



University of Colorado
Boulder

Human-Robot Interaction

Measuring in HRI Research III

Professor **Dan Szafir**

*Computer Science & ATLAS Institute
University of Colorado Boulder*

Subjective Measures (continued)

Question Design

Main challenge is **response bias**

Principled question design can help avoid response bias

Designing Good Questions

Principle

Avoid “leading” or “loaded” questions

Example

“Don’t you agree that social workers should earn more money than they currently earn?”

Yes, they should earn more

No, they should not earn more

Don’t know / No opinion

Designing Good Questions

Principle

Avoid double negatives

Example

“Do you agree or disagree with the following statement?

Teachers should not be required to supervise their students during recess.”

Designing Good Questions

Principle

Always aim at capturing firsthand experiences and beware of asking about information that is acquired only secondhand

Example

People are very good at describing criminal activity directed at them but terrible at describing how much crime happens in their neighborhood

Designing Good Questions

Principle

Beware of asking hypothetical questions – people are not good at predicting what they will do as they have limited direct experience with future situations

Example

“How would you feel crossing the street in a pedestrian walkway in front of a driverless, autonomous car?”

Designing Good Questions

Principle

Beware of asking about causality – events mostly have more than one reason and people are not good at describing why they do the things they do

Example

“Were you limited in your daily activities because of your back problem?”

“What is the main reason why you did not vote?”

“Were you homeless because of the high cost of housing?”

Designing Good Questions

Principle

Beware of asking about solutions to complex problems – people in general do not have informed opinions about complex issues

Example

Issues commonly seen on the news (climate change, gun violence, healthcare, etc.)

Designing Good Questions

Principle

Avoid asking more than one question at a time – the answers to two questions can be dramatically different

Example

“Would you like to be rich and famous?”

“Are you physically able to do things like swim and run without difficulty?”

Designing Good Questions

Principle

Avoid asking questions that impose unwarranted assumptions: double-barreled or one-and-a-half-barreled questions

Example

“Would you vote for or against a candidate who supports reducing federal spending on education and welfare?”

“With the economy the way it is, do you think investing in the stock market is a good idea?”

Figure 4.5. One-and-a-Half-Barreled-Questions Related to the SALT II Treaty.

- The United States is now negotiating a strategic-arms agreement with the Soviet Union in what is known as SALT II. Which one of the following statements is closest to your opinion on these negotiations?
- I strongly support SALT II
 - SALT II is somewhat disappointing, but on balance I have to support it.
 - I would like to see more protection for the United States before I would be ready to support SALT II.
 - I strongly oppose the SALT II arms agreement with the Russians.
 - I don't know enough about the SALT II to have an opinion yet.

Designing Good Questions

Principle

Beware of questions that include hidden contingencies – questions must apply to the majority of your sample

Example

To measure social activity:

“How often did you attend religious services or participate in church-related activities during the past month?”

Designing Good Questions

Principle

The words in questions should be chosen so that all respondents understand their meaning and all respondents have the same sense of what the meaning is

Example

“Would you ever support biblioclasm?”

“How many years have you suffered from acute asthma?”

Designing Good Questions

Principle

When words or terms that have meanings that are likely not to be shared, definitions should be provided to all respondents

Example

In the past 12 months, how many times have you seen or talked with a medical doctor about your health?

If asked: “Include visits to psychiatrists, ophthalmologists, and any other professional with a medical degree.”

Designing Good Questions

Principle

If definitions are provided, they should be given before the question itself is asked

Example

“How many days in the past week have you exercised? When you consider exercise be sure to include walking, work around the house, or work on a job, if you think they constituted exercise.”

Better: “The next question is going to ask you about how often you’ve engaged in exercise. We want you to include walking, work around the house, or work on a job, if you think they constituted exercise. Using this definition, in the last week, on how many days did you exercise?

Designing Good Questions

Principle

The time period referred to by a question should be unambiguous. Questions about feelings or behaviors must refer to a period of time.

Example

“Are you able to run half a mile without stopping?”

“How many drinks do you usually have on days when you drink any alcoholic beverages at all?”

Designing Good Questions

Principle

A question should end with the question itself. If there are response alternatives, they should constitute the final part of the question

Example

“Would you say that you are very likely, fairly likely, or not likely to move out of your house in the next year?”

Better: “In the coming year, which one of these categories best describes how likely you think you are to move out of your house: very likely, fairly likely, or not very likely?”

Designing Good Questions

Principle

Clearly communicate to all respondents the kind of answer that constitutes an adequate answer to a question

Example

“When did you move to this community?”

Possible answers:

When I was sixteen.

Right after I was married

In 1953.

Better: “In what year did you move to this community?”

Designing Good Questions

Principle

Specify the number of responses to be given to questions for which more than one answer is possible

Example

“What was it about the brand you bought that made you buy it rather than some other brand?”

Designing Good Questions

Principle

Design survey instruments to make the tasks of reading questions, following instructions, and recording answers as easy as possible for interviewers and respondents

Example

Break survey into clearly defined sections

Group similar questions

Demographics should generally come last

Exception: if you are screening participants based on some aspect of demographics

Designing Good Questions

Principle

Measurements will be better to the extent that people answering questions are oriented to the task in a consistent way — training respondents

Categorical Questions

Principles

Mutually exclusive response categories

Exhaustive list of categories

Don't know / No answer category

Example

“What is your current age?”

10 or less

10 to 20

20 to 30

30 to 40

Designing Good Questions

Principle

If what is to be covered is too complex to be included in a single question, ask multiple questions

Principle

Use multiple questions to measure the same thing

Enables estimates of reliability

Scale construction -> next lecture!

Designing Questions

What is wrong with these questions?

How can we fix them?

Exercise 1

“How many cups of coffee or tea do you drink in a day?”

Principle: Ask for an answer in only one dimension

Fix: Separate the question into two:

“How many cups of coffee do you drink during a typical day?”

“How many cups of tea do you drink during a typical day?”

Exercise 2

“What brand of computer do you own?”

- IBM PC Apple

Principles: Avoid hidden assumptions. Make sure to accommodate all possible answers

Fix 1: Make each response a separate dichotomous item:

“Do you own an IBM PC?” Yes No

“Do you own an Apple computer?” Yes No

Fix 2: Add necessary response categories and allow for multiple responses:

“What brand of computer do you own?” (Check all that apply)

- Do not own computer IBM PC Apple Other

Exercise 3

“Have you had pain in the last week?”

- Never Seldom Often Very often

Principle: Make sure question and answer options match

Fix: Reword either question or answer to match:

“How often have you had pain in the last week?”

- Never Seldom Often Very often

Exercise 4

“Where did you grow up?”

- Country Farm City

Principle: Avoid questions having non-mutually exclusive answers

Fix: Design the question with mutually exclusive options:

“Where did you grow up?”

- House in the country Farm in the country City

Exercise 5

“Are you against drug abuse?”

- Yes No

Principle: Write questions that will produce variability in the responses

Fix: Eliminate the question

Exercise 6

“Which one of the following do you think increases a person’s chance of having a heart attack the most?”

- Smoking Being overweight Stress

Principles: Encourage the respondent to consider each possible response to avoid the uncertainty of whether a missing item may represent either an answer that does not apply or an overlooked item.

Fix: Which of the following do you think increases the chance of having a heart attack?

Smoking: Yes No Don’t know

Being overweight: Yes No Don’t know

Stress: Yes No Don’t know

Exercise 7

“Do you currently have a life insurance policy?”

- Yes No – *If no, skip the next question*

Principle: Avoid branching as much as possible to avoid confusing respondents.

Fix: If possible, write as one question

“How much did you spend last year for life insurance? (Write ‘0’ if none)”

Questions?

Types of Answers

Open Questions

Advantages

- More free response – less bias due to assumptions of investigator
- Can lead to entirely new interpretations / unanticipated answers
- Allows study of respondent thought process

Disadvantages

- Often requires coding afterwards (e.g., content analysis)
- Resource intensive (cost, time, etc.)

Closed Questions

Advantages

Easy and quick to fill in

Easy to code for quantitative analysis

More consistent across respondents

More precisely related to concept of interest

Disadvantages

Respondent might feel constrained or frustrated

Respondents with no answer tend to still give one

Ordering of Questions

From general to particular

From easy to difficult

From factual to abstract

Start with closed format

Start with questions relevant to the main subject

Do not start with demographic and personal questions

Sensitive Questions

Affects nonresponse

National Survey of Families
(females)

Total household income* 8.15%

Lifetime male sexual partners†
3.05%

Ever received public assistance*
2.22%

Highest grade completed† 0.04%

Affects answers

10th graders' smoking past month:

2 surveys at school = 29.8%,
35.3%

1 survey at home = 23.4%

Marijuana past month:

2 surveys at school = 20.5%, 25%

1 survey at home = 12.5%

*ACASI = Audio Computer-Assisted Self Interview

†CAPI = Computer-Assisted Personal Interview

Standardizing Responses

Rating scales

Allows respondents to specify their level of agreement to a statement

Mostly used as 5-, 7-, and 9-point scales

“How well do you think that Ralphie the Buffalo represents CU Boulder?”

Very poorly 1 2 3 4 5 6 7 Very well

Can also be used in forced-choice tests

2- or 4-point scales

“Neither agree nor disagree” not provided

Likert Scales

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I feel good about my work on the job.					
On the whole, I get along well with others at work.					
I am proud of my ability to cope with difficulties at work.					
When I feel uncomfortable at work, I know how to handle it.					
I can tell that other people at work are glad to have me there.					
I know I'll be able to cope with work for as long as I want.					
I am proud of my relationship with my supervisor at work.					
I am confident that I can handle my job without constant assistance.					
I feel like I make a useful contribution at work.					
I can tell that my coworkers respect me.					

Semantic Differential Scales

“This part consists of a number of words that describe different characteristics of your partner. Read each item and mark the appropriate rating for it. Indicate to what extent you feel this way right now, that is, at the present moment.”

Sociable		2	3	4	5	Unsociable
Kind		2	3	4	5	Cruel
* Hard		2	3	4	5	Soft
Successful		2	3	4	5	Unsuccessful
* Foolish		2	3	4	5	Wise
Strong		2	3	4	5	Weak

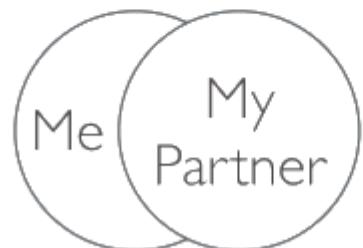
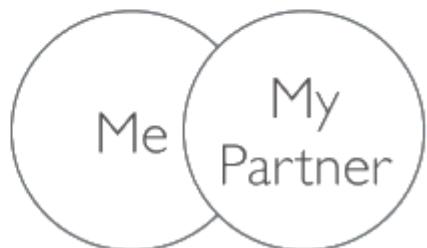
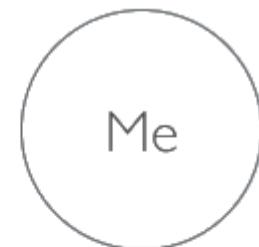
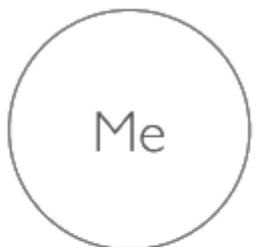
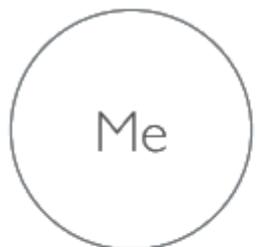
Rating Scales

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers:

1 very slightly or not at all	2 a little	3 moderately	4 quite a bit	5 extremely
interested			afraid	
distressed			irritable	
excited			alert	
upset			ashamed	
strong			inspired	
guilty			nervous	
scared			determined	
hostile			attentive	
enthusiastic			jittery	
proud			active	

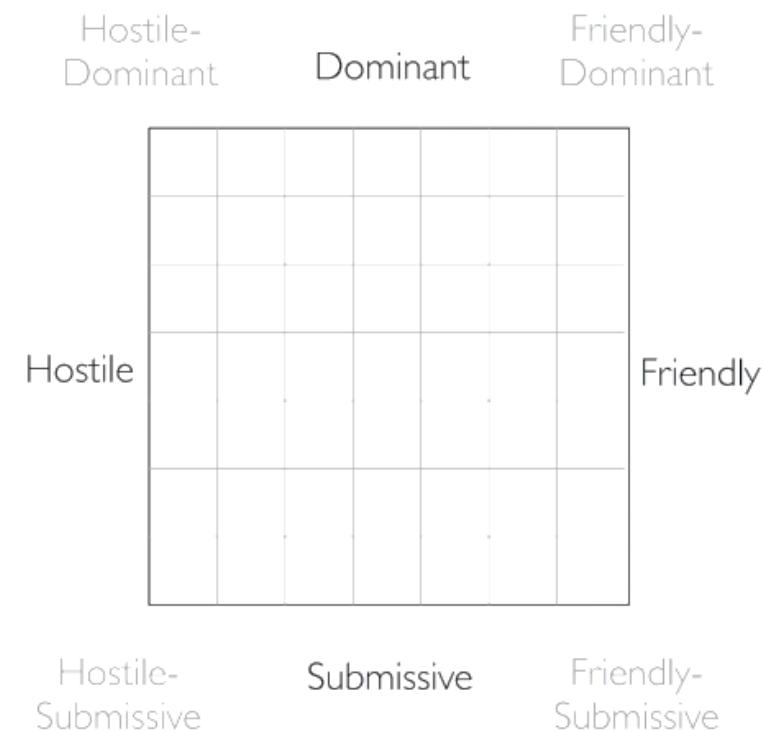
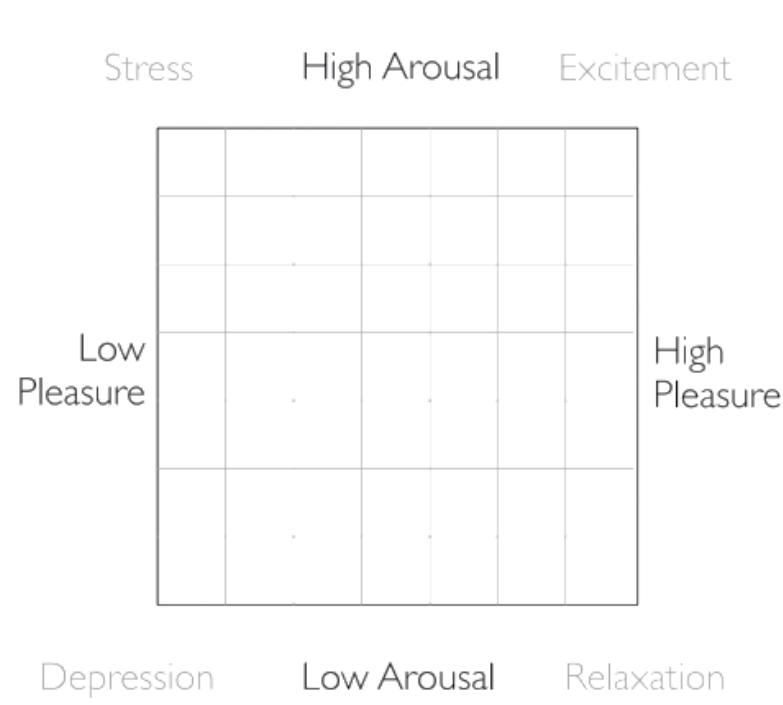
Rating Scales

Please circle the picture below which best describes your relationship with your partner:



Rating Scales

Below are two 2-dimensional grids representing a scale of how you perceived your partner's attitude. Please mark where on the grid you think closest represents your partner's attitude.



Online Surveys

Questions to ask yourself

Do all people in your population have access to the online survey?

Can all people in your population use an online survey with similar proficiency?

What are the risks if the information is intercepted?

Are you providing an incentive?

How will you implement quality control measures?

Common Issues

Question complexity increases errors and nonresponse

Navigation complexity increases errors and nonresponse

Small field for open-ended questions reduces words

No follow-up for nonresponse increases nonresponse

Payment amount has little impact but equalizes groups

Analyzing Ordinal Rating (“Likert”) Items

Can ordinal data, converted to numbers, be treated as interval?

i.e., can means, standard deviations, and parametric statistics be used to analyze Likert items?

Controversial; statisticians have argued for and against

Primer: Jamieson, Susan. "Likert scales: how to (ab) use them." *Medical education* 38.12 (2004): 1217-1218.

Most recent evidence provides support that parametric tests may be used with ordinal data (and can even be used when assumptions such as normal distribution of data are violated)

See: Norman, Geoff. "Likert scales, levels of measurement and the “laws” of statistics." *Advances in health sciences education* 15.5 (2010): 625-632.

However, not all reviewers agree

Analyzing Ordinal Rating (“Likert”) Items

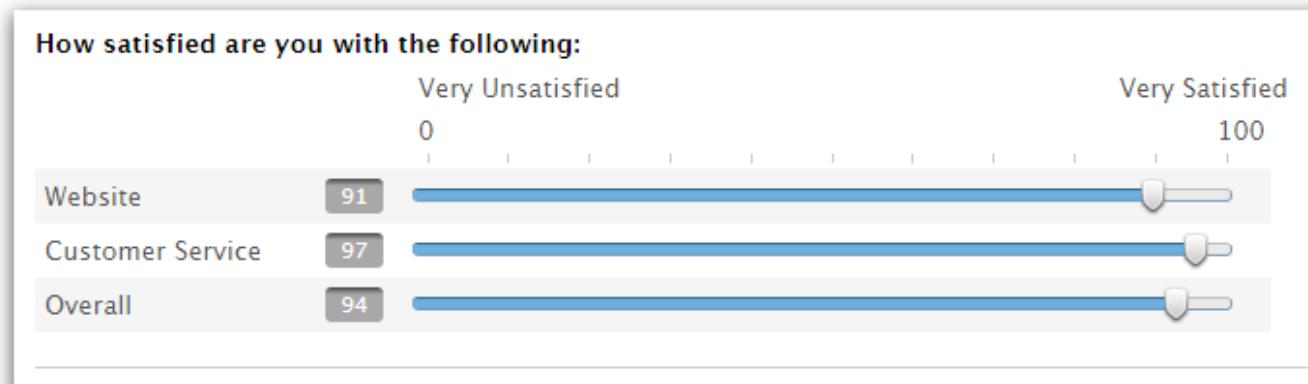
Recommendation: if analyzing a single item, make it continuous (or use nonparametric analysis)

This is also still a matter of controversy. See:

Kero, P., & Lee, D. (2015). Slider scales and web-based surveys: A cautionary note. *Journal of Research Practice*, 11(1), Article V1. Retrieved from <http://jrp.icaap.org/index.php/jrp/article/view/513/414>

Roster, C. A., Lucianetti, L., & Albaum, G. (2015). [Exploring slider vs. categorical response formats in web-based surveys](#). *Journal of Research Practice*, 11(1), Article D1.

Better: use multiple items to form a scale using **scale construction**



Questions?

Physiological Measures

Common Physiological Measures

Eye tracking

Pupil dilation

Electrodermal Activity (EDA) / Galvanic skin response (GSR)

Muscle activity – Electromyogram (EMG)

Brain activity

Electroencephalogram (EEG)

Functional Magnetic Resonance Imaging (fMRI)

Functional Near-infrared Imaging (fNIR)

Eye Tracking

Head or desk mounted equipment tracks position of the eye

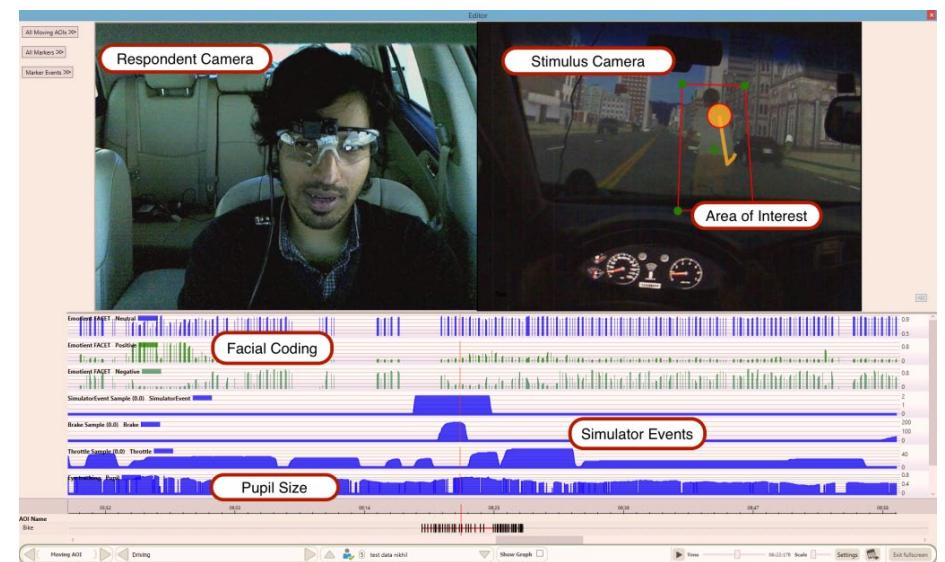
Eye movement might indicate the amount of cognitive processing a display requires

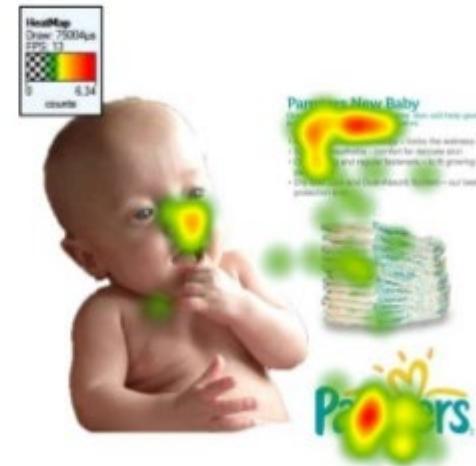
Measurements:

Fixations: Eye maintains stable position. Number and duration may indicate level of difficulty of display or level of interest in subject

Saccades: Rapid eye movement from one point of interest to another

Scan paths: Patterns of eye movements combining saccades and fixations





Pupil Dilation

Pupils dilate in response to:

- Extreme emotional situations (fear, pain, contact with nerves)

- Loads on working memory, increased attention, sensory discriminations, cognitive load

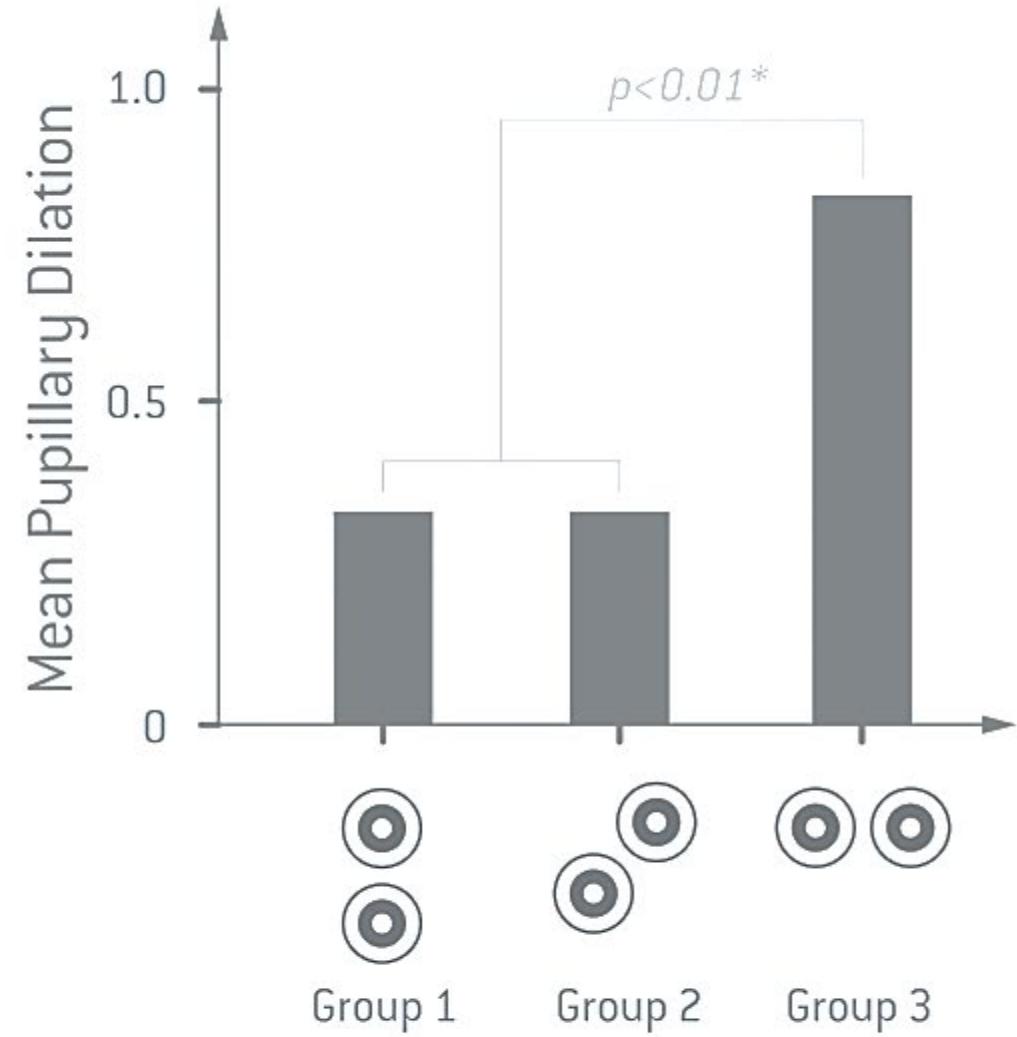
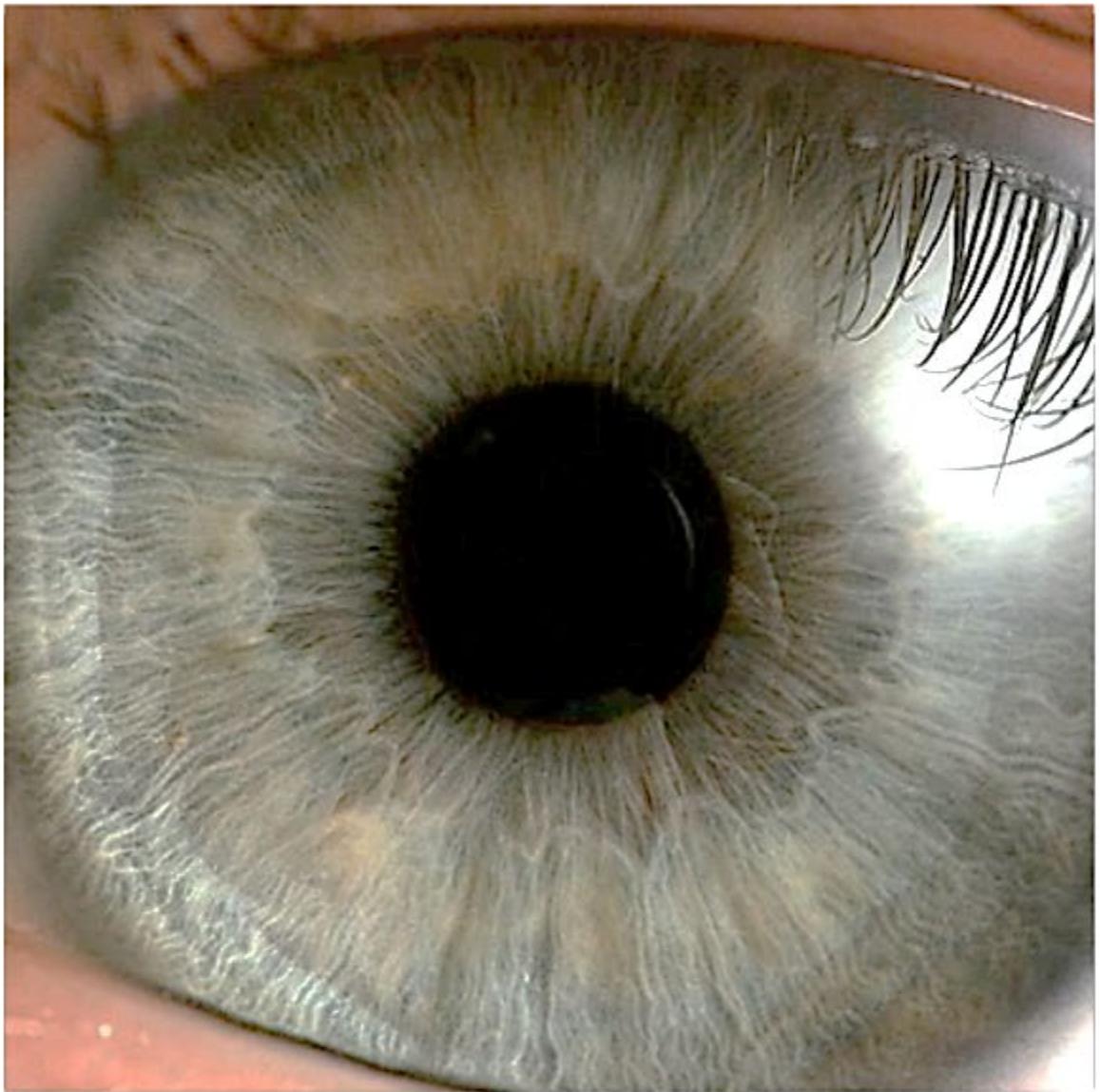
- Changes in ambient light

Intensifies perception of emotional states (e.g., sadness)

Mirror response to other people's pupil sizes

- Might indicate empathy





Pupillary dilation in response to
eye-shaped patterns
(Coss, 1972)

Affect

Emotional responses linked to physical changes

May help determine a user's reaction to a robot

Measurements:

- Heart activity, including blood pressure, volume, and pulse

- EDA

- EEG

- fNIRS

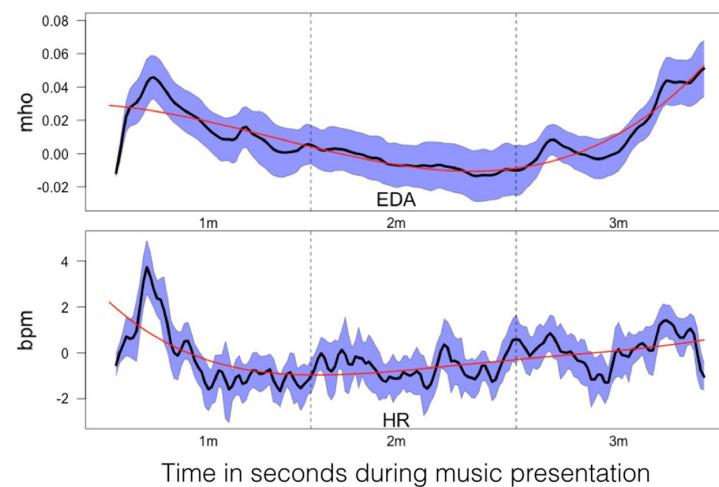
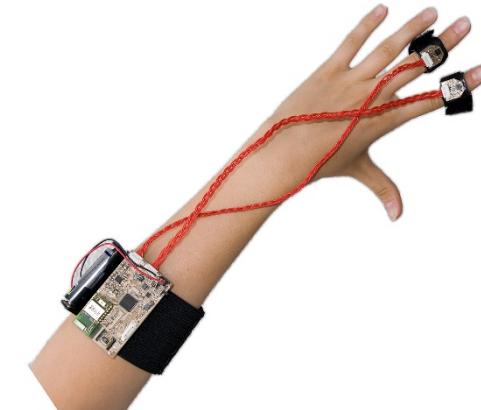
Difficult to interpret some physiological responses

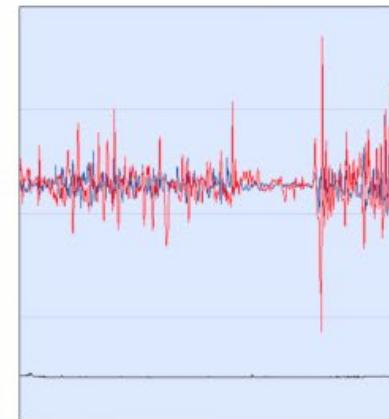
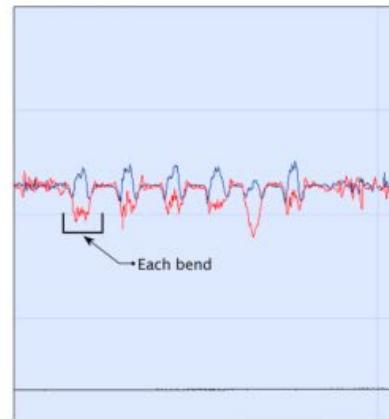
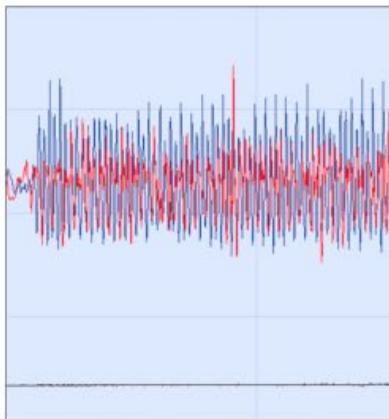
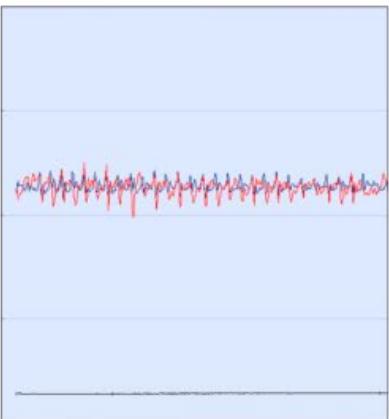
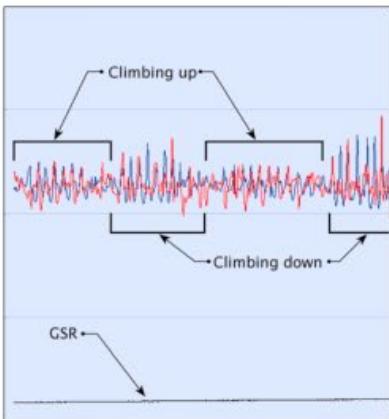
EDA / GSR

Measures variation of electrical properties of the skin in response to sweat secretion

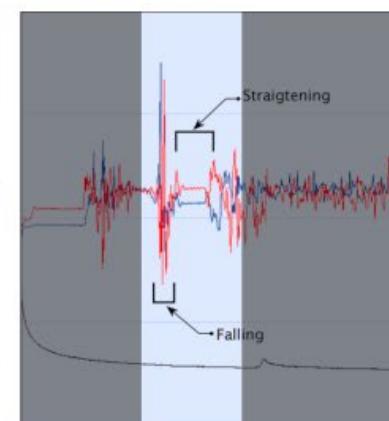
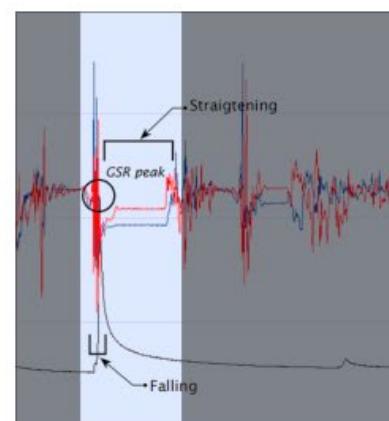
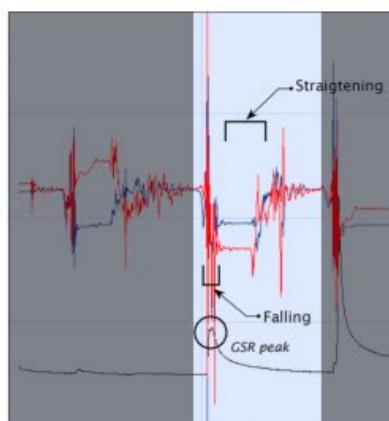
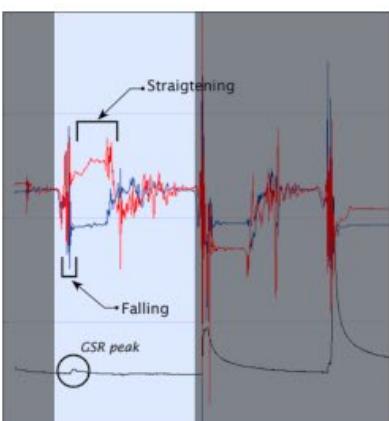
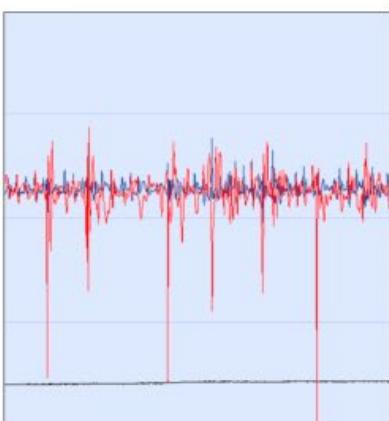
Sensitive to affective stimulation and attention

May measure long-term recall and affect





~ 30 seconds



- Longitudinal acceleration (average)
- Transverse acceleration (average)
- GSR (average)



Brain Activity

Electroencephalography (EEG)

Typically use FFT to identify frequencies (“bands”)

Common bands: alpha, beta, theta, delta, gamma, mu

Can look for signature “event-related potentials”

Example: P300

Hemispheric lateralization

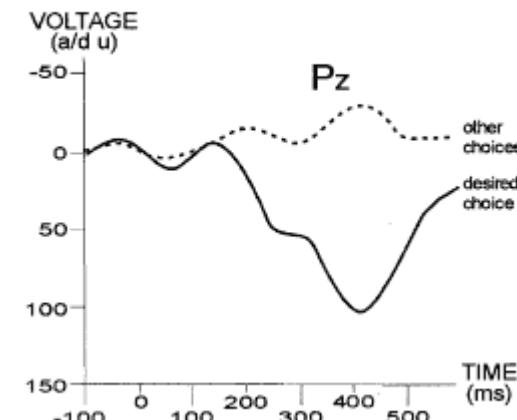
Distinguishes between activity in the:

Left hemisphere of the brain – processes visual stimuli

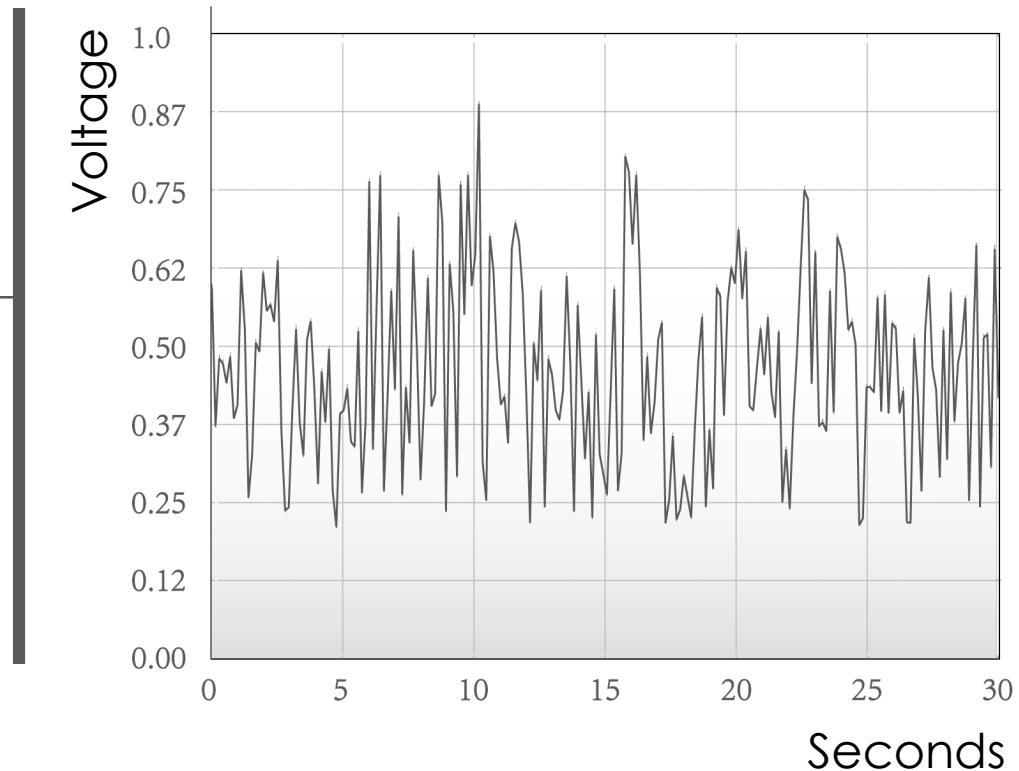
Right hemisphere of the brain – processes verbal stimuli



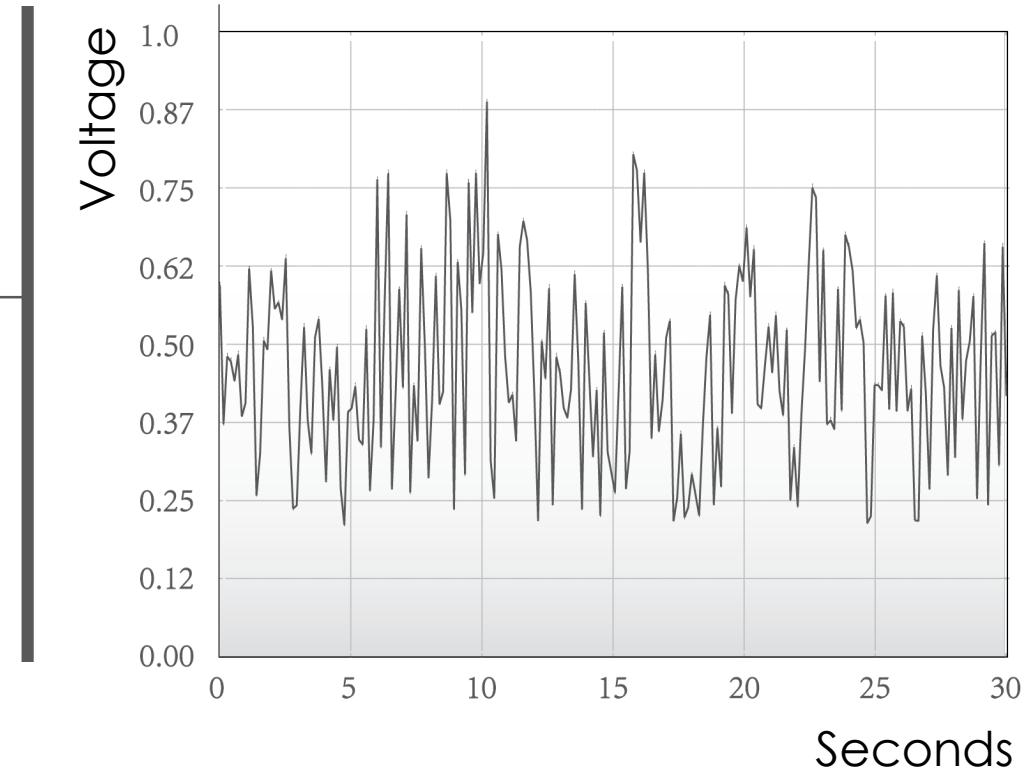
P300 EVOKED POTENTIAL

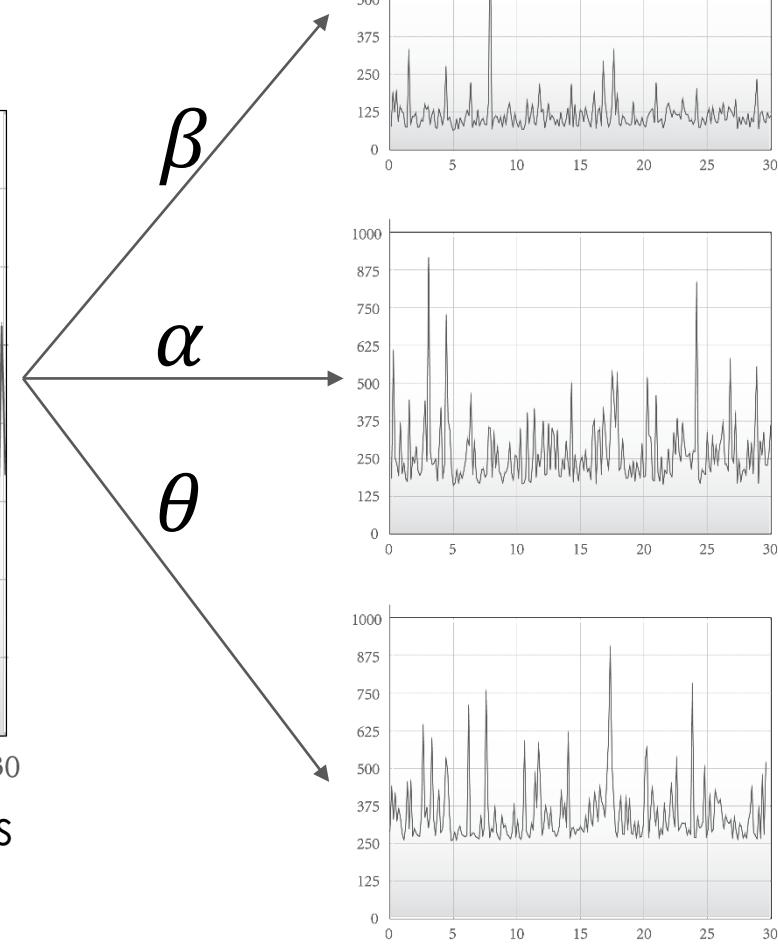
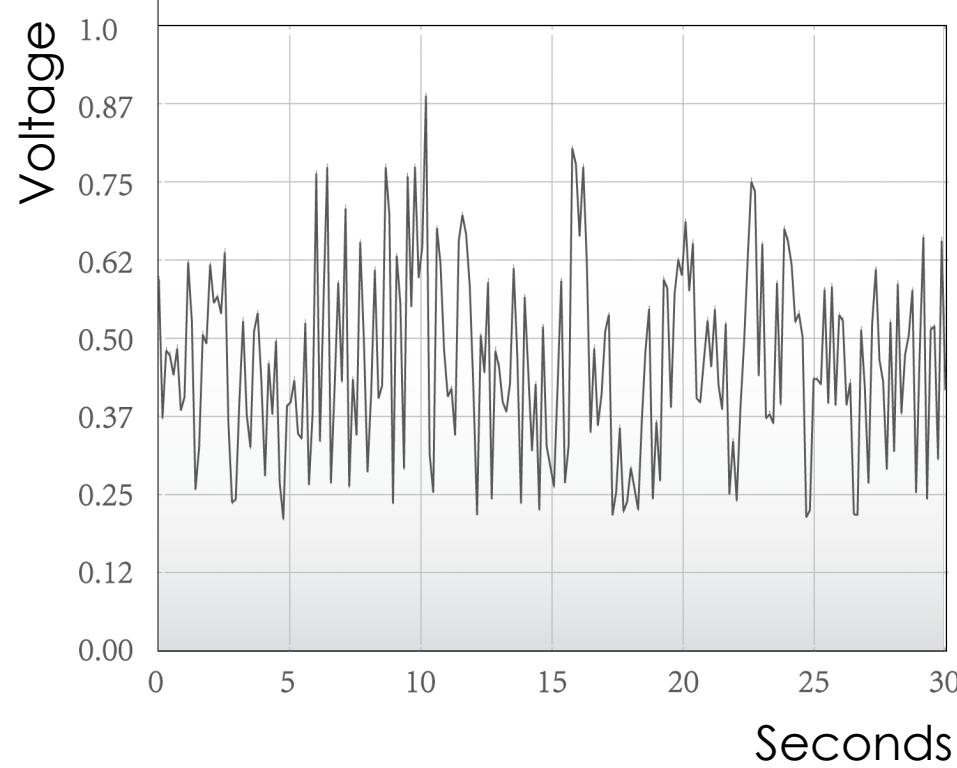


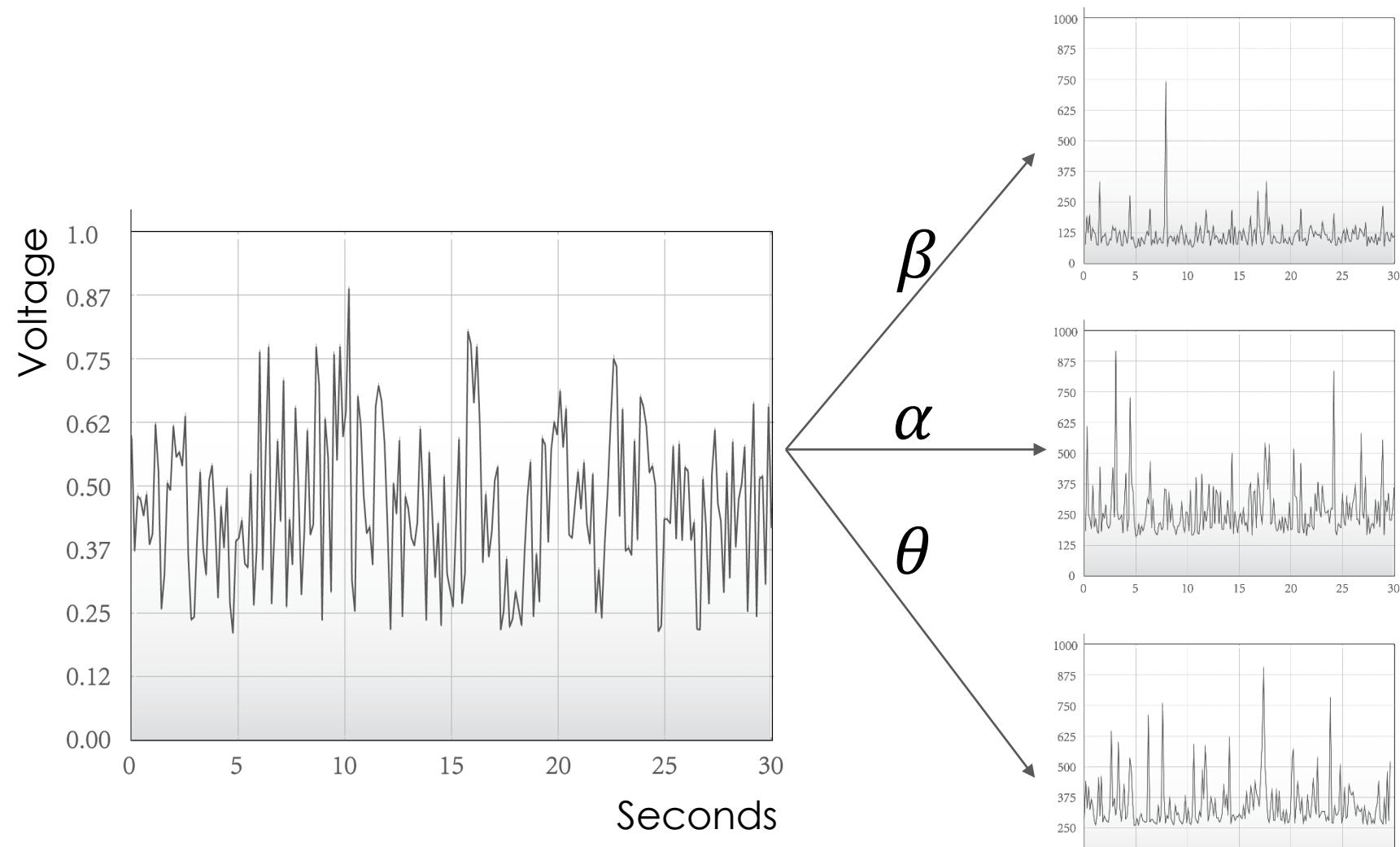
From EEG to Attention

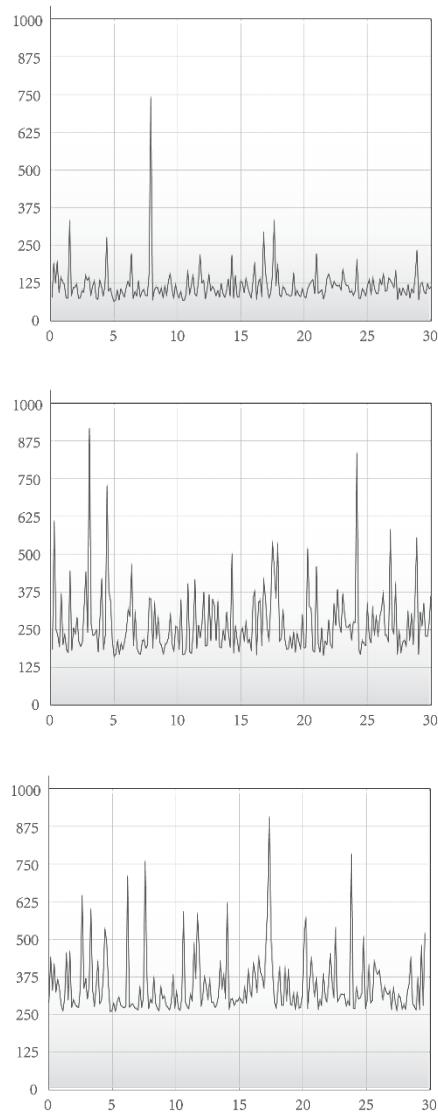


From EEG to Attention





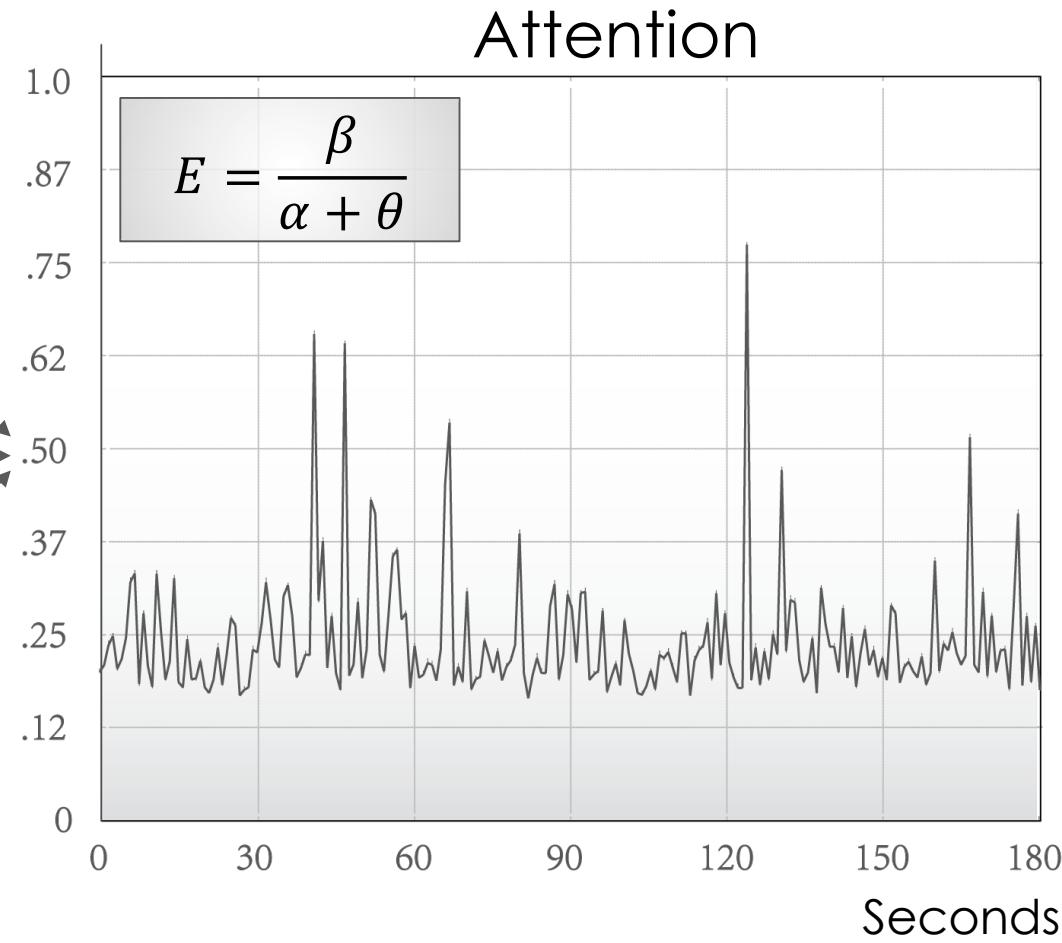




β

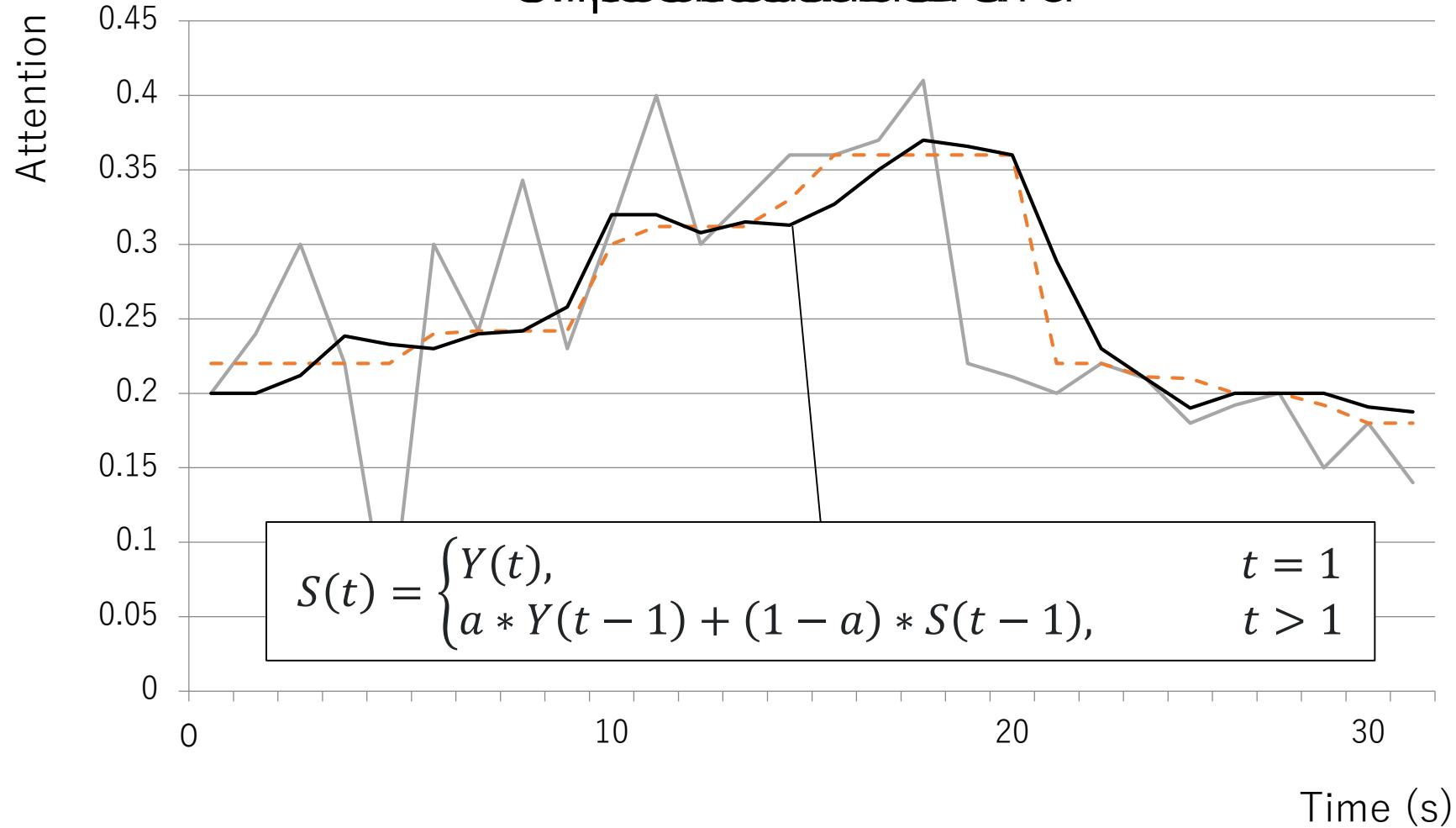
α

θ



Processing

Smoothed Data

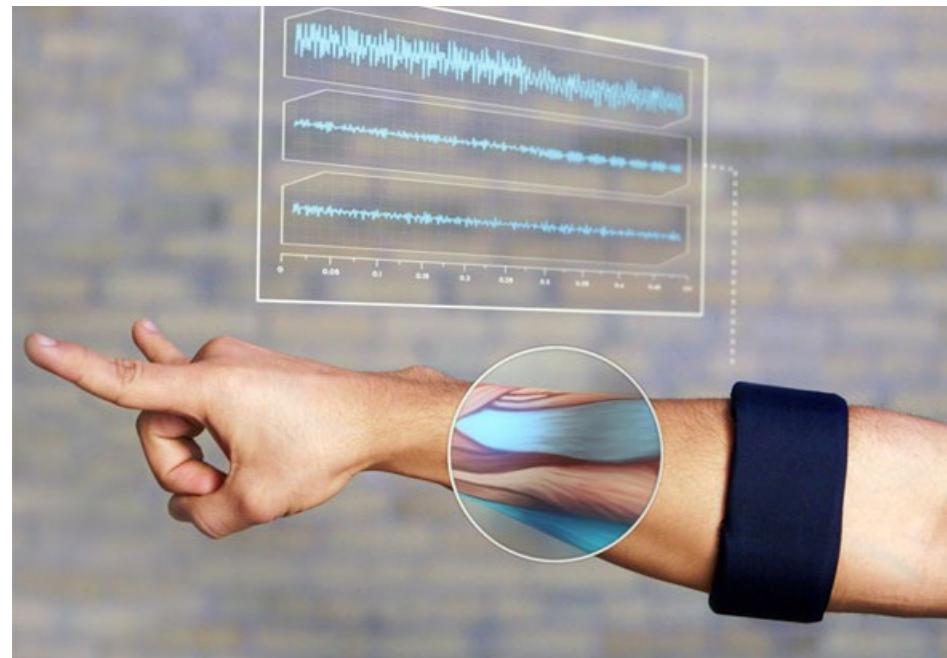


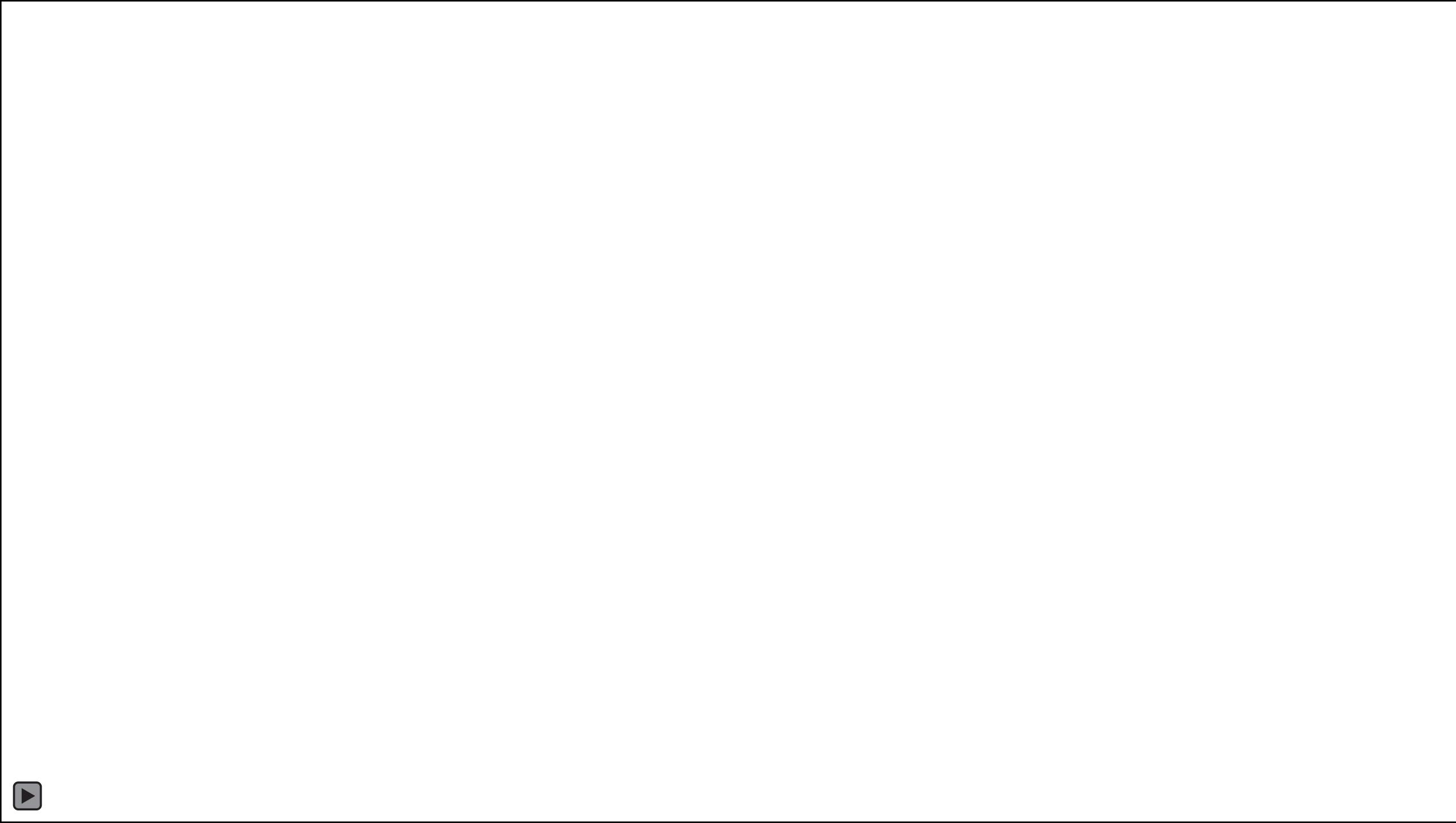
EMG

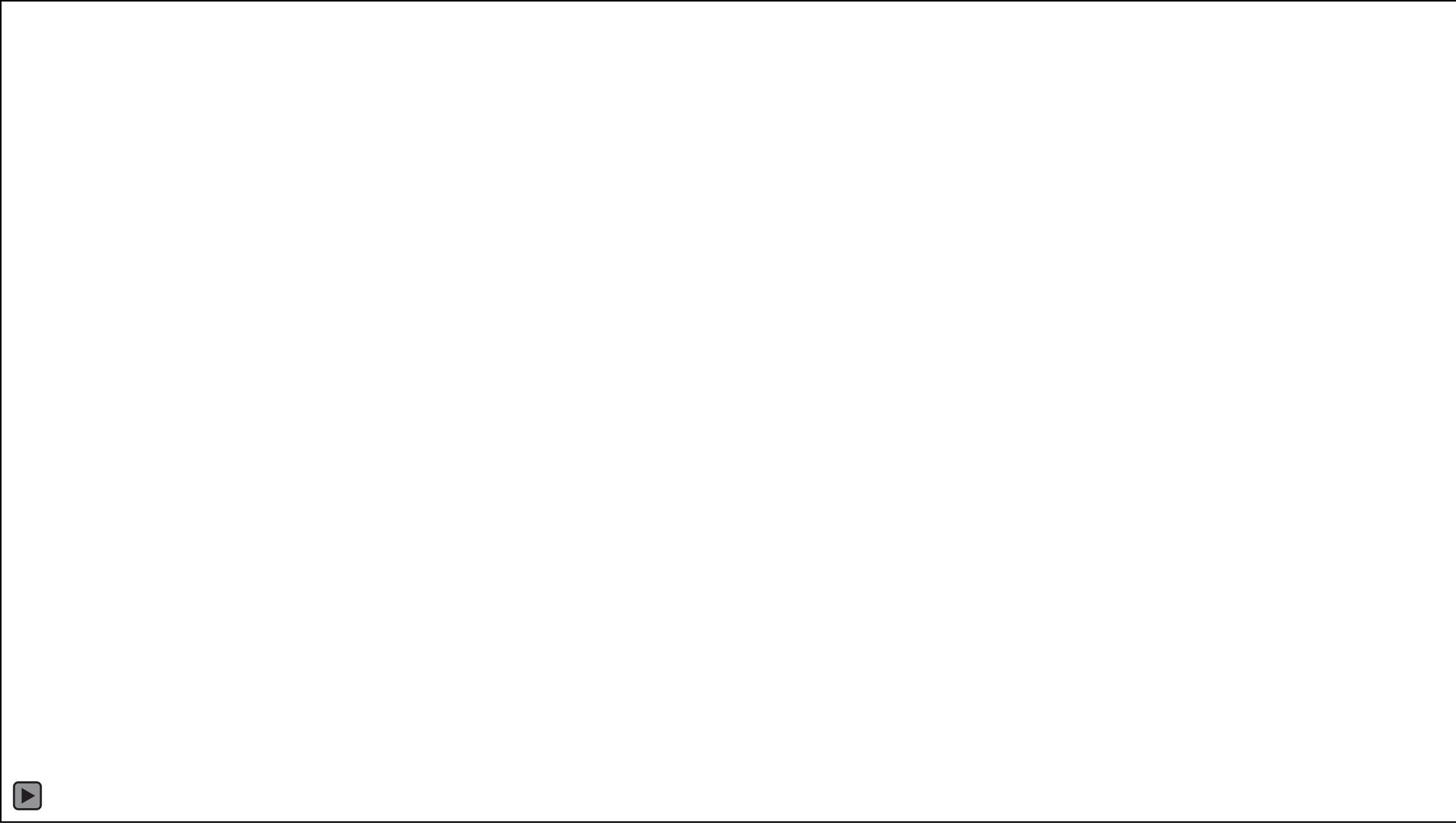
Measures electrical activity in the muscles

Often used to “ground” EEG

Can be used to gauge facial expressions, gestures, etc.







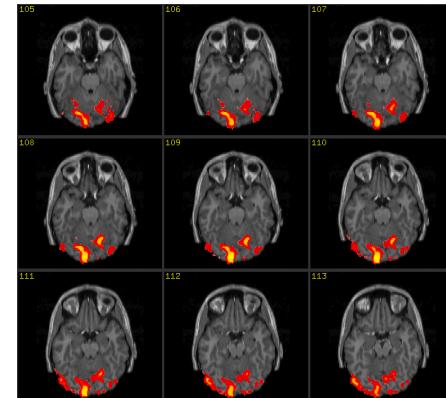
fMRI

Measures changes in blood flow related to neural activity

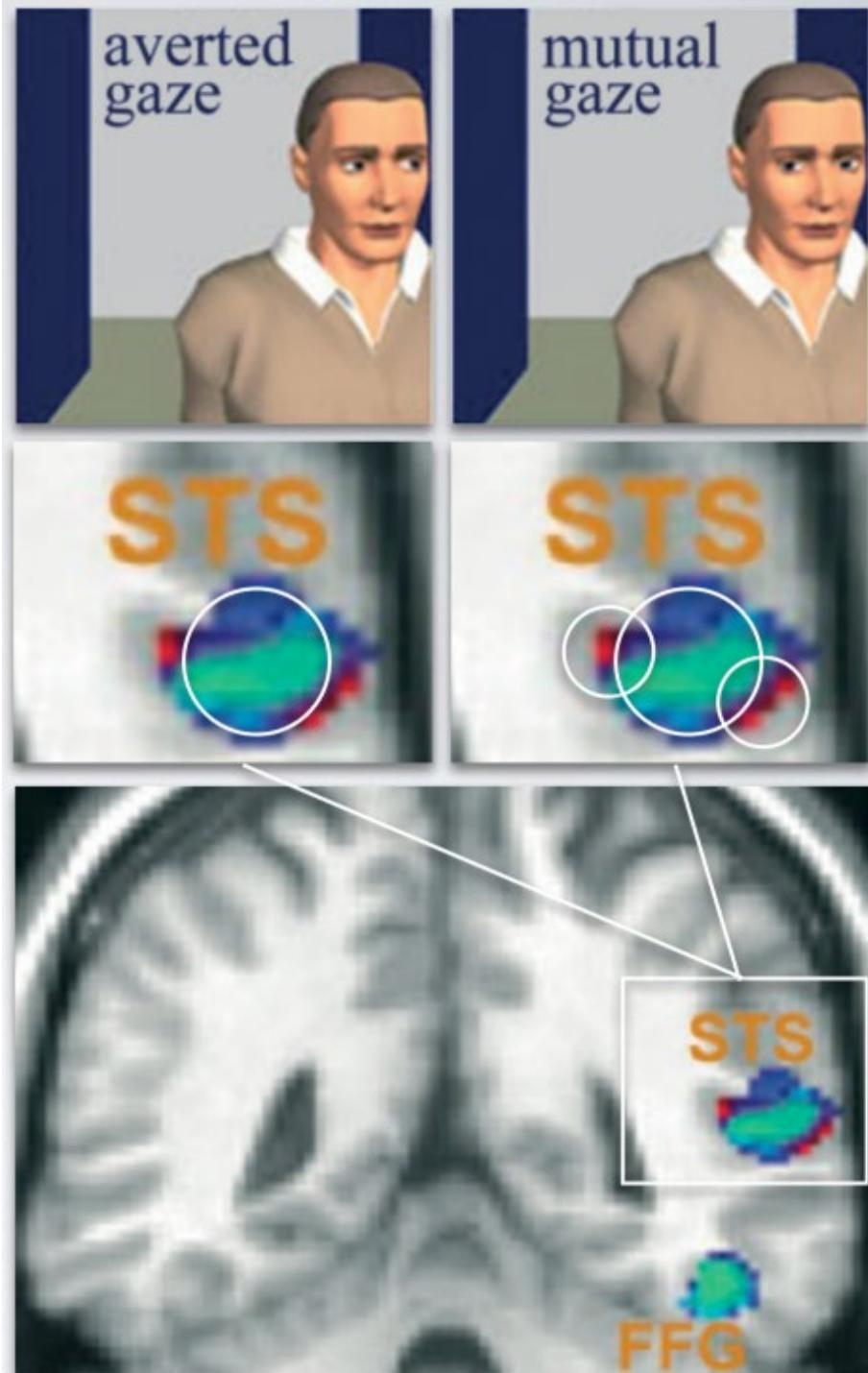
Blood-oxygen-level dependence (BOLD)

High spatial resolution

Useful in emotion research



Mutual and averted gaze stimuli
& response in the brain
(Pelphrey et al., 2004)



fNIRs

Measures blood oxygenation
(hemodynamic responses)

Indicates cognitive activity

High spatial, low temporal
resolution



Physiological Measures

Advantages

- More direct measures of emotion, cognitive activity, etc.

- Enables new research

Disadvantages

- Expensive, requires mastery, technicians

- Low signal-to-noise ratio

- Best used with other measures

Questions?

Next

Reading #8:

To be posted

Due **Monday 4/1**

Individual Assignment #3: Scale Construction

Due **Wed 4/10**

See guidelines on Moodle

Upcoming Lectures: ROS + Scale construction

Download and install ROS, R, and JMP before class!

ROS – see guidelines on Moodle

JMP

R (I also recommend downloading RStudio)



University of Colorado
Boulder

THANKS!

Professor **Dan Szafir**

*Computer Science & ATLAS Institute
University of Colorado Boulder*