How informative is dialect about vowel distributions?

Dave F. Kleinschmidt, Kodi Weatherholtz, and T. Florian Jaeger, University of Rochester
dkleins2@ur.rochester.edu

Introduction

Listeners need to cope with talker variability
Can think of this as a statistical inference problem: infer current talker's cue distributions.
Combine experience with current talker with prior experience with other talkers
Structure in talker variability determines how best to use prior experience.

Question

How much do socio-indexical grouping variables tell us about a talker's phonetic cue distributions?
Focus on dialect. Compare to sex and talker identity
Structure could vary based on how cues are represented, so also compare raw formant frequencies (Hz) and Lobanov normalized formants (z-score within talker) which removes differences in overall average F1 and F2 across all vowels (e.g., due to differnt vocal tract sizes)

Data

F1 and F2 measurements from isolated hVd words from Nationwide Speech Project
Data from 48 talkers: 4 male and 4 female from 6 dialect regions.
5 repetitions per vowel (on average: a few have 6, a few less).

Approach

Extract group-conditioned cue distributions for each vowel:

p(cue | vowel, group)

Maximum likeilhood normal distribution (mean and covariance)
Based on raw formant frequencies in Hz, or Lobanov-normalized (z-scored within talker).
Conditioned on **talker**, **sex**, **dialect**, **dialect+sex** together, or nothing (**marginal**)

Answers

Knowing a talker's dialect is informative about their vowel distributions
But less so than talker identity, and sex (for raw formants)
Also useful for vowel recognition, but only for some dialect-vowel combinations

Quantified how much listeners can benefit from tracking cue distributions conditioned on different socio-indexical variables

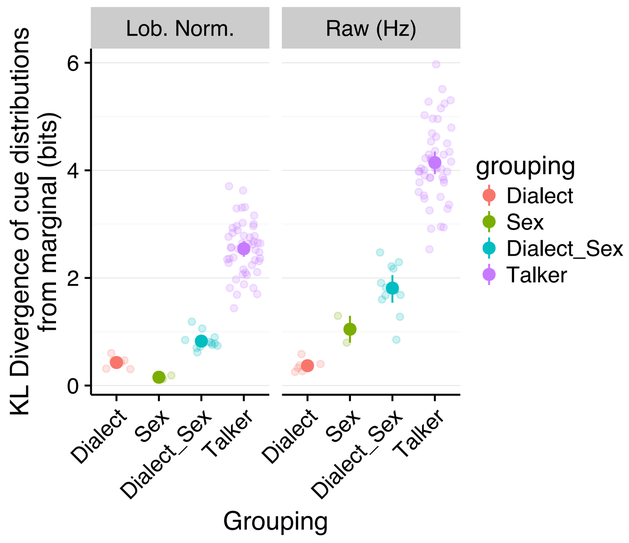
Dialect is informative and useful, at an intermediate level, depending on how cues are represented.

Informativity about **cue distributions**

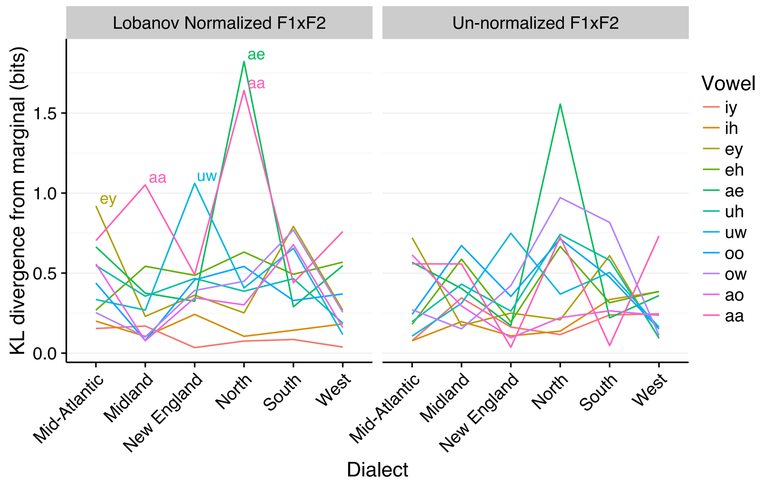
If a variable is **informative** about cue distributions, then distributions **conditioned on** that variable are **different** from each other and hence from the **marginal** distribution.

Quantify by average **information gain** (KL divergence) for conditional distributions over marginal. How much does knowing the value of the socio-indexical variable tell you about the cue distributions, relative to not knowing it?

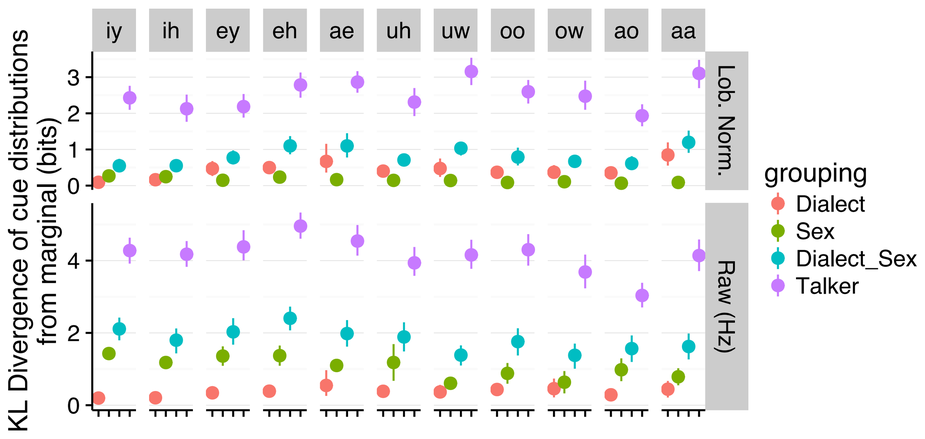
Overall



By dialect



By vowel



Utility for **speech recognition**

By **utility** of a socio-indexical variable, we mean the advantage for speech recognition from

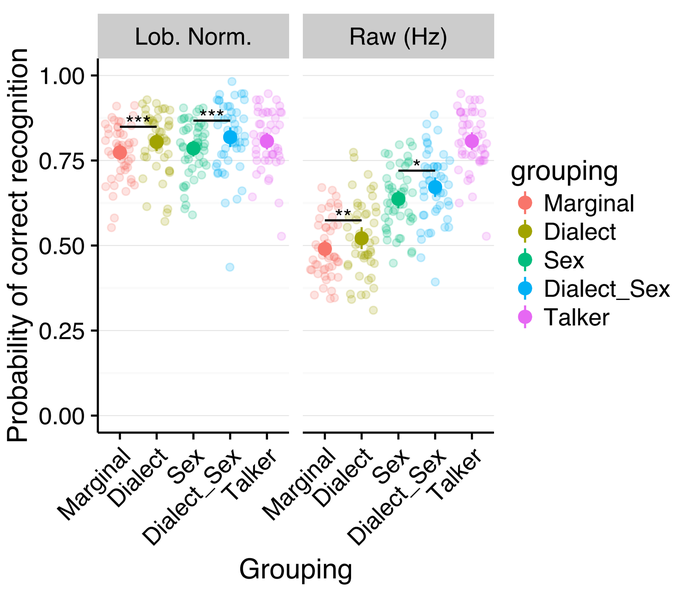
1. **tracking** cue distributions conditioned on that variable
2. **knowing** the value of that variable

Quantify this using probability of correct recognition by an **ideal listener** model.

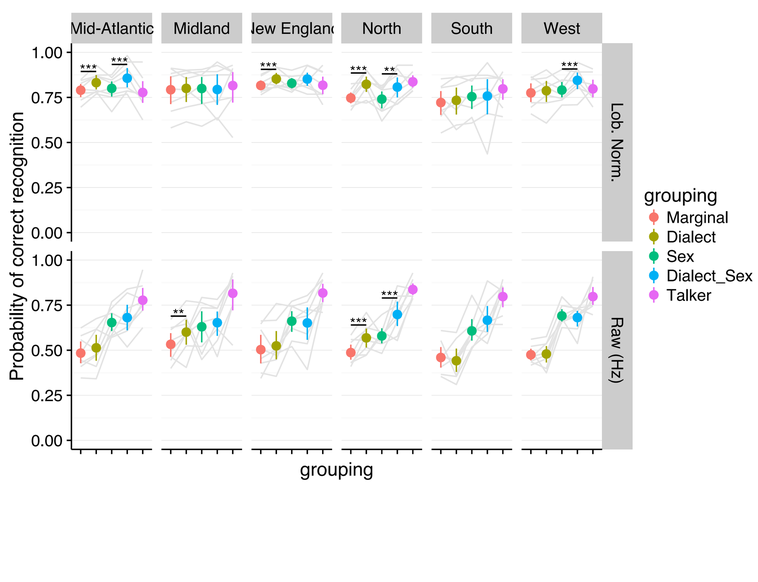
Classify observations based on conditional cue distributions using Bayes rule:

p(category | cue, group) ∝ p(cue | category, group) p(category)

Overall



By dialect



By vowel