

# Affective Computing

Olga Perepelkina  
HSE, 2020

# **Тема 8. Мультимодальное распознавание эмоций**

# План лекции

- Мультиmodальное распознавание эмоций: люди
- Мультиmodальное распознавание эмоций: машины

# Multimodal emotion recognition

**В чем проблема?**

# Теория базовых эмоций

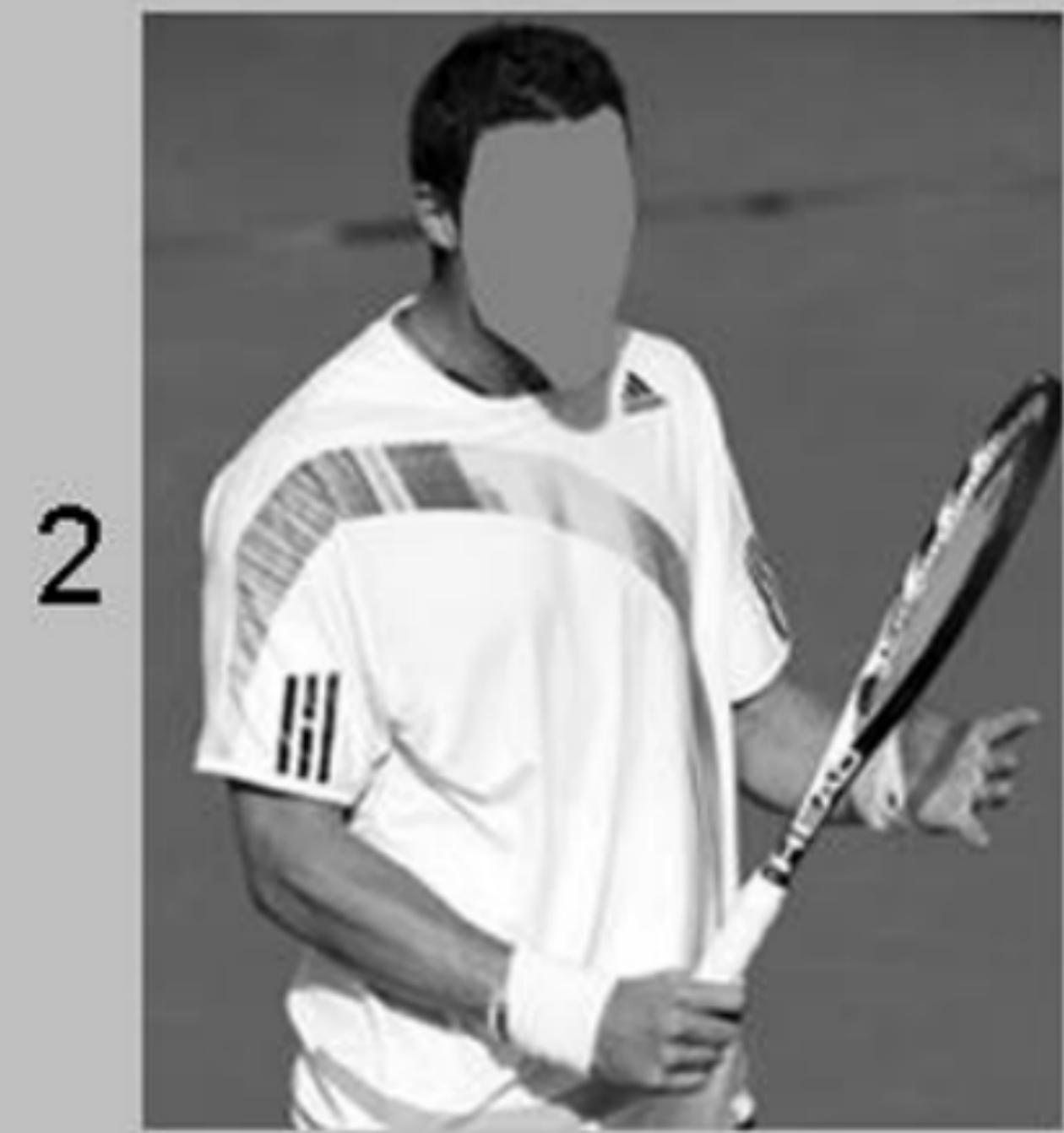
- Лицевые экспрессии соотносятся с определенными эмоциями.
- Эти конфигурации являются общими для всех людей, независимо от их пола, расы, образования, культуры и т.д.
- Эта теория оказалась очень влиятельной на протяжении последних **50 лет**.



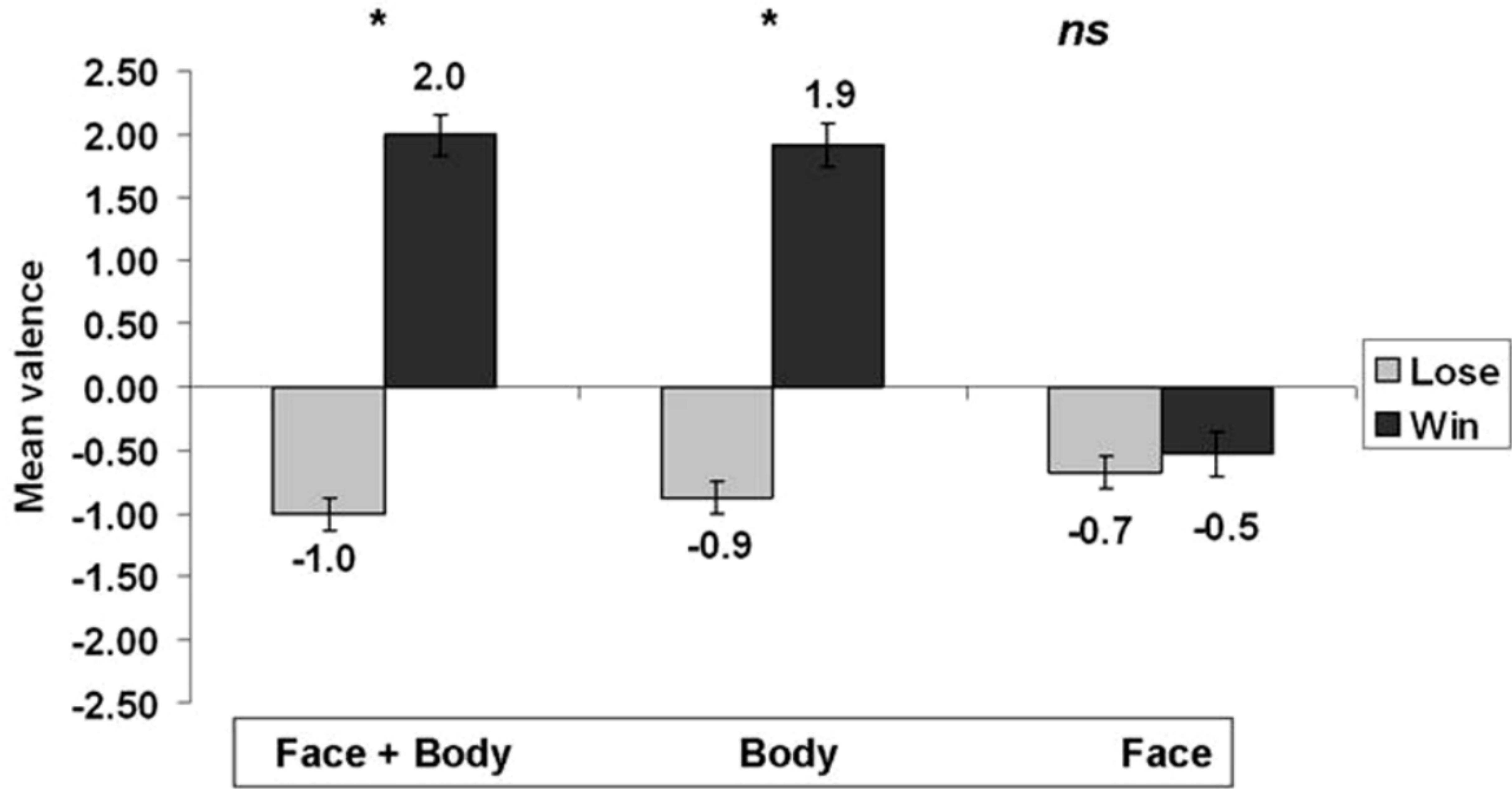




A



2



# Предложения от Лизы Барретт

- В идеальном случае должно быть записано много видео человека в естественных условиях. И алгоритмы должны автоматически вытащить проявления, такие как движения лица, походку, тон голоса. + записанные движения глаз, ЭЭГ и др.
- Добавить еще туда физического контекста (где человек находится, есть ли другой человек рядом, как он реагирует и тд). Это она называет **Multimodal observations**.
- Люди должны разметить видео (self-reports). Предлагает категории: **valence**, **arousal**, **appraisals** (описания, как воспринимается ситуация) и **emotion-related goals**.
- Должна быть вариативность в ситуациях и социальных проявлениях.

# Multimodal emotions – Humans

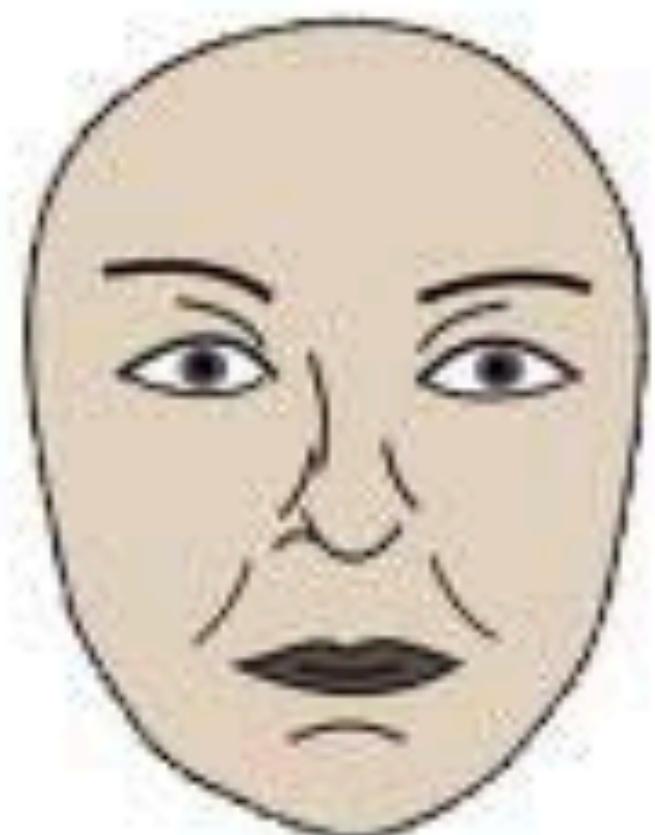
# Human emotion perception

- Люди распознают эмоции по изолированным каналам: **лучше всего по лицу**, но также по голосу, языку тела, даже прикосновениям.
- Зрительная и слуховая модальности влияют друг на друга (движения лица меняют звук, речь меняет форму лица), но, например, не влияют на тактильную модальность.
- Для разных модальностей – разная временная динамика (для лица может быть достаточно короткое предъявление, для голоса – последовательность).

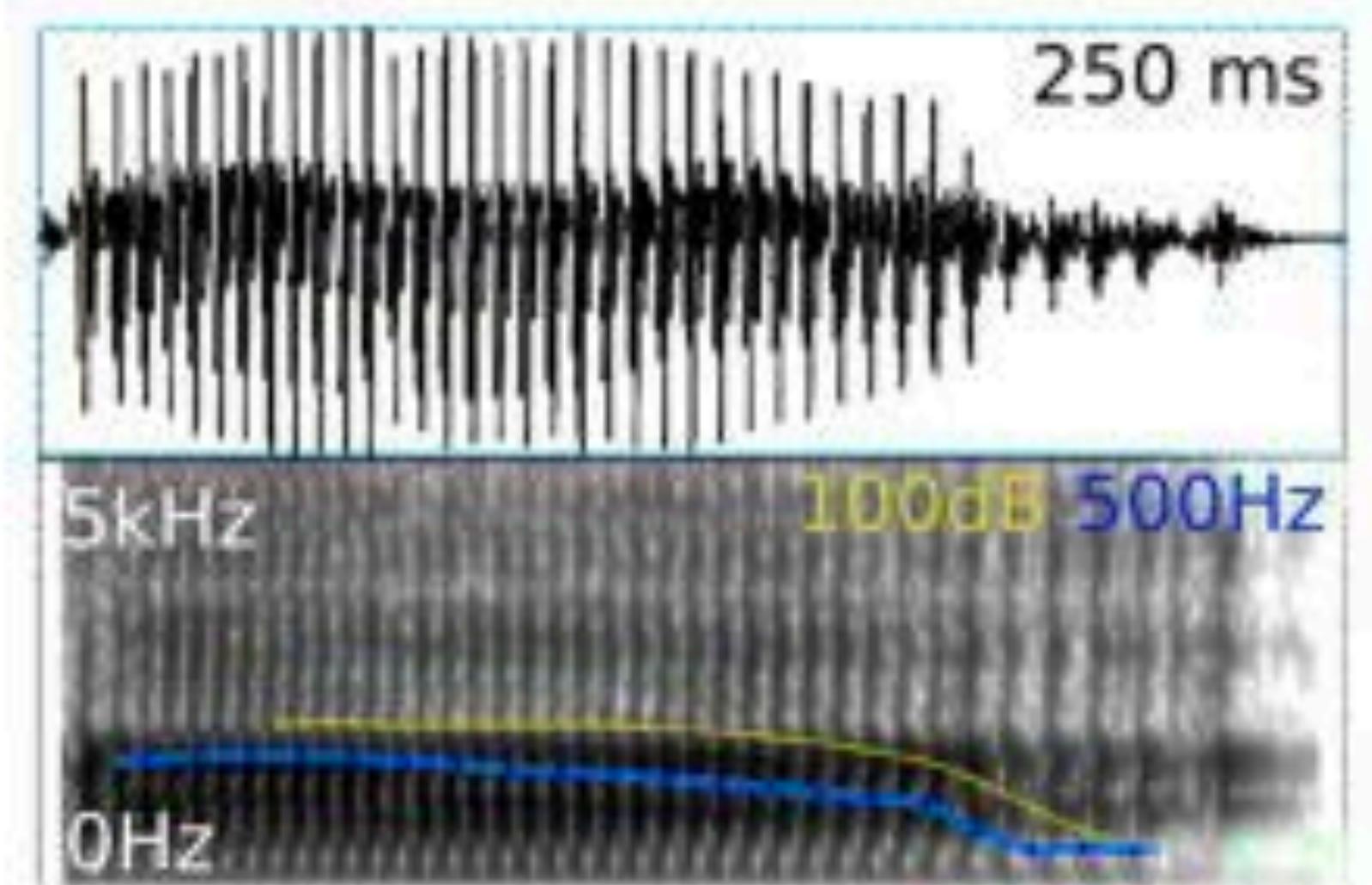
Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).

## Face

**Neutral State**



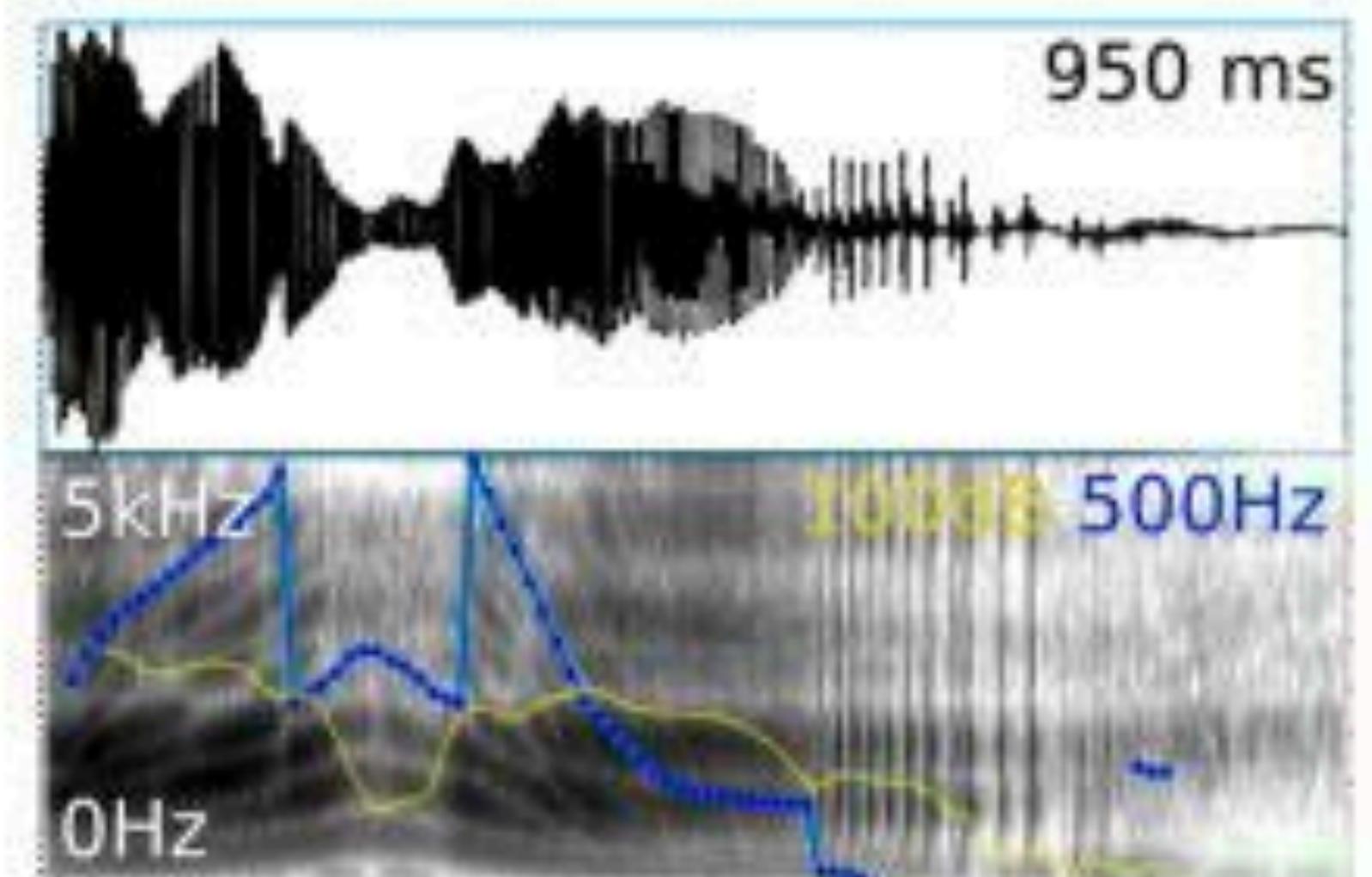
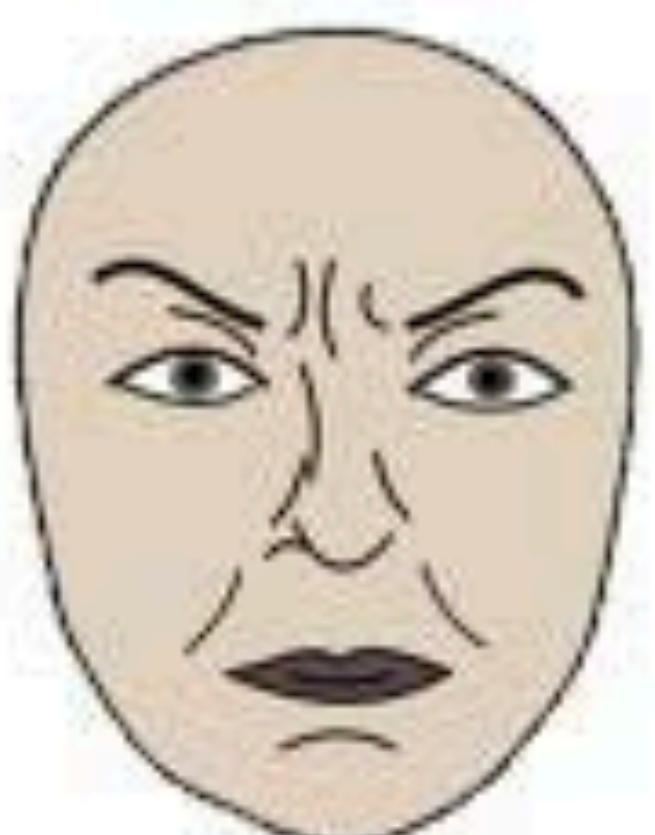
## Voice



## Touch

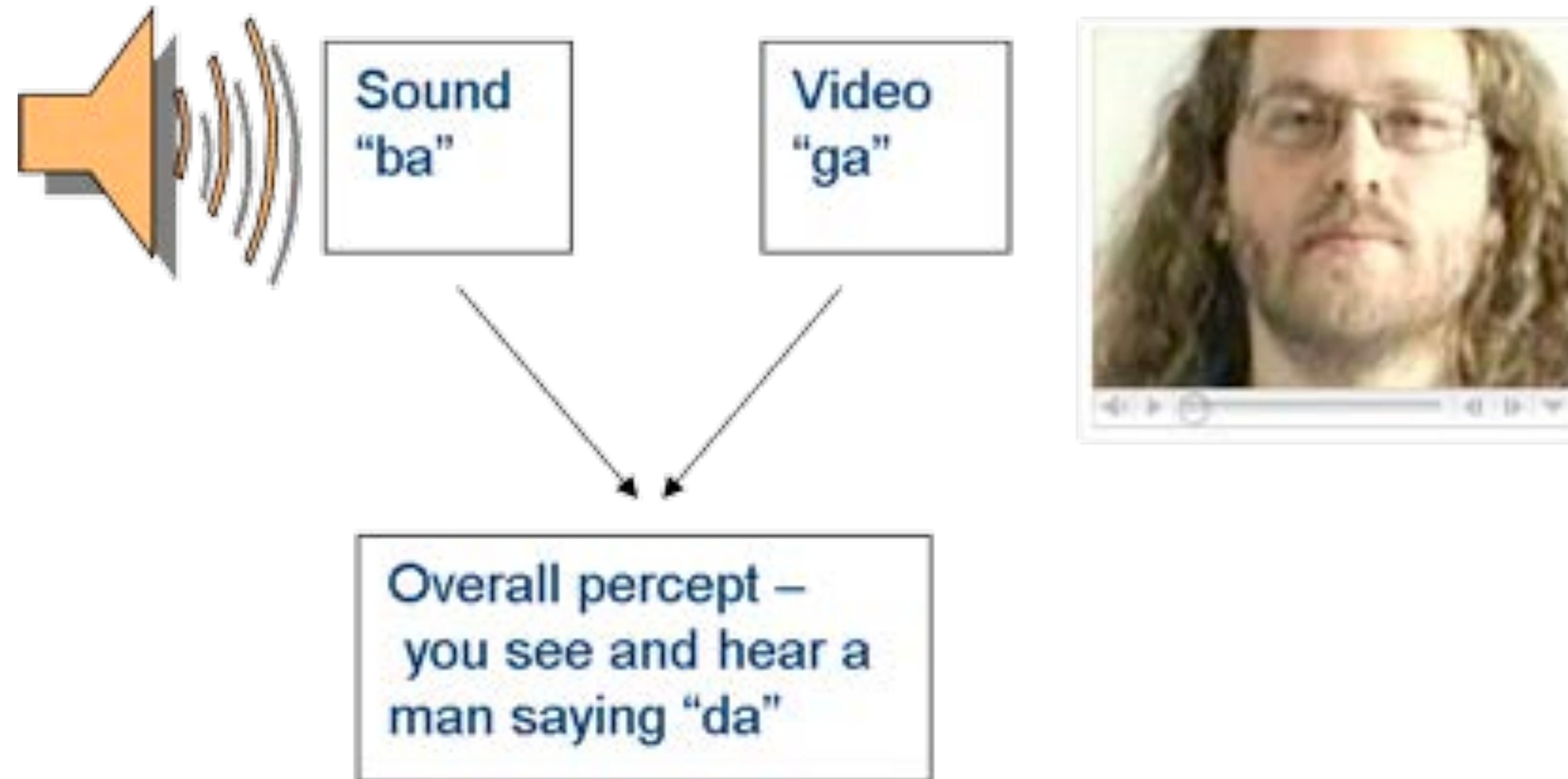


**Emotionally Expressive State**



# McGurk Effect

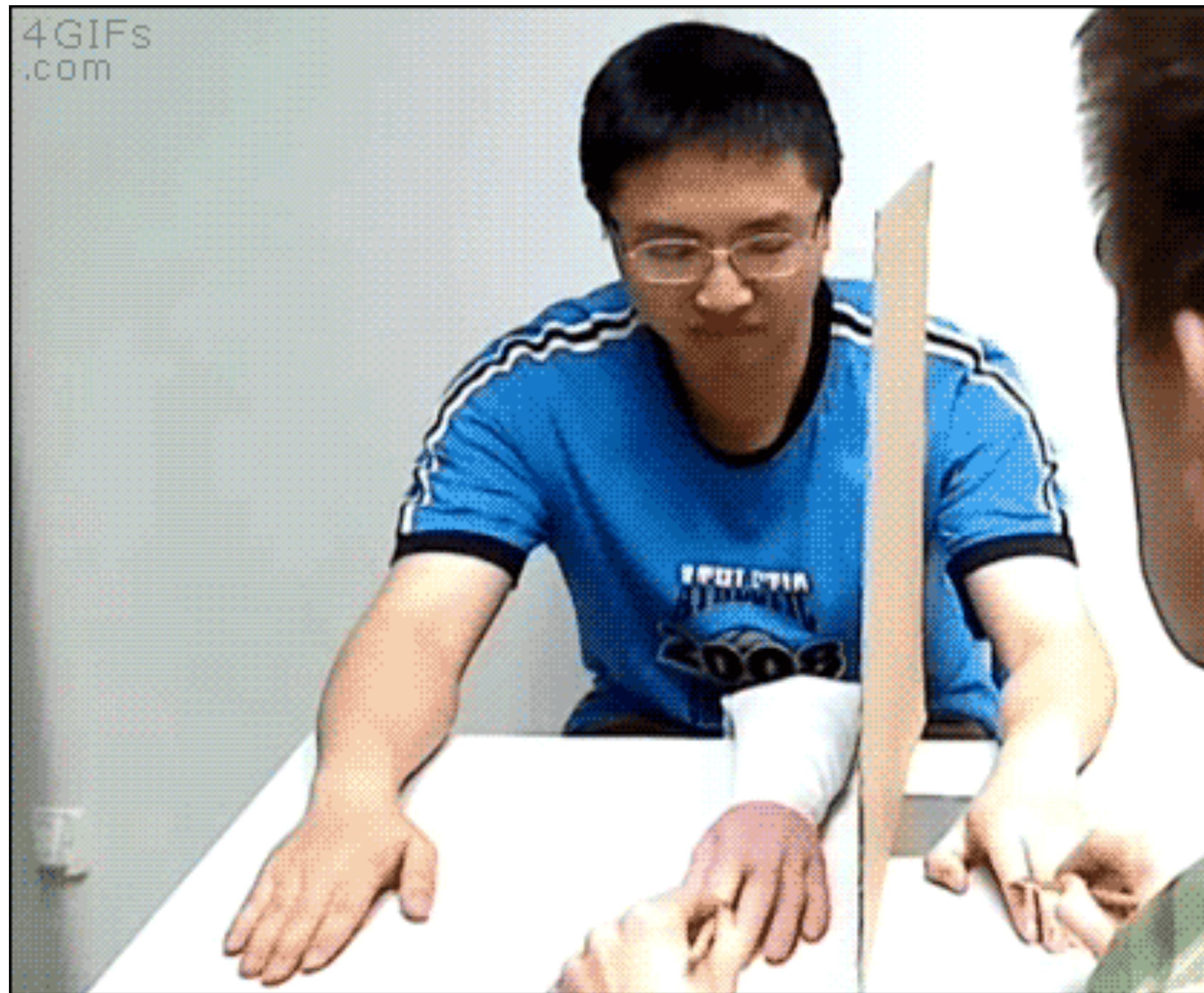
# McGurk Effect



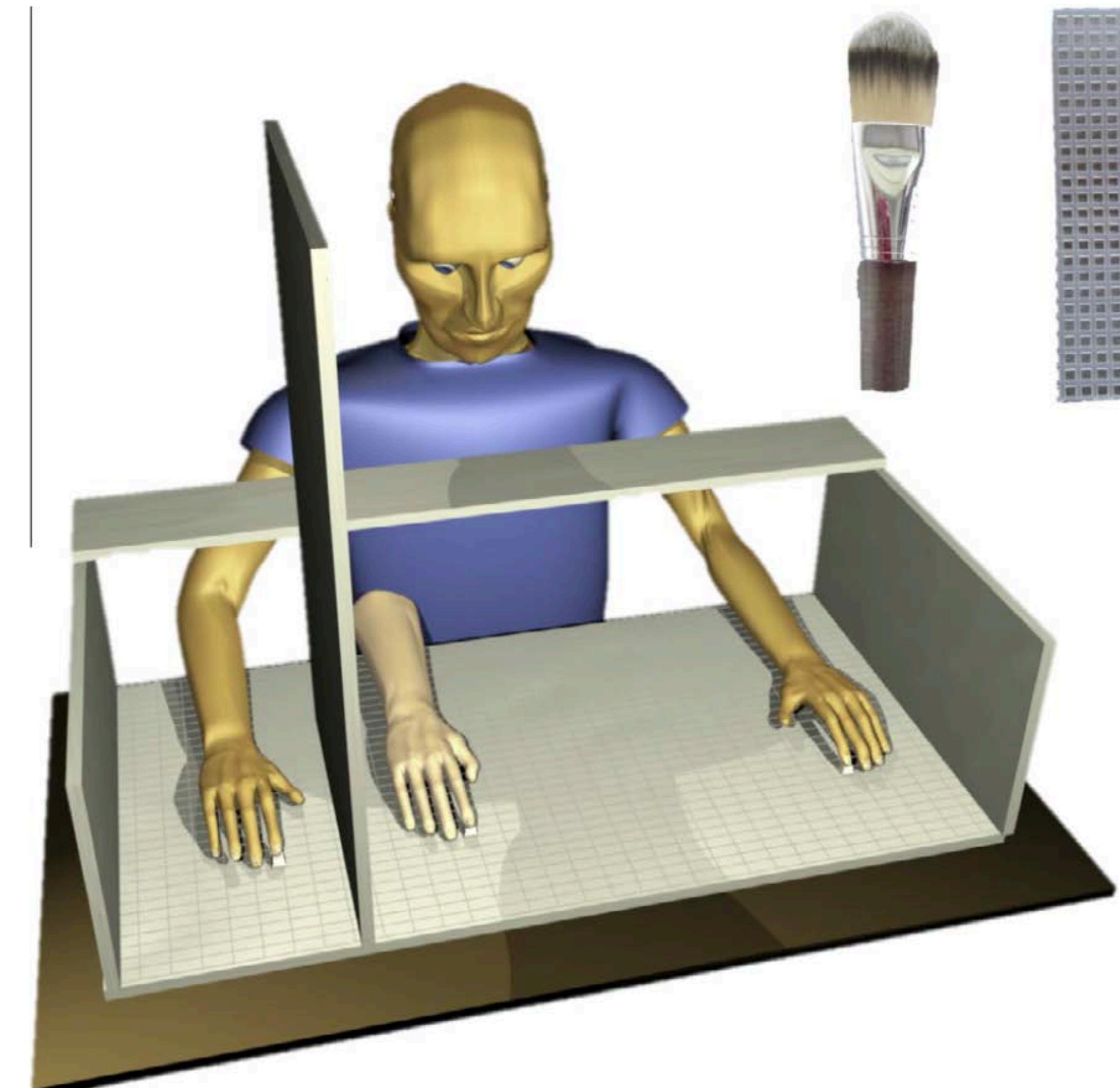
# Emotional McGurk Effect

intended emotion audio	visual	<i>content</i>	perceived emotion		
			<i>happy</i>	<i>sad</i>	<i>angry</i>
<i>content</i>	<i>happy</i>	38	<b>56</b>	4	2
	<i>sad</i>	39	0	<b>48</b>	13
	<i>angry</i>	3	0	<b>78 (**)</b>	19
<i>happy</i>	<i>content</i>	15	<b>85 (*)</b>	0	0
	<i>sad</i>	20	<b>51</b>	27	2
	<i>angry</i>	20	<b>45</b>	20	15
<i>sad</i>	<i>content</i>	<b>49</b>	25	24	2
	<i>happy</i>	35	<b>52</b>	13	0
	<i>angry</i>	2	0	<b>98 (*)</b>	0
<i>angry</i>	<i>content</i>	23	<b>74 (**)</b>	0	3
	<i>happy</i>	40	<b>42</b>	3	15
	<i>sad</i>	4	0	24	<b>72 (*)</b>

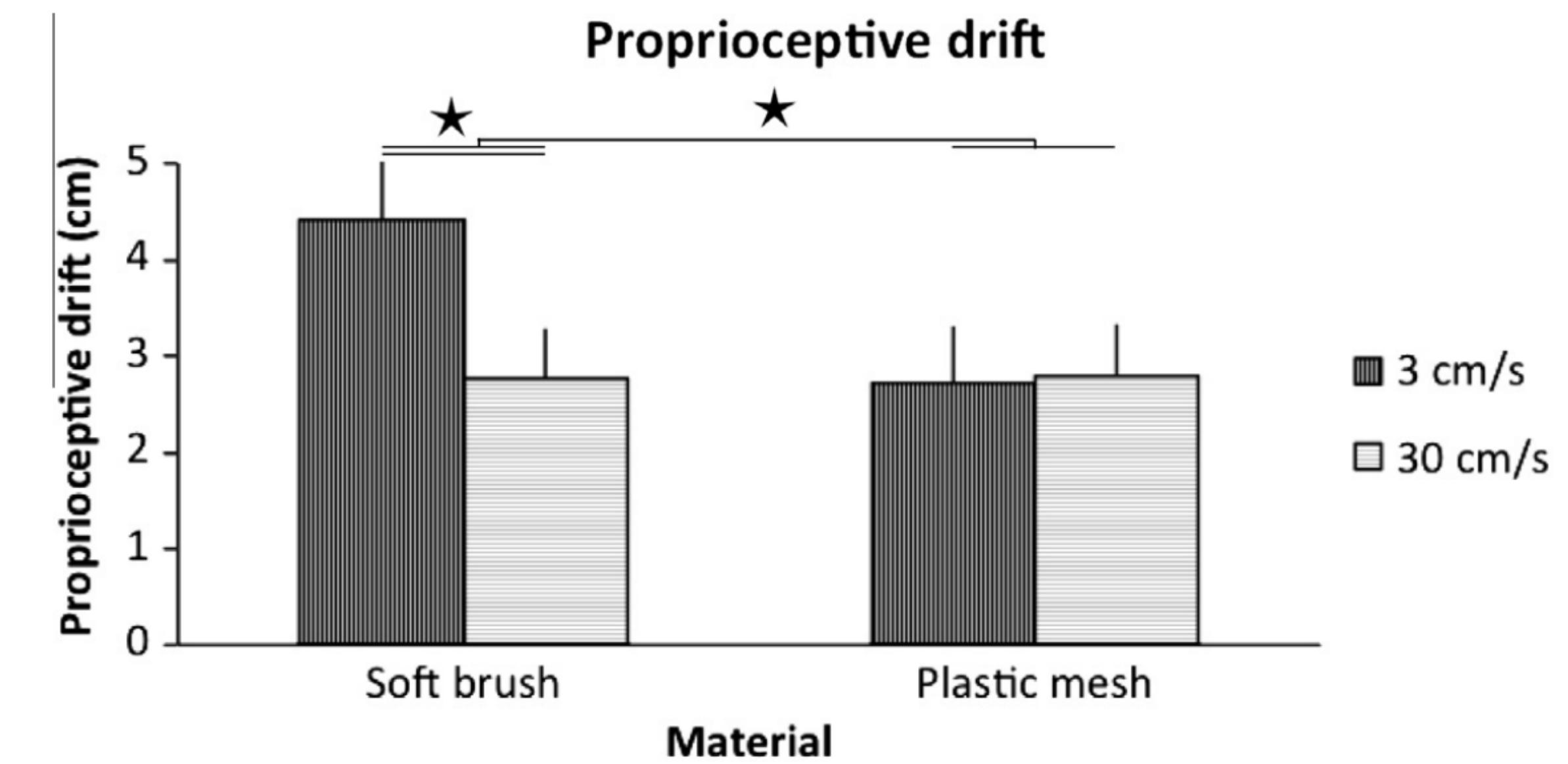
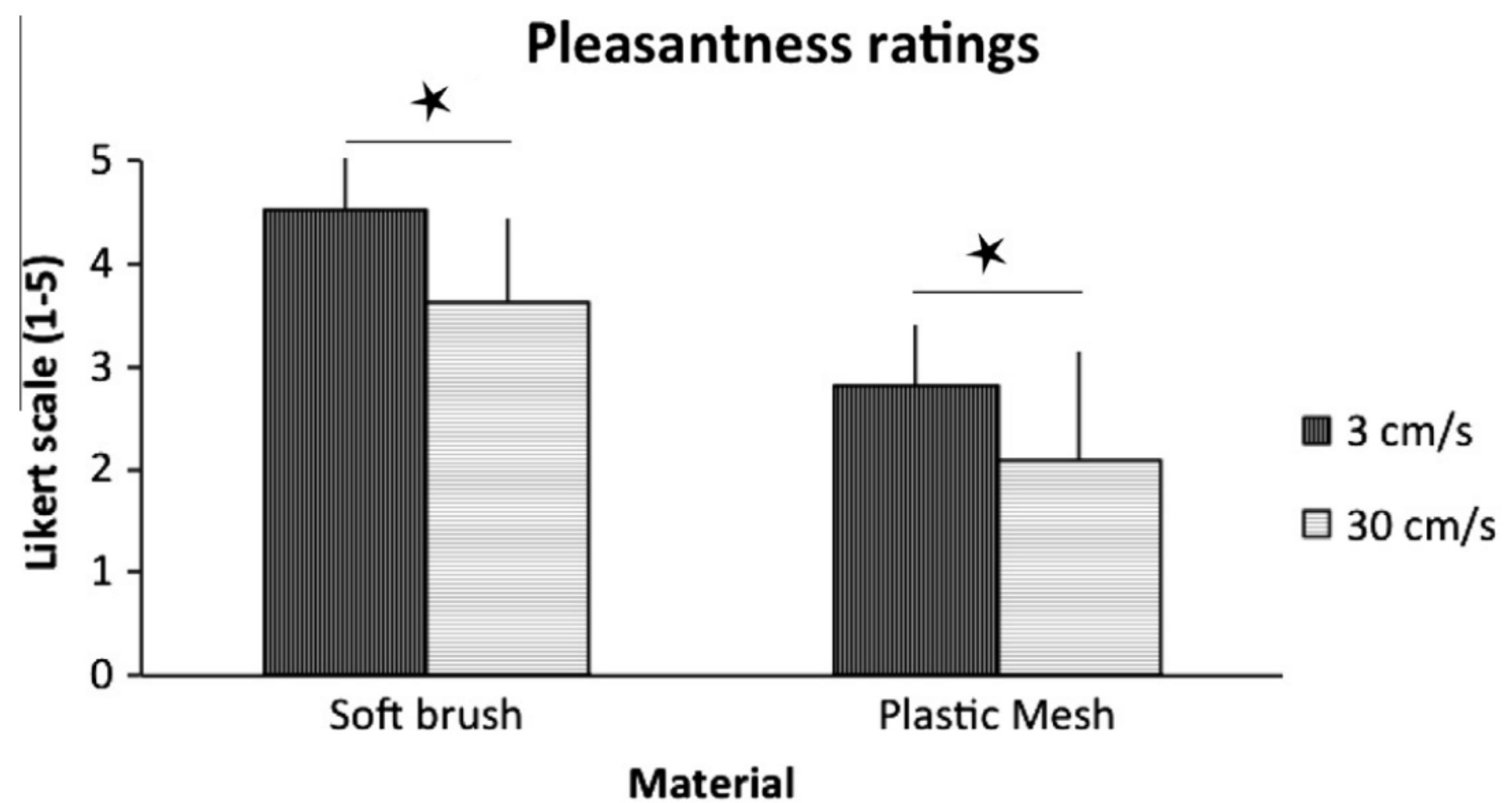
# Rubber Hand Illusion



# Affective Rubber Hand Illusion



# Affective Rubber Hand Illusion



# Multisensory integration

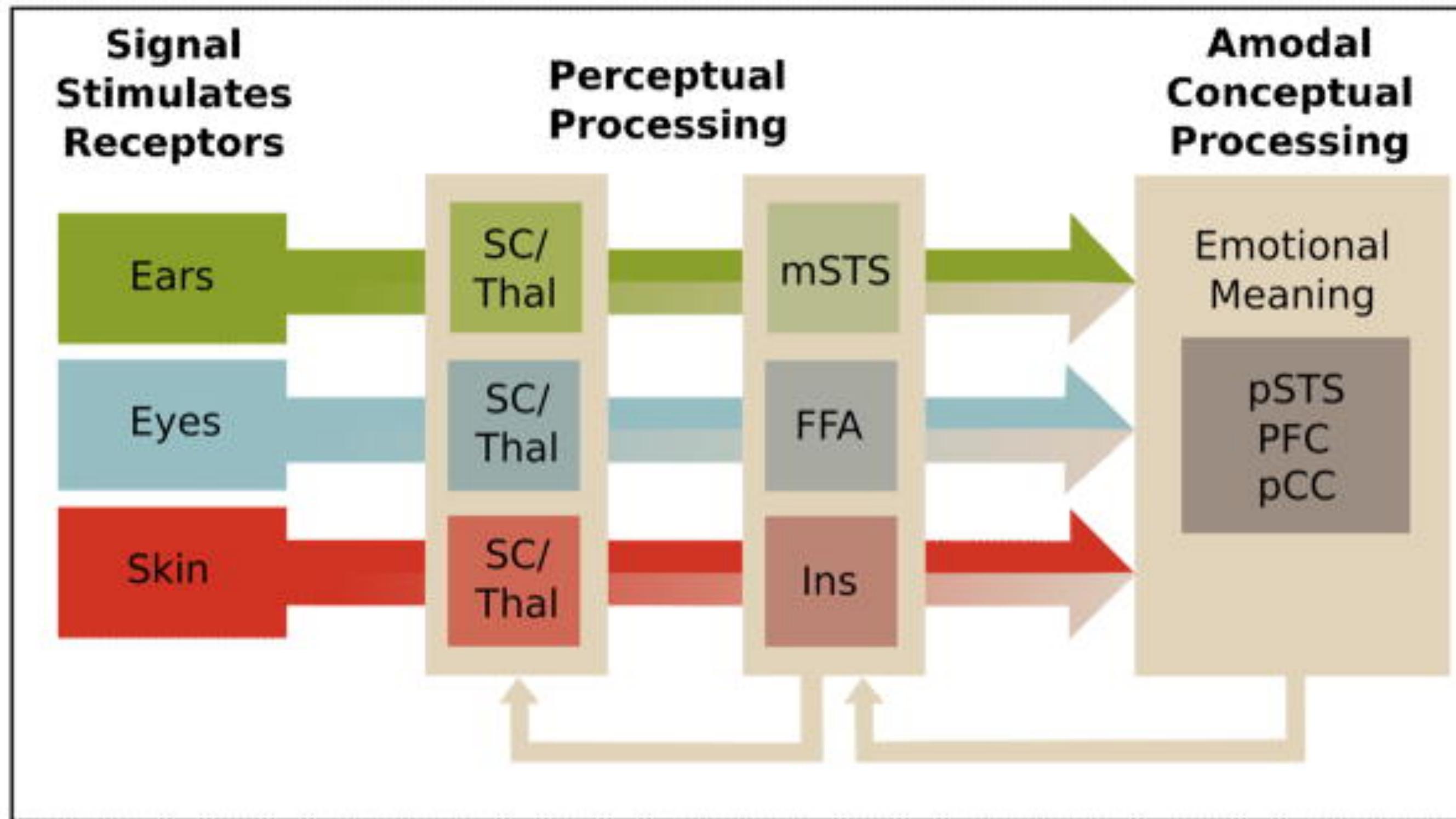
- Мультисенсорная интеграция эмоций в мозге подчиняется законам описанным для МСИ в целом:
  - Между сигналами разной модальности не должно быть рассогласования больше, чем 100 мс, чтобы они воспринимались как единое событие
  - Рассогласованность в пространстве не так критична, и иногда вообще не влияет на восприятие (напр., в кинотеатре, эффект чревовещания и т.д.).
  - Более качественное восприятие бимодальных стимулов (по сравнению с унимодальными) заметно сильнее на стимулах более низкой интенсивности, чем на стимулах высокой интенсивности (**principle of inverse effectiveness**).

Klasen, Martin, Yu-Han Chen, and Klaus Mathiak. "Multisensory emotions: perception, combination and underlying neural processes." (2012): 381-392.

# Multisensory integration: emotion

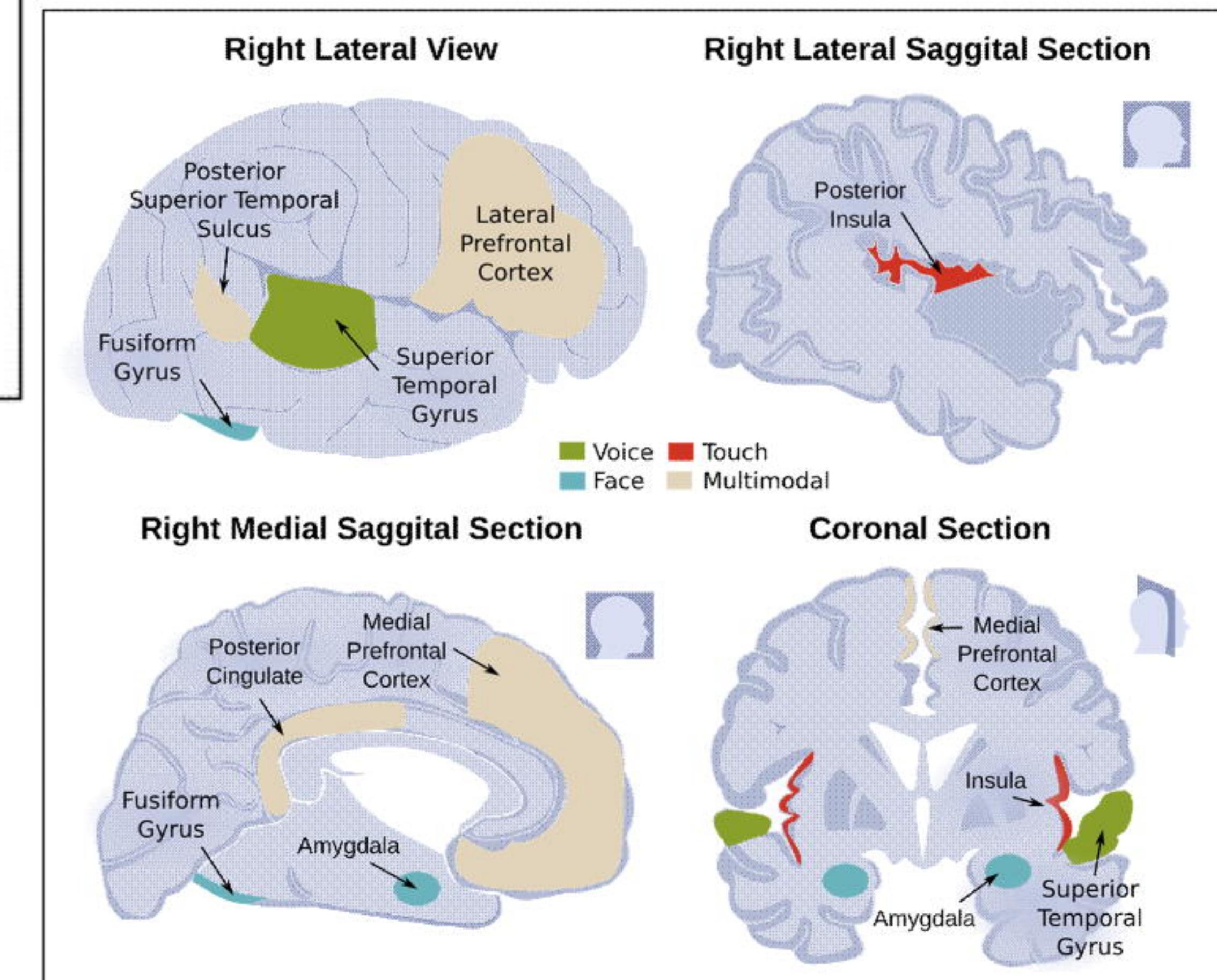
- Мультимодальная интеграция эмоций это **не просто поздний процесс обработки информации**, который возникает после того, как информация от всех модальностей обработана.
- Наоборот, существуют мультимодальные нейроны, которые действуют на самых **ранних этапах обработки** сигнала (например, в таламусе и верхнем двухолмии).
- Эти ранние ММ процессы обработки информации обеспечивают temporal binding, кросс-модальное взаимодействие (напр., снижение порога УМ восприятия).

Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).



SC: Superior Colliculus  
Thal: Thalamus  
mSTS: medial Superior temporal sulcus  
FFA: Fusiform Face Area

Ins: Insula  
pSTS: Posterior Superior Temporal Sulcus  
PFC: Prefrontal Cortex  
pCC: Posterior Cingulate Cortex



# Multisensory integration: emotion

- На более поздних этапах ММ обработки эмоций (различные отделы коры – РFC, поясная извилина etc.) возникают процессы амодальной, концептуальной обработки информации.
- В результате этого возникает интегрированный целостный образ (holistic percept).
- Higher-level representations can feed back and **modulate** lower-level representations.

# Multisensory integration: emotion

- Согласованные мультимодальные сигналы распознаются быстрее и проще, чем рассогласованные (как на уровне мозга, так и на уровне поведения).

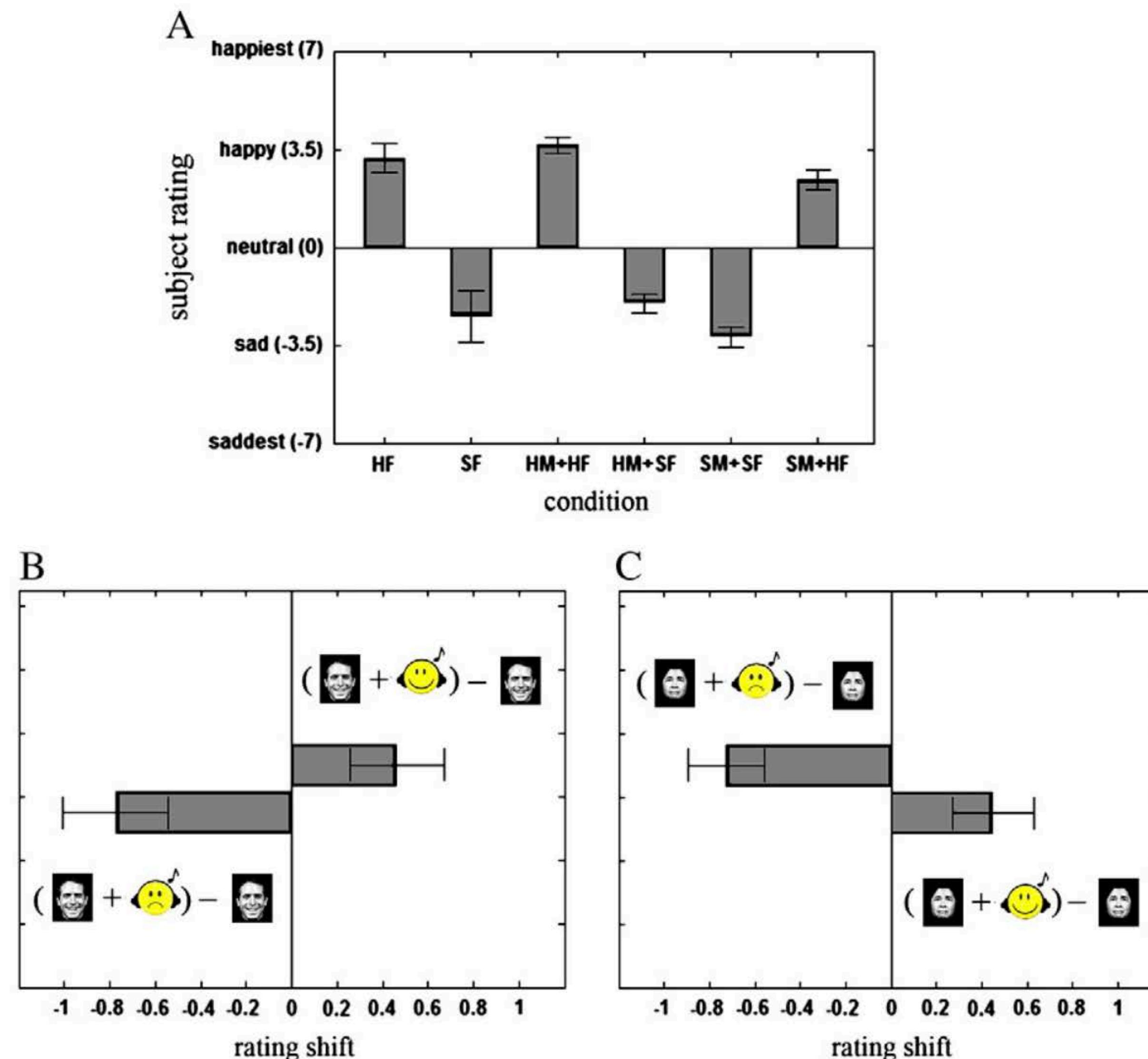
Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).

Klasen, Martin, et al. "Neural processing of emotion in multimodal settings." Frontiers in human neuroscience 8 (2014).

# Multisensory integration: emotion

- В эксперименте Gelder & Vroomen предъявляли **неоднозначные** стимулы лица и давали слушать эмоциональный голос (и просили не обращать на него внимание). Восприятие эмоции лица **зависело** от эмоции в голосе.
- То же самое для ситуации наоборот: неоднозначный голос и стимулы лица (на которые не нужно было обращать внимание), восприятие эмоции голоса зависело от эмоций на лицах.

De Gelder, Beatrice, and Jean Vroomen. "The perception of emotions by ear and by eye." *Cognition & Emotion* 14.3 (2000): 289-311.



**Fig. 4.** Change in face emotion ratings by simultaneous presentation of music. (A) Effect of happy and sad music on rating of happy and sad faces. Bars represent mean ratings with standard deviation. (B) Effect of happy and sad music on rating of happy faces. Ratings for faces alone were subtracted from ratings of faces under each music condition for each subject. Bars represent mean rating differences with standard deviation. Happy music increased rating (more happy,  $p=0.051$ ) and sad music decreased the rating (less happy,  $p=0.008$ ) of emotion on happy faces. (C) Effect of happy and sad music on rating of sad faces. Happy music increased rating (less sad,  $p=0.030$ ) and decreased rating (more sad,  $p=0.002$ ) of emotion on sad faces.

# Multisensory integration: emotion

- Лицо, похоже, является главным каналом при восприятии эмоций.
- В исследованиях, где подавали конфликтную эмоциональную информацию (голос и лицо), лицо несло больший вес в принятии решения об эмоции.

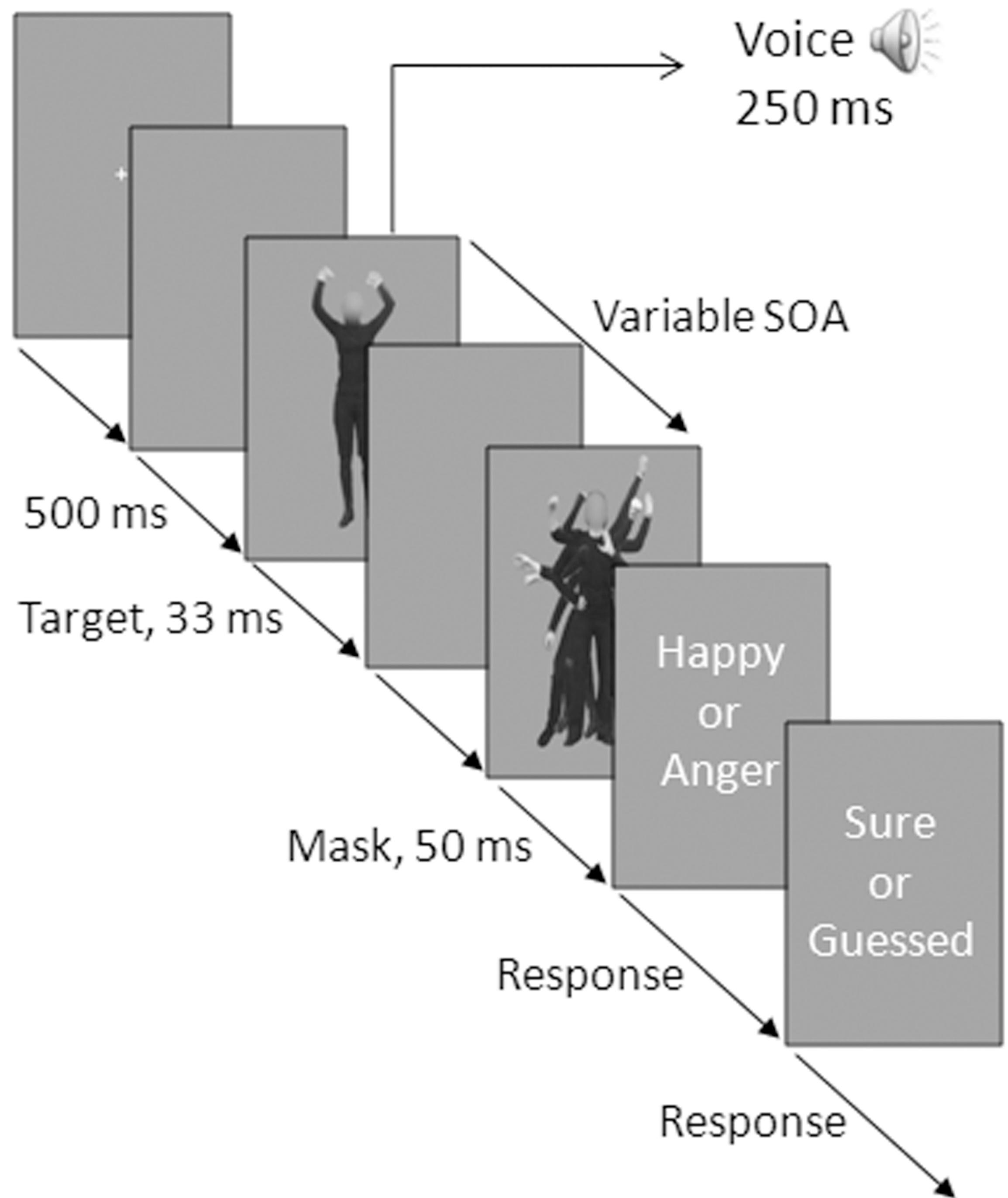
Argyle, Michael, Florisse Alkema, and Robin Gilmour. "The communication of friendly and hostile attitudes by verbal and non-verbal signals." European Journal of Social Psychology 1.3 (1971): 385-402.

Mehrabian, Albert, and Susan R. Ferris. "Inference of attitudes from nonverbal communication in two channels." Journal of consulting psychology 31.3 (1967): 248.

# Multisensory integration: voice -> body

- Показывали angry/happy bodily expressions и эмоционально окрашенный angry/happy голос.
- Должны были распознать эмоцию по позе, игнорируя голос
- Точность распознавания эмоции была **выше**, когда стимулы были **конгруэнтными**.
- Зрительные стимулы могли быть настолько короткими, что сознательно не распознавались, но эффект сохранялся

Stienen, Bernard MC, Akihiro Tanaka, and Beatrice de Gelder. "Emotional voice and emotional body postures influence each other independently of visual awareness." PLoS One 6.10 (2011): e25517.



Angry

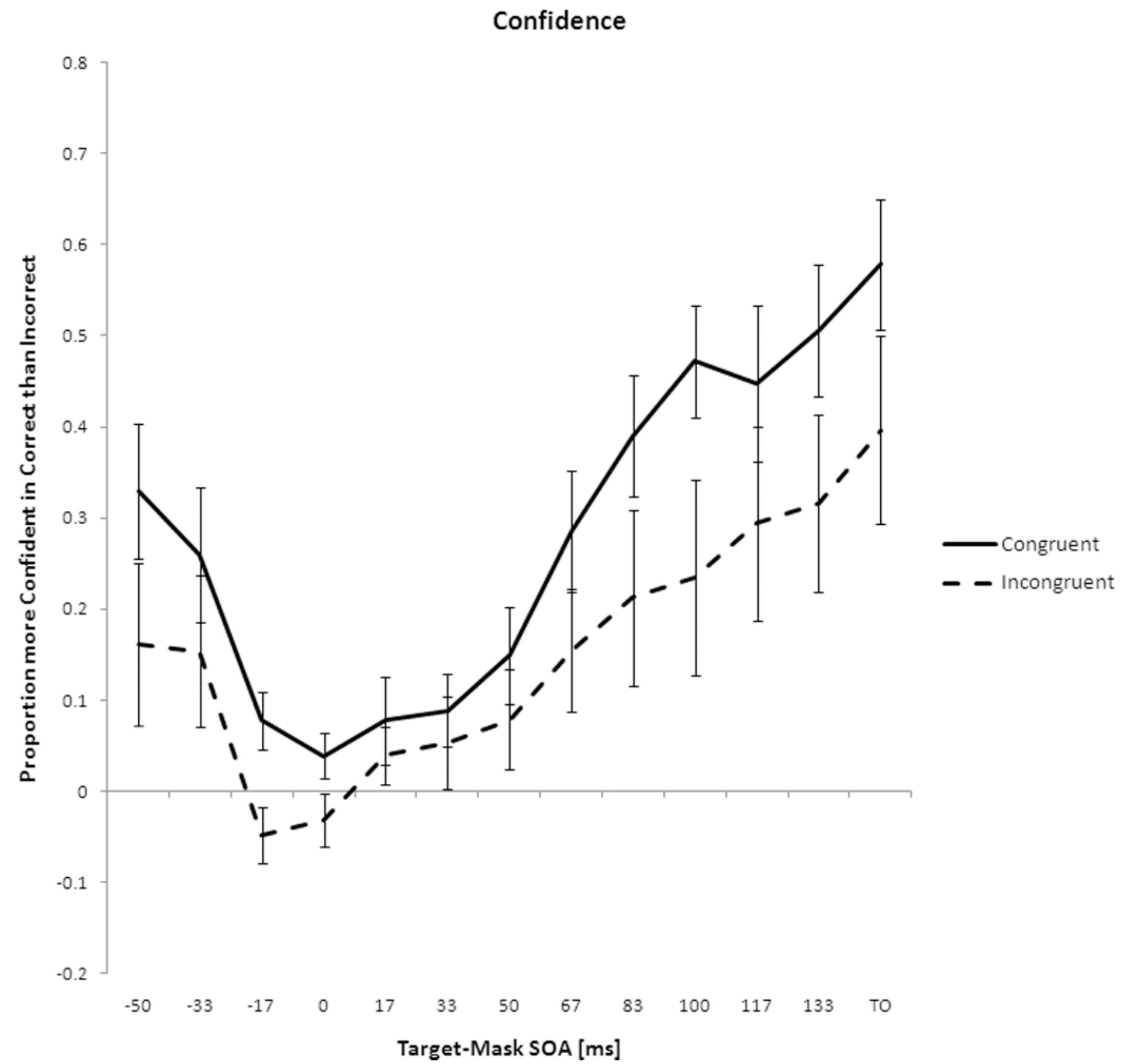
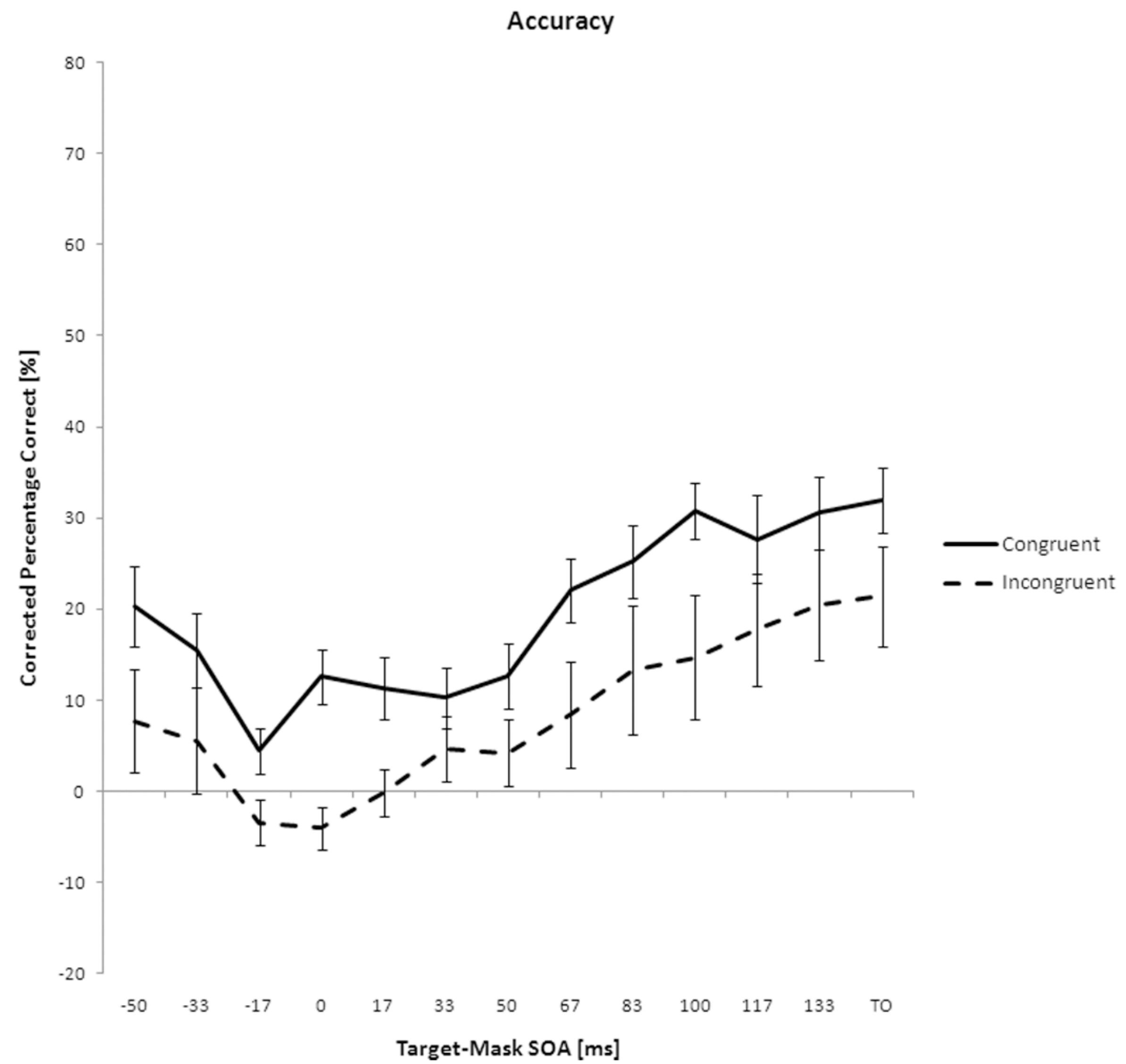


Happy



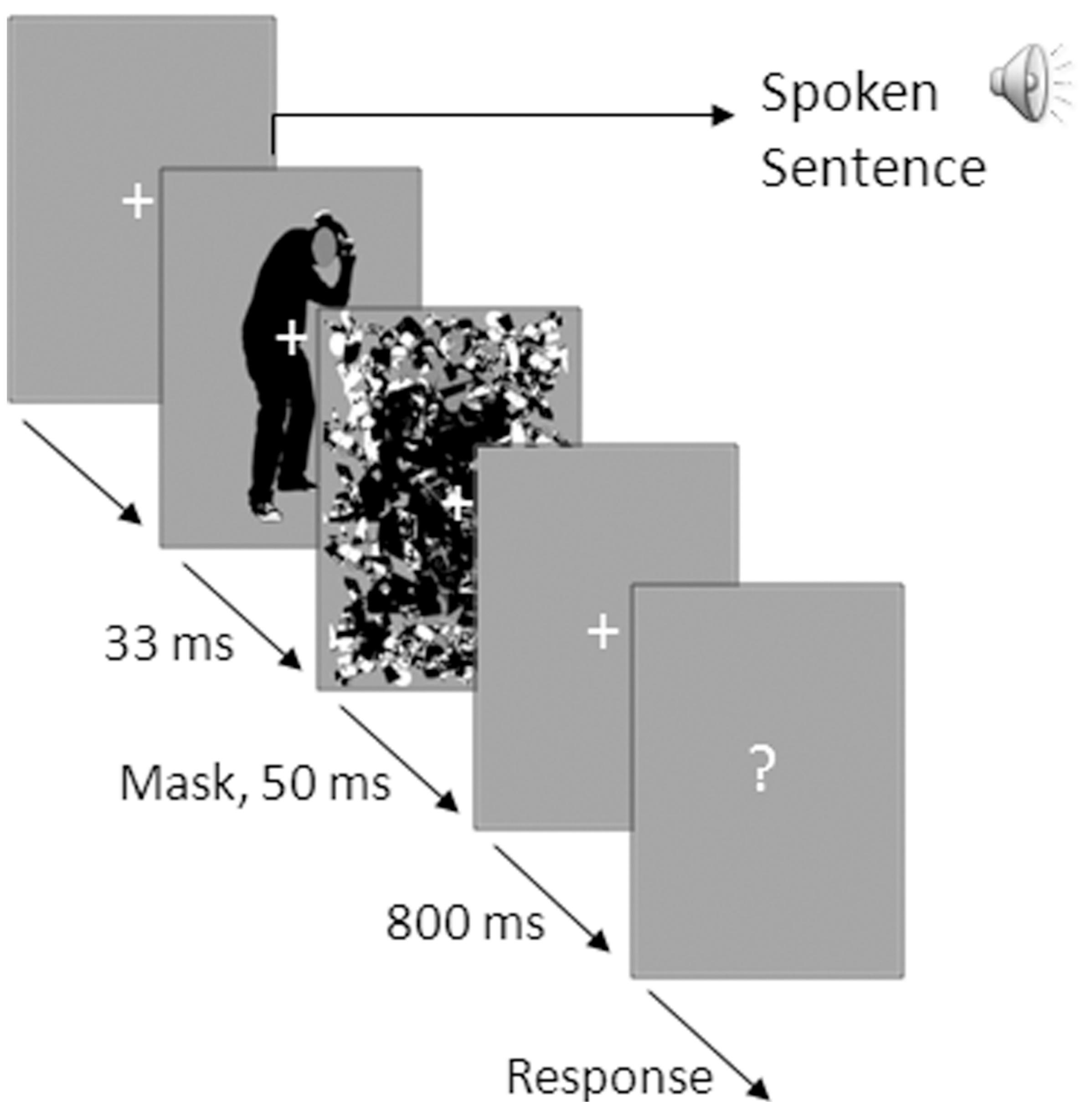
Mask





# Multisensory integration: body -> voice

- Предъявили **неоднозначный** голос (50/50 morphed between fearful and happy). Распознавали эмоцию голоса в **зависимости** от эмоции по языку тела.
- Stienen, Bernard MC, Akihiro Tanaka, and Beatrice de Gelder. "Emotional voice and emotional body postures influence each other independently of visual awareness." PLoS One 6.10 (2011): e25517.



Fear

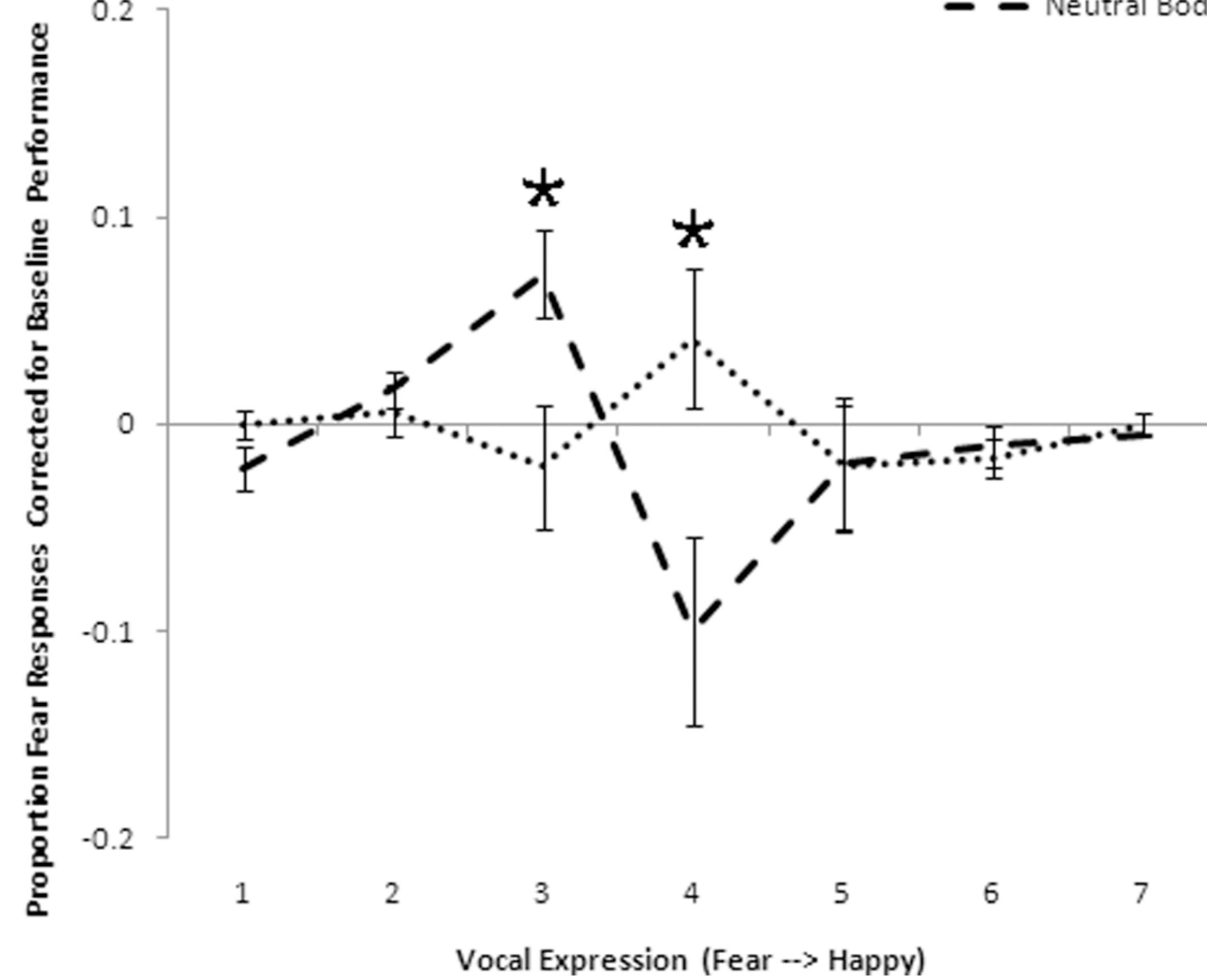
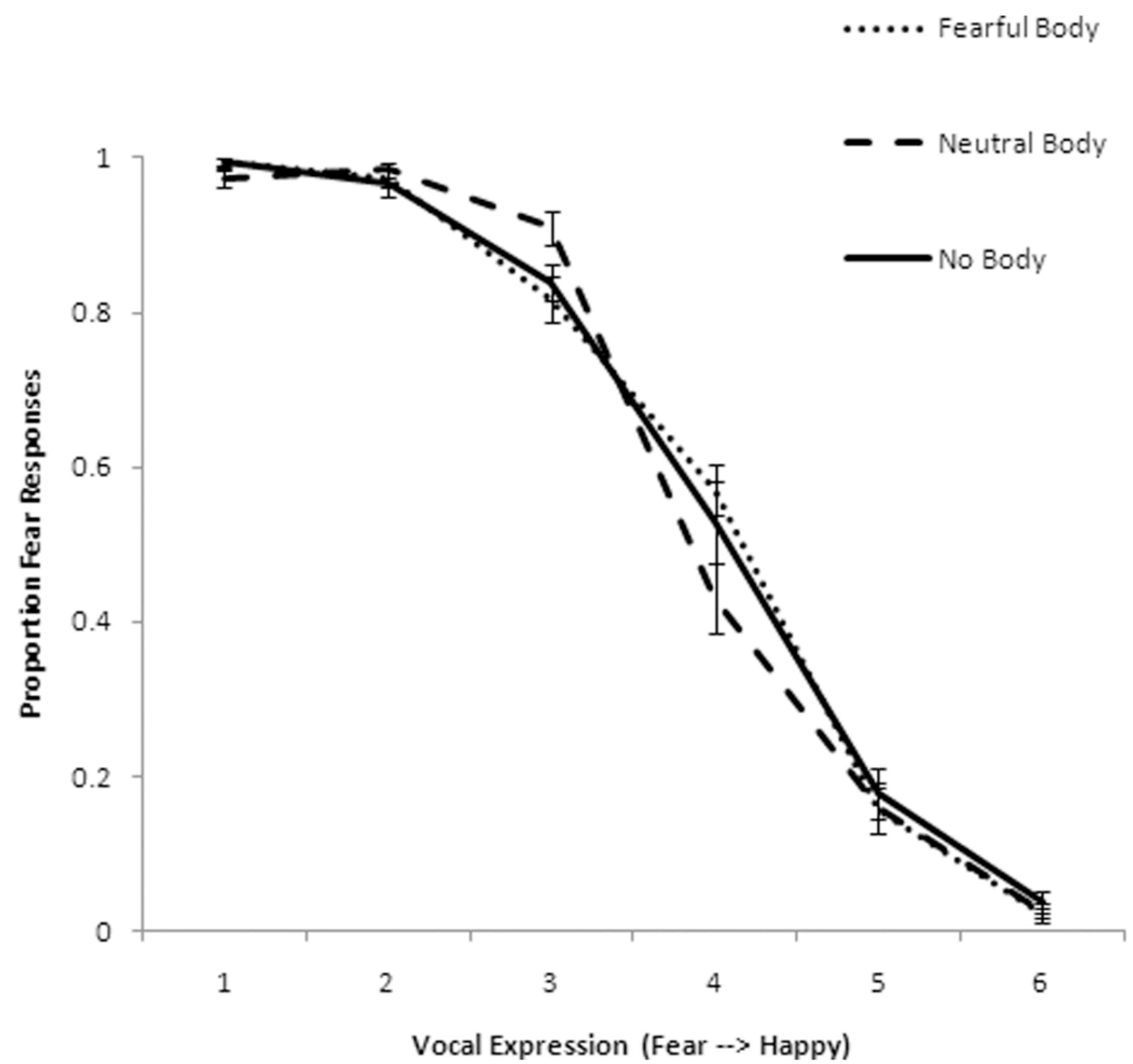


Neutral



Mask





# Trustworthiness Perception

- Предъявили комбинацию правдивых/ложных лиц и голос:
  - Trustworthy face + Trustworthy voice
  - Trustworthy face + Nontrustworthy voice
  - Nontrustworthy face + Trustworthy voice
  - Nontrustworthy face + Nontrustworthy voice
- А также каналы по-отдельности (Trustworthy/Nontrustworthy face, Trustworthy/Nontrustworthy voice)

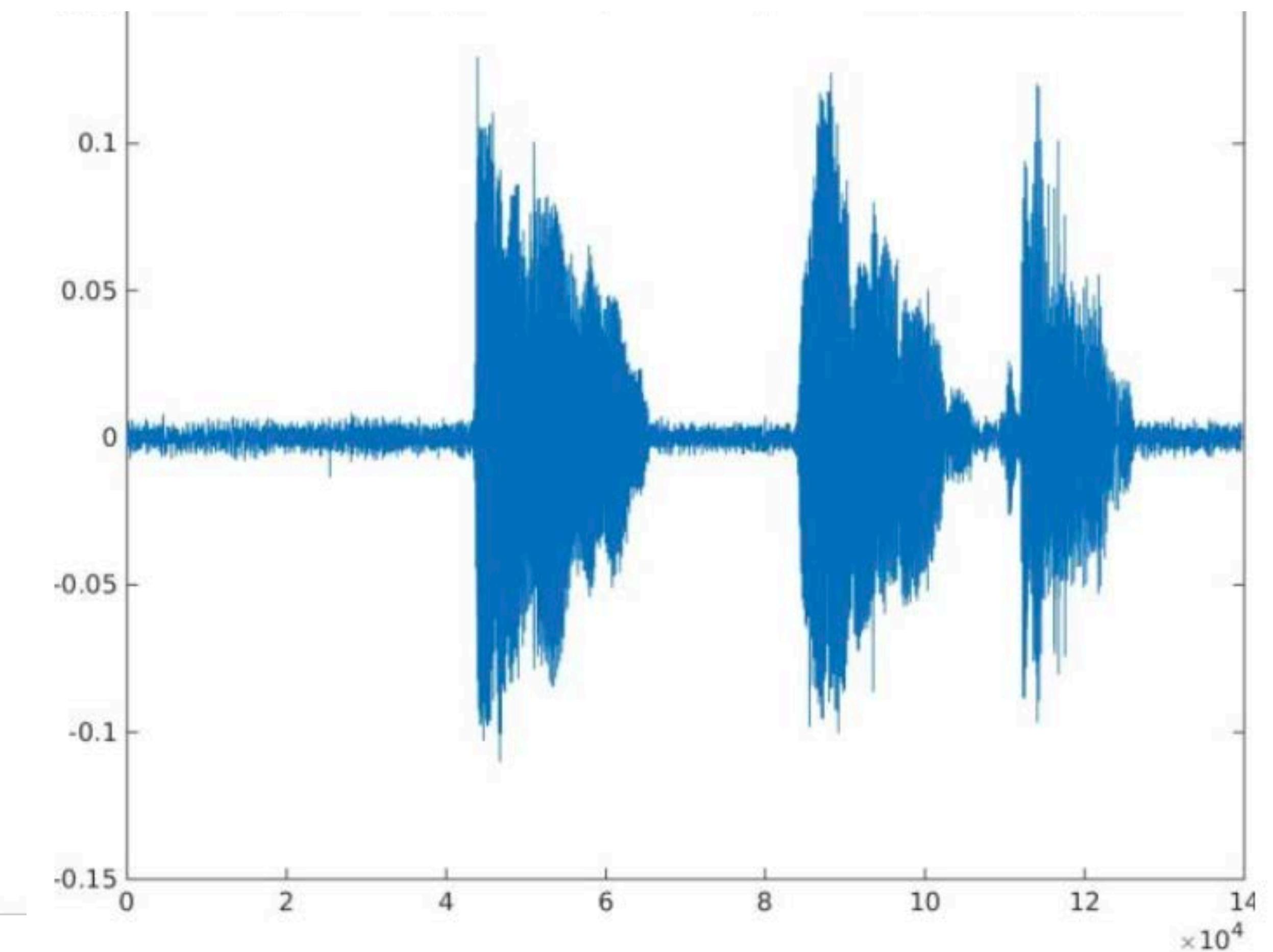
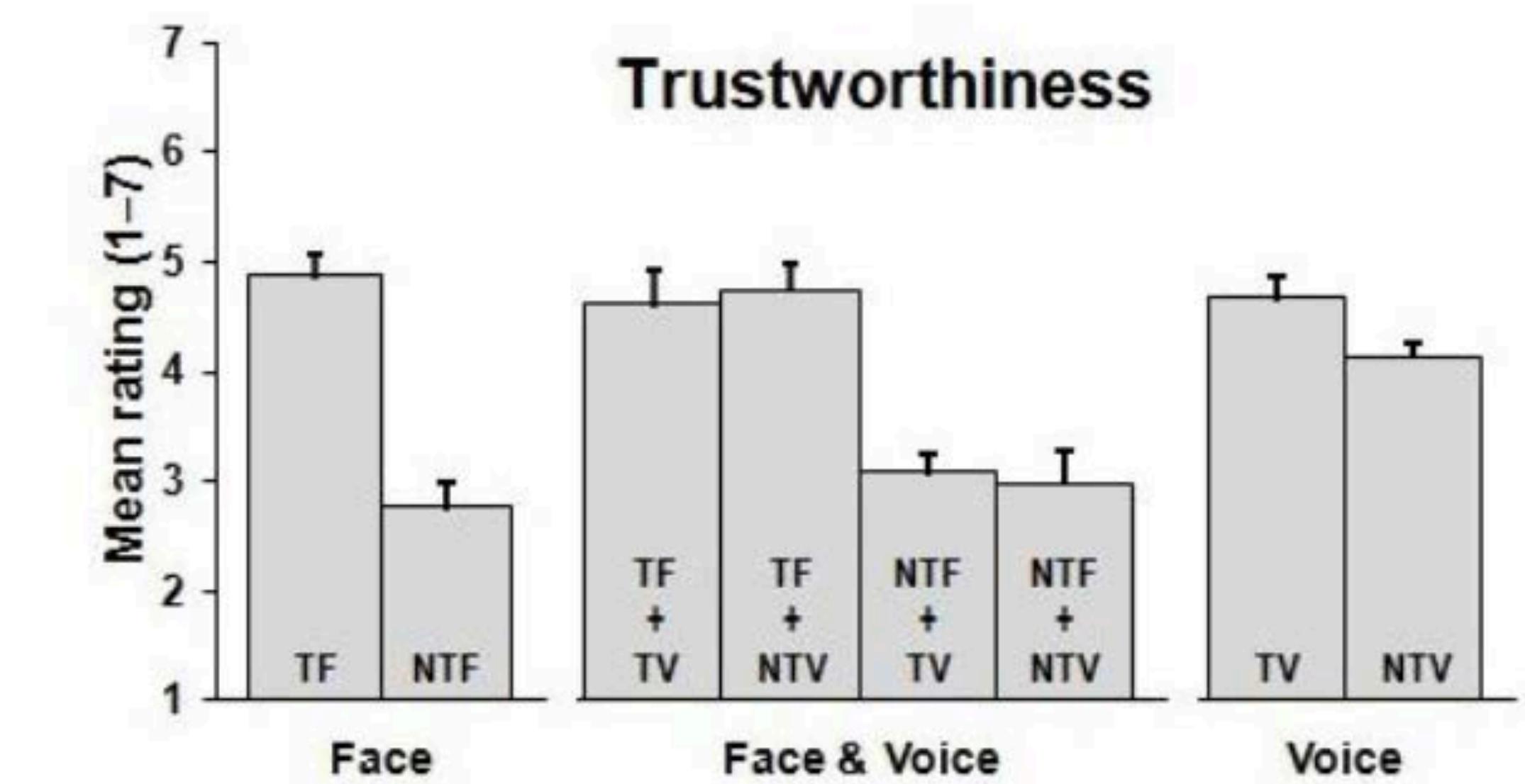


Figure 1: *Alignment of a non-trustworthy visual and a trustworthy auditory channel (NTF + TV).*

# Trustworthiness Perception

- Восприятие правдивости/лжи по лицу было значительно точнее, чем по голосу
- Дисперсия восприятия правдивости объяснялась 27,2% по лицу и 1,6% по голосу.



Tsankova, Elena, et al. "The multi-modal nature of trustworthiness perception." Proceedings of the International Speech Communication Association (ISCA). 2015.

# Вербальное и невербальное поведение в политических дебатах

- В поведении на дебатах описаны два типа поведения: **challenge** (несогласие, протест, попытки умалить заслуги соперника и т.д.) и **support** (описание достоинств своей партии, сильных позиций, стремление продемонстрировать свои лучшие черты и создать красивую социальную картину).
- Оба типа поведения могут проявляться как вербально, так и невербально.

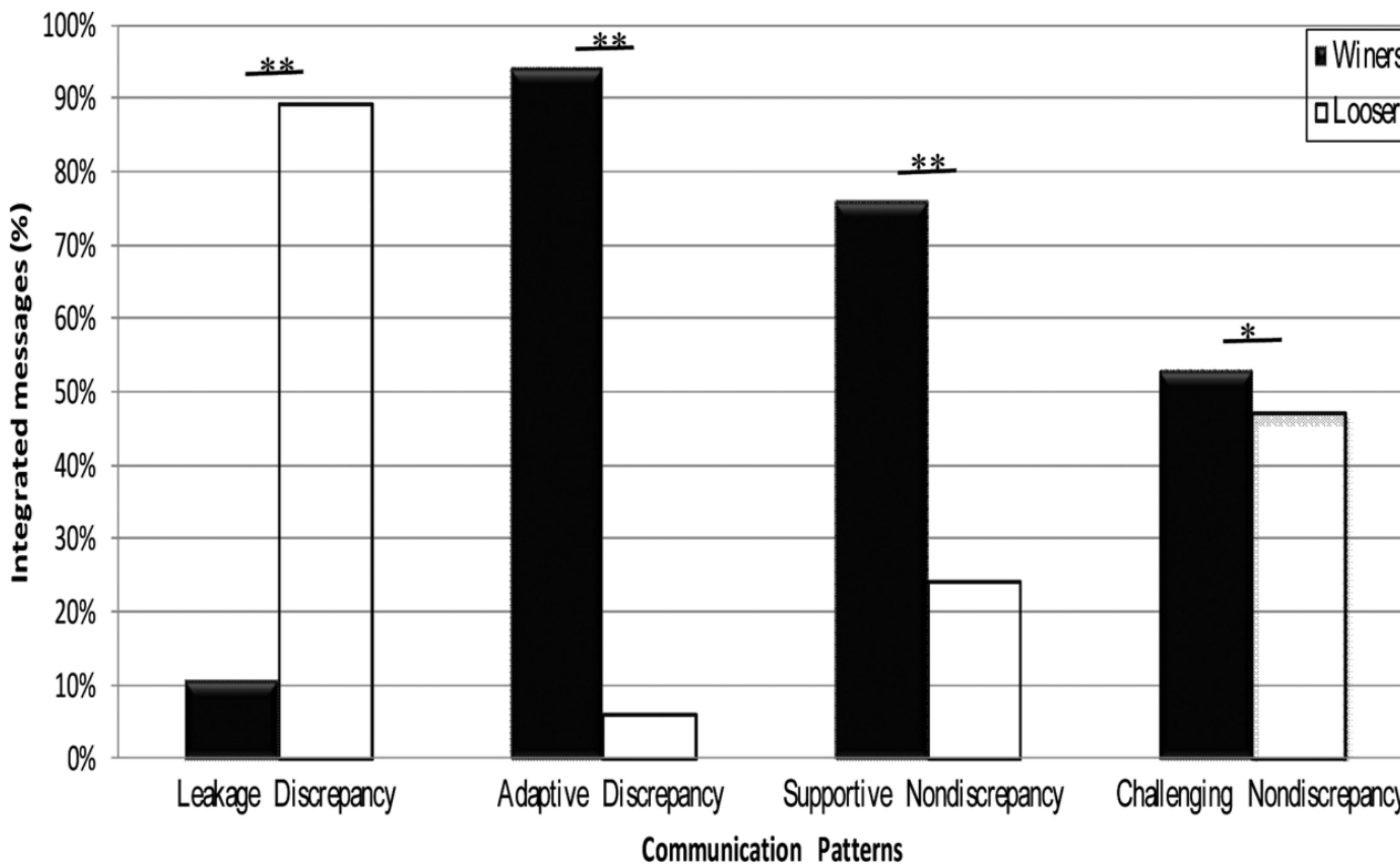
Grebelsky-Lichtman, Tsfira. "The role of verbal and nonverbal behavior in televised political debates." *Journal of Political Marketing* 15.4 (2016): 362-387.

# Вербальное и невербальное поведение в политических дебатах

- Возможны 4 типа комбинаций верbalного и неверbalного challenge/support:
  - Supportive nondiscrepancy: вербальная и невербальная поддержка
  - Challenging nondiscrepancy: вербальный и невербальный вызов
  - Leakage-type discrepancy: вербальная поддержка, невербальный вызов
  - Adaptive discrepancy: вербальный вызов, невербальная поддержка
- Первая, возможно, вторая и последняя стратегии считаются эффективными в дебатах и ведут к выигрышу.

# Вербальное и невербальное поведение в политических дебатах

- В исследование включили дебаты с 7 участниками
- Определили победивших и проигравших
- Проанализировали вербальное и невербальное (голос, лицо, тело) поведение (не автоматически): соотнесли с категориями support и challenge
- Соотнесли стратегии поведения с победами и поражением
  - Для победивших были типичными стратегии: Supportive nondiscrepancy (верbalная и nonverbalная поддержка) и Adaptive discrepancy (вербальный вызов, невербальная поддержка)
  - Для проигравших типичной была стратегия Leakage-type discrepancy (вербальная поддержка, невербальный вызов)



Multimodal emotions –  
computers



# Human sensing modalities



Behavior

Physiological  
response

Audio

Visual

Peripheral

Central

Language

Prosody

Face

Body

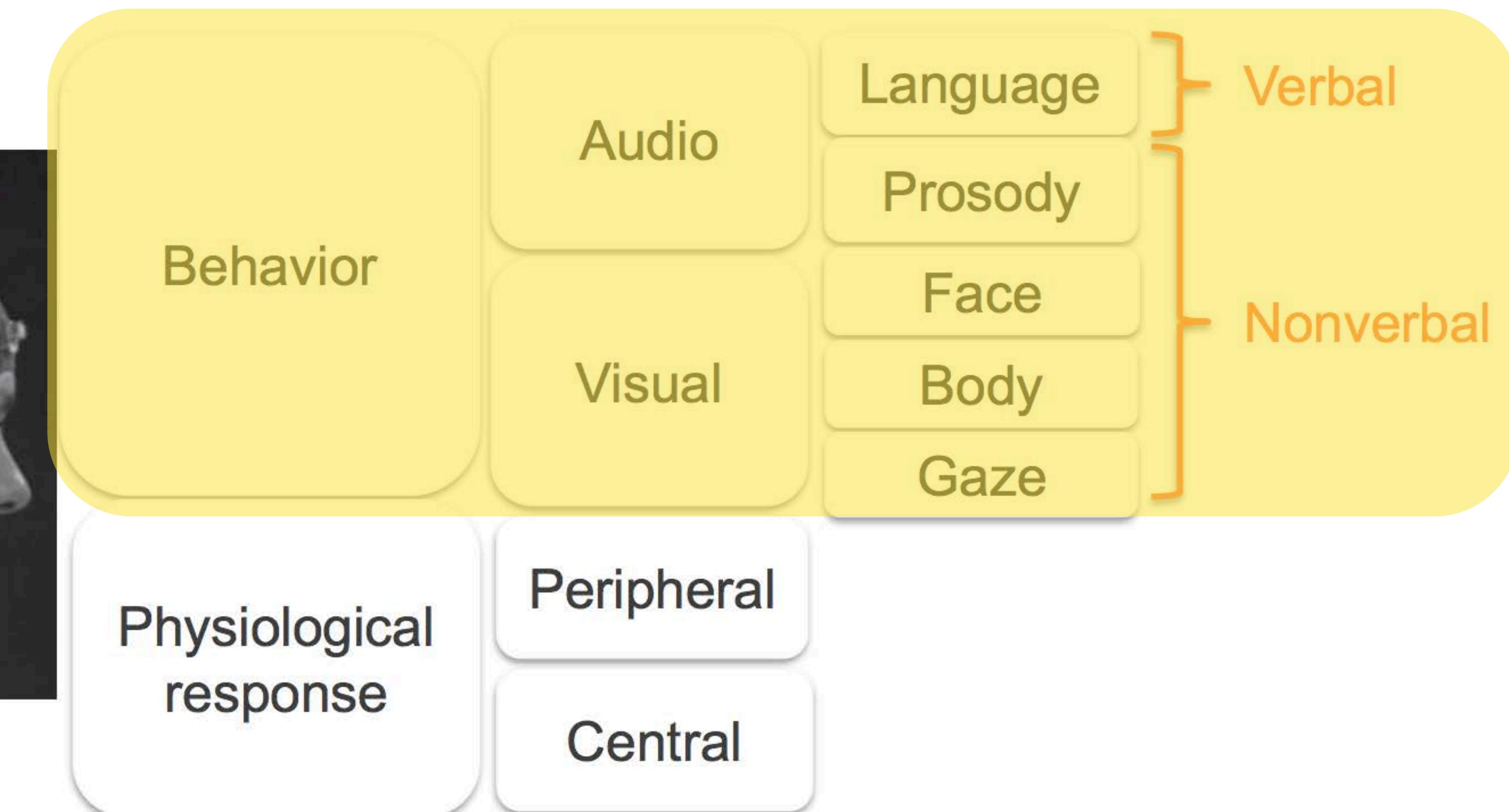
Gaze

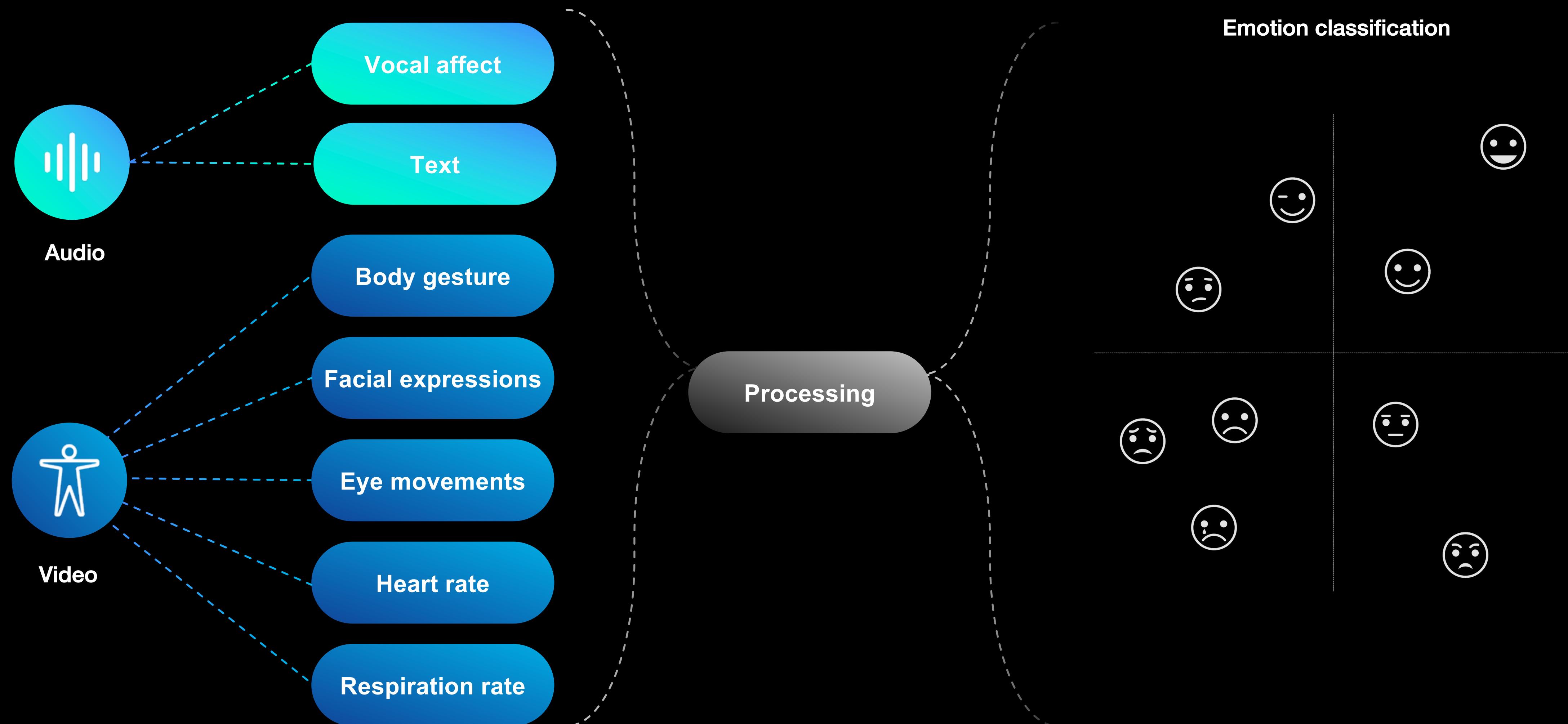
Verbal

Nonverbal



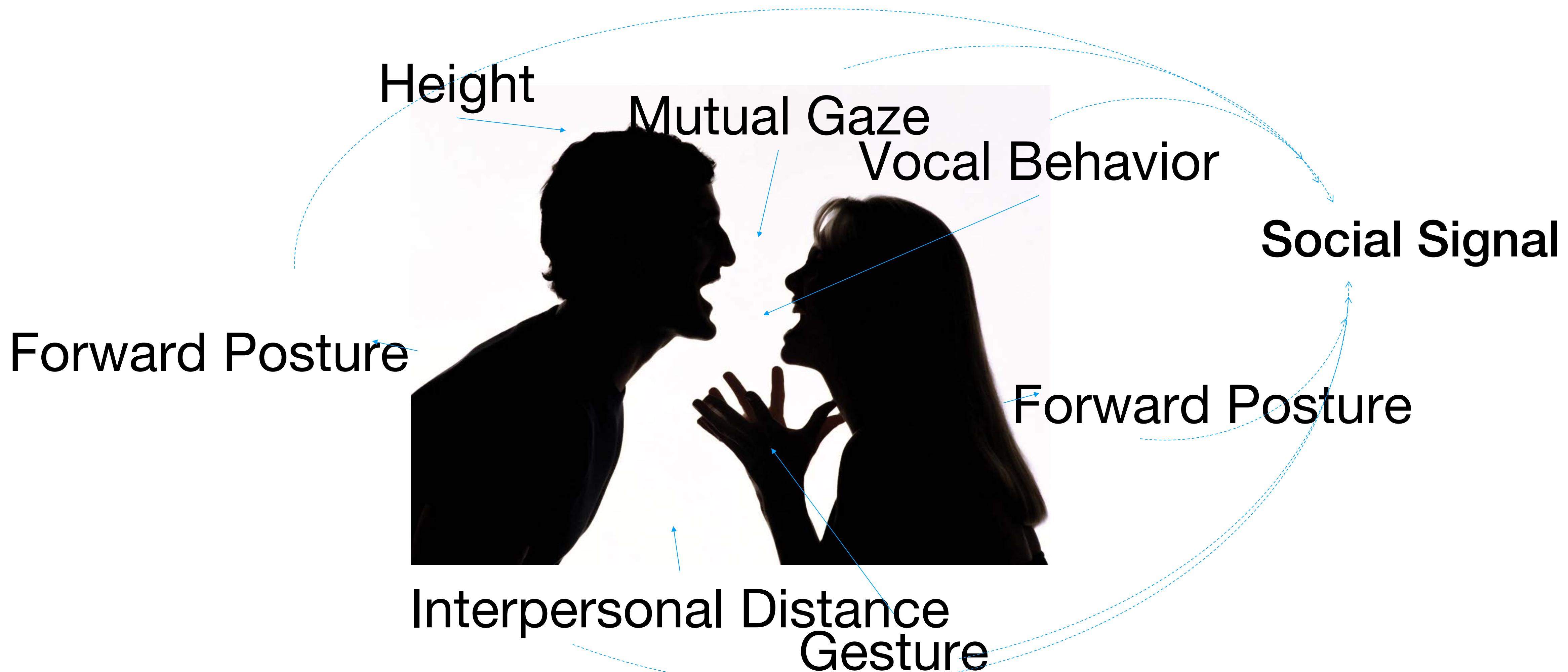
# Human sensing modalities



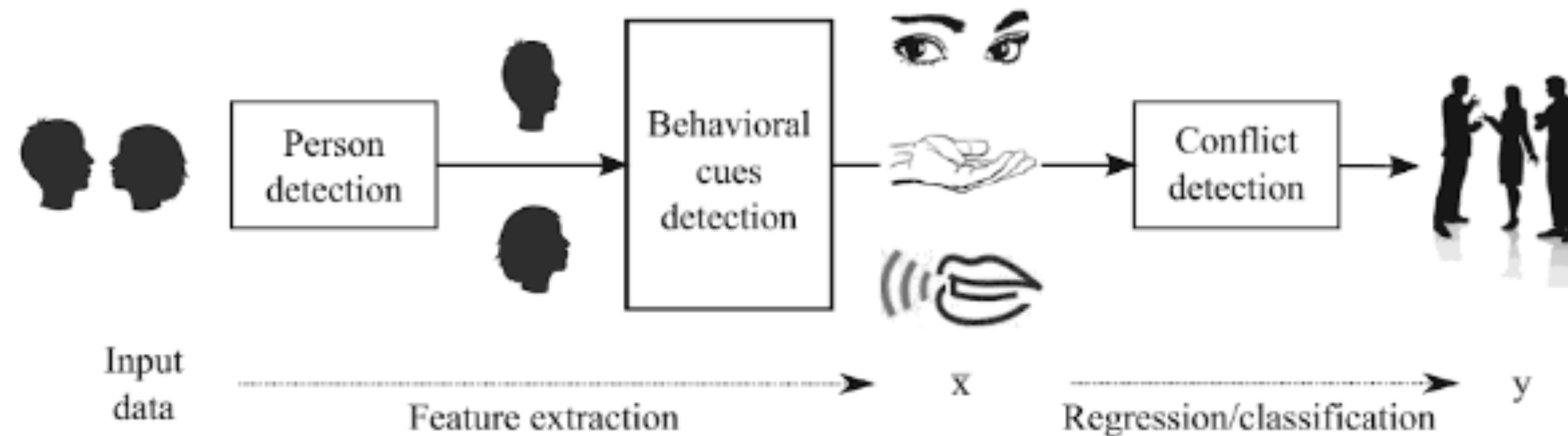


# Social Signal Processing

Nonverbal behavior cues



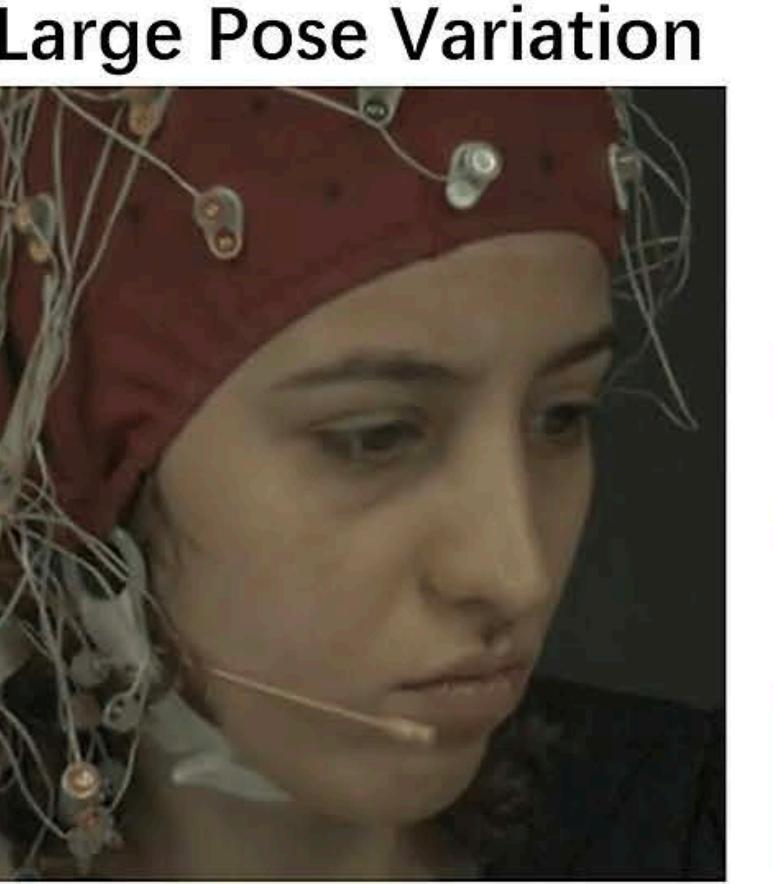
# Conflict detection



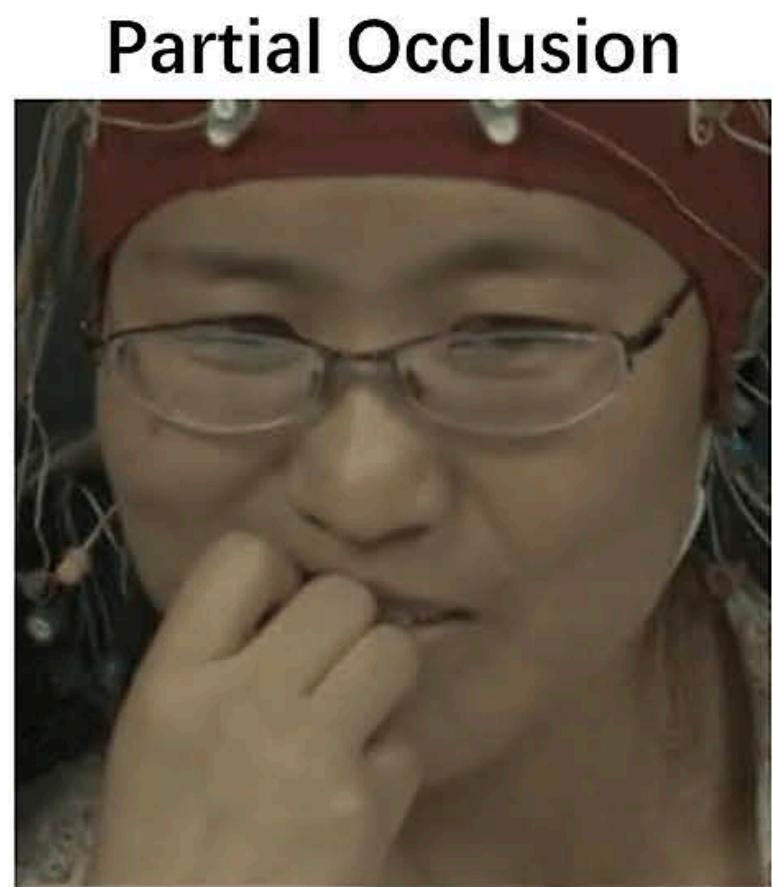
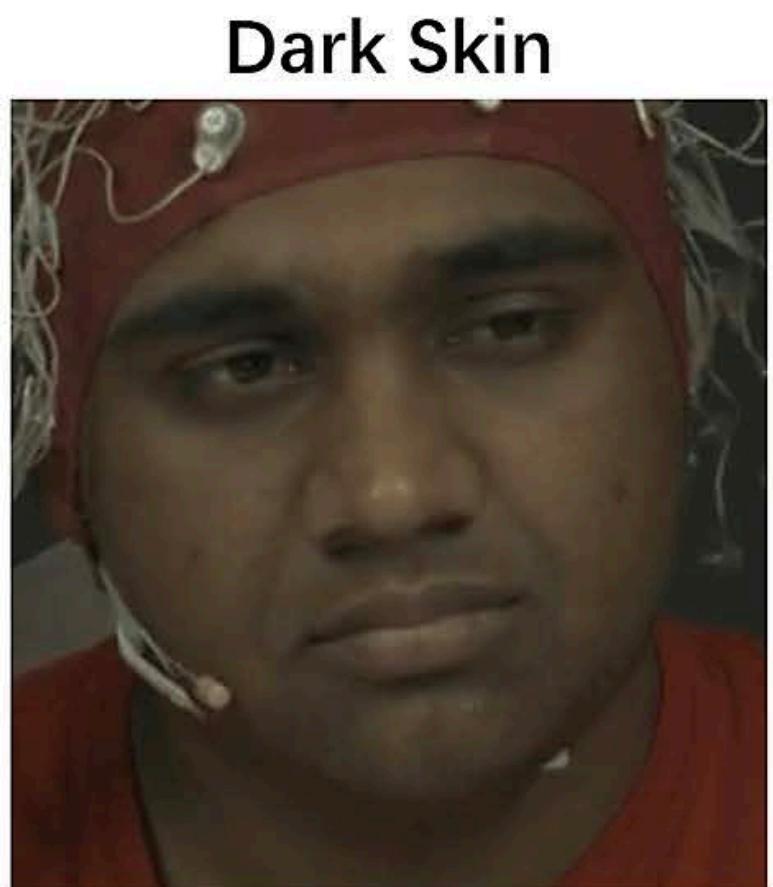
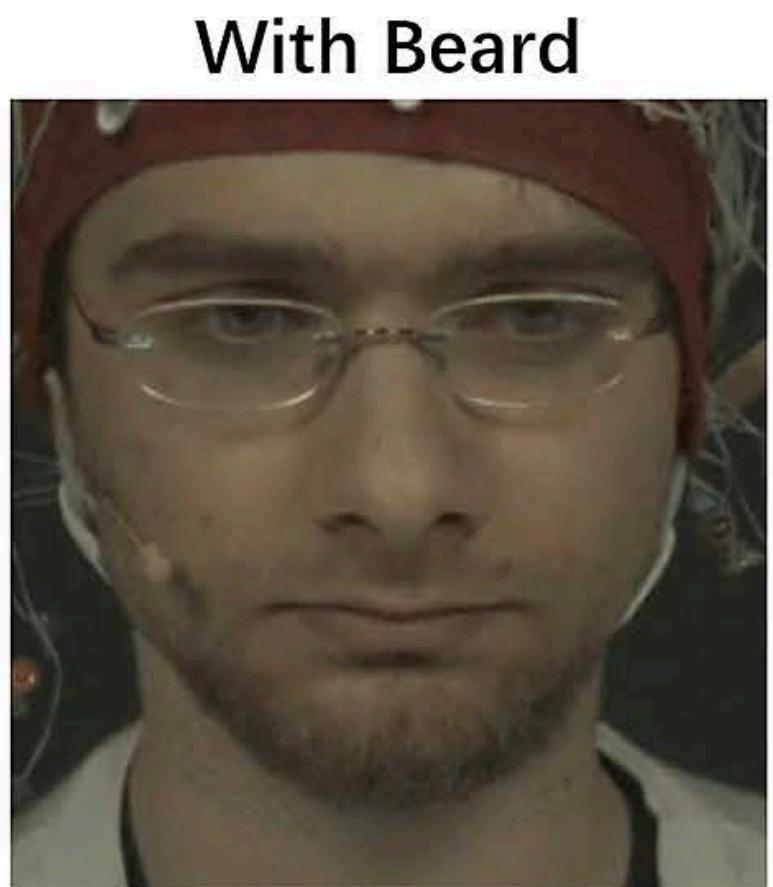
IEMOCAP (Busso et al., 2008)	<ul style="list-style-type: none"> <li>• Motion capture data</li> <li>• Audio</li> <li>• Video</li> </ul>	10 actors <b>11.5 hours</b>	Happiness, anger, sadness, frustration and neutral state
MAHNOB (Soleymani et al., 2011)	<ul style="list-style-type: none"> <li>• Video</li> <li>• Audio</li> <li>• Eye Tracking</li> <li>• EEG</li> <li>• ECG</li> <li>• Respiration</li> <li>• Skin temperature</li> </ul>	30 actors <b>9 hours</b>	Valence, Arousal, Dominance
RECOLA (Rengeval et al., 2013)	<ul style="list-style-type: none"> <li>• Video</li> <li>• Audio</li> <li>• EDA</li> <li>• ECG</li> </ul>	46 participants <b>5 hours</b>	Valence, arousal, social dimension
CMU-MOSEI (Zadeh et al., 2018)	<ul style="list-style-type: none"> <li>• Video</li> <li>• Audio</li> </ul>	1000 speakers <b>65 hours</b>	Happiness, sadness, anger, fear, disgust, surprise
EmotionMiner (Neurodata Lab)	<ul style="list-style-type: none"> <li>• Video</li> <li>• Audio</li> </ul>	<b>140 hours</b>	22 categories

# IEMOCAP





**MAHNOB**



# RECOLA

[arousal - SEQUENCE 2.mp4](#)

[valence - SEQUENCE 2.mp4](#)

[social dimension annotation](#)

[arousal - SEQUENCE 3.mp4](#)

[valence - SEQUENCE 3.mp4](#)

[social dimension annotation](#)

[arousal - SEQUENCE 4.mp4](#)

[valence - SEQUENCE 4.mp4](#)

[social dimension annotation](#)

[arousal - SEQUENCE 5.mp4](#)

[valence - SEQUENCE 5.mp4](#)

[social dimension annotation](#)

[arousal - SEQUENCE 6.mp4](#)

[valence - SEQUENCE 6.mp4](#)

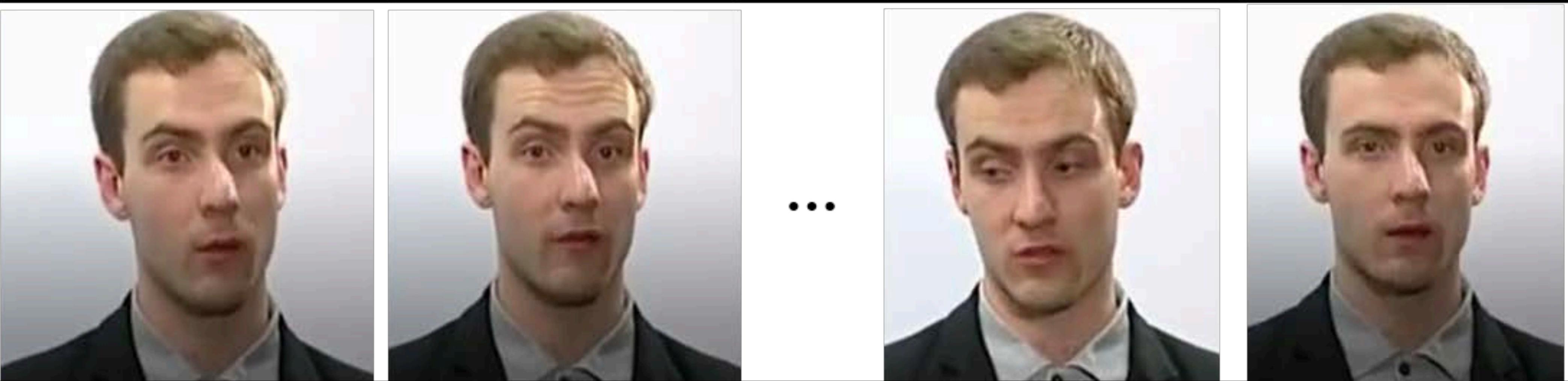
[social dimension annotation](#)



very negative

very positive

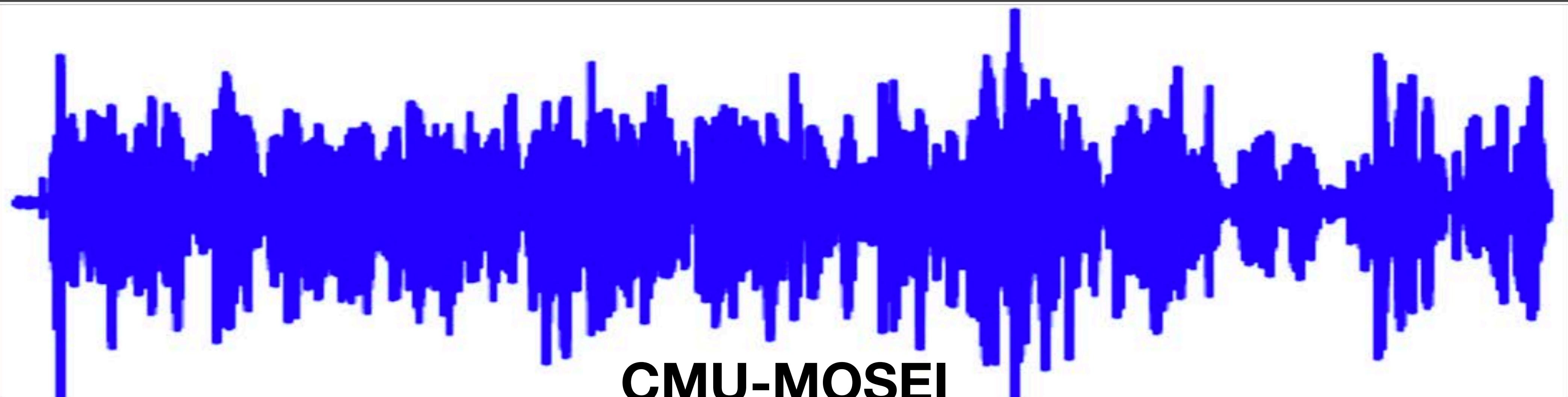
**Face**



**Text**

... the teaching assistantship lasted for 9 months, uhhh ... Interacting in English, not my language ...

**Speech**

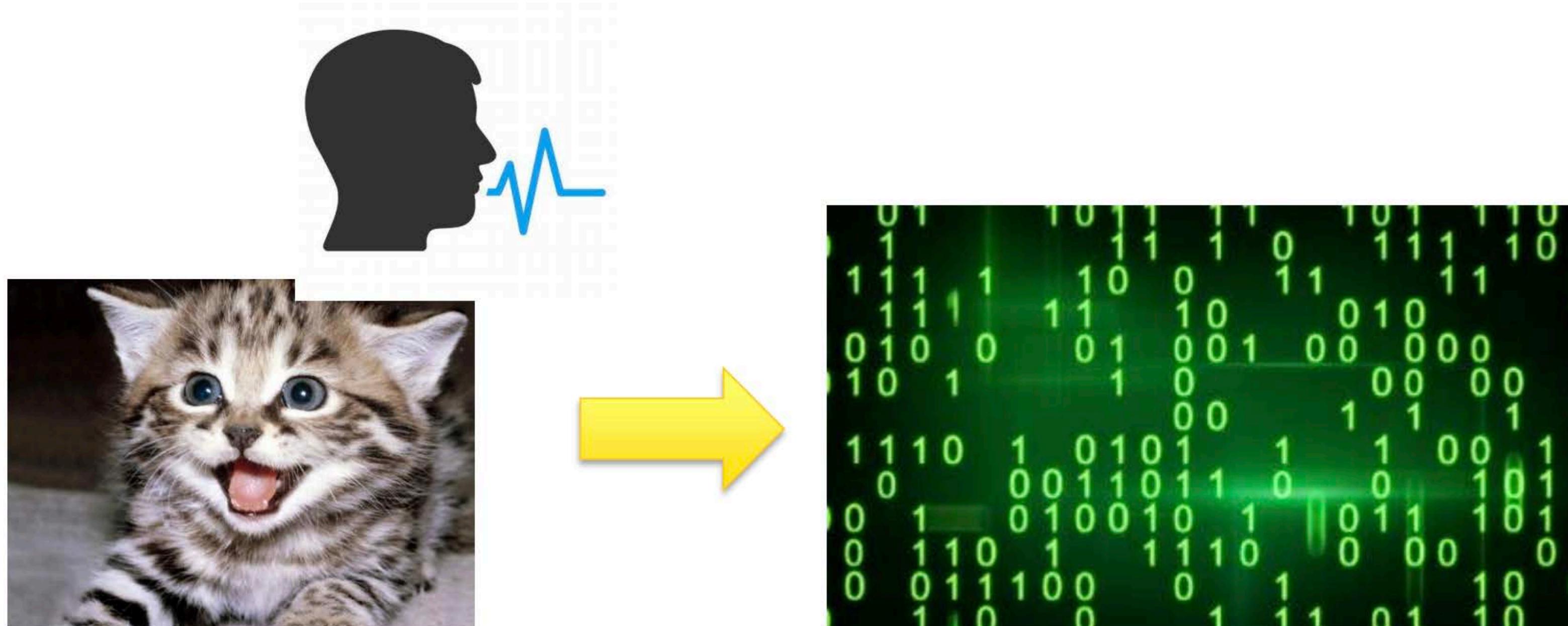


# Multimodal affect detection systems

- По обзору 2015 г. 90 мультимодальных (ММ) систем – большинство ММ систем являются:
  - person-dependent (62,2%)
  - Используют аудио+видео каналы (55,6%)
  - Используют датасеты с сыгранными эмоциями (52,2%)
  - Детектируют базовые эмоции и простые непрерывные измерения arousal-valence (64,5%)

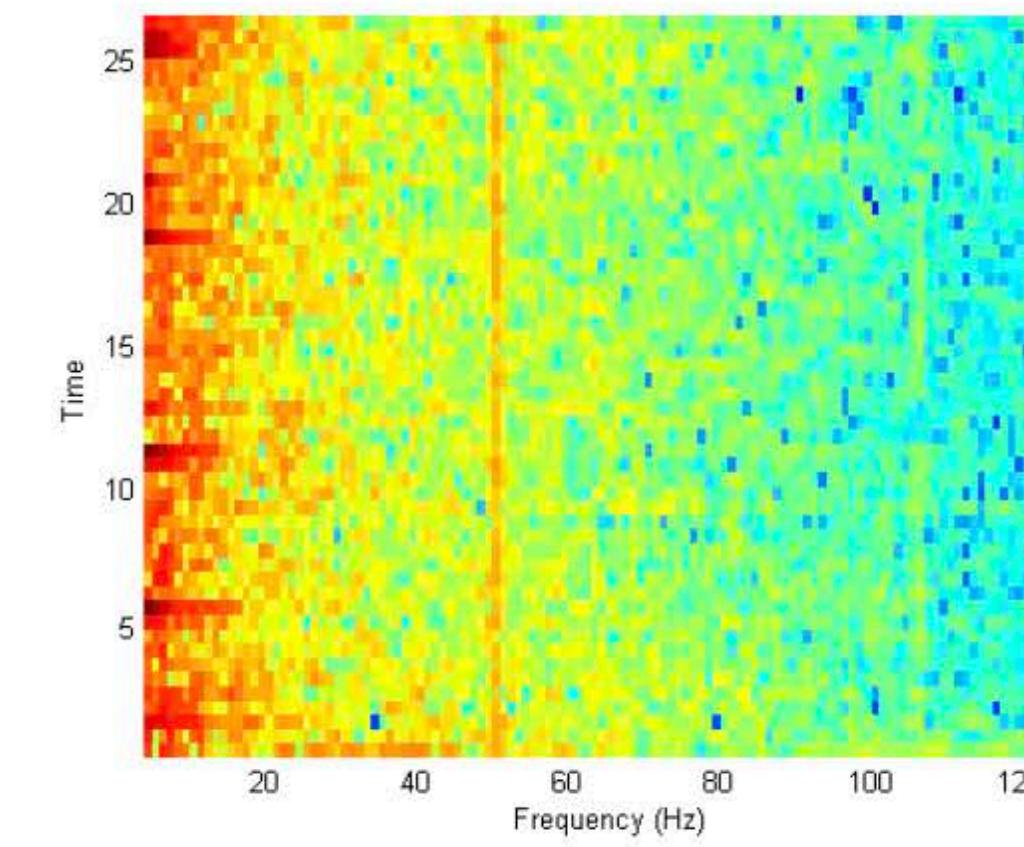
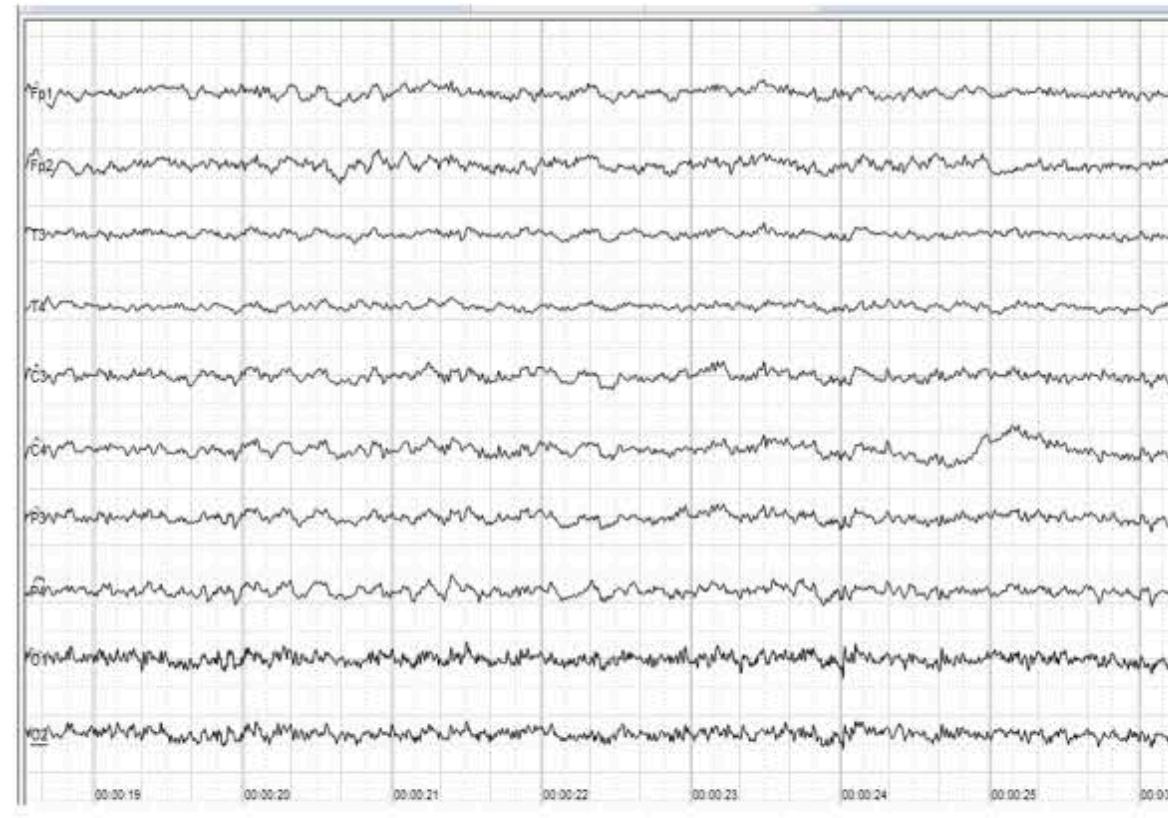
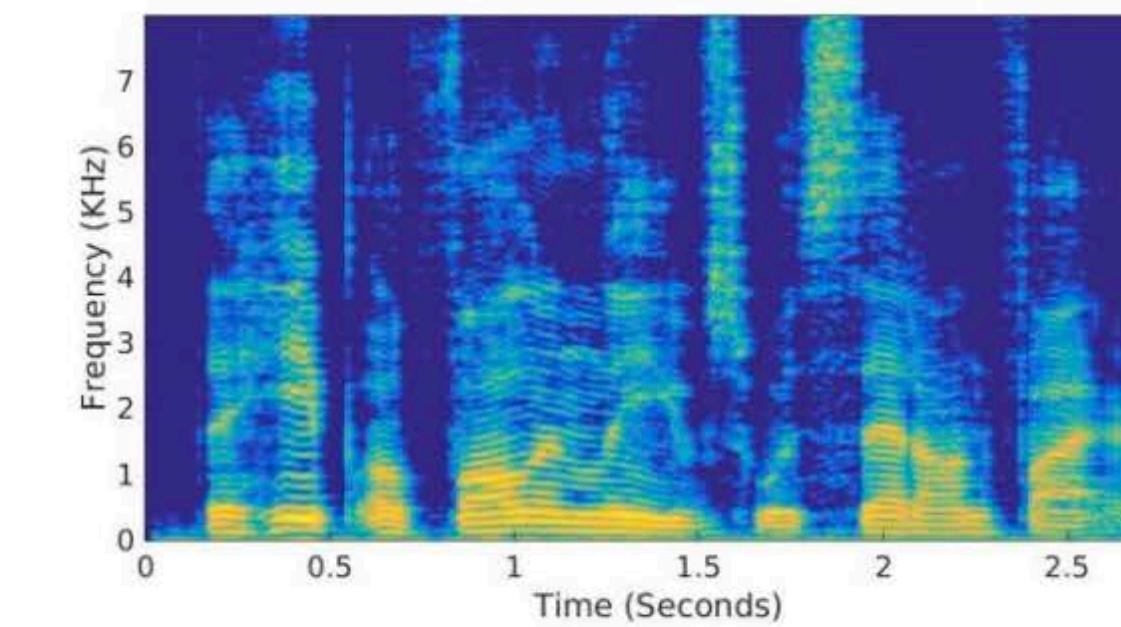
D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." ACM Computing Surveys (CSUR) 47.3 (2015): 43

# Representation Learning



! ? punctuation  
! ? adjective ;  
Language  
word

# Features

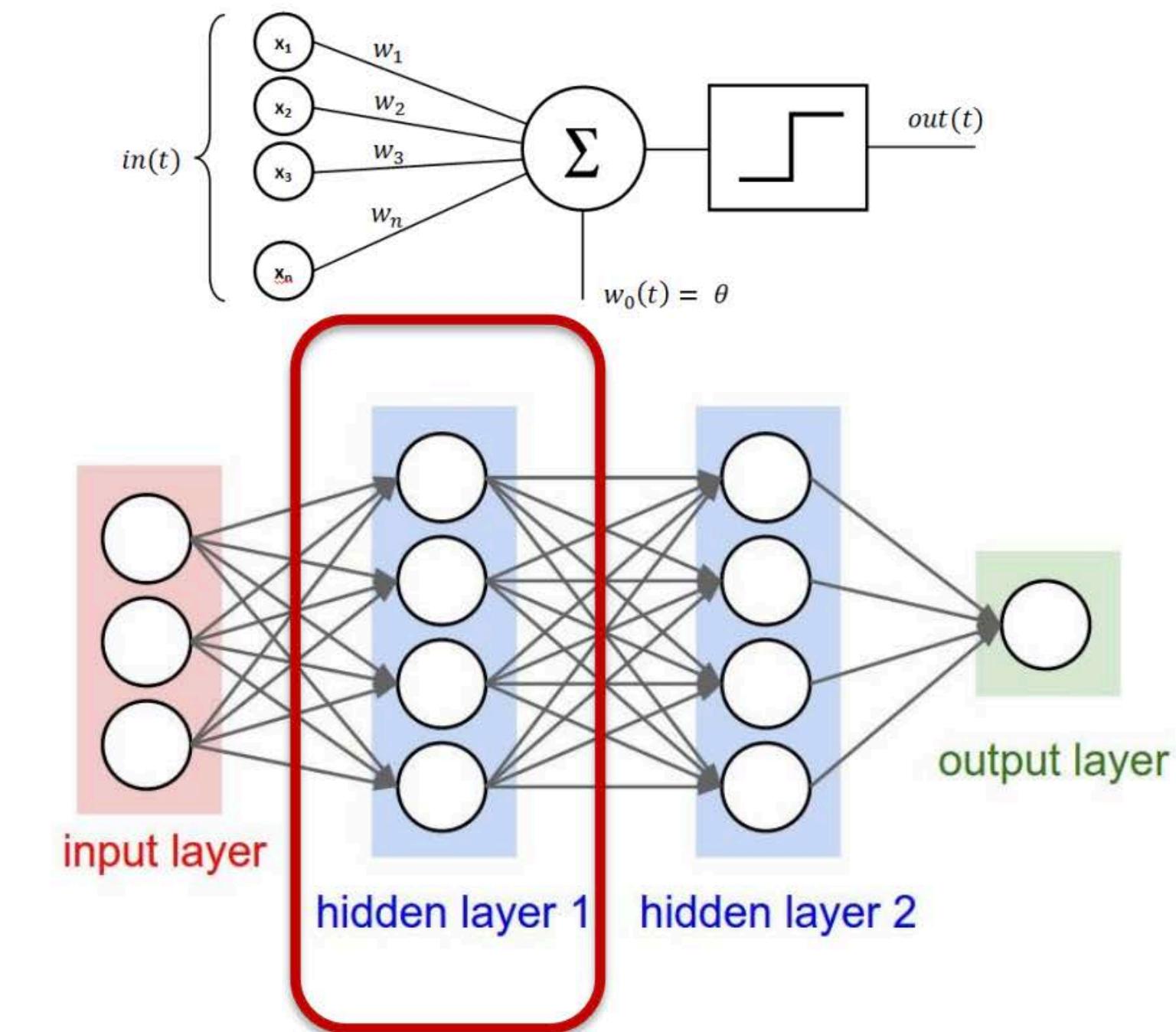


# Features



## Learning representations – neural networks

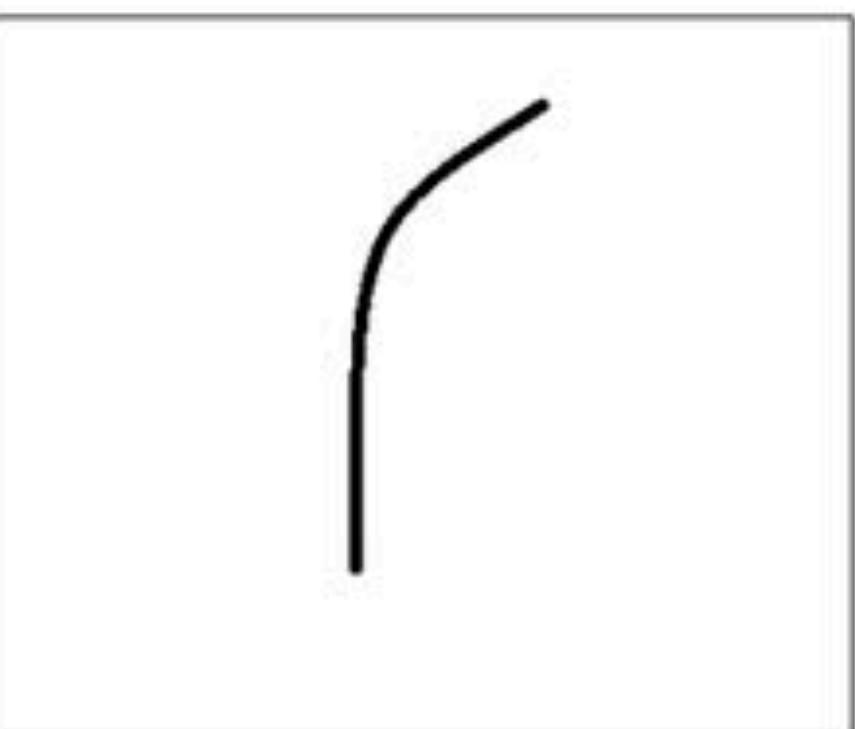
- Perceptron
- Multi-layer perceptron



Learns representation

## Пример ядра с обученным признаком

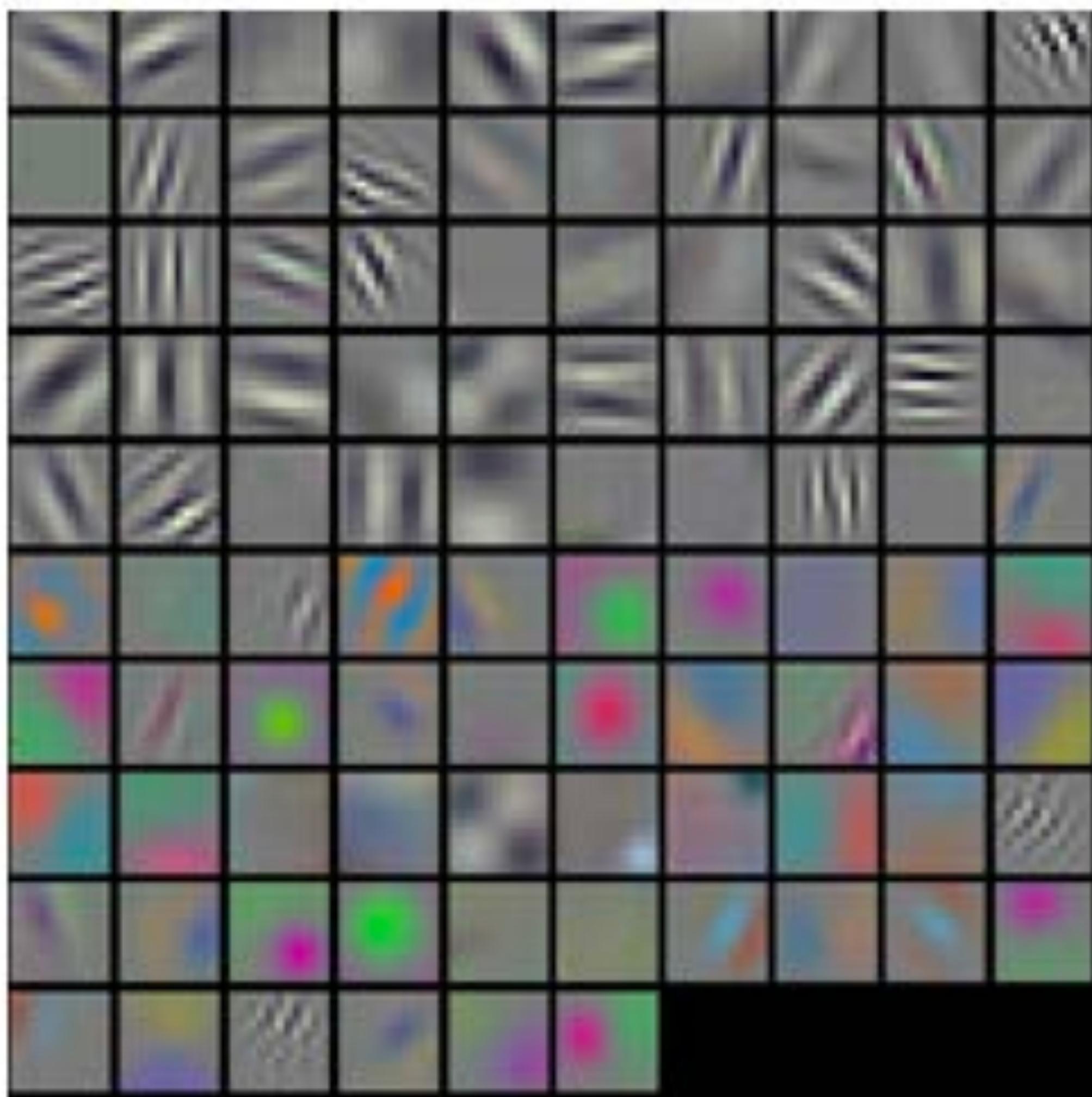
0	0	0	0	0	30	0	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0



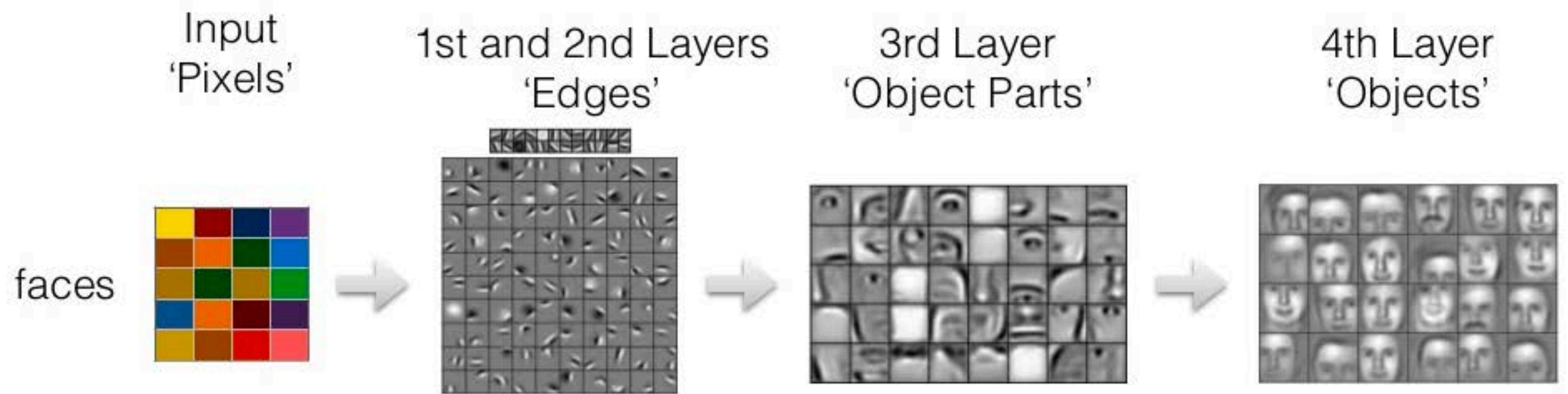
## Входное изображение



В этом месте, где желтое окно, будет большой отклик (сигнал), что говорит о наличие этого признака на изображении.

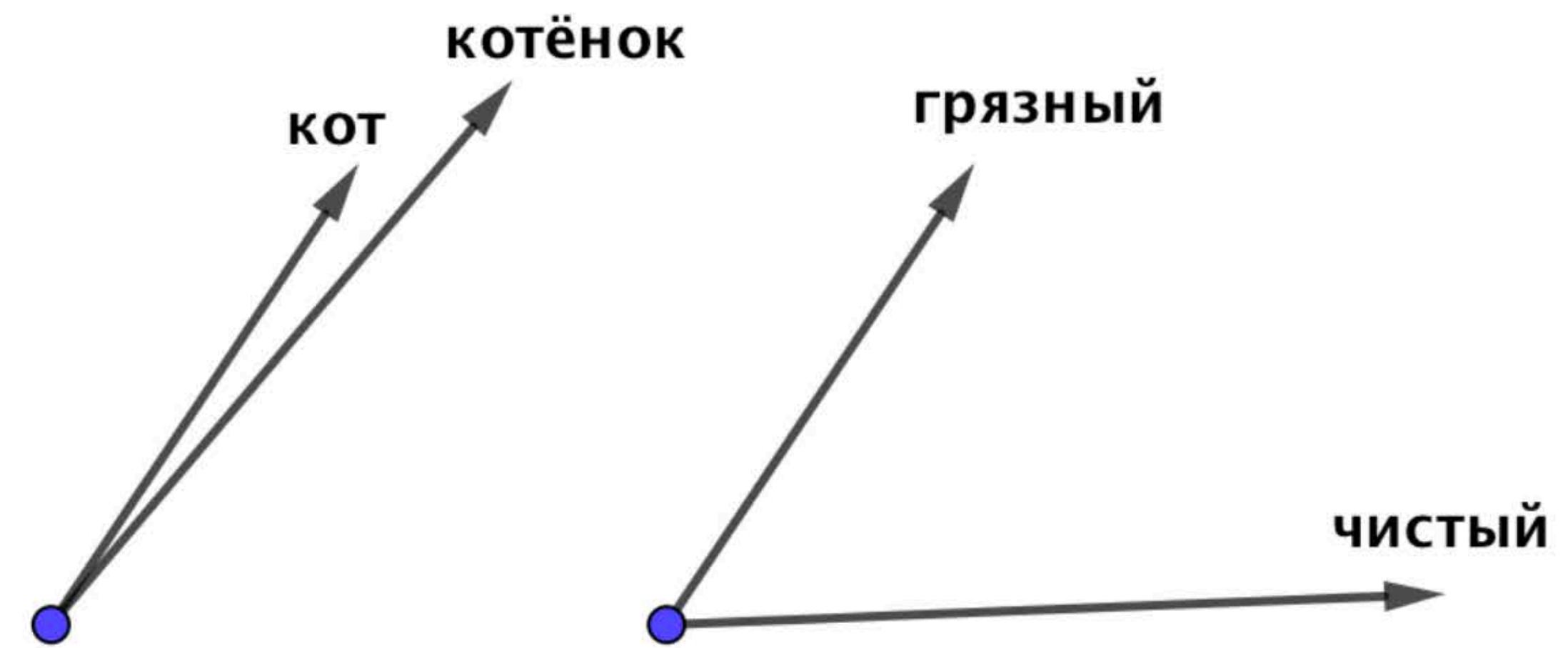


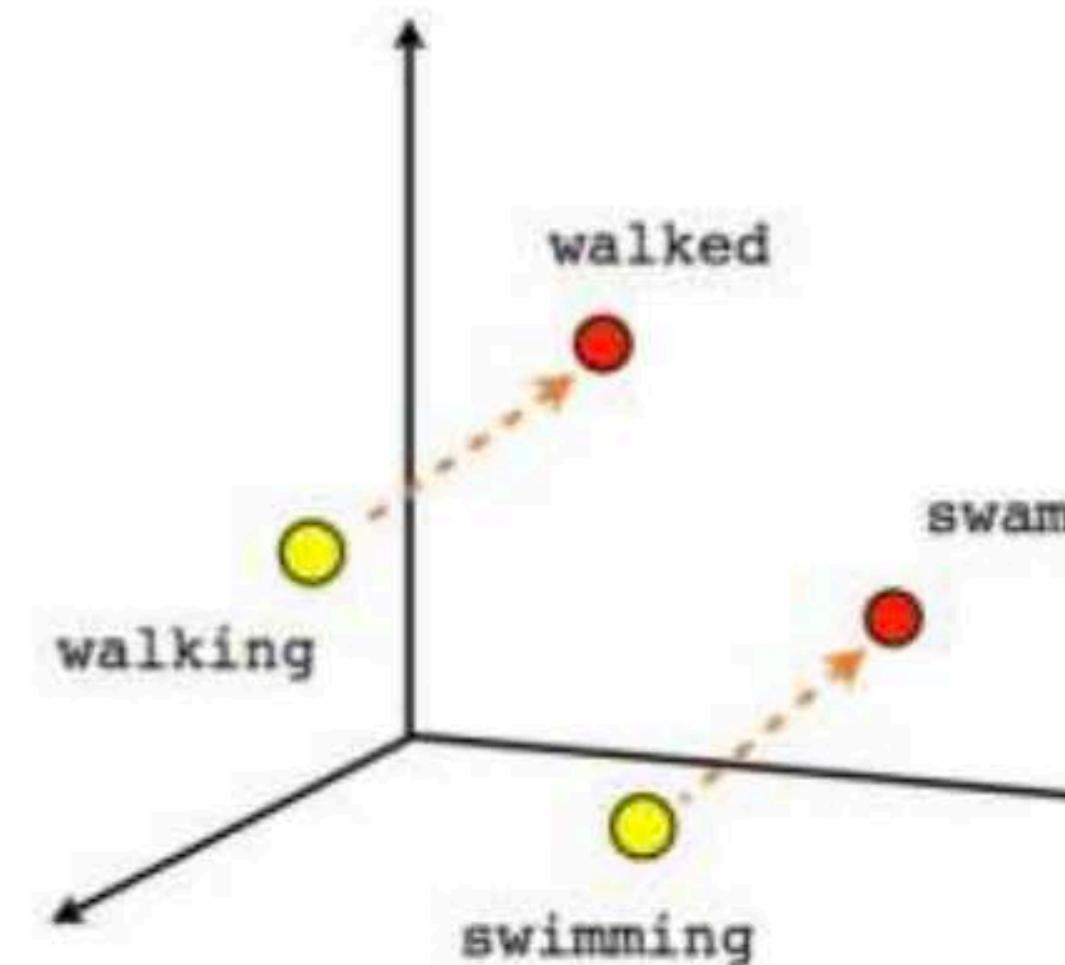
Visualizations of filters



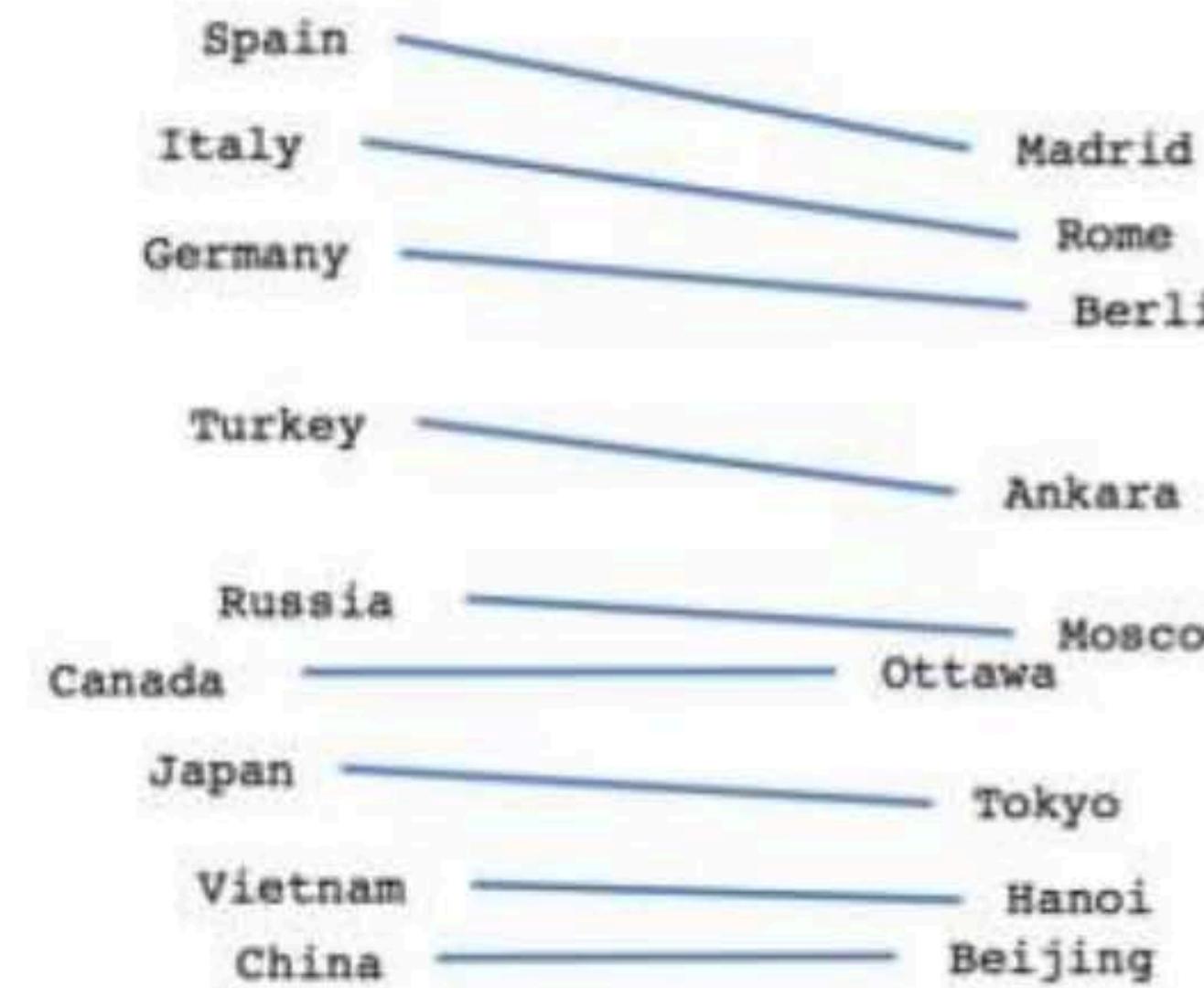
# Word2Vec

- Опишем каждое слово вектором так, чтобы похожие слова были близкими векторами
- Давайте оценивать вероятность встретить слово i в контексте слова j
- Обучаем модель на корпусе текстов естественной речи (Google news, Tweets...)

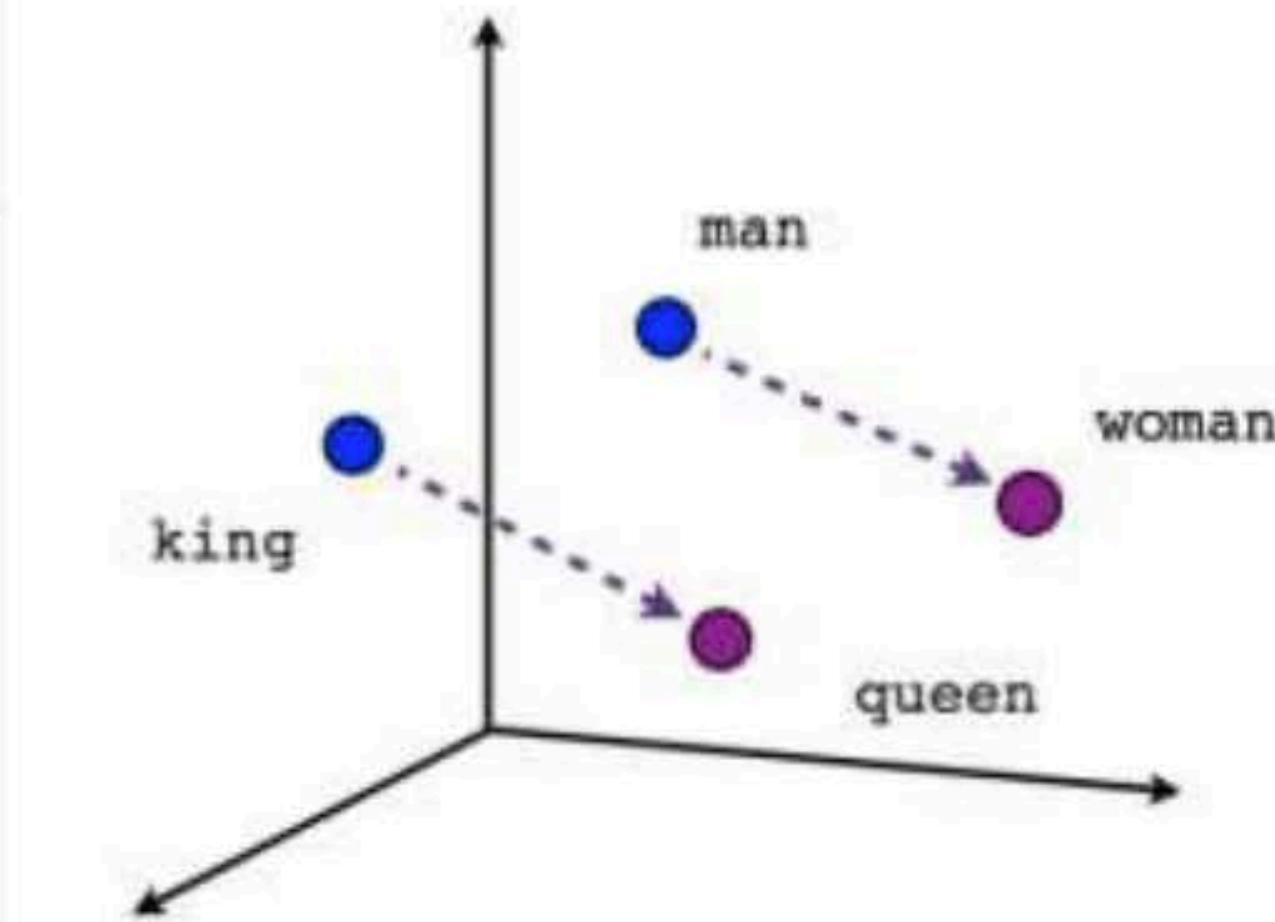




Verb tense



Country-Capital



Male-Female

$$\text{vector[King]} - \text{vector[Man]} + \text{vector[Woman]} = \text{vector[Queen]}$$

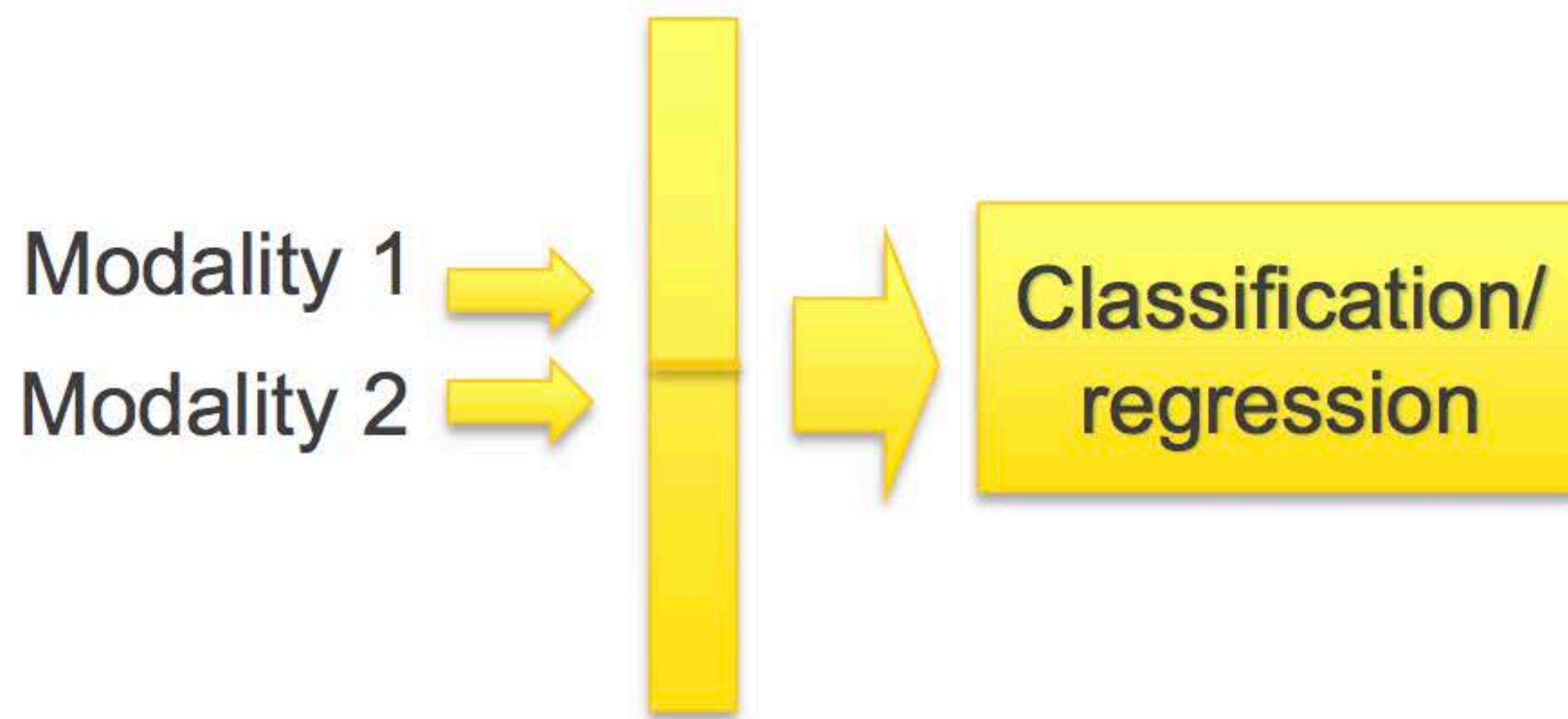
# Multimodal Fusion

- Feature-based (early)
- Decision-based (late)
- Hybrid fusion

# Feature-based fusion

- Integrates features immediately after they are extracted.
- Training of a single model.
- Проблемы:
  - фичи из разных каналов имеют разные time scales
  - Много фич от кучи каналов => нагрузка на вычислительную мощность

- Concatenate features from different modalities
- Challenges
  - Requires the same sampling rate
  - Sensitive to the dimensionality of each modality



# Decision-based fusion

- Performs integration after each of the modalities has made a decision.
- Averaging, voting schemes, weighting sum, etc.
- Проблемы:
  - Decision-based fusion ignores the low level interaction between the modalities.

- Requires two-step training
- Do not inter-modality interactions
- Fusion techniques
  - Majority vote
  - Linear combination of confidence scores
  - Product of confidence scores



# Hybrid fusion

- Combines outputs from early fusion and individual unimodal predictors.
- Например, объединение в два шага: 1) feature-level fusion от аудио+видео сигналов, 2) decision-level fusion от результатов первого классификатора и классификатора по физиологическим данным.



Table 3  
Descriptive Statistics on Study Features

<b>Dimension</b>	<b>Prop.</b>	<b>Dimension</b>	<b>Prop</b>
<b>Data type</b>		<b>Measure. model</b>	
Acted	.522	Discrete	.644
Induced	.278	Dimensional	.356
Natural	.200		
<b>Detection model</b>		<b>Affect detected</b>	
Classification	.922	Disc. basic	.367
Regression	.078	Disc. non-basic	.078
		Discrete mixed	.178
<b>No. of modalities</b>		Dim. simple	.278
Bimodal	.867	Dim. complex	.100
Trimodal	.133		
<b>Fusion method</b>			
<b>Modality</b>		Feature	.389
Face	.767	Decision	.356
Voice	.822	Hybrid	.056
Text	.167	Model	.200
Body	.133		
<b>Eye Gaze</b>		<b>Validation method</b>	
Peri. Physio.	.011	Person indep.	.378
Central Physio.	.111	Person dep.	.622
Content	.056		
	.067		

Notes. Prop. = Proportion; Peri. = Peripheral; Physio. = Physiology; Content = Content/Context; Measure. = Measurement; Disc. = Discrete; Dim. = Dimensional; Indep. = Independent; Dep. = Dependent.

# Multimodal recognition

- Тип датасета влияет на точность классификации: сыгрынные датасеты дают более высокую точность, чем натуральные.
- Нет четких данных о том, какой тип и количество модальностей нужно для достижения наиболее точной классификации. Неясно, как каждая модальность вносит вклад в общую производительность системы (таких исследований по сравнению вклада модальностей мало).

Al Osman, Hussein, and Tiago H. Falk. "Multimodal Affect Recognition: Current Approaches and Challenges." Emotion and Attention Recognition Based on Biological Signals and Images. InTech, 2017

# Multimodal fusion

- Мета-анализ 90 исследований Unimodal (UM) vs Multimodal (MM) affect detection показал, что мультимодальный подход дает большую точность (в среднем на 9,83%, для 85% систем).
- Повышение точности при использовании мультимодального подхода на натуральных датасетах более скромное (4,59%) по сравнению с сыгранными датасетами (12,7%)

D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." *ACM Computing Surveys (CSUR)* 47.3 (2015): 43

# Multimodal fusion

- Точность лучшего УМ классификатора в значительной степени объясняет дисперсию ММ классификатора (80,6%).
- Мультимодальная точность существенно выше второго по производительности УМ классификатора (среднее улучшение на 29,4%).

D'Mello, Sidney, and Jacqueline Kory. "Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies." Proceedings of the 14th ACM international conference on Multimodal interaction. ACM, 2012

# Multimodal fusion

- В более позднем исследовании было также показано, что между точностью лучшего УМ и ММ классификаторами наблюдается сильная положительная корреляция ( $r(88)=0.87$ ,  $p<0.01$ )
- Лучший УМ классификатор был значимым предиктором в линейной регрессии, а второй по качеству УМ классификатор – уже не значим.
- => Существенную долю дисперсии ММ классификатора объясняет лучший УМ классификатор.
  - D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." *ACM Computing Surveys (CSUR)* 47.3 (2015): 43

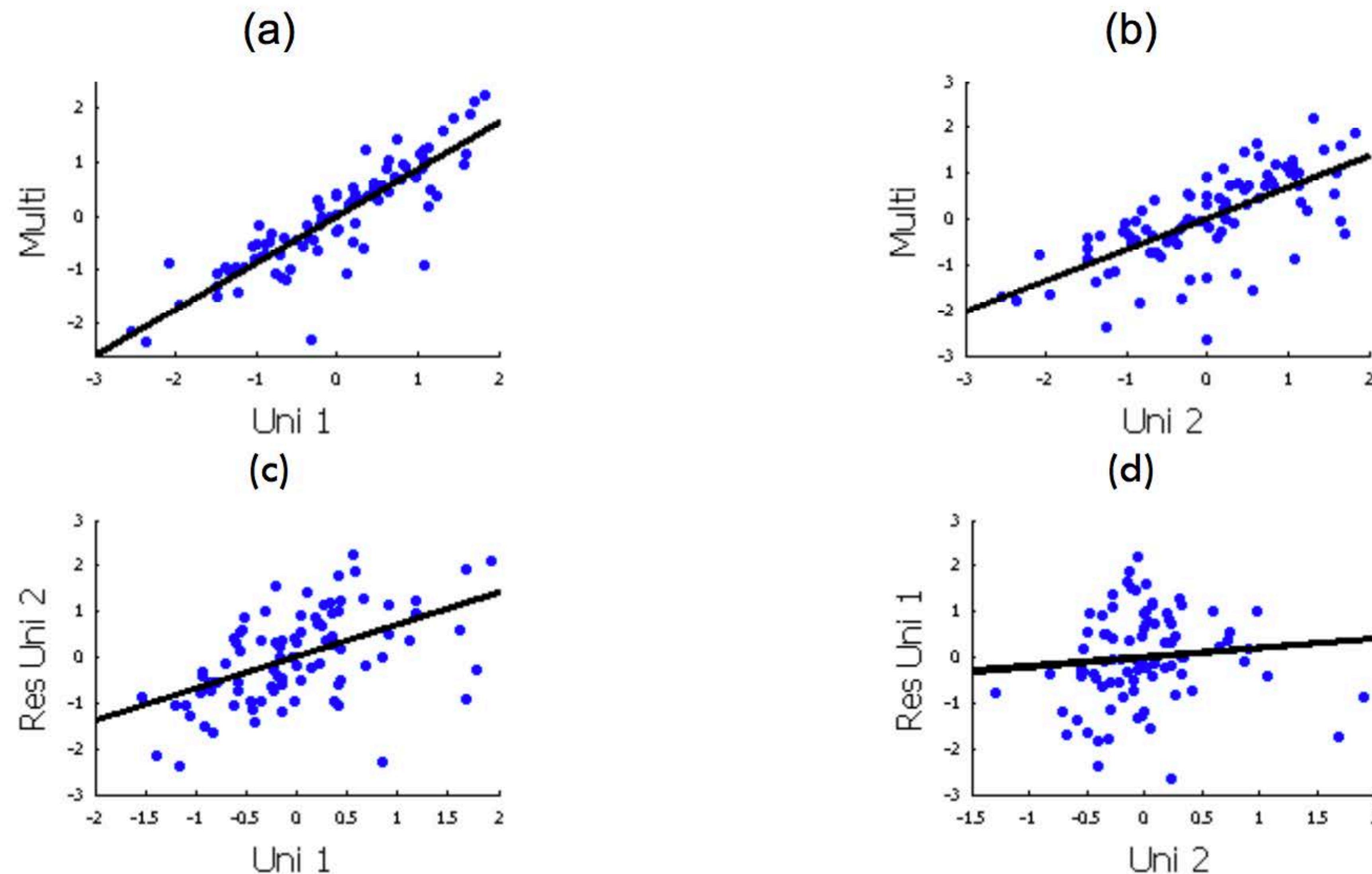


Figure 3. Scatter plots denoting relationships between MM and UM accuracy along with regression line for: (a) regression of MM (Multi) on best UM (Uni 1) accuracy; (b) regression of MM (Multi) on second-best UM (Uni 2) accuracy; (c) same as (a) but after controlling for second-best UM accuracy; and (d) same as (b) but after controlling for best UM accuracy.

# Multimodal fusion

- В некоторых работах однако лучший УМ классификатор справляется лучше, чем ММ система (на RECOLA)
- Вклад отдельных каналов также может быть разным: например, модели, основанные только на аудио, лучше распознавали страх, а модели на зрительные признаках – лучше отвращение (EmotiW challenges).
- Kaya, Heysem, Furkan Gürpinar, and Albert Ali Salah. "Video-based emotion recognition in the wild using deep transfer learning and score fusion." *Image and Vision Computing* (2017).

# Вклад каналов: аудио

- CALLAS Expressivity Corpus.
- 4 эмоции (positive-low, positive-high, negative-low, negative-high).
- 3 канала: аудио, видео (лицо), жесты.
- Отдельные классификаторы по каждому из каналов: **аудио 51%, лицо 48%, жесты 42%**.
- Fusion: 49-55% (в зависимости от техники).
  - Wagner, Johannes, et al. "Exploring fusion methods for multimodal emotion recognition with missing data." IEEE Transactions on Affective Computing 2.4 (2011): 206-218.

# Вклад каналов: жесты

- Датасет по типу GEMEP (собранный на студенческой школе)
- Эмоции: Anger, despair, interest, pleasure, sadness, irritation, joy and pride
- Каналы: Visual (face, body) and audio
- Отдельные классификаторы: лицо 48.3%, **жесты 67.1%**, голос 57.1%.
- Fusion: **Feature-level 78,3%** (все кроме despair выше 70%), Decision-level 74,6%
  - Castellano, Ginevra, Loic Kessous, and George Caridakis. "Emotion recognition through multiple modalities: face, body gesture, speech." Affect and emotion in human-computer interaction (2008): 92-103

# **Выводы**

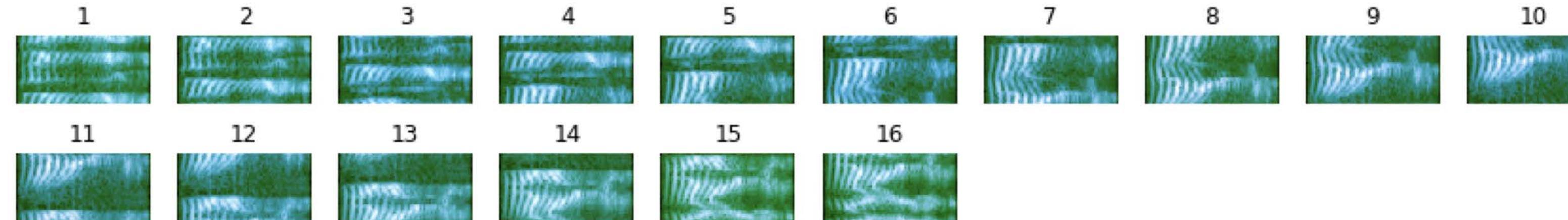
- Люди лучше всего распознают эмоции и ложь по лицу (вклад этого канала в восприятие эмоций и лжи наибольший).
- Каналы влияют друг на друга (лицо, голос, тело):
  - конгруэнтные мультимодальные сигналы воспринимаются быстрее и точнее унимодальных;
  - неоднозначные эмоции в одном канале воспринимаются в зависимости от эмоции в другом канале;
- При **рассогласованной** информации от каналов при восприятии эмоций и лжи люди доверяют **лицу**.

- Мультимодальные подходы к автоматическому распознаванию эмоций дают преимущество по сравнению с унимодальными (в среднем на 9,83%).
- Самый лучший унимодальный классификатор в значительной степени объясняет точность ММ классификатора (80,6% дисперсии).
- Сравнение техник объединения каналов (feature-level, decision-level и hybrid-fusion) не показало преимуществ одних над другими.
- Вклад разных каналов – разный, зависит от эмоции (лучше всего по лицу и голосу).

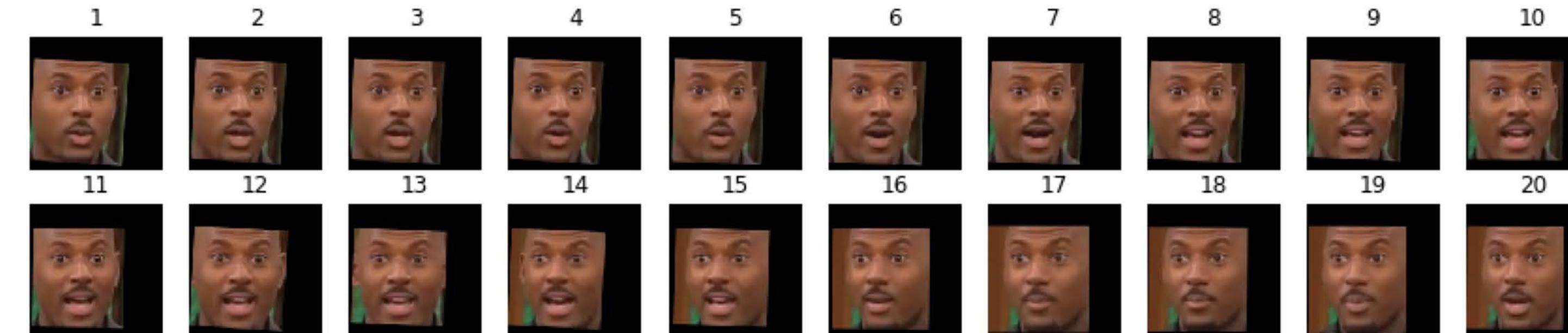
# Исследование

# Emotions: Happiness, Anxiety, Disgust, Surprise, Anger, Sadness, Neutral

## • Audio



## • Video



## • Text

in your head when you get em up people like that you think  
oh no oh no i i got to be on my best behaviour

# Features (признаки)

- A - audio: пересекающиеся фрагменты спектrogramмы + предобученная на MOSEI датасете нейросеть
- F - face: кадры лица + предобученная на EmotionMiner нейросеть
- Т - text:
  - speech2text: получаем наборы слов по имеющимся аудио
  - фиксированные признаки на каждое слово, полученные BERT'ом
- В - body: фиксированные признаки на весь видеоряд, полученные от обучения на классификацию движений

## Классификация по одной модальности:

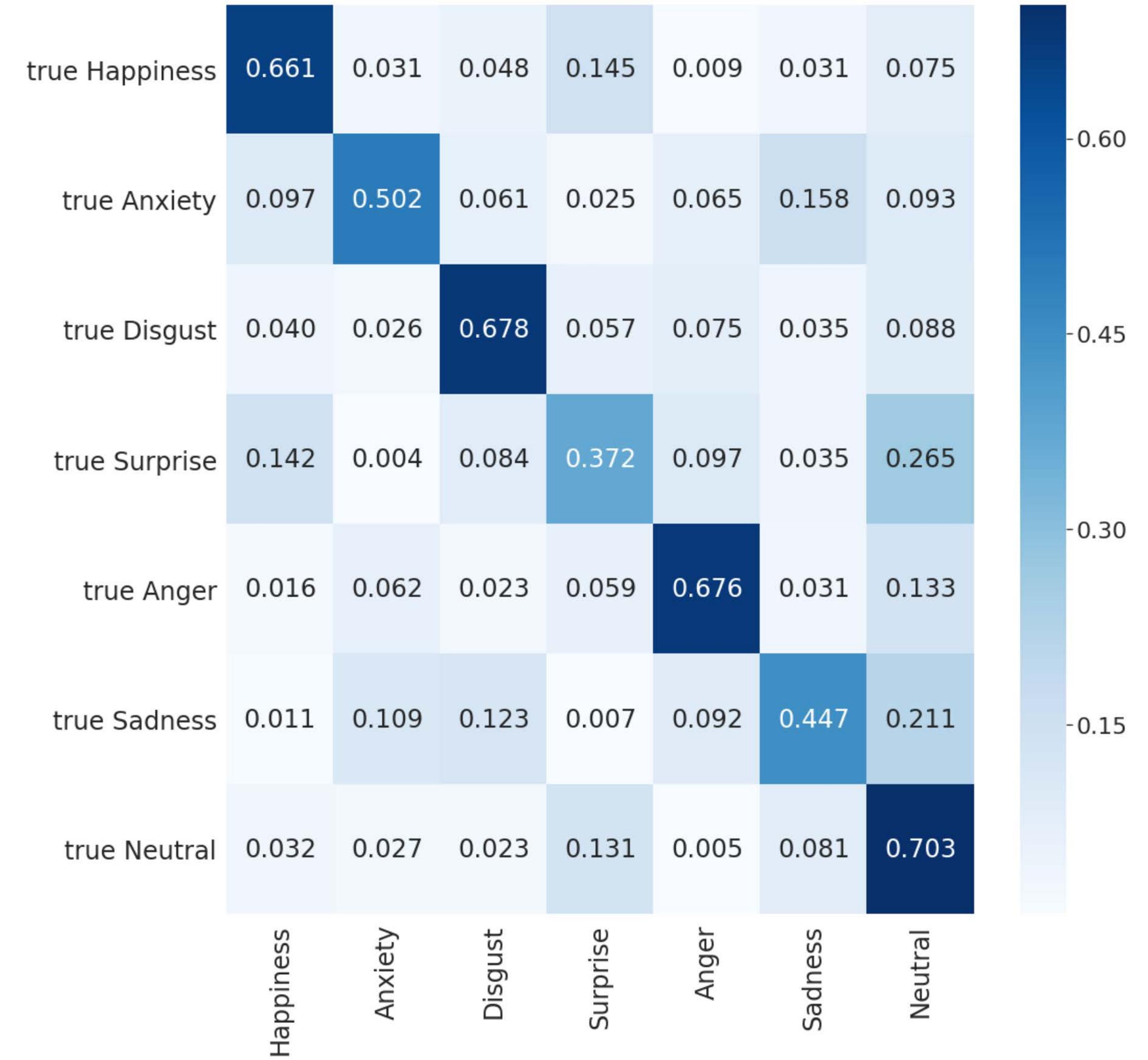
	A	T	F	B
Precision	0,358	0,262	0,497	0,062
Recall	0,400	0,308	0,519	0,181

## Комбинации с конкатенацией признаков после LSTM:

	A + T	A + V	F + T	F + B	All
Precision	0,369	0,512	0,525	0,476	0,559
Recall	0,388	0,524	0,539	0,492	0,540

### Обозначения:

- M + N – конкатенация признаков для модальностей M и N
- All – конкатенация 3-ёх модальностей A + V + T



# HW\_5

- Доделать практические задания (Elan)
- Прислать отчет на почту [o.perepelkina@neurodatalab.com](mailto:o.perepelkina@neurodatalab.com)
- Дедлайн: 21 февраля
- !!! 22 февраля – защита групповых проектов

# Защита проектов

- Постановка проблемы (2 балла)
- Адекватность используемых методов (2 балла)
- Адекватность полученных выводов (2 балла)
- Логичность и последовательность изложения текста отчета (2 балла)
- Качество устной презентации (2 балла)
- Максимум 10 баллов

# Защита проектов

- 15 минут на выступление (презентация) + 10 мин на вопросы
- Название, авторы
- Вклад каждого автора в проект
- Проблема/практическая задача/мотивация проекта
- Гипотезы, методы, эксперименты, результаты, статистические тесты
- Выводы, дальнейшие шаги и идеи