



Assessing Self-awareness and Transparency when Classifying a Speaker's Level of Certainty



Heather Pon-Barry

School of Engineering and Applied Sciences, Harvard University
Cambridge, MA 02138, ponbarry@eecs.harvard.edu

Stuart Shieber

School of Engineering and Applied Sciences, Harvard University
Cambridge, MA 02138, shieber@seas.harvard.edu

Overview

- Is prosody helpful in automatically classifying a speaker's internal state?
- We examine one aspect of internal state: level of certainty.
- Past work focuses on classifying the *perceived* level of certainty, but this quantity often differs from a speaker's *actual* level of certainty.
- Our data is a corpus of single-sentence utterances that are annotated with:
 - The perceived level of certainty
 - The speaker's self-reported level of certainty
 - Whether the statement is correct or incorrect
- We say a speaker is *transparent* if their self-reported level of certainty is aligned with their perceived level of certainty.
- We say a speaker is *self-aware* if their self-reported level of certainty is aligned with the correctness of their utterance.
- Our models, trained on prosodic features, correctly classify a speaker's self-reported level of certainty 75% of the time.
- Intelligent systems can use this information to make inferences about a user's internal state, e.g., whether someone has a misconception, makes a lucky guess, or needs encouragement.

Uncertainty Corpus

- 20 speakers
- 600 utterances
- Method of elicitation:
 - Speakers are presented with a sentence containing one or more gaps
 - Options for filling in the gap are displayed
 - Upon hearing a beep the speaker reads the sentence aloud
- Five annotators rate the perceived level of certainty on a 5-point scale (average $\kappa = 0.43$)
- Speakers rate their *own level of certainty* on the same 5-point scale

Transportation

Q: What is the best way to get to North Station from the Harvard T-stop?

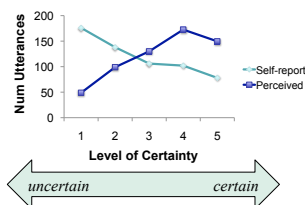
- A: Take the red line to _____
- Park Station
 - Downtown Crossing
- and transfer to the _____.
- green line
 - orange line

Vocabulary

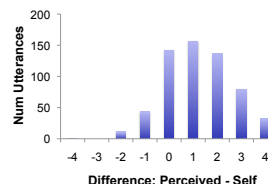
Only the _____ workers in the office laughed at all of the manager's bad jokes.

- pugnacious
- craven
- sycophantic
- spoffish

Self vs. Perceived Level of Certainty



Self-reported levels of certainty were consistently lower than perceived levels of certainty



"Perceived" level of certainty = average of 5 annotators' ratings.

Self-Awareness & Transparency

Self-awareness

	Answer: INCORRECT	Answer: CORRECT	
Self-report: UNCERTAIN	Self-aware	Non-self-aware	User makes a lucky guess or lacks confidence
Self-report: CERTAIN	Non-self-aware	Self-aware	

Transparency

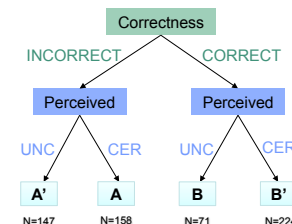
	Perceived as: UNCERTAIN	Perceived as: CERTAIN	
Self-report: UNCERTAIN	Transparent	Opaque	System may give inappropriate feedback
Self-report: CERTAIN	Opaque	Transparent	

Acknowledgements

This work was supported in part by a National Defense Science and Engineering Graduate Fellowship.

Classification Experiment

- Decision tree classifier
 - C4.5 algorithm with pruning
- Prosodic features:
 - Pitch (F0): min, max, mean, stdev, range, absolute slope
 - Intensity (RMS): min, max, mean, stdev
 - Temporal: silence, total duration, speaking duration, speaking rate
- Divide utterances into four subsets



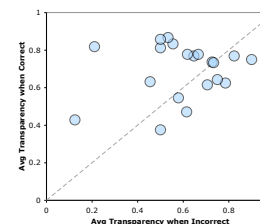
Without Subsets (Baseline)	
Accuracy Majority Class	Accuracy DT Classifier
52.30	66.33

Subset	Accuracy Subset Majority	Accuracy DT Classifier
A	65.19	68.99
B	53.52	69.01
A'	84.35	84.35
B'	75.89	75.89
Overall	72.49	75.30

Informative Features

- High information gain for two features
 - Percent Silence** (expected, based on past work)
 - Speaking Rate** (unexpected, speaking rate was not strongly correlated with perceived level of certainty in past work)

Speaking Personalities



- Each dot corresponds to an individual speaker.
- Speakers who are equally transparent regardless of correctness fall along the dashed line.
- Outliers may indicate presence of different speaking personalities