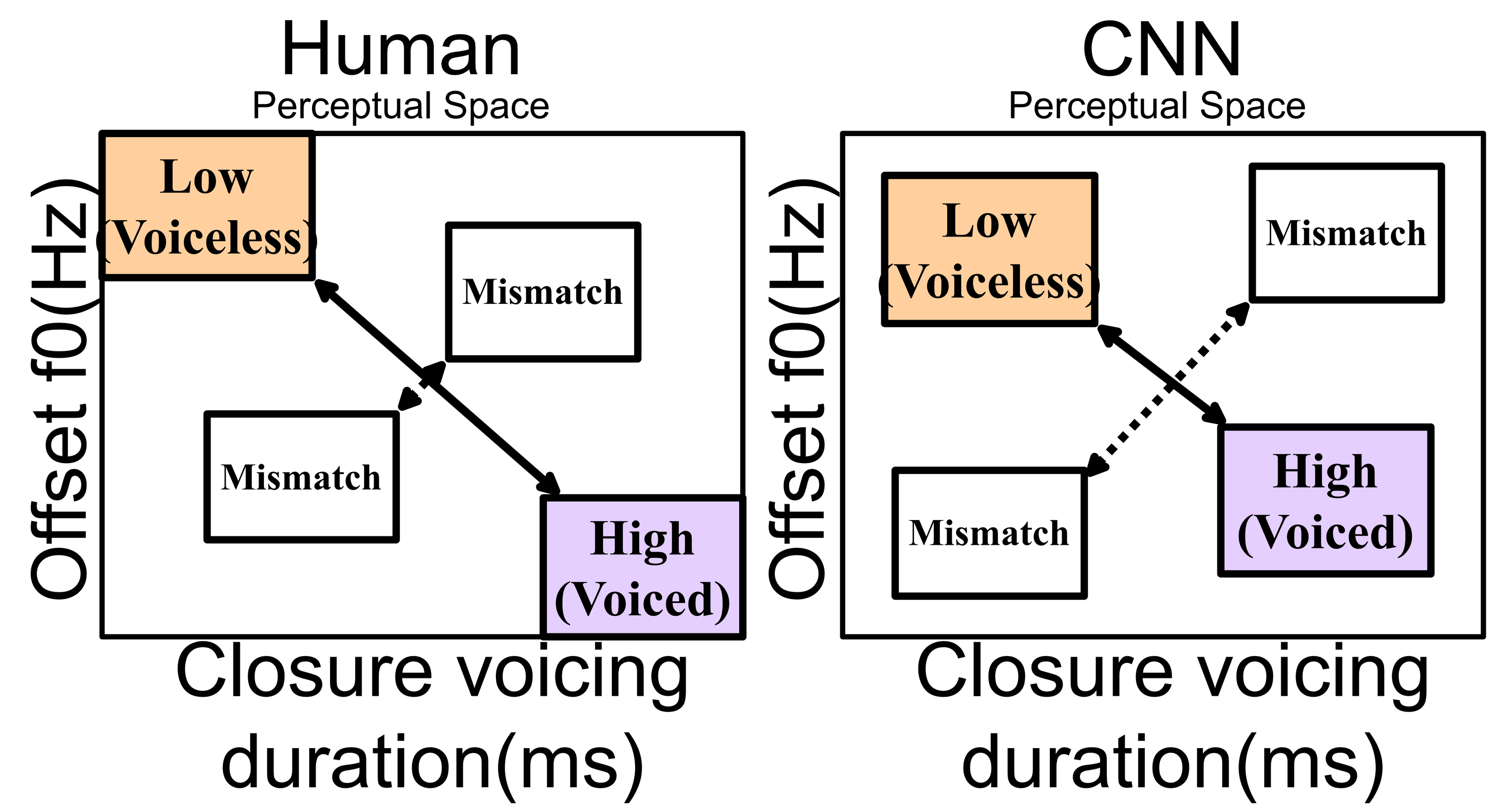

Fig 8. Visual interpretation of human and CNN perceptual spaces for one pair of cues.

DISCUSSION

- Not yet applicable to implementing Yang (2019)'s IPP account
- Possible explanations for the discrepancy:
  - How correlated each pair of cues is in the training data
  - Usefulness of each cue for classification task