# A Methodology for Using Crowdsourced Data to Measure Uncertainty in Natural Speech

Lara Martin[1], Matthew Stone[2], Florian Metze[1], and Jack Mostow[1]
IEEE Spoken Language Technologies Workshop, December 7th-10th 2014
[1]Carnegie Mellon University, [2]Rutgers University

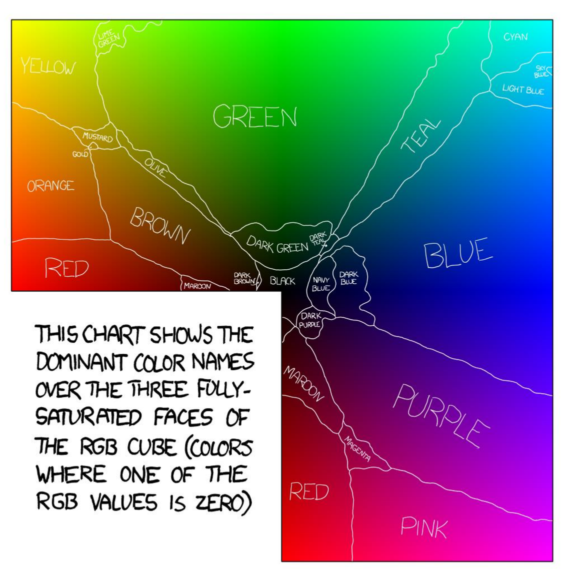
## Introduction

**Prosody:** the melodic structure of speech → meaning in speech
**Our focus:** uncertainty in prosody→ entropy
**Question:** Does the amount of uncertainty a person expresses correlate with the ambiguity of the color one is trying to describe?
**Hypothesis:** If so, there will be prosodic information that distinguishes utterances in low-entropy contexts from utterances in high-entropy contexts.

## Methods

### Data:

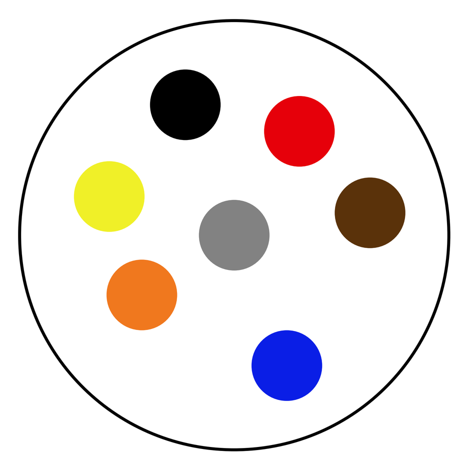
- XKCD color internet survey
- 222,500 participants typed color names

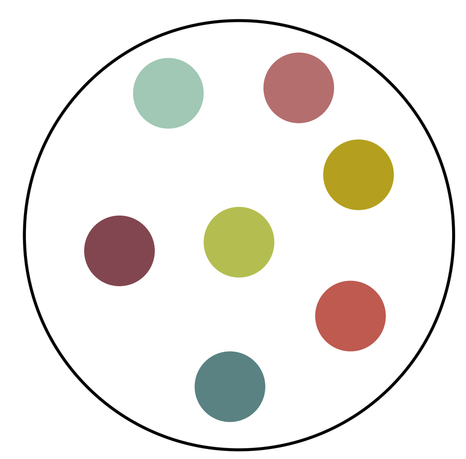


http://imgs.xkcd.com/blag/satfaces_map_1024.png

selection →

- 60 RGB colors (30 high and 30 low entropy)
- Presented one at a time in random order in a 2x2-inch square


Low Entropy


High Entropy

collection →

- 18 American native-English speakers (10 female, 8 male)
- Not colorblind
- Given 6 seconds to verbally state the name of the color (while being recorded)
- 488 low-entropy recordings
- 468 high-entropy recordings

### Classification:

**Features of Uncertainty**


Praat

automatic:
# of syllables, # of pauses (>= 0.3 s), total duration, length of phonation, speech rate, articulation rate, ASD (phonation length/ # syllables)
(Wempe & de Jong 2008)

by hand:
ToBI final tones (L-L%, H-L%, L-H%, H-H%)

# filled-pauses (words like “um” and “uh”)
&
Gender of participant

**Classes**

Low Entropy Color

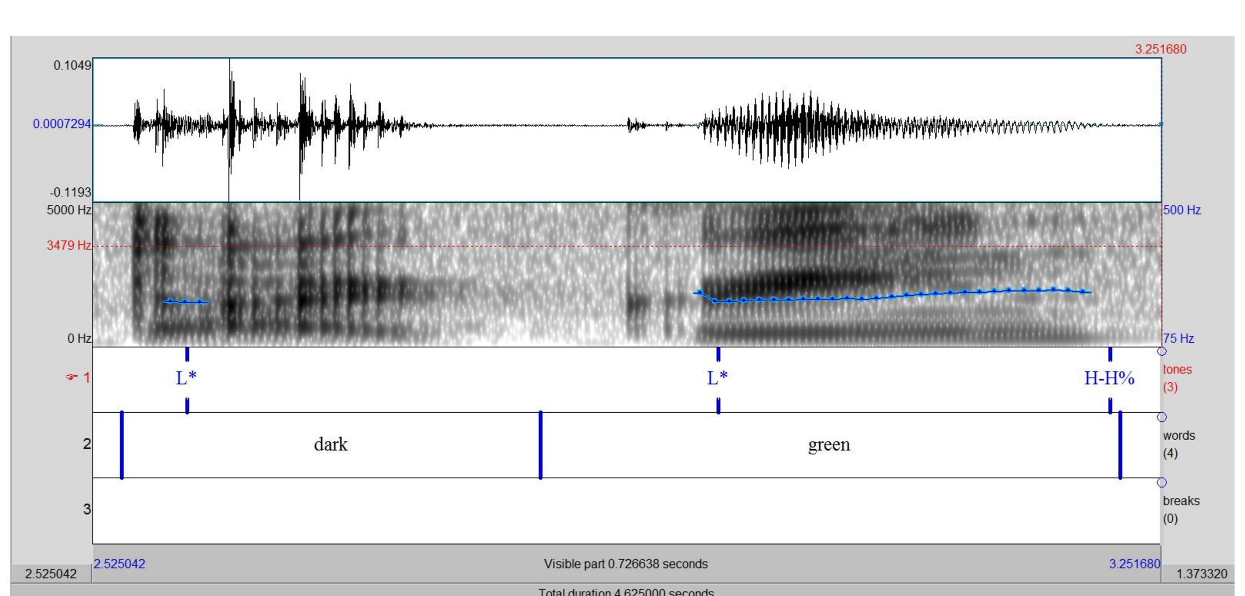
High Entropy Color

**Details**

- Binary Classification
- 10-fold CV using Weka
- Naïve Bayes classification tree

### Example:


2x2-inch square



Participant saying “dark green” with try tone (H-H%).

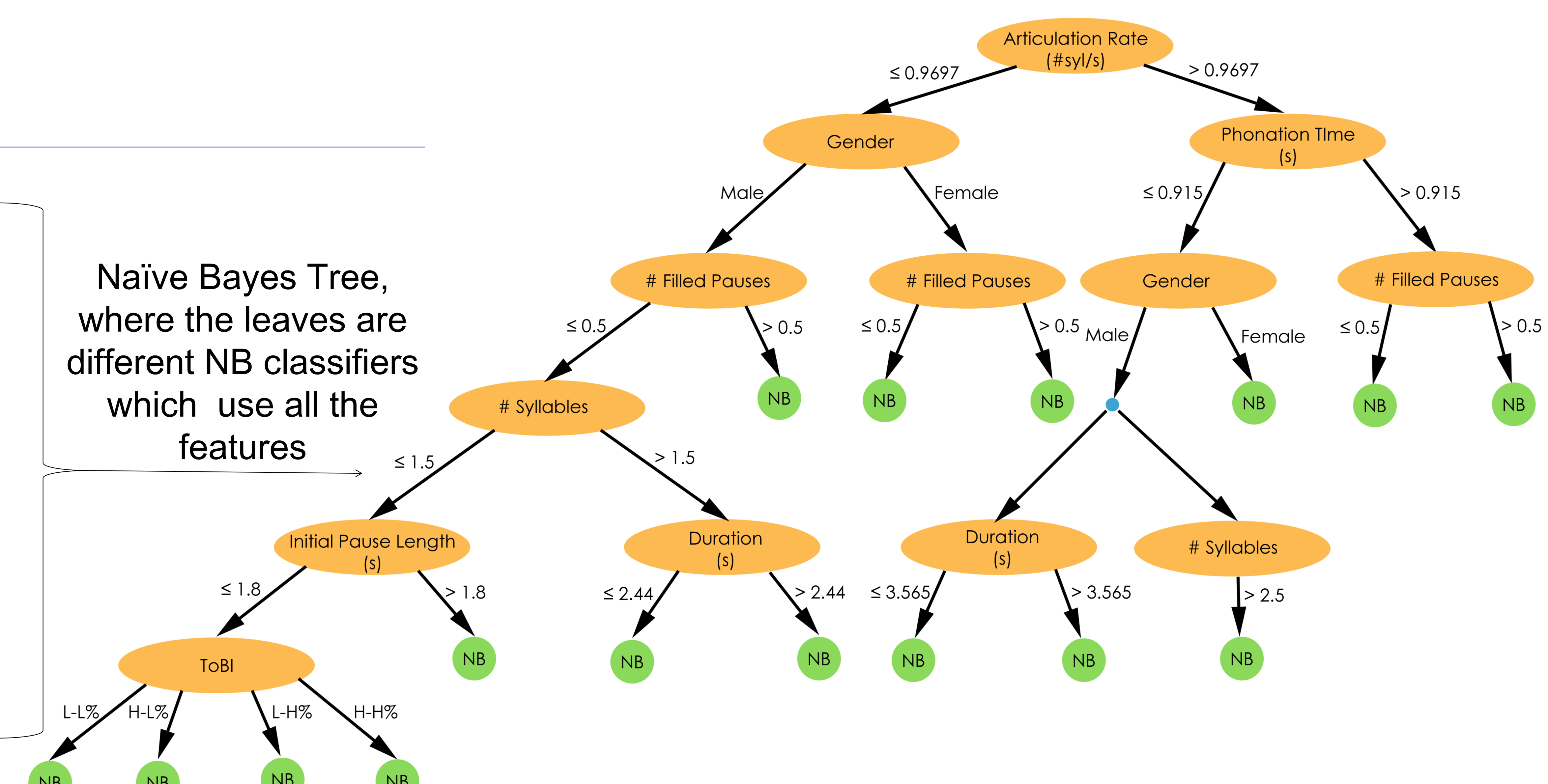
## Results

65.27% accuracy with gender
62.34% accuracy without gender
P < 0.0001

| | Hyp-Low | Hyp-High | Total |
|---|---|---|---|
| Ref-Low | **346** | 142 | 488 |
| Ref-High | 190 | **278** | 468 |
| Total | 536 | 420 | |

Confusion matrix of high & low entropy classifications produced from Naïve Bayes Tree

Naïve Bayes Tree, where the leaves are different NB classifiers which use all the features




## Questions for Future Work

- Were these prosodic cues really measuring uncertainty or just the difficulty of the task?
- Why didn’t intonation play a bigger role? Why were people mostly using H-L% monotone or L-H%? Could we make the task more like a conversation?
- Why was there a gender difference? Cultural? Biological?
- Can this be used in applications? For example, if people used more uncertainty while using voice search, would it improve precision and/or recall?
- Will this work with other media besides color?
- Are these features cross-lingual?

## Conclusions

- Entropy **can** correlate with prosodic cues thought to predict uncertainty in speech
- We **can** use crowdsourced data to calculate entropy values, despite being from a non-verbal task

For references, please see paper.