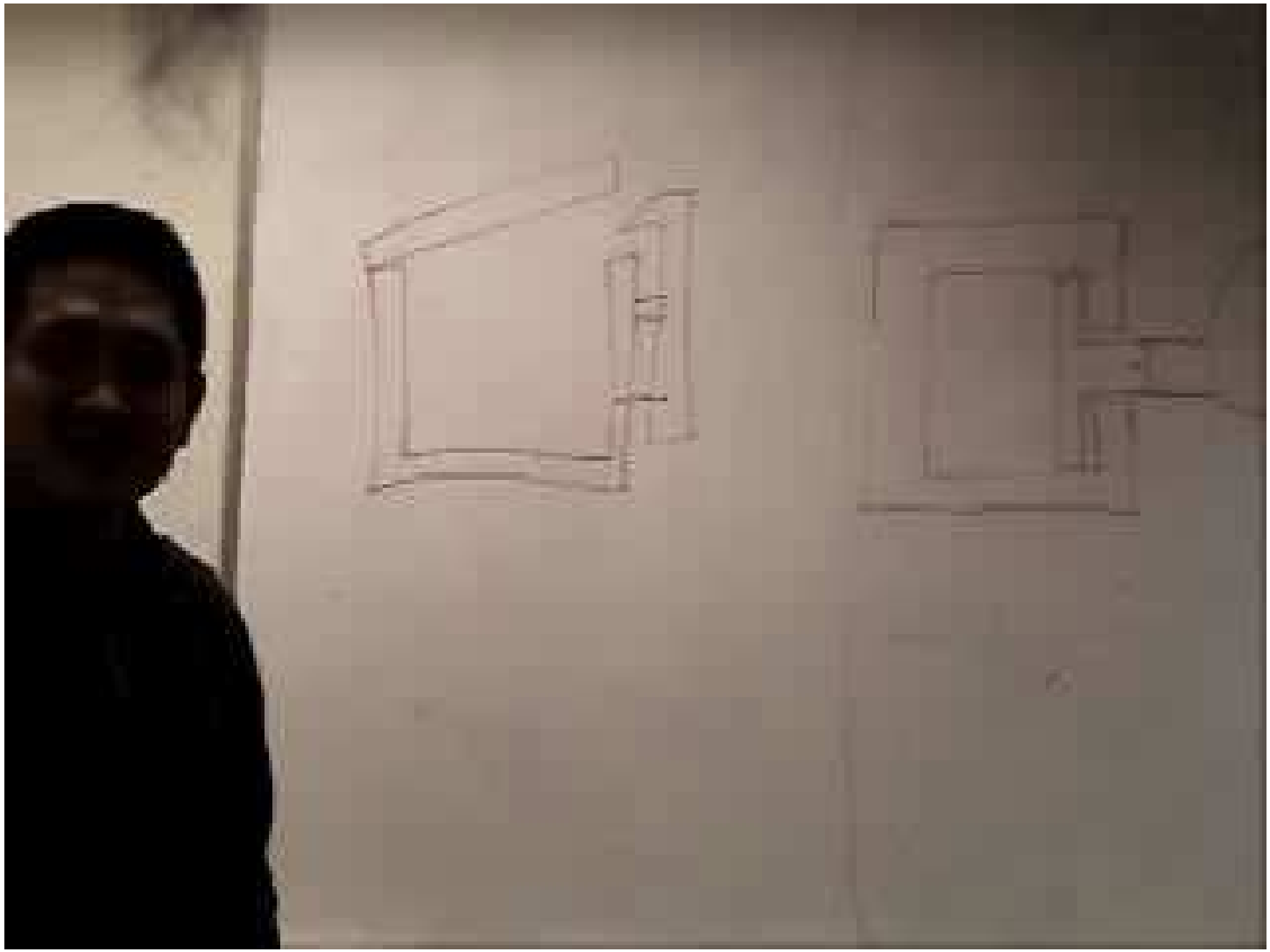
Gesture and Language

Jacob Eisenstein

Introduction

- I’m a student with Randy Davis and Regina Barzilay
- I study gesture and natural language processing
  - Gesture is a modality that accompanies speech
  - It communicates some things far more efficiently than speech
  - Yet we have almost no idea how it works or what it does (it’s like the dark side of the moon)
- Please don’t make fun of the speaker in this video (or others)

An Example



Notes about this video

- K: “What’s your research about?”
- Me: “Gestures”
- K: “I’m sorry, I don’t think I made any.”

- Lesson – there’s a lot we don’t know about our own gestures
- What he’s saying is incomprehensible without gestures
  - There’s a huge part of communicative behavior that we don’t understand
- Before we can build gesture into multimodal interfaces, we should understand how it works.

Four Questions about Gesture

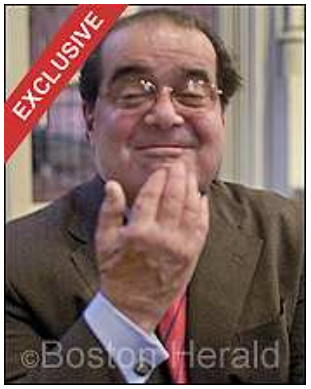
- What is gesture?
- How is gesture linked to speech?
- (How) do gestures communicate?
- What can we (engineers) do with gestures?

A Continuum of Gestures

speech absent ⟷ speech present

Sign Language

Emblems



Gesticulation

Deictic

Representational

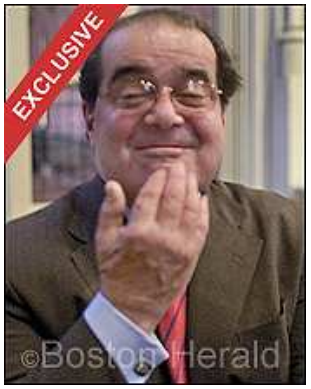
Beat

A Continuum of Gestures

conventionalized ⟷ idiosyncratic

Sign Language

Emblems



Gesticulation

Deictic

Beat

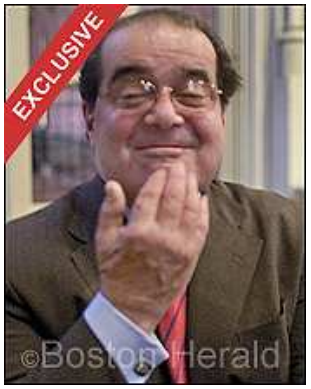
Representational

A Continuum of Gestures

Structured, bottom-up → Unstructured, top-down

Sign Language

Emblems



Gesticulation

Deictic

Beat

Representational

Notes about the Cassell Paper

- "Propositional" gestures
  - To me this seems like an attempt to differentiate from existing systems
  - I don't think the distinction of conscious / unconscious is crucial
- Metaphorics and iconics: this distinction is not critical

Deictic Gestures

Deictic Gestures

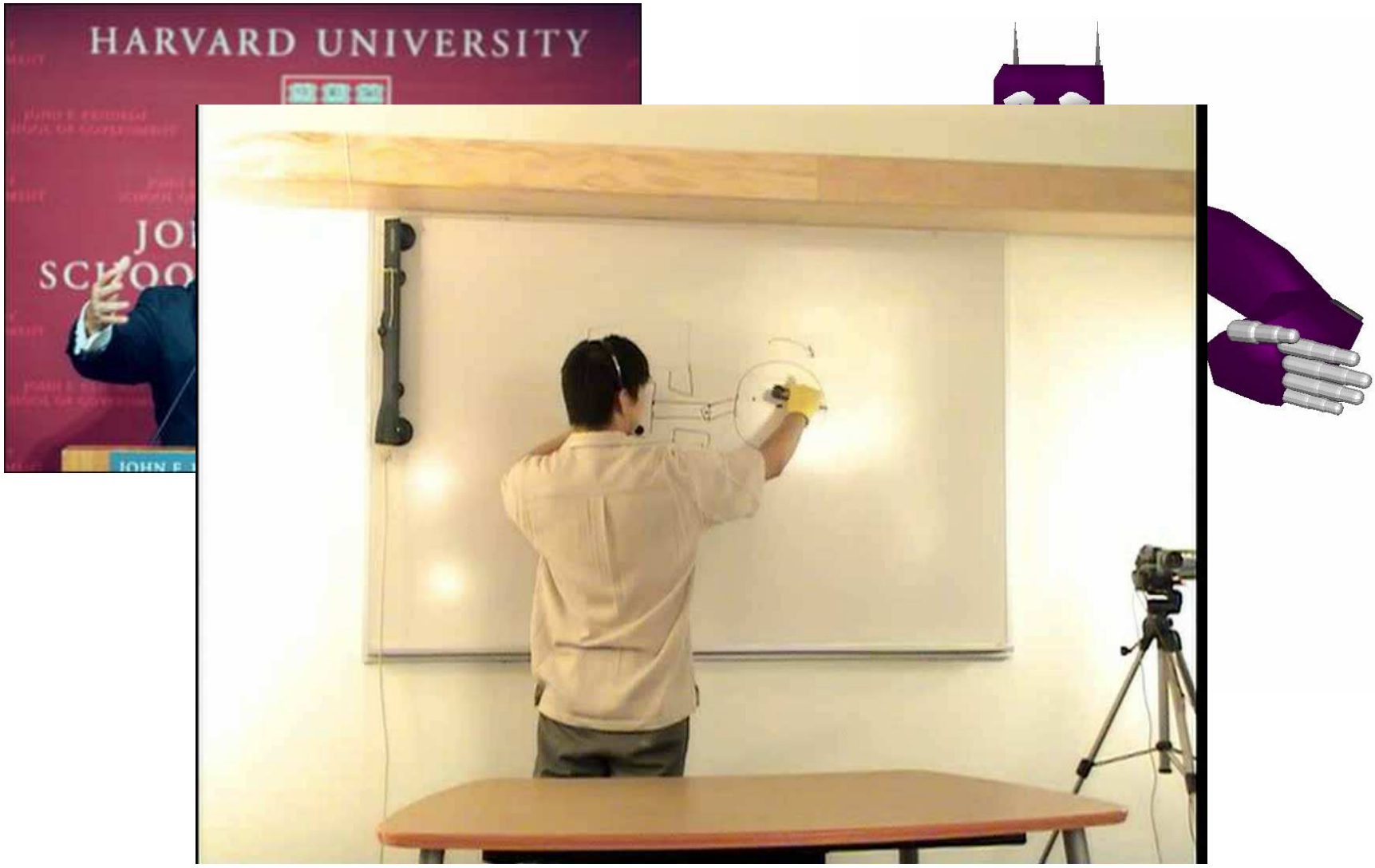
- Create meaning by spatial reference
  - To a physical thing (age 2)
  - Or to an abstract idea (age 12)
- Often a “pointing” handshape, but not always
  - In some cultures: lips, middle finger
  - In some contexts: open hand

Beat Gestures

Beat Gestures

- Beat gestures are baton-like movements
- They add emphasis to speech items that are significant (Cassell & McNeill, 1990)
- Fluent use of beats does not appear until around age 11 (Cassell 1989)

Representational Gestures



Representational Gestures

- Two types
  - Iconic: expressing concrete physical properties
  - Metaphoric: expressing a physical metaphor
- These gestures communicate meaning through their spatial properties
- The extent to which these gestures communicate is one of the most hotly-debated topics

So, what is gesture?

- We've provided a list of types -- what's the unifying theme?
- It's hand movements *that are associated with communication*
  - This may seem like an unacceptably loose definition, but it isn't.
  - Goodwin and Goodwin (1986) report that if you just ask people to record what they see, they restrict their descriptions to a consistent subset of hand movements.
    - This subset excludes fidgeting, self-touching, etc.

Sidebar: Is all gesture cultural?

- This is the single most common thing people say about my research.
  - "What about <insert ethnic group>?" (this is almost always Italians)
- Things that are cultural
  - Rate of gesture
    - But not as much as you might think -- Italians think about gestures more, and so are more cognizant of them than the English, but don't actually gesture that much more (Kendon)
  - Emblems
  - Handshapes
    - Kendon notes that using the thumb to point backwards or to the side is normal, but pointing directly at someone is inappropriate and nearly obscene.
- Things that are not cultural
  - Temporal synchrony with speech
  - Use of similarity / differences to explain semantics
  - …

Four Questions about Gesture

- What is gesture?
- How is gesture linked to speech?
- (How) do gestures communicate?
- What can we (engineers) do with gestures?

Gestures and Speech

- Why think of gestures as part of language?
- Some links between gestures and speech
  - Temporal synchrony
  - Mental representation
  - Gesture is produced by the congenitally blind, while speaking to the blind
  - Aphasia
  - Gestural “motherese” (adults mimic infants gesturing patterns)

Temporal Synchrony of Gestures and Speech

- Stuttering (Mayberry & Jaques 2000)
  - Stuttering never begins when a gesture is beginning
  - If stuttering begins during a gesture, the gesture stops immediately
- Gesture and speech remain synchronized despite delayed audio feedback (McNeill 2004)
- The fastest part of the gesture correlates very tightly with the highest pitched part of the speech (Kendon 1999; Kettebekov, Yeasin, and Sharma 2005)

Mental Representation

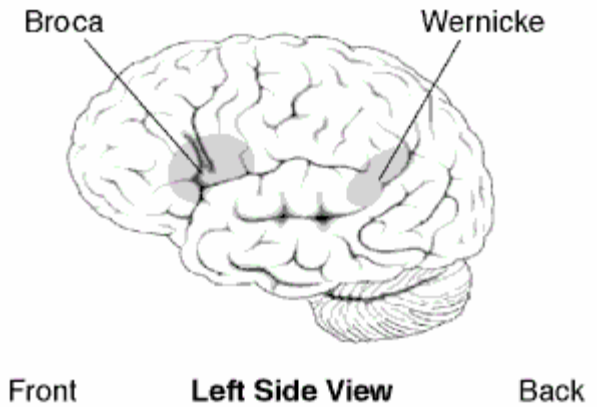
- We'll discuss this later, but several studies have shown that listeners do not remember whether information was communicated through gesture or speech
- This suggests that such information is stored at a higher level of representation, encompassing different linguistic modalities

Gesture and Speech: Neuropathology

- I.W. lost sense of touch and proprioception (movement and position sense at 19, due to an infection
  - Not paralyzed, but because of the loss of motor feedback, I.W. was unable to walk, write, feed himself
  - After 30 years, he has taught himself to do this using only visual feedback
  - Without vision, he cannot perform instrumental actions
  - Yet he gestures very naturally
    - Timing, shape (morphokinesis), observer viewpoint unaffected
    - Relative locations of hands (topokinesis) and character viewpoint damaged
      - When asked to indicate the shape of a triangle, his hands end up misaligned
- I.W.'s ability to gesture suggests a dedicated gesture-language mental structure, independent of instrumental control

Asphasia

- Broca's area
  - Damage causes difficulty in producing grammatically correct sentences
  - Sufferers use gesture successfully as a crutch to express themselves
- Wernicke's area
  - Damage impairs semantic processing
    - Speech is often grammatical but meaningless
    - Speech processing is impaired
  - Recovered asphasics report that they found speech unintelligible, and were not able to control or understand their own words
  - Gestures recur with speech, but without apparent meaning



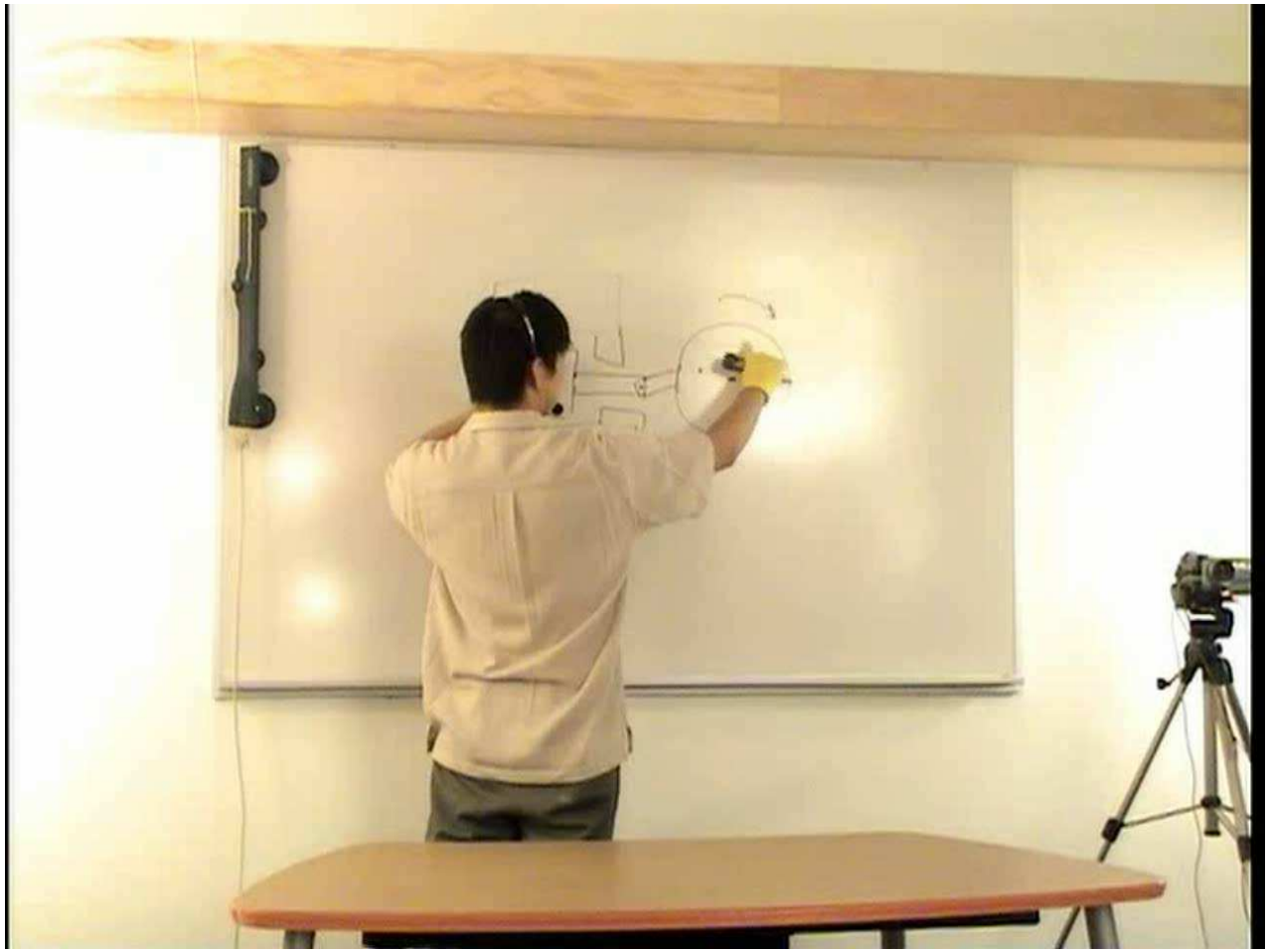
Four Questions about Gesture

- What is gesture?
- How is gesture linked to speech?
- (How) do gestures communicate?
- What can we (engineers) do with gestures?

How do gestures communicate?

- The “catchment” model (McNeill 1992)
  - Gestures provide a system of similarities and contrasts in visual form that supplement speech
  - Similarities in gestural form imply semantic similarities
  - Semantic contrasts are expressed with gestural contrasts
    - Tweety / sylvester example

Catchment Example



Do Gestures Communicate?

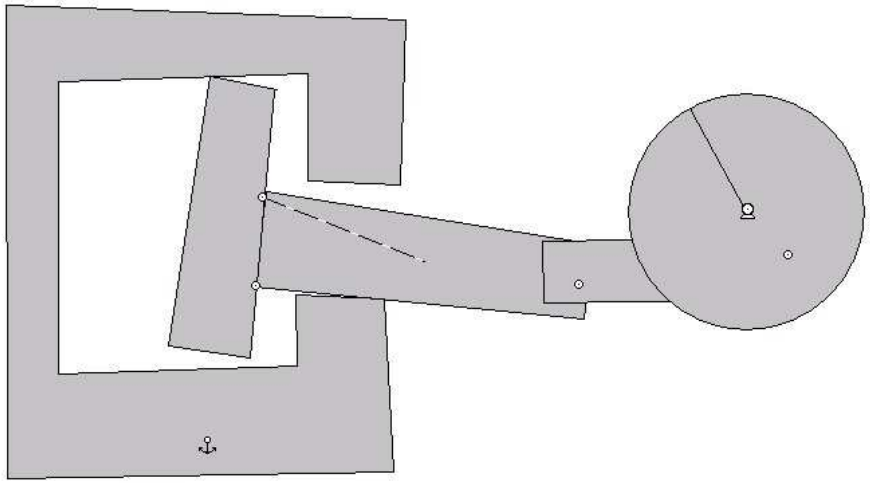
- Speaker benefits
  - Word recall
  - Cognitive Load
- Listener benefits
  - There is less evidence for this than one would think
  - (Some) gestures seem to improve comprehension
  - “Mismatched” gestures impair comprehension

Do Gestures Communicate?

- “This circle is rotating clockwise and this piece of wood is attached at this point and this point but it can rotate. So as the circle rotates, this moves in and out. So this whole thing is just going back and forth.”

Do Gestures Communicate?

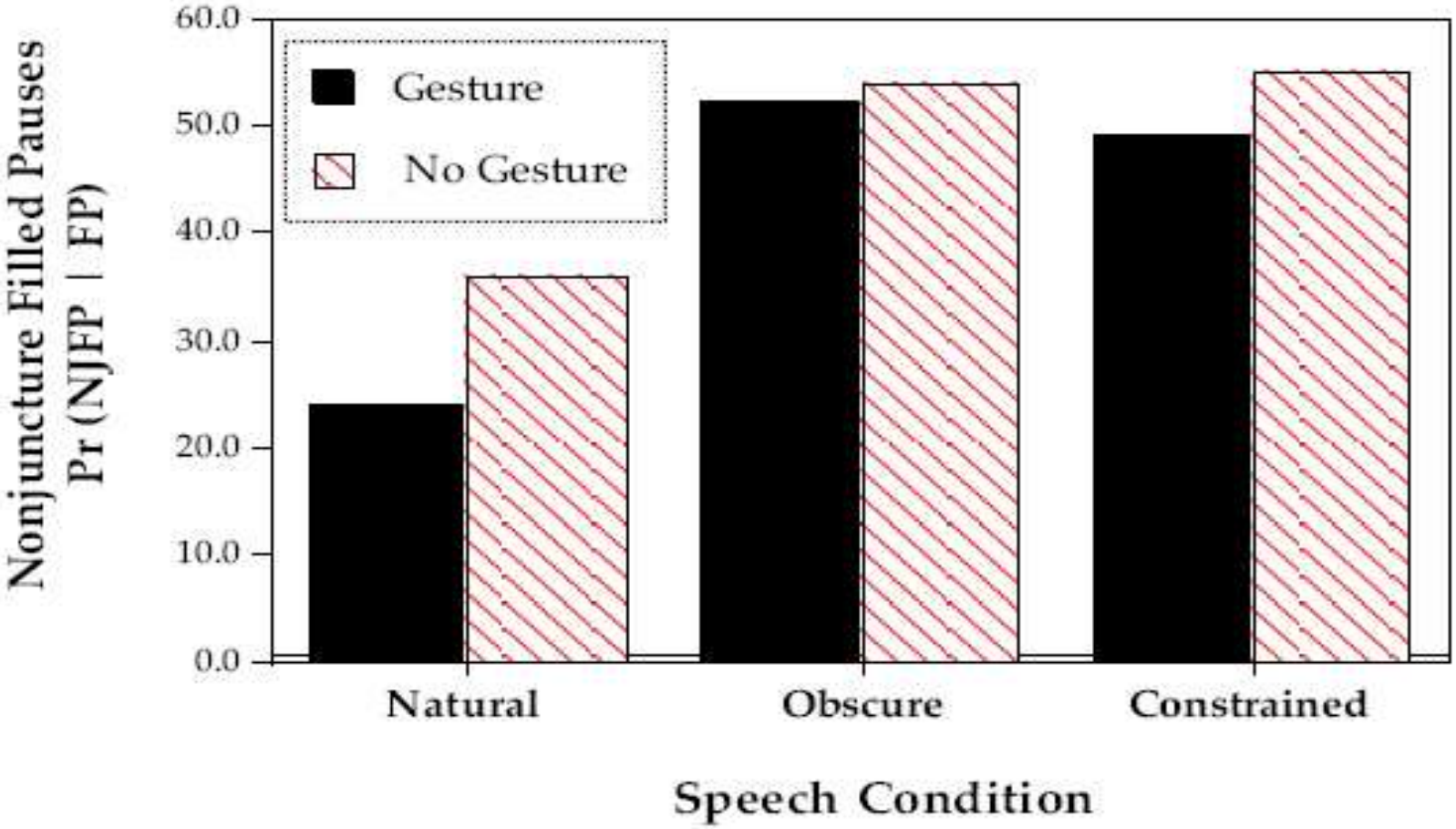
- “This circle is rotating clockwise and this piece of wood is attached at this point and this point but it can rotate. So as the circle rotates, this moves in and out. So this whole thing is just going back and forth.”



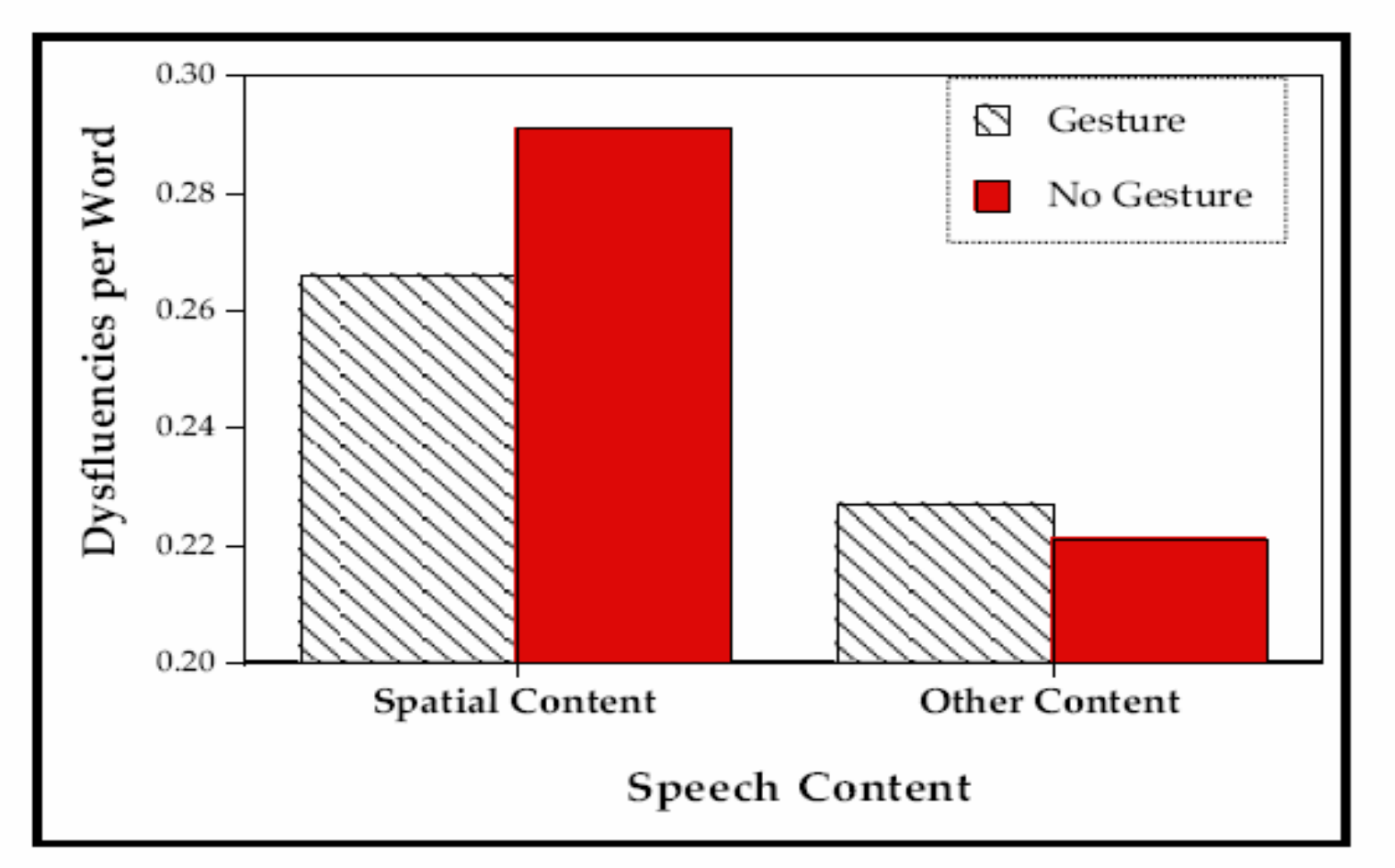
Gesture aids speaker's word recall

- Robert Krauss argues that the primary function of gesture is to enhance word recall.
- Evidence:
  - Gestures reduce filled pauses, and disfluency
  - They do so more for spatial terms than for non-spatial ones
- Setup
  - Speakers were videotaped defining 20 common English words
  - This dataset was collected by a high school student, Flora Fan Zhang
  - Conditions
    - "Natural", "Constrained" (no C), "Use obscure words whenever possible"
    - Gesture permitted, not permitted

Gestures Reduce Filled Pauses



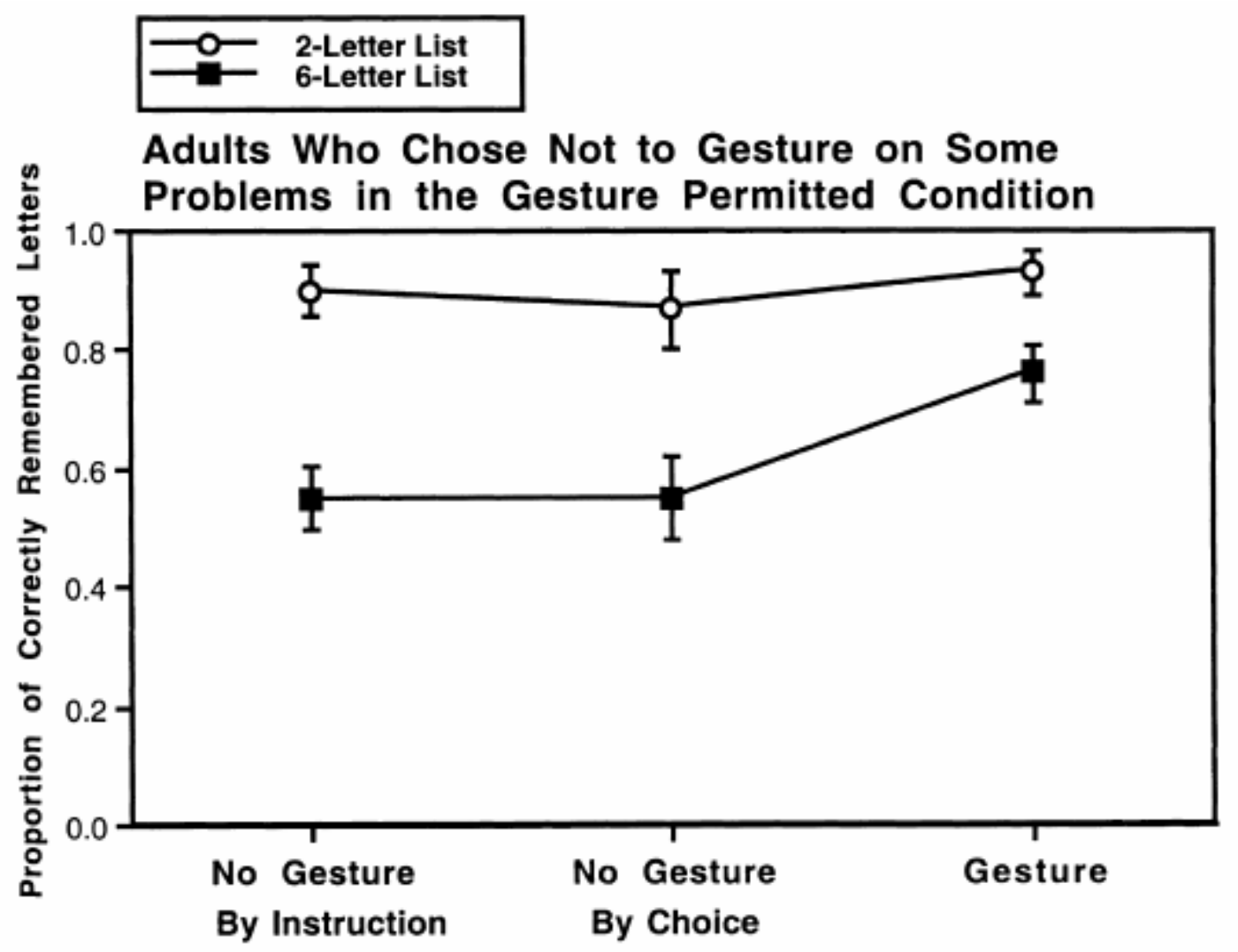
Gestures Reduce Dysfluency



Gesture reduces speaker's cognitive load

- Goldin-Meadow et al. (2001)
- Gestures require motor planning and execution
  - So gesturing should increase mental load?
- Task: explain arithmetic while remembering a list of words
- Participants remember more words when they gesture

Results from Goldin-Meadow et al. (2001)



Semantics

- Gesture improves language comprehension -- Kelly et al. (1999)
- Tested peoples' comprehension of scenarios in which meaning was deliberately communicated using gesture and speech
  - Professional actors in the videos
- Experiment 1: Deictic gestures are used to indicate a reference
  - Example: "Go get the burgers" + gesture pointing at a bike
- Experiment 2: Iconic gestures are used to describe an event
  - Example "I told my friend about the party" + gesture indicating phone next to ear

Results from Kelly et al. (1999)

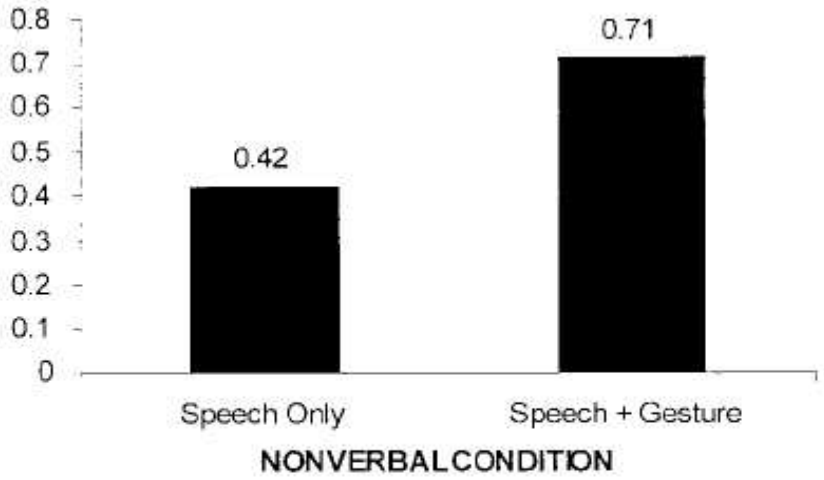


FIG. 1. Intended Action responses by nonverbal condition, Experiment 1.

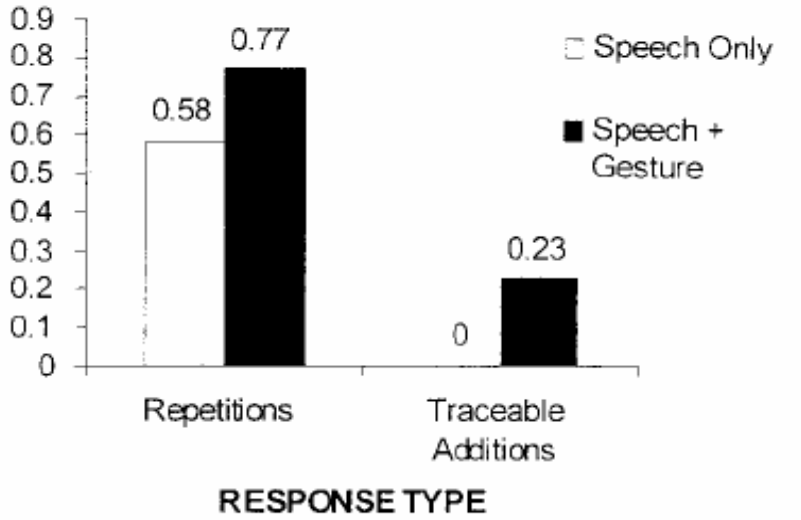


FIG. 4. Repetitions and Traceable Additions by nonverbal condition, Experiment 4.

Listeners failed to remember whether the information was communicated via gesture or speech

Gestures in Education

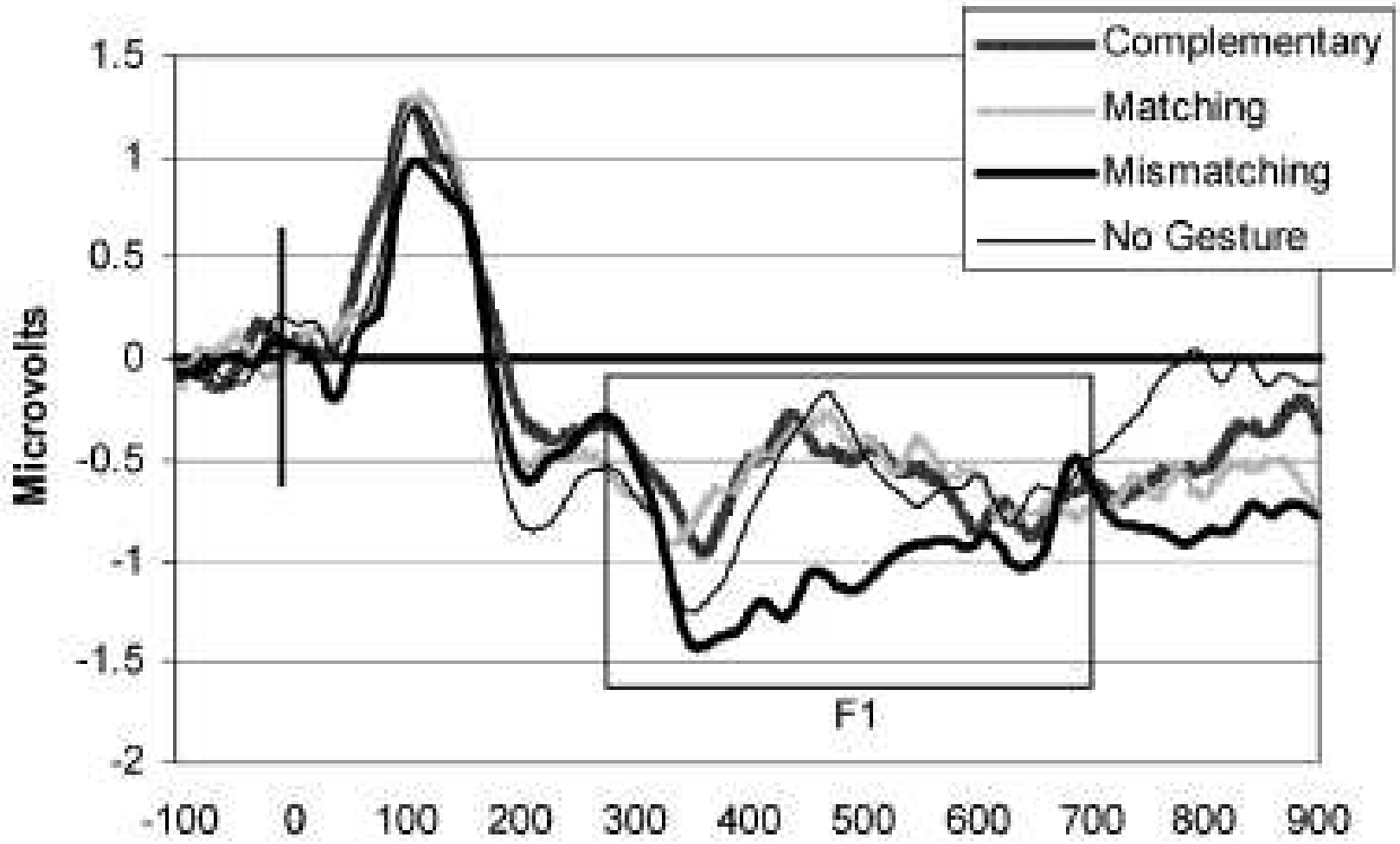
- 4^{th} grade teachers explain solutions to arithmatic problems such as: 3 + 4 + 5 = __ + 5
- When teachers used gestures that matched their verbal explanations, performance was best
- When gestures mismatched verbal explanations, performance was worse than with no gestures at all.

Goldin-Meadow, Kim, and Singer, 1999

Bad Gestures: EEG Study

- Kelly et al., 2004
- Electroencephalograph (EEG) measures electrical signal of brain activity at high temporal resolution
- N400 is a signal indicating surprise or confusion, e.g., “The man likes cream and sugar in his socks.” Typically this signal is right lateralized
- When gestures are semantically incompatible with speech, the same signal is observed.
- Keep this in mind when you attend Patrick Winston’s lecture on public speaking.

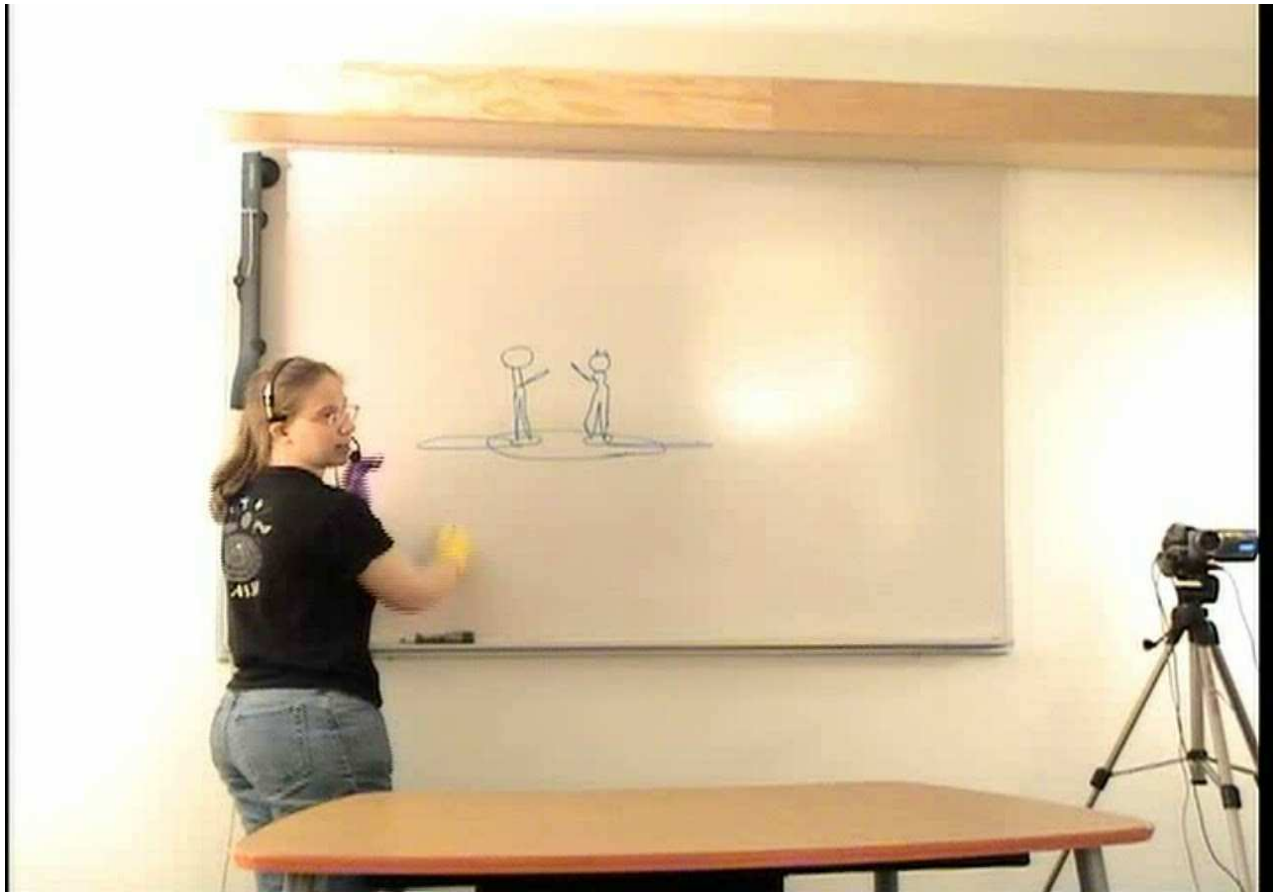
EEG Data from Kelly et al. (2004)



Four Questions about Gesture

- What is gesture?
- How is gesture linked to speech?
- (How) do gestures communicate?
- What can we (engineers) do with gestures?

Why “recognition” is not the right approach for gesticulation



Why understanding gesture is not like understanding speech

- In speech, meaning is built "bottom-up"
  - Morphemes à words à phrases à sentences à paragraphs à …
    - "ed" à gestured à he gestured à he gestured all the time à …
- Gestures are atomic
  - They have a beginning, middle, and end, but these parts are meaningless on their own

Gesture: some applications

- Who's talking?
  - Since movement correlates with tightly with pitch, you could ID a speaker in a room by picking the one whose gestures line up with the audio.
  - This has been done using lips (e.g. Siracusa et al 2003) – what about gestures?
- What are they saying?
  - Identify gestural similarities, use these to inform your semantics.
  - Some possible similarity features: position, trajectory, handshape…
  - Applications: coreference (Eisenstein 2007) & topic segmentation
- Are they done yet?
  - Turn, sentence, and topic boundaries rarely co-occur with gestural strokes
  - Incorporate stroke detection into a linguistic model for detecting turn, sentence, and topic boundaries (Chen 2006)
- When should I look?
  - Generate keyframe summaries of vides by identifying meaningful gestures (Eisenstein 2007)

Potential Future Work

- Is this important?
  - Are there beats? These may indicate important keywords.
- Was that a mistake?
  - Spoken restarts are accompanied by gestural restarts

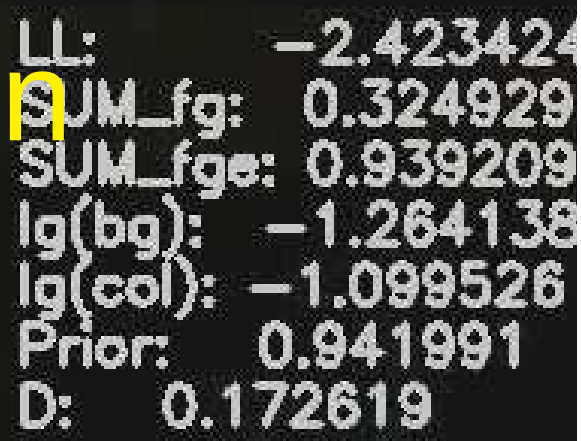
Gesture and Coreference

- Coreference is when two noun phrases mean the same thing
- If gestures are similar, coreference should be more likely
- If gesture conveys non-redundant information, it should improve coreference resolution when added as a feature

Dataset

- Goal: natural, but tractable
- Participants explain the behavior of mechanical devices
- Available modalities were varied:
  - ***Printed diagram***
  - Whiteboard
  - No visual aids

Pose estimation



- Articulated upper body model
- Align model with foreground, hands with expected color blobs
- Search model space with particle filter
- Tracks non-occluded hand correctly ~90% of the time
  - Struggles with occlusion, depth cues

Gesture and coreference

- Noun phrase coreference: do two noun phrases refer to the same thing?
- Two observations from psychology
  - Gestural similarity à semantic similarity
  - Meaningful gesture is more common when references are ambiguous
- Goal: learn coreference and gesture salience jointly
- Without labels for gesture salience

Hidden-variable conditional model

$$p(y|\mathbf{x};\mathbf{w}) = \sum_m p(y,m|\mathbf{x};\mathbf{w}) = \frac{\sum_m exp(\psi(y,m,\mathbf{x};\mathbf{w}))}{\sum_{y',m} exp(\psi(y',m,\mathbf{x};\mathbf{w}))}.$$

- Psi is a potential function on the features, class labels, and the hidden variable
- P(m|x) quantifies gesture salience

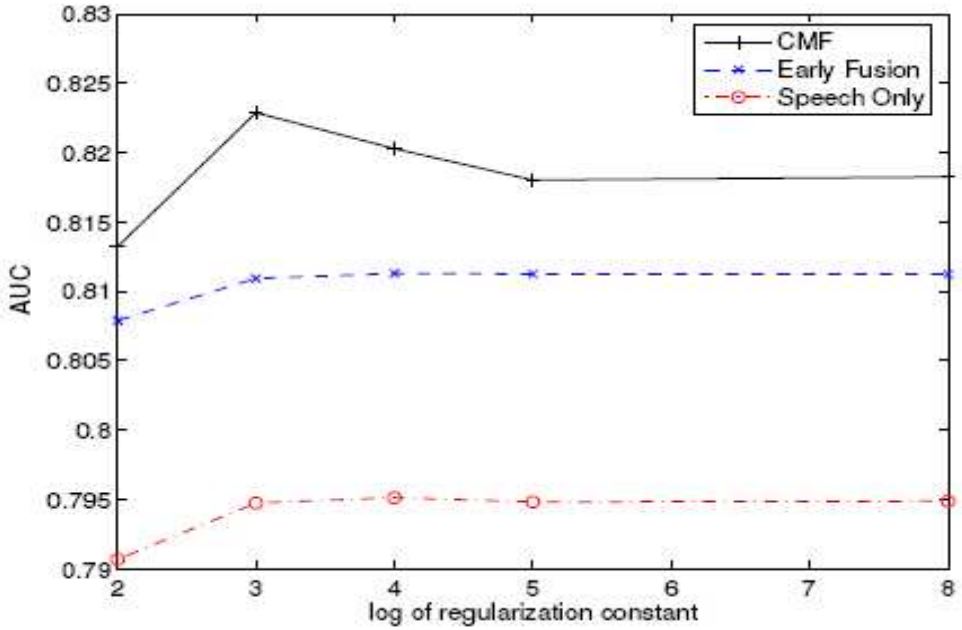
$$\psi(y,m,\mathbf{x};\mathbf{w}) \equiv \begin{cases} y(\mathbf{w}_{v,1}^T\mathbf{x}_v + \mathbf{w}_{nv}^T\mathbf{x}_{nv}) + \mathbf{w}_m^T\mathbf{x}_m & m=1 \\ y\mathbf{w}_{v,2}^T\mathbf{x}_v - \mathbf{w}_m^T\mathbf{x}_m & m=-1 \end{cases} \quad (3)$$

Implementation

- Gradient-based optimization, with L2-norm regularization
- Features
  - Linguistic similarity: edit distance, string match, number agreement…
  - Gestural similarity: Euclidean distance, dynamic time warping
  - Gesture salience features: NP-type, count, hand speed, distance to rest position

Evaluation

- Corpus:
  - 16 short videos
  - 1141 noun phrases
- Effectiveness of gesture features increases by 73%

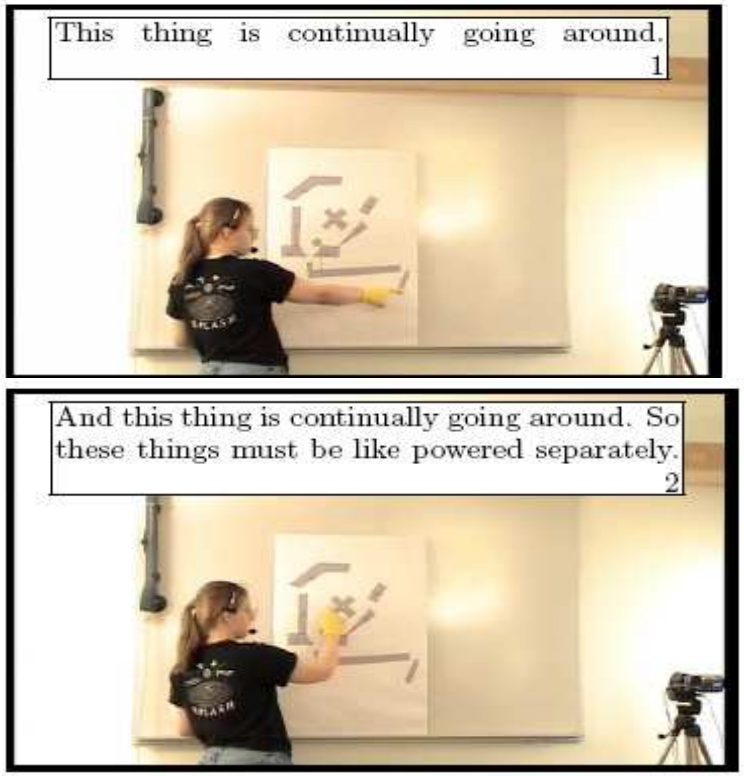


Is this really gesture salience?

- Coreference performance improves
- But does p(m|x) *really* capture gesture salience?
- Application 2: keyframe selection
  - Idea: select keyframes with high p(m|x)
  - Are these the same keyframes human raters would select?

Keyframe selection results

- Using only labeled data for coreference, we learn gesture salience as a hidden variable
- Outperforms standard image-only and text-only approaches



Model	F-Measure	Recall	Precision
GESTURE-SALIENCE	**.404**	**.383**	**.427**
POSE-CLUSTERING	.290	.290	.290
NP-SALIENCE	.239	.234	.245
RANDOM-KEYFRAME	.120	.119	.121

Appendix: What you need to do empirical research on gestures

- Recognition
  - Fairly high barrier to entry
    - Videos of people gesturing
    - A way to track their hands
    - Transcripts, with time-alignment
  - There's some existing data out there, but it's incomplete (AMI)
- Hand-labeling
  - Time-consuming, although less so than writing a tracker
  - Make sure you fully understand what you're trying to label
- Generation
  - There may some toolkits available for this