

# Perception Experiment

June 20, 2023

Morgan Moyer

---

Some possible research questions:

1. Is there a reliable prosodic difference for each speech act? (Experiment 1)
  2. (How much speaker variability is there?)
  3. Is there a perceptible prosodic difference across the three speech acts?
  4. Can hearers reliably use prosodic information to disambiguate the speaker's intended message (speech act)?
  5. Does prosody carry meaning?
  6. Can prosody override context?
- 

There are several ways to think about how to answer 3–6, though they may seem like the same question, they affect how we would design the experiment.

## 1. How do we turn “Speech Act” into a dependant measure?

### (a) Paraphrase of the Speech Act

- Which paraphrase is the closest to what the speaker intended? What do you think the speaker intended?
- Pro: by placing the participant in the hearer's shoes, it might more naturally get at the intition that we want, namely, how is prosody used as a cue to determine speaker meaning?  
We can quantify uncertainty in the speech signal if we have the three possible answers as a probability distribution, using a relative entropy measure (KL divergence).
- Con: ?

Texte

### (b) Response to the Speech Act

- Which is the best response to this statement?expected
- Which response do you think the speaker intended the hearer to give?
- Uses the responses from Experiment 1 as the DM.
- Pro: Puts the participant in the position of reasoning about the speaker's intention/goal

- Con: Maybe it removes the participant from the conversational situation because they're not being asked

(c) **Speech Act Categorization Task (i.e., Hellbern & Sammler 2016)**

- Participants choose the category they think the speech act is.
- Pro: Directly tests what we want?
- Con: Not the most naturalistic task, and would involve somewhat intensive training on categorization

## 2. Forced Choice versus Sliders

(a) Forced Choice between three options

- Makes trickier stats, Multi variate logistic regressions, but in Bayesian equivalent?

(b) Sliders:

- Three options each rated on a sliding scale that adds to 1 (so we can make a probability distribution)
- This aligns with our assumptions about the ambiguity of 'Tu peux V' phrases in French:
- I.e., interpretation as the process of inference under uncertainty about the intended (speaker) meaning
- The ambiguity can be modeled as a probability distribution over the three meanings (SA interpretations)
- Prosody and Context change the probability distribution over the three meanings
- With this method we can capture changes in the probability distribution

## 3. What exactly are we interested in testing?

(a) Are we interested in quantifying the role of prosody alone?

Call this **Signal - Context**

- Yes, so we can present the utterances without the preceding contexts and see if prosody alone is sufficient to determine
- We can also quantify Accuracy (cf. Trott et al 2019) by comparing participants responses to what was meant

(b) Are we interested in quantifying the interaction between prosody and context?

- i. Items with the contexts they were originally tested in Experiment 1.  
Call this, **Signal + Context**

- Comparison with the results from the test without context (I did something like this with my questions database study)
- Quantify relative entropy using Kullback-Leibler Divergence from the flat distribution (Whichever task shows the least divergence)
- We can also test whether participants' responses match with our a priori intuitions of what each SA context was meant (creating an Accuracy measure)

- ii. Cross Items x Contexts (addressing Question 6)  
Call this, **Signal x Context**

- In this way we can pit the two cues against each other
- Con: factorial explosion in the Latin Square, the minimal Latin Square is the 3 x 3, the worst (if we want to quantify by-item variability) then we have 36 x 36
- Then we have to consider how to deal with by-speaker variability.

In my opinion, I think we should do the Signal - Context and the Signal + Context experiments, and conduct the comparative analysis. This is also a novel analysis that the other studies looking at similar things haven't done.

While the Signal x Context study is interesting, I think it would require much more time to design. (Plus, it would be much more expensive)

### **Some other questions worth looking into:**

1. How reliable (unvaried) is the prosodic marker for each speech act?  
Is there more than one marker for each SA? Or does there seem to be one (main)?
  - This is essentially our analysis for Experiment 1, and with it we can try to understand also Question 5 (the more variation, the lower the reliability of the prosodic signal alone)
  - Variation
2. How reliably can each unique prosodic signal disambiguate the speech act ?
  - For each acoustic correlate, what is its contribution to disambiguation?
  - This would be harder to analyze because it requires further isolating the individual components (assuming there are multiple)

---

### **Some thoughts about Ruytenbeek & Trott**

- They did a different acoustic analysis: mean f0, f0 slope and duration
- Perception task only looked at Assertion vs. Question as predictor
- Intent Predictor = something like the original SA context, what I was thinking would factor as Accuracy
- we might directly put acoustic values into our model? They didn't exactly directly test the effect of the signal on responses