

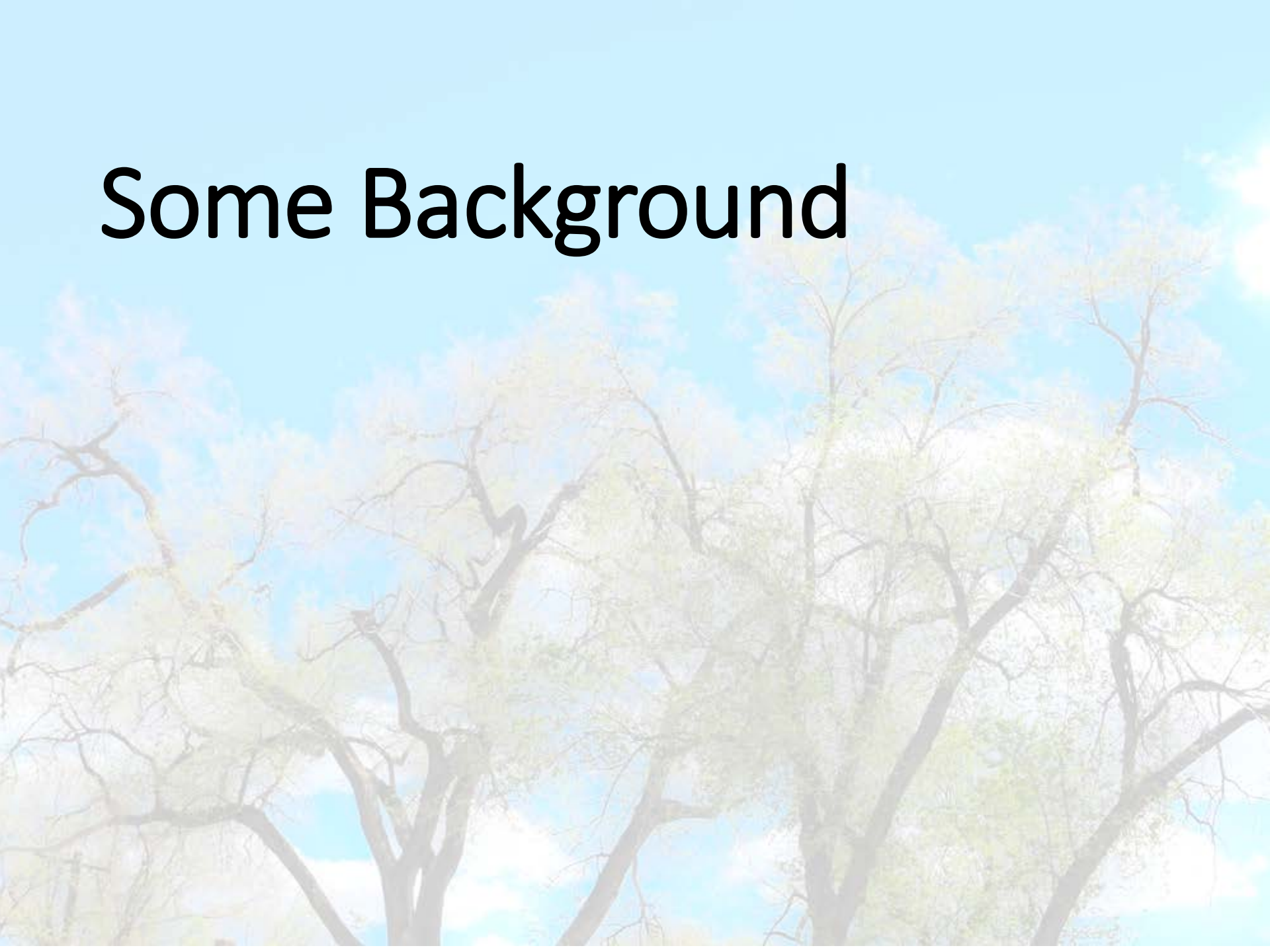
# Comparing categorical and dimensional ratings of emotional speech

Noam Amir<sup>1</sup>, Reut Rubinstein<sup>1</sup>, Adi Shlomov<sup>1</sup>,  
and Gary Diamond<sup>2</sup>

<sup>1</sup> Dept. of Communication Disorders, Tel Aviv University, Tel Aviv, Israel

<sup>2</sup> Psychology Dept., Ben-Gurion University, Beer Sheba, Israel

# Some Background









# Two basic approaches to labeling emotional speech:

## 1. Categorical:

- Anger, Sadness, Fear, Disgust, Joy, Surprise etc.

## 2. Dimensional:

- Valence, Arousal, Dominance, etc.

### Question:

- How do these two relate?
- In other words – how would the same stimuli be judged when using **both** methods?



# Theoretical studies

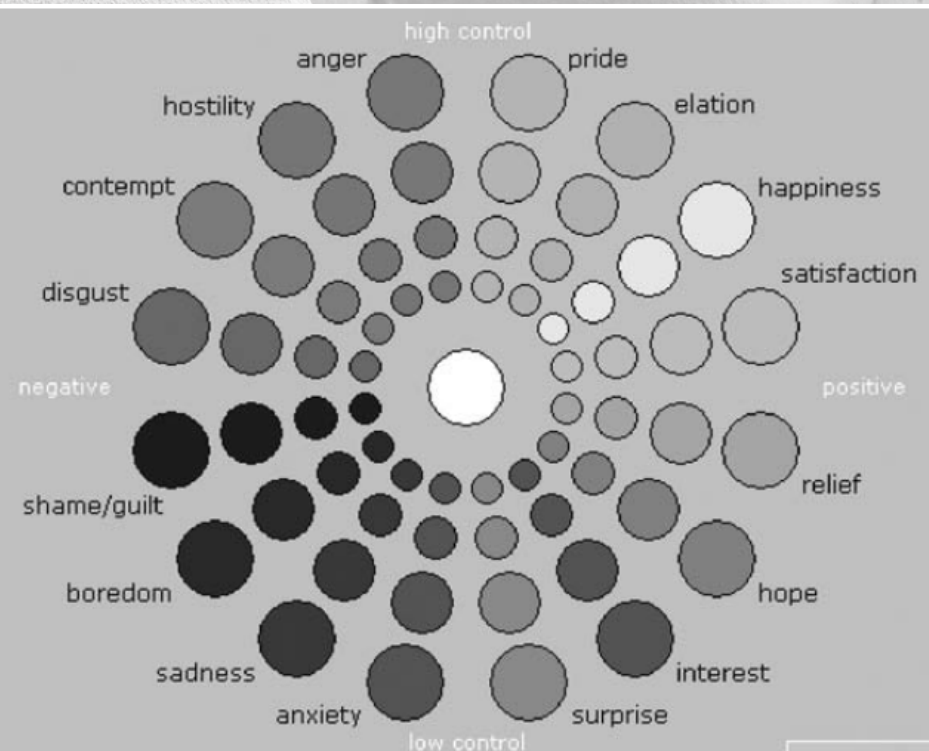


# Scherer (2005):

need not coincide). Even more importantly, it is difficult to differentiate the aspect of intensity of feeling from bodily excitation.

Thus, extremely intensive anger is likely to be characterized by high arousal whereas intense sadness may be accompanied by very low arousal.

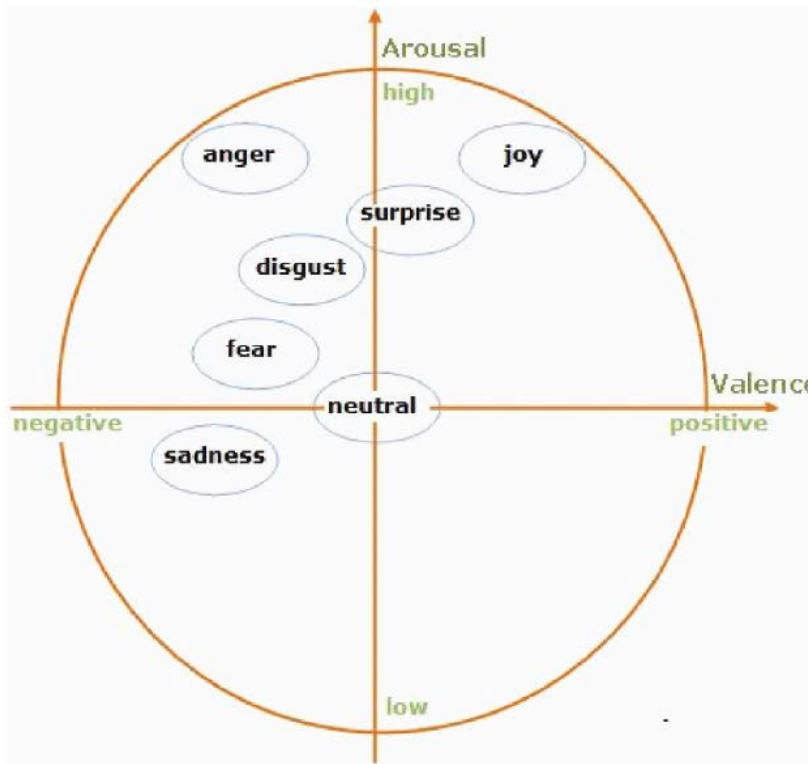
## The Geneva emotion wheel:



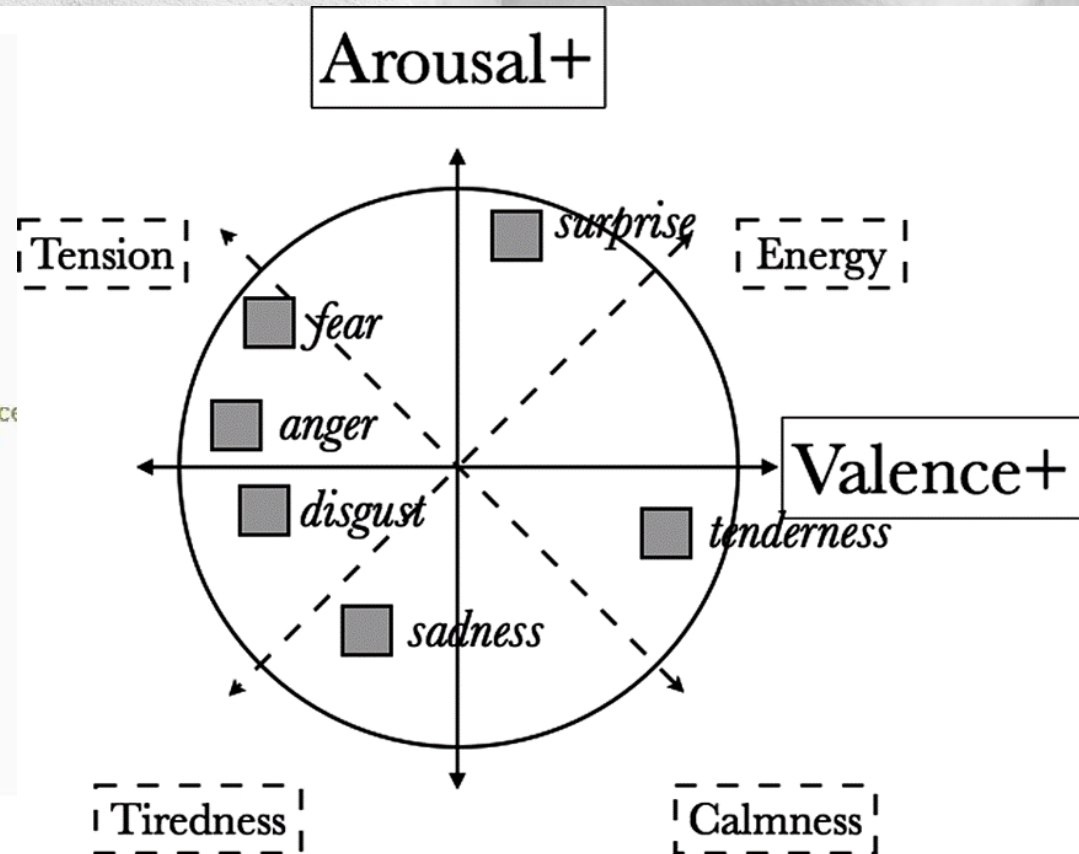


Gunes & Pantic  
2010 quoting  
Russell:

Schimmack and  
Grob 2000:



(b)



# Empirical studies

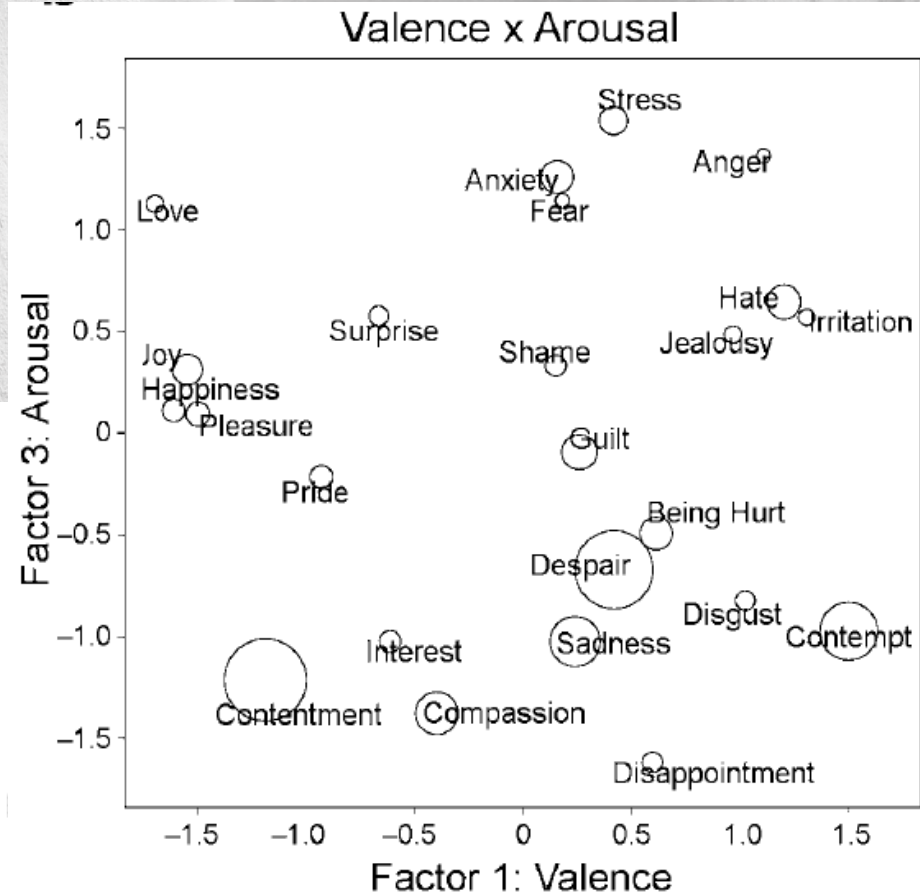
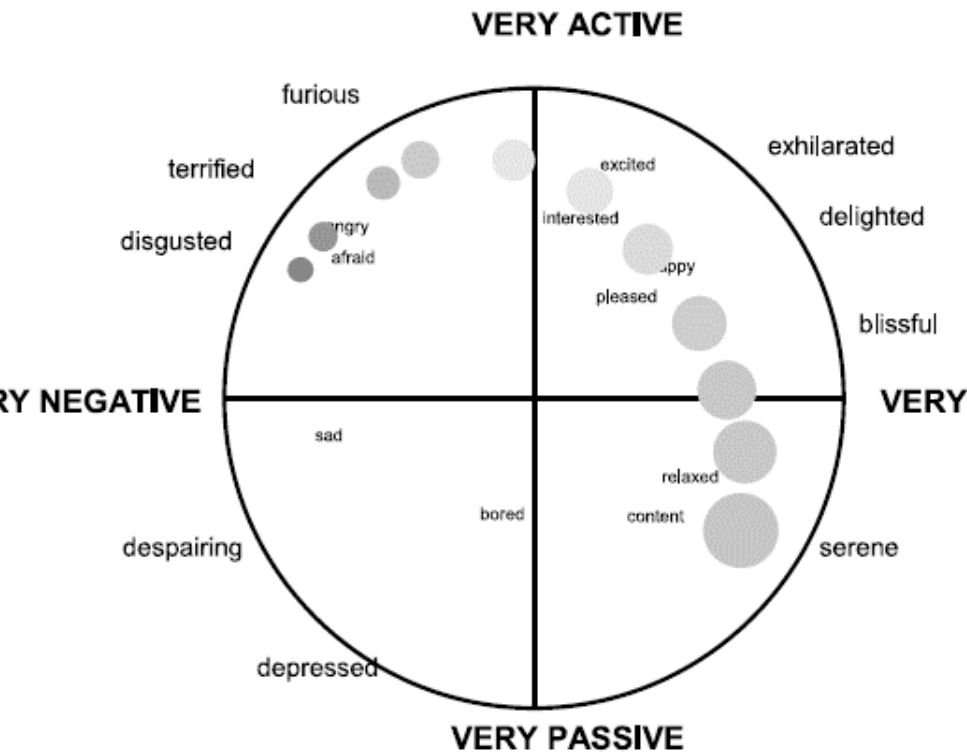
Based on **evaluating terms** for emotions





# Fontaine et al., 2007

- Cowie et al 1999



# Empirical studies

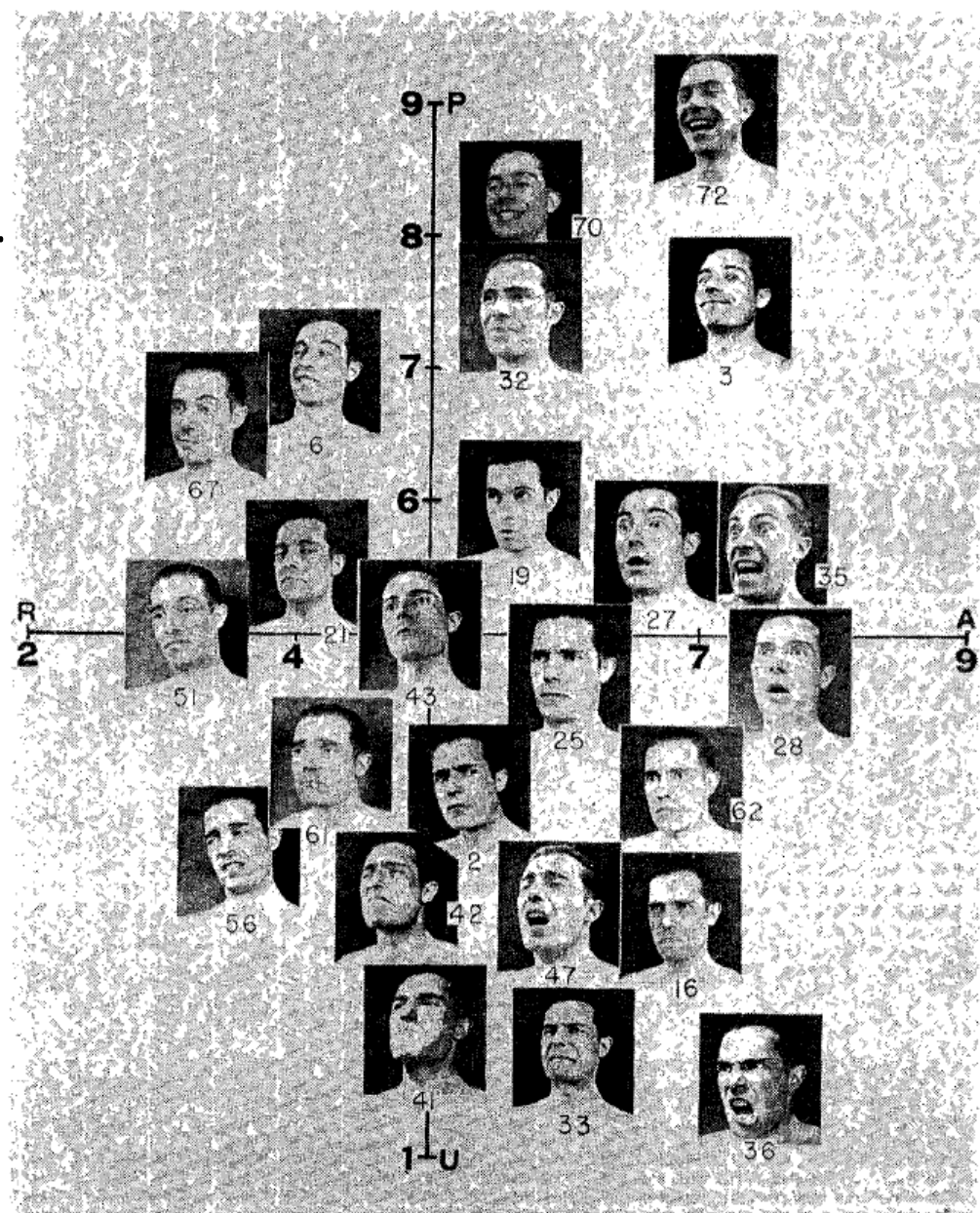
Based on **evaluating emotional stimuli**





# Schlosberg 1952

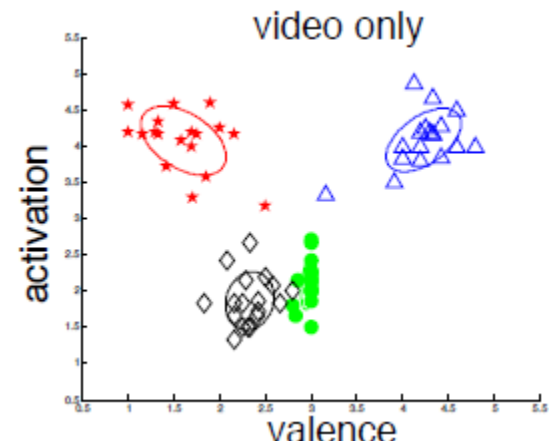
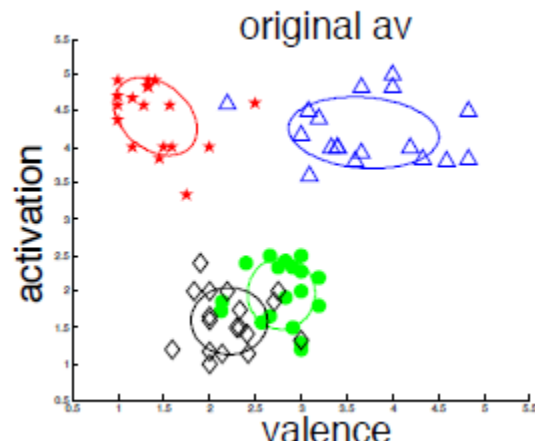
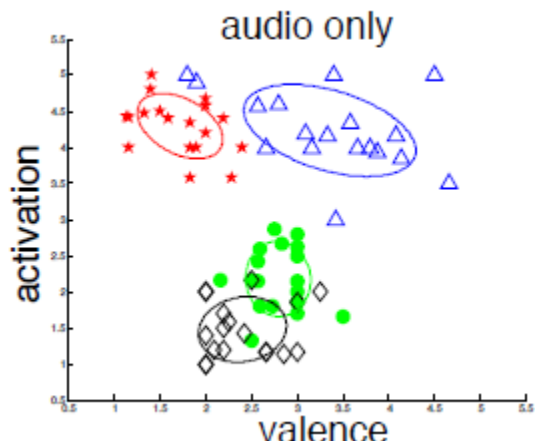
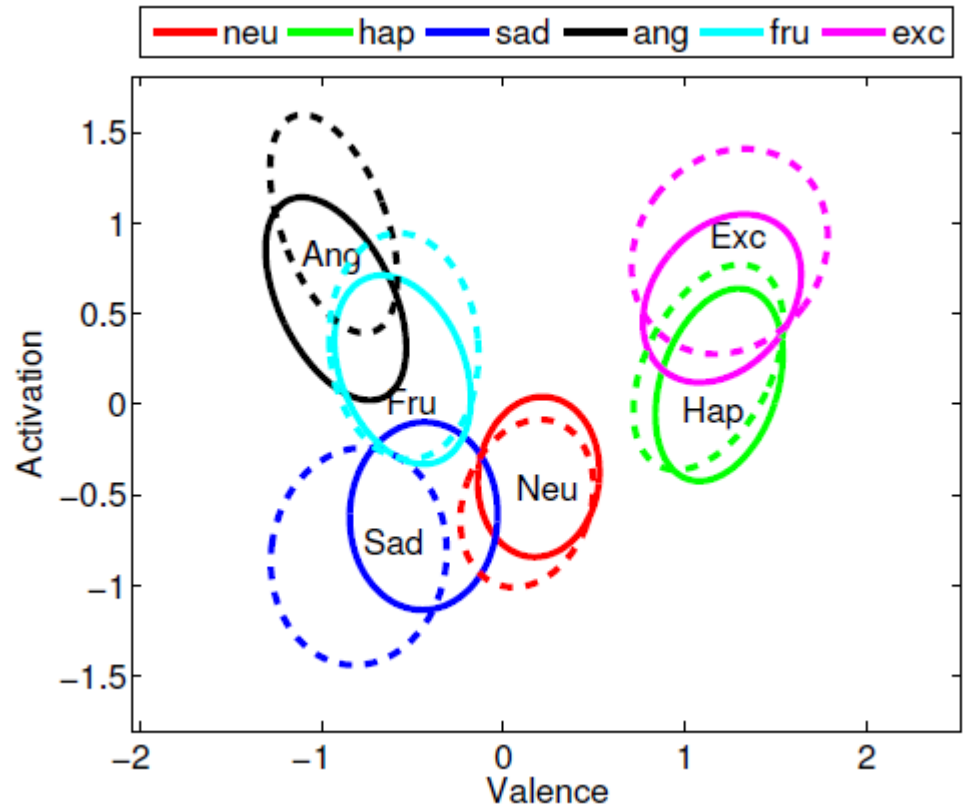
- Note different orientation of axes



Busso & Narayanan,  
2008

Mower Provost et  
al., 2013

*(note location of  
neutral!)*





# To summarize:

- Theoretical descriptions appear to be quite schematic and not clearly based on empirical evidence
  - ... *and they prefer circles*
- Relatively few empirical studies based on judgement of audio/video stimuli exist
  - Paradigms are different
  - Results don't always agree
  - We found it very interesting to pursue this further



# The present study

**Disclaimer:** the reviewers say it has many flaws!



# Speech stimuli

- Based on **three** corpora of emotional speech:
  1. **Acted emotional speech in Hebrew**
    - 4 professional actors
    - Portrayed emotions: Joy, Anger, Sadness and Fear.
    - Sentences, words and nonsense words
    - 191 stimuli
  2. **Young German adults playing a multi-participant computer game**
    - 43 stimuli
  3. **Young Israeli women in a psychotherapy session**
    - “empty chair” therapy
    - 103 stimuli
- 337 stimuli overall
- Clearly each corpus has its limitations!
  - *This is not “the” definitive study*



# Judges:

- 20 female students aged 21-28
- No known hearing problems

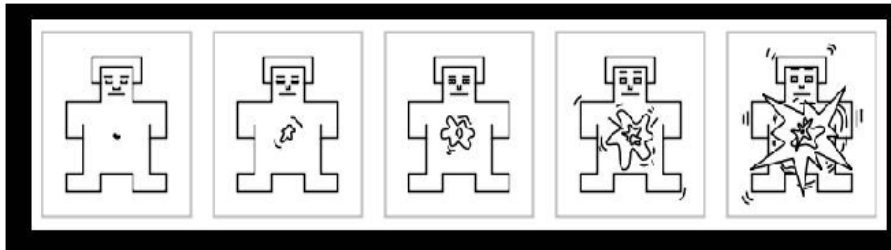


# Experiment was administered via the following GUI:

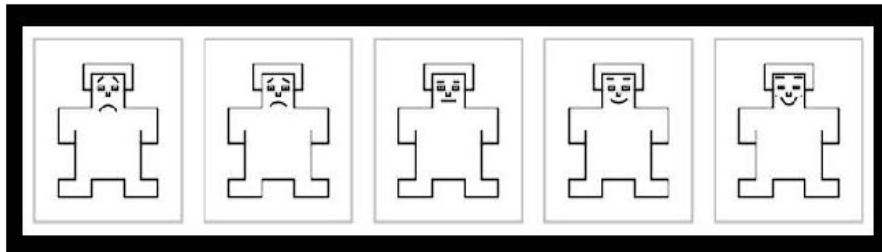
Enter name:

START

Arousal



Valence



Emotion category:

- Anger
- Joy
- Sadness
- Fear
- Neutral
- Other

Emotion Intensity

- Weak
- Medium
- Strong

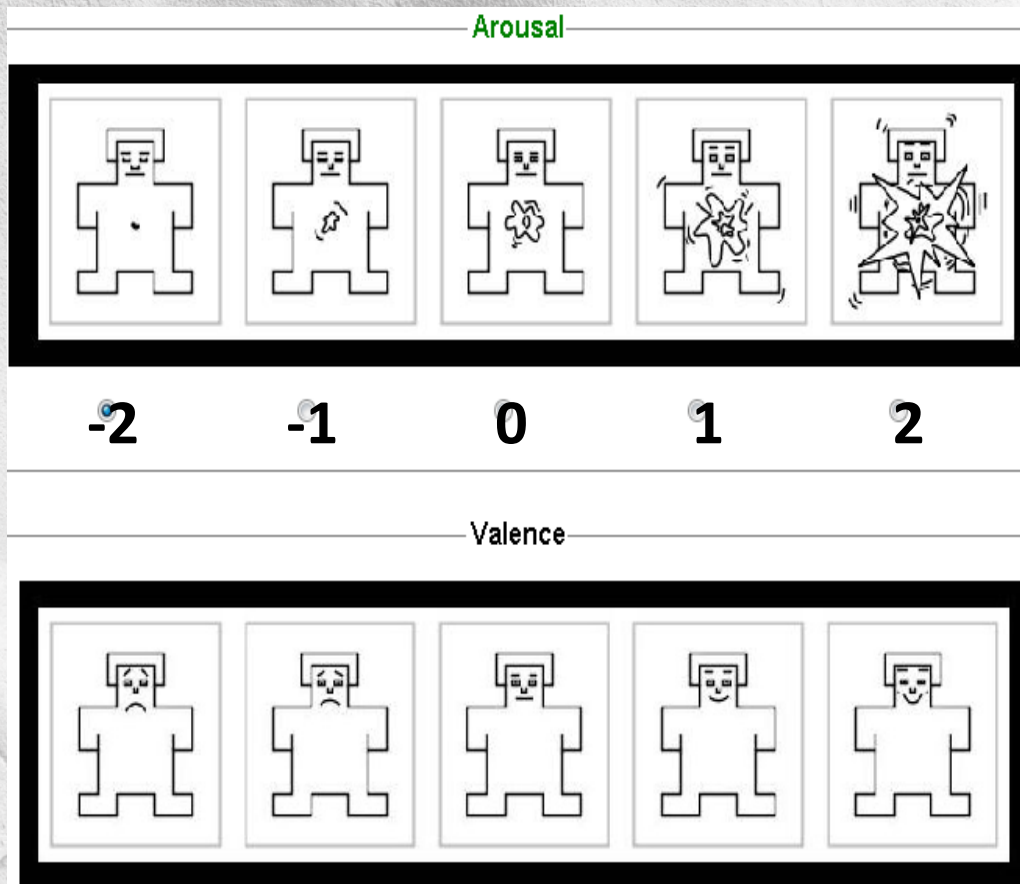
Utterance

1

out of

Next

# Ratings were transferred to numbers:



## Emotion Intensity

1  Weak

2  Medium

3  Strong



# Stages in analysis of results:

1. Determine category for each stimulus
2. Perform statistics per **stimuli**
3. (Perform statistics per **subject**)

No time for this in  
the presentation  
today...



# Categorical “sleight of hand”

- Question: when 20 judges perform categorical judgment of a given stimulus, choosing between 5 emotions, how do we set a threshold for deciding what category this stimulus “truly” belongs to?
- Our answer: **binomial calculation**
  - The probability of **8** or more judges deciding on the same category by chance: 0.03
  - Only 309 of the 337 stimuli had a 8 or more judgments in the same category
  - Only these stimuli were retained

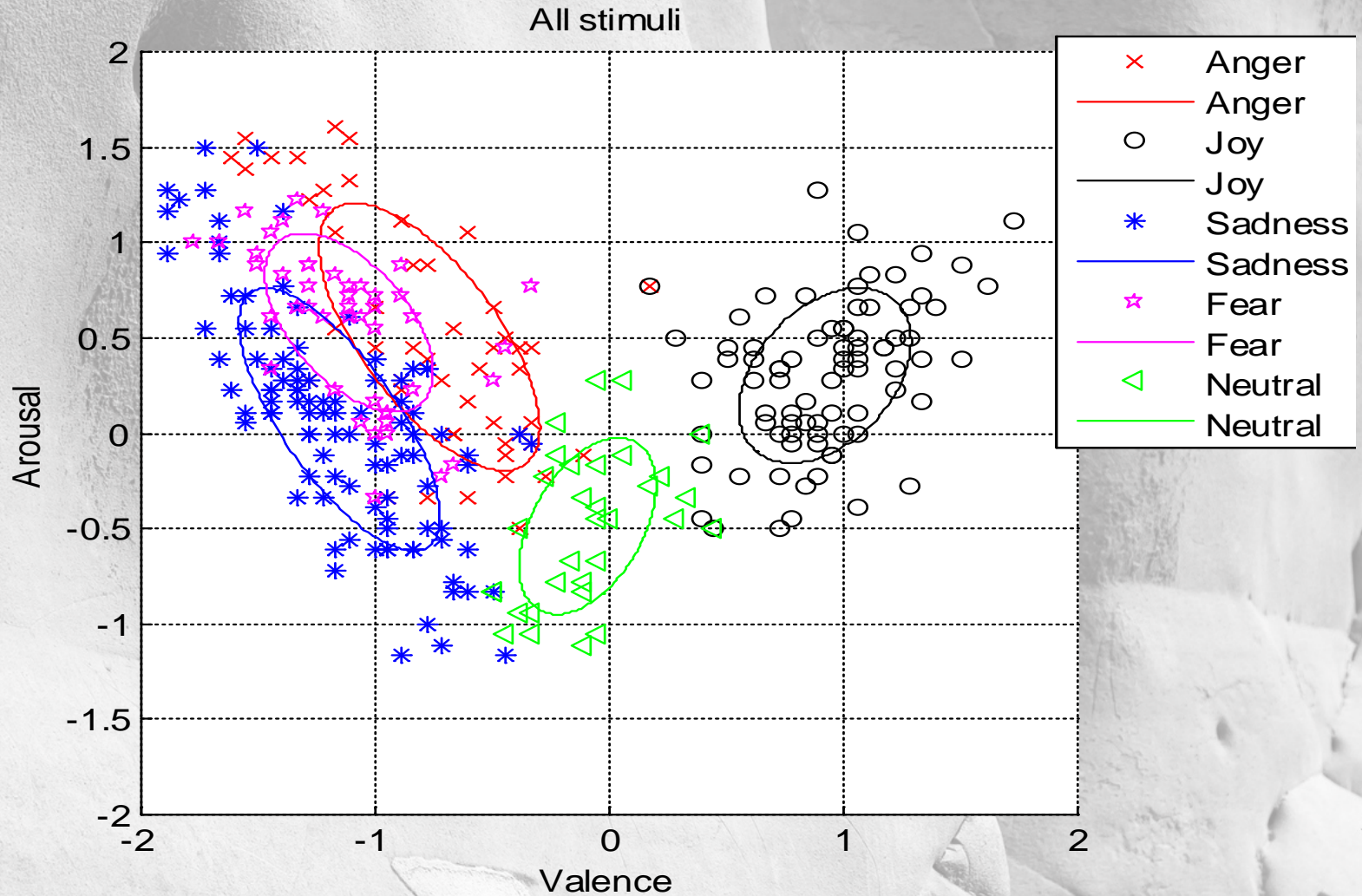


Results -  
*analyzed per*  
*stimulus*

N=309



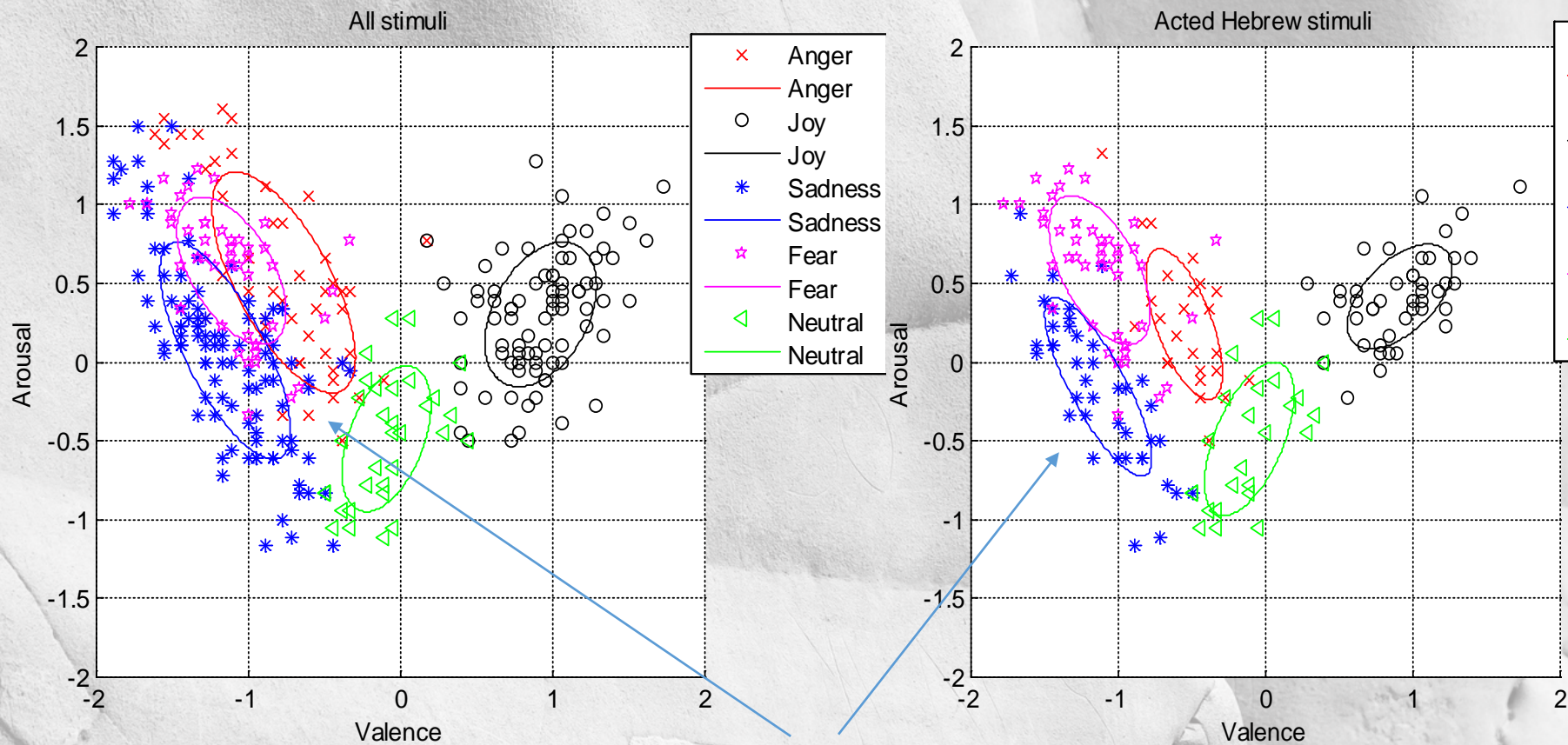
# All stimuli:



Notice: 1) large overlap of negative emotions; 2) Sad has both + and - Arousal



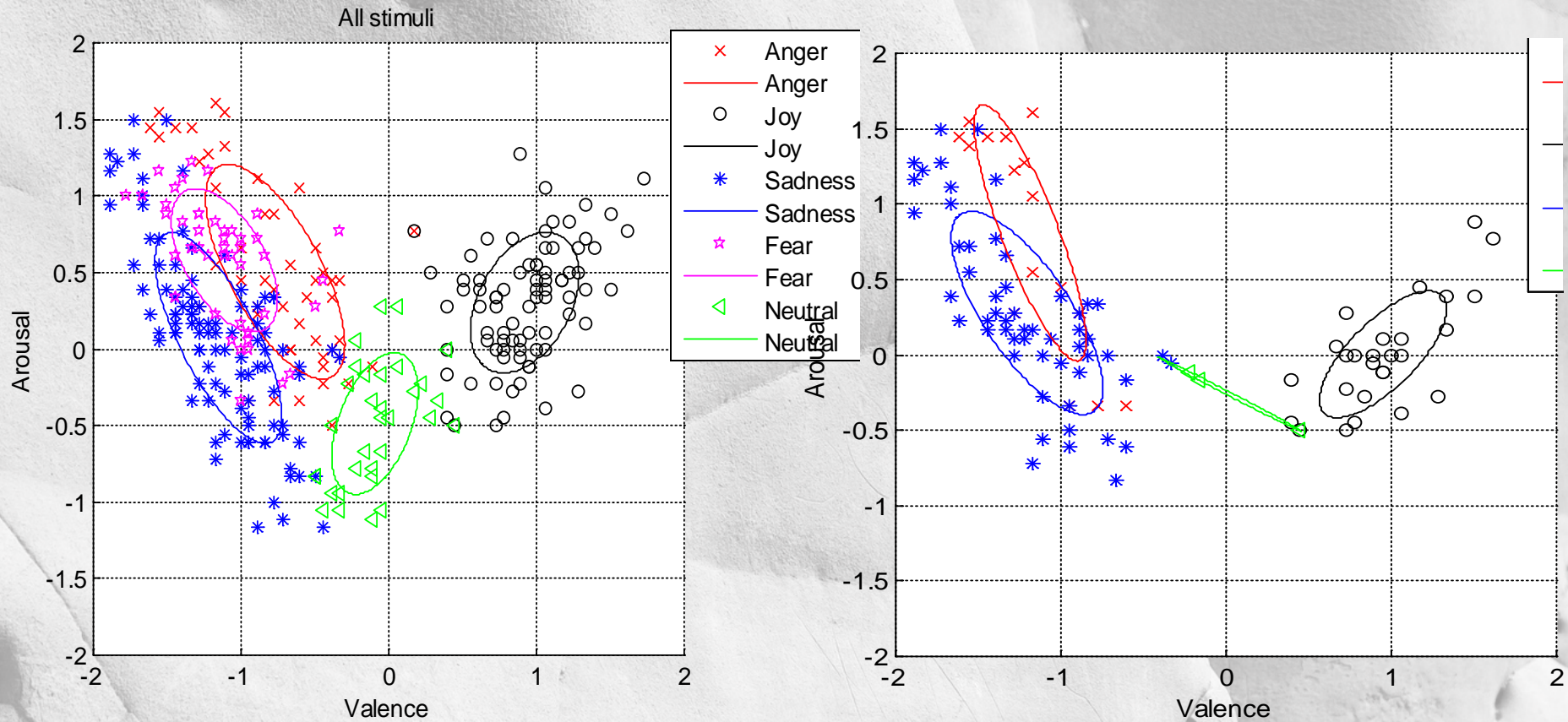
# All vs Hebrew acted only:



Notice: 1) the long ellipses; 2) Neutral is centered on negative Arousal

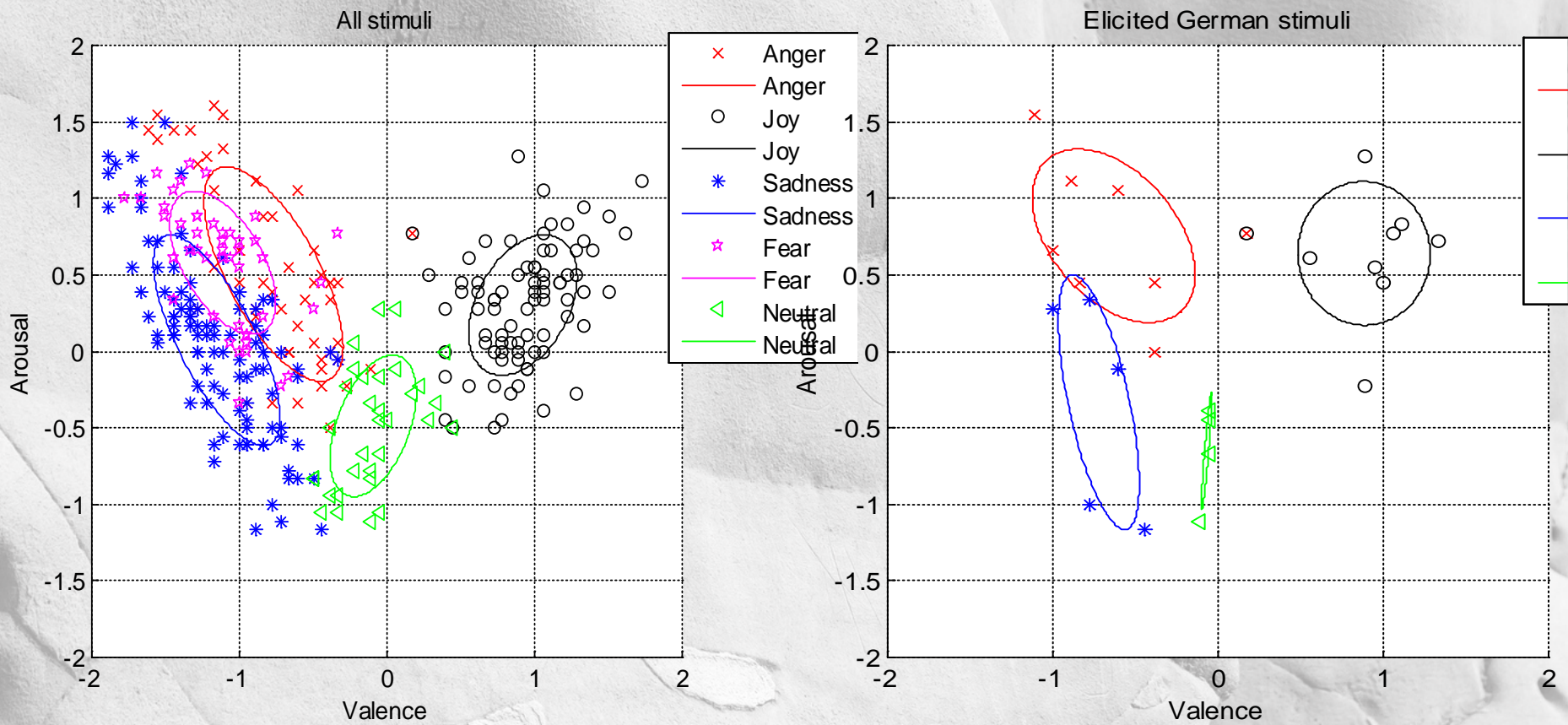


# Hebrew – acted vs. naturalistic:

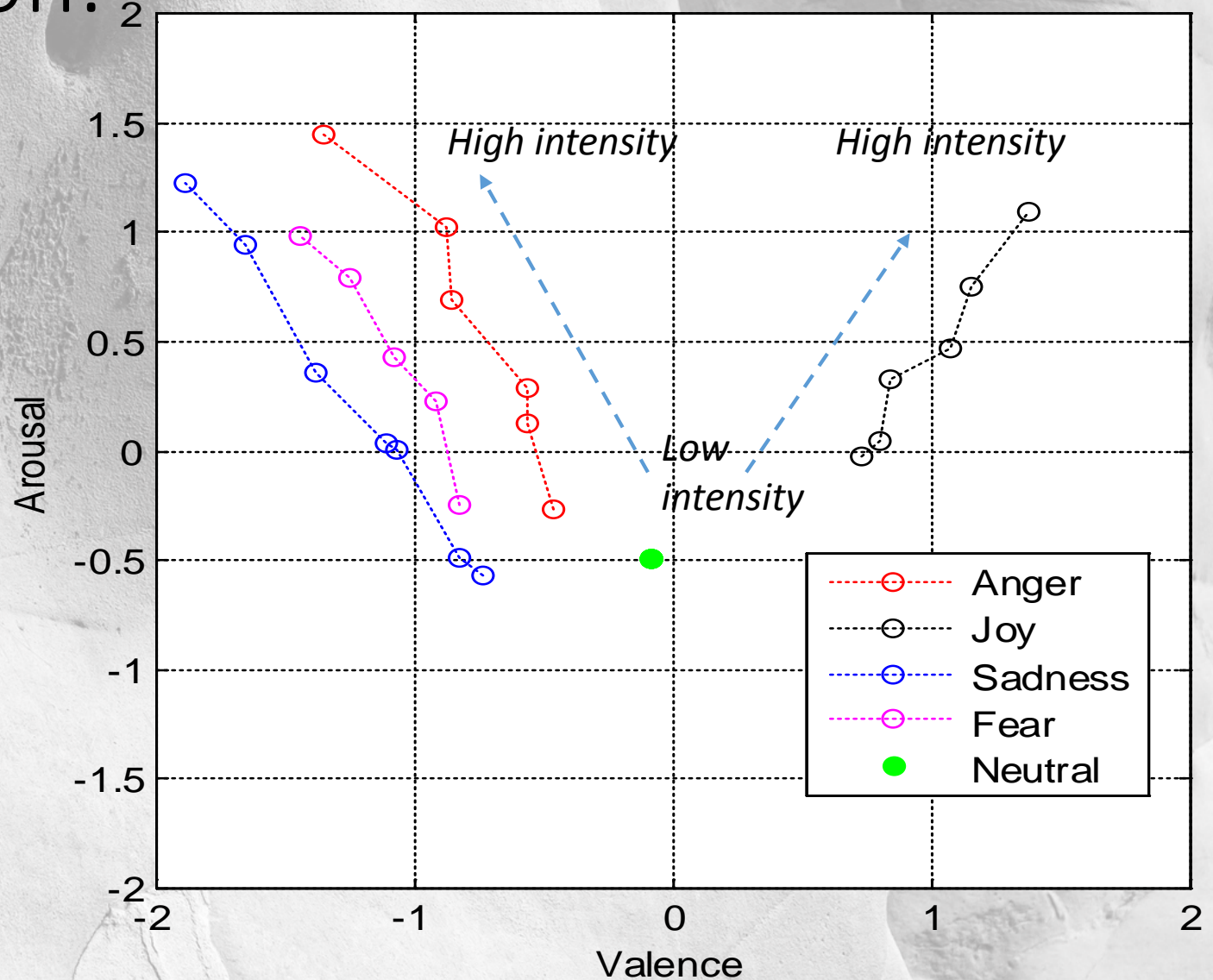




# All vs German:

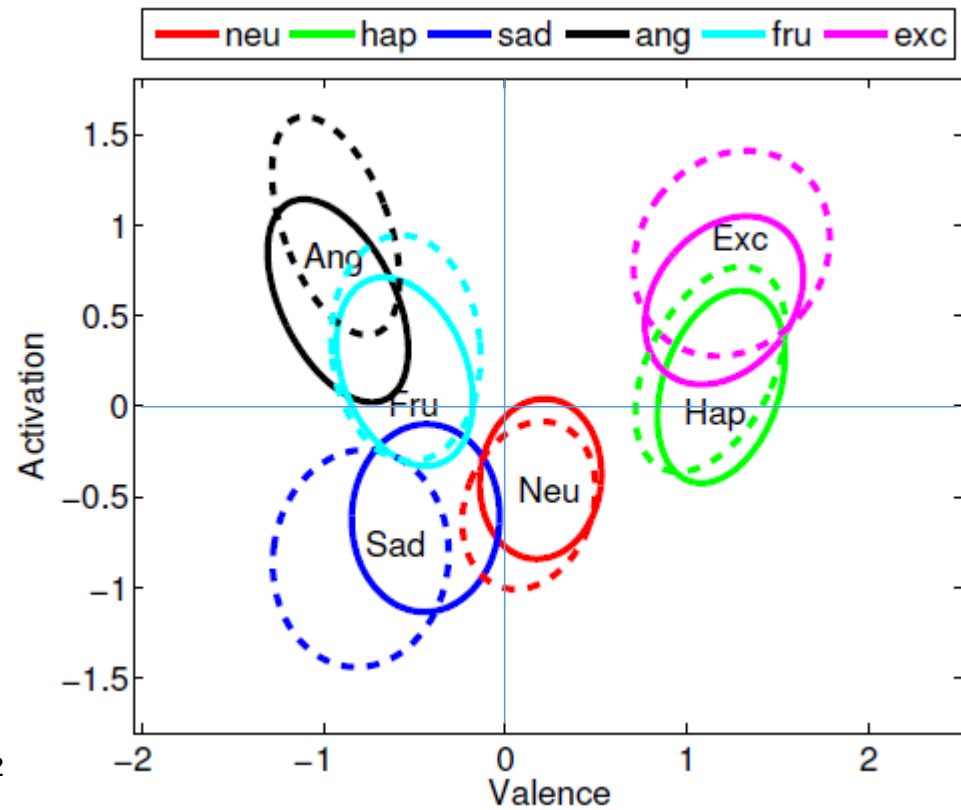
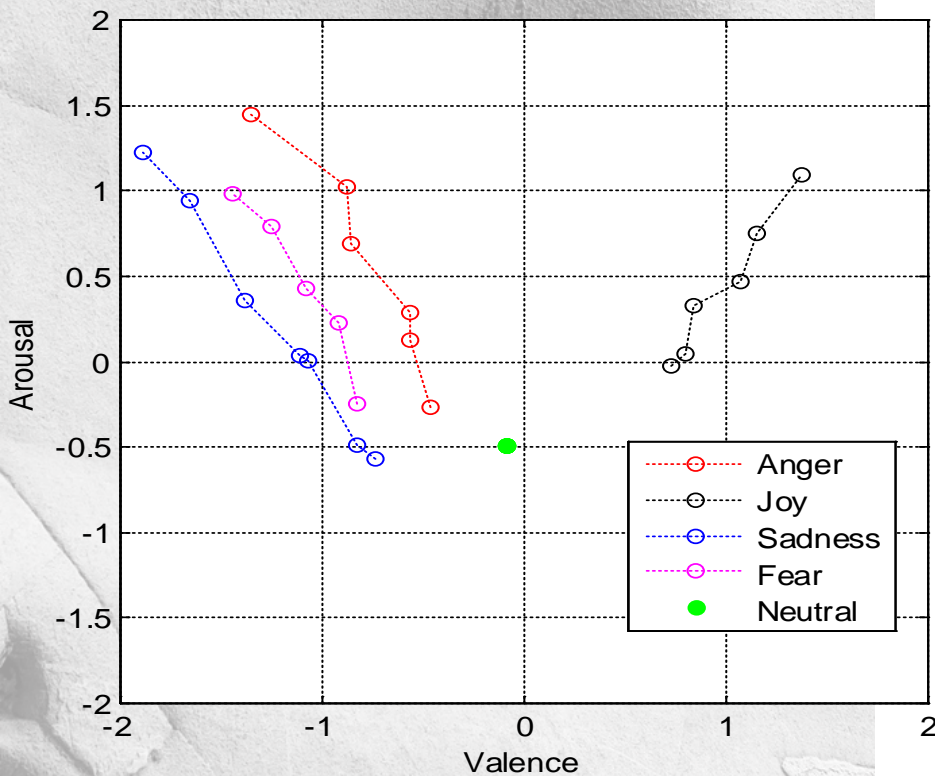


# Average rating vs. intensity of emotion:

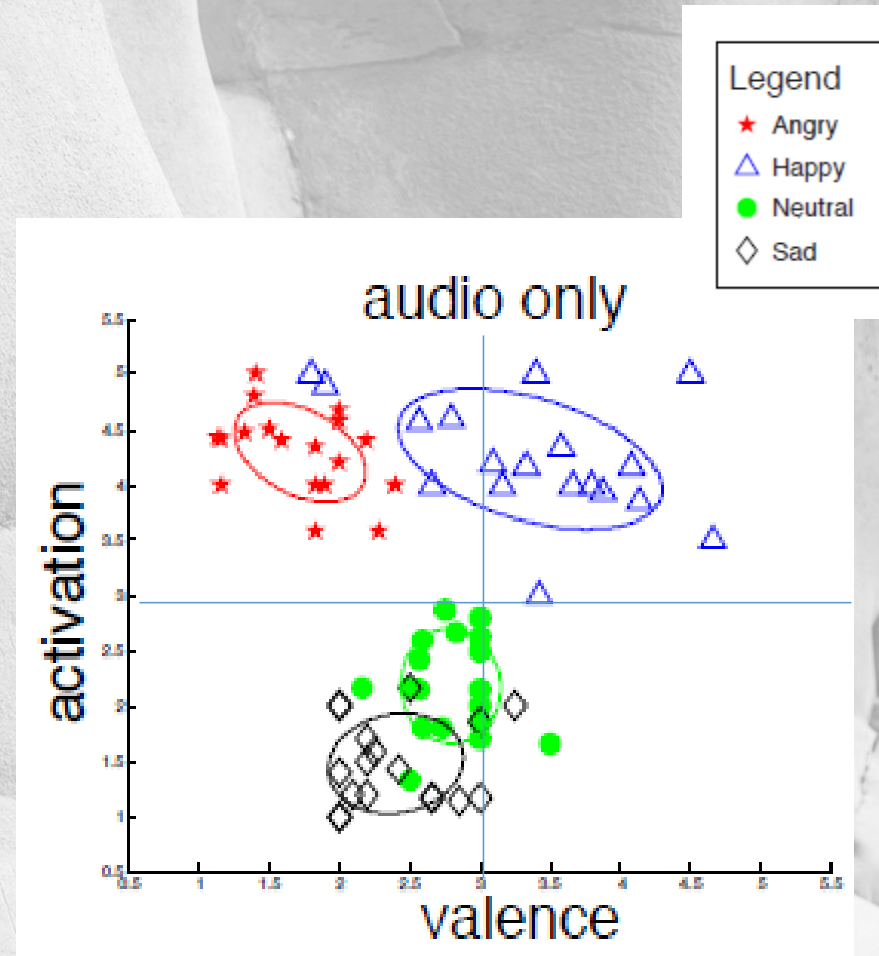
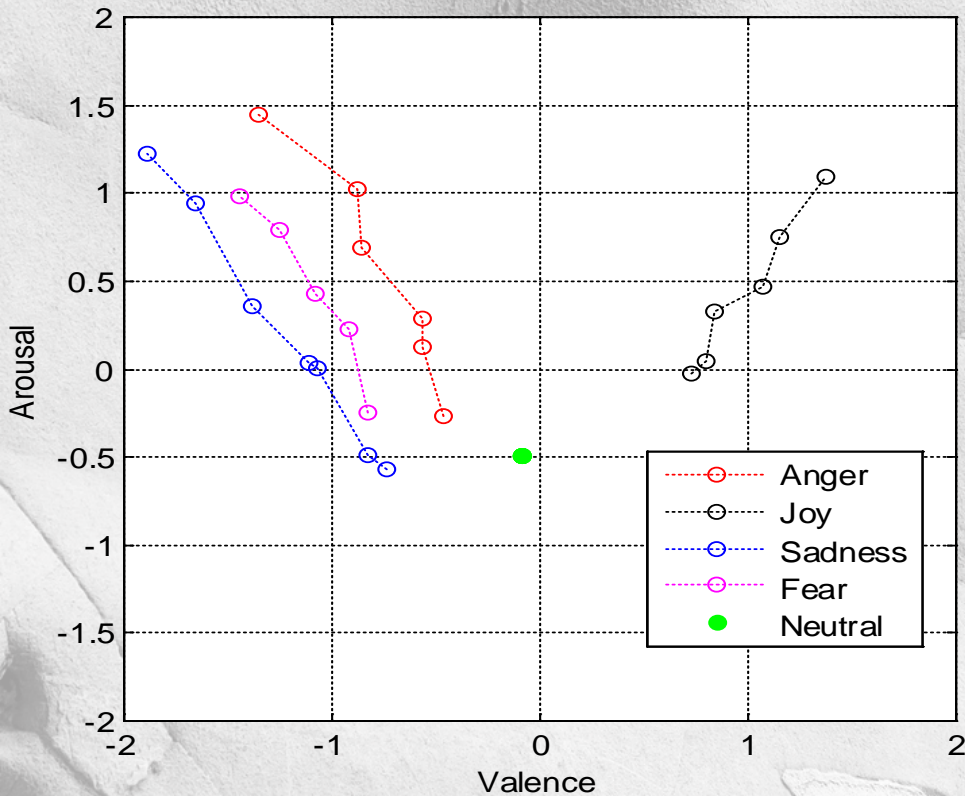




# Our results compared to Busso & Narayanan:



# Our results compared to Mower-Provost et al.:





# Discussion

- Clearly, results depend to a certain extent on the type of data
- Acted data appears to give the most concentrated “clouds”
- There appears a certain bias towards low activation
  - This is most marked in the neutral category, which is not located at the origin, but centered at negative activation
  - Only the truly empirical studies (ours and others) found this
- The relationship between emotional intensity and the centroid in dimensional space is interesting and has not been remarked upon previously

# Outlook

- This type of experiment should be repeated across more types of data, languages and cultures
- It appears quite important to take a closer look at empirical results vs. the theoretical studies



# Thank you!

- And don't forget that there are further results – analysis by listener (N=20) in the written paper

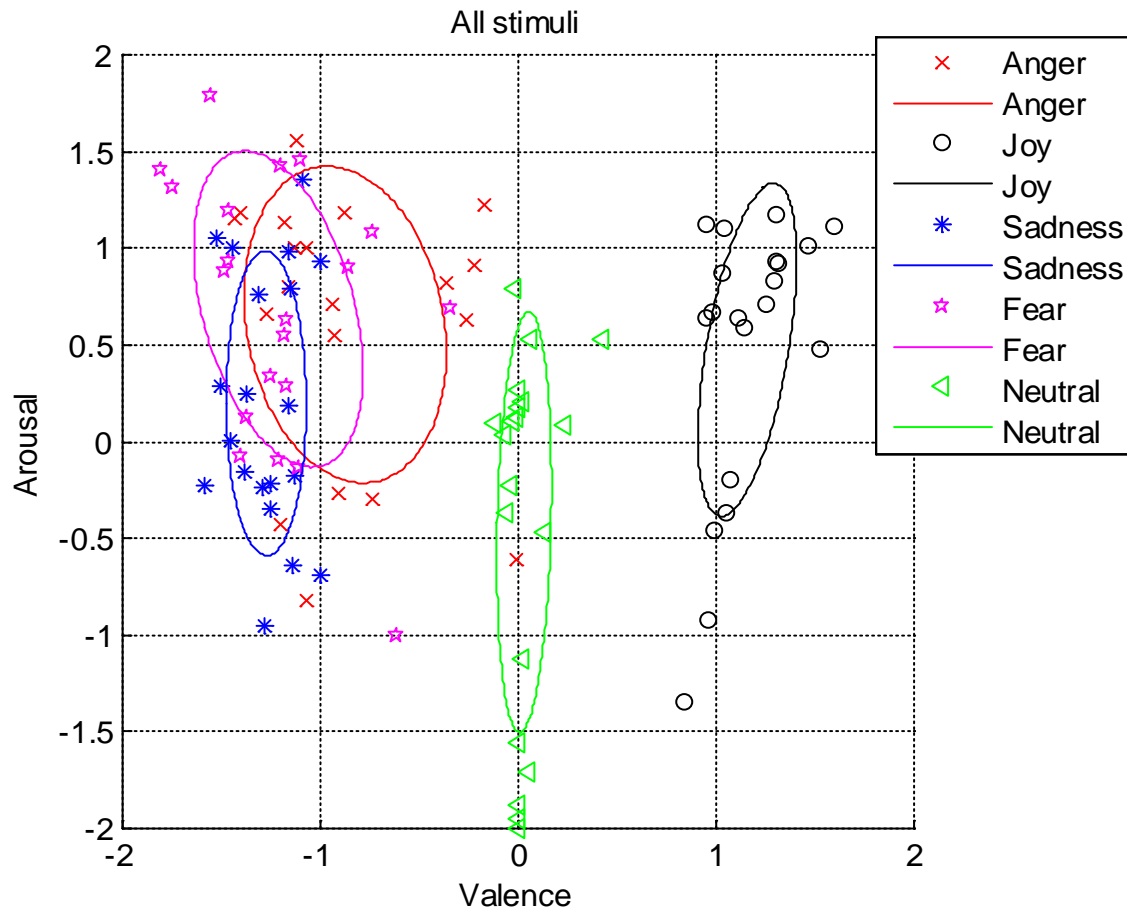
# ניתוח לפי מאזין

- $N=20$
- ביצענו מספר ניתוחים סטטיסטיים:
  - ANOVA למדידות חוזרות
  - מתאם פירסון- הסכמה בין שופטים



# תוצאות כלל הגירויים

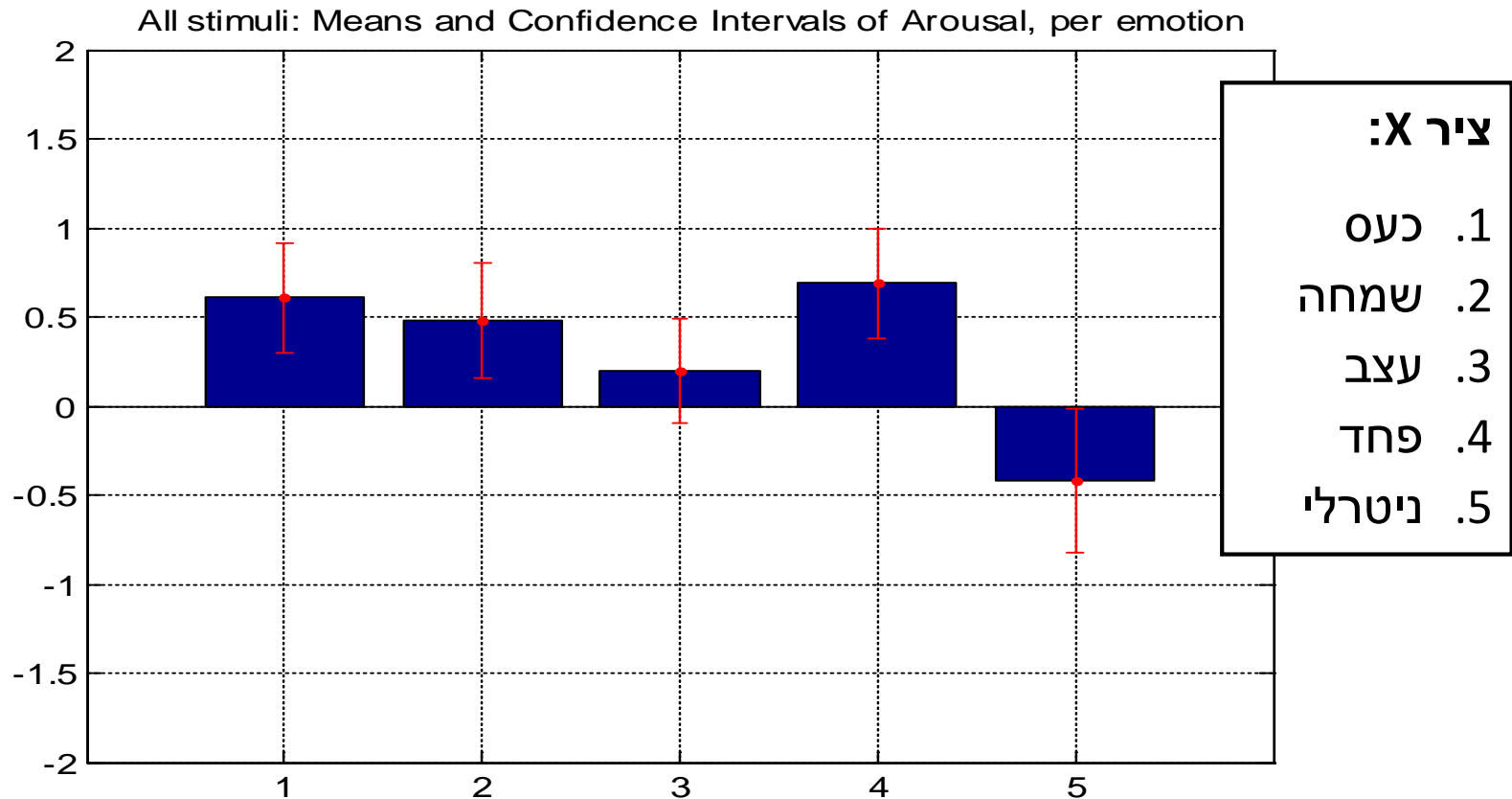
כלל הגירויים: ממוצעי ממדי הרגשות, עבור כל הנבדקות





# תוצאות - ממוצע וס"ת של מידת העוררות:

כלל הגירויים: ממוצעים ומרווח בר סמך של מידת עוררות הרגשות

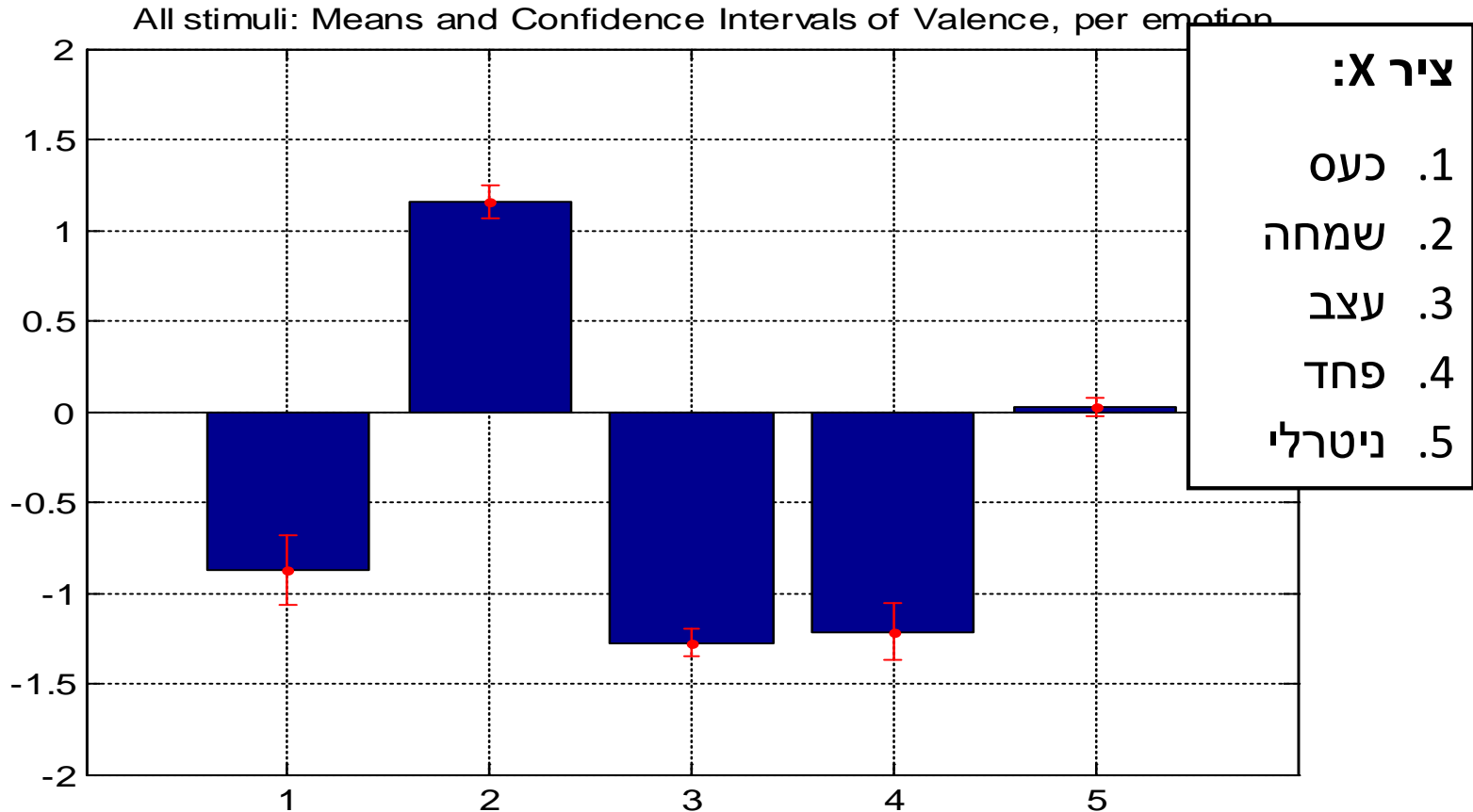






# תוצאות - ממוצע וס"ת של מידת הערכיות:

ממוצע ומרווח בר סמך של מידת ערכיות הרגשות של כל הגירויים



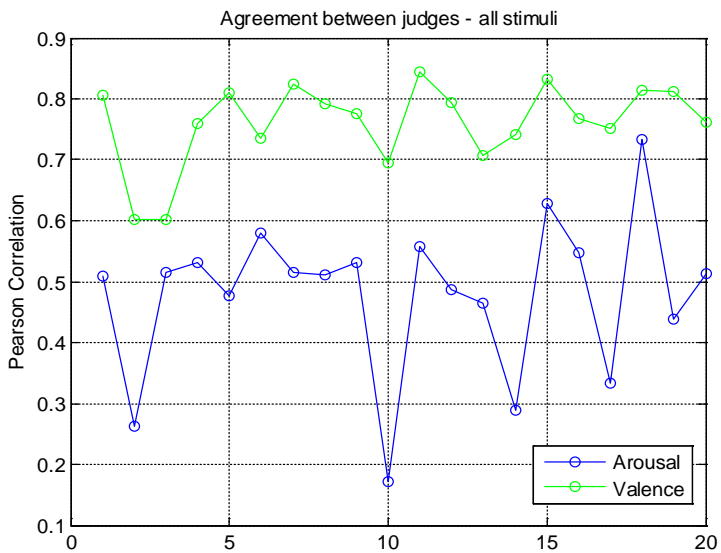
# דיון – כלל הגירויים במחקר:

- למדידות חוזרות תומך ANOVA ניתוח שונות מסוג בממצאי הסטטיסטיקה התיאורית:
  - אפקט מרכזי **לסוג הרגש** בשני המימדים
    - הבדלים מובהקים בין הרגשות השונים במרחב
  - **בניתוח השוואות מרובות**
    - **במימד הערכיות** הרגשות השונים מובחנים זה מזה
    - **במימד העוררות** הרגשות מובחנים, אולם אין הבדל מובהק בין פחד לכעס ובין עצב לשמחה ובין פחד לשמחה

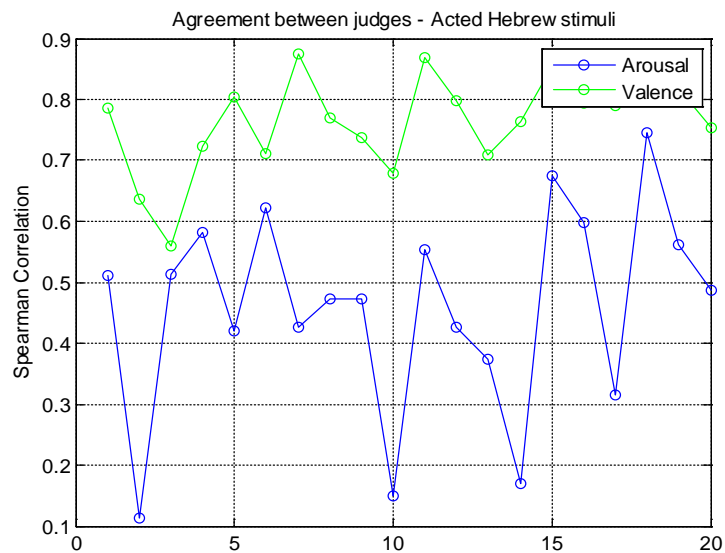


# תוצאות - הסכמה בין שופטים (פירסון):

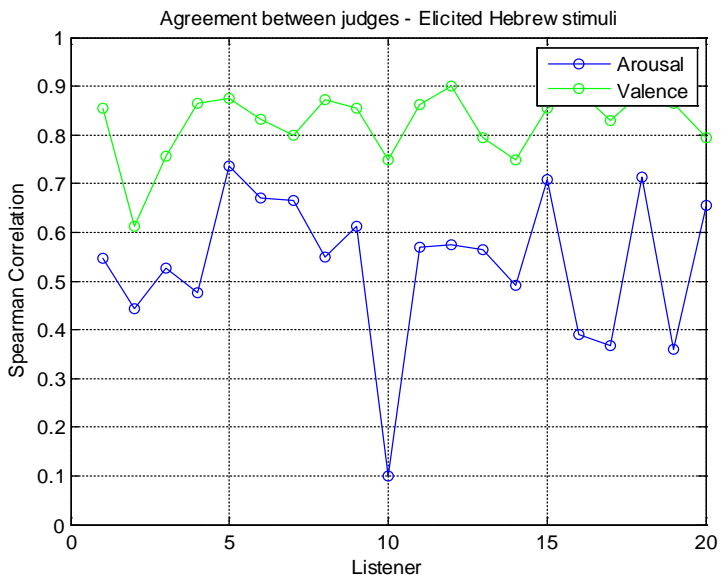
כלל  
הגירויים



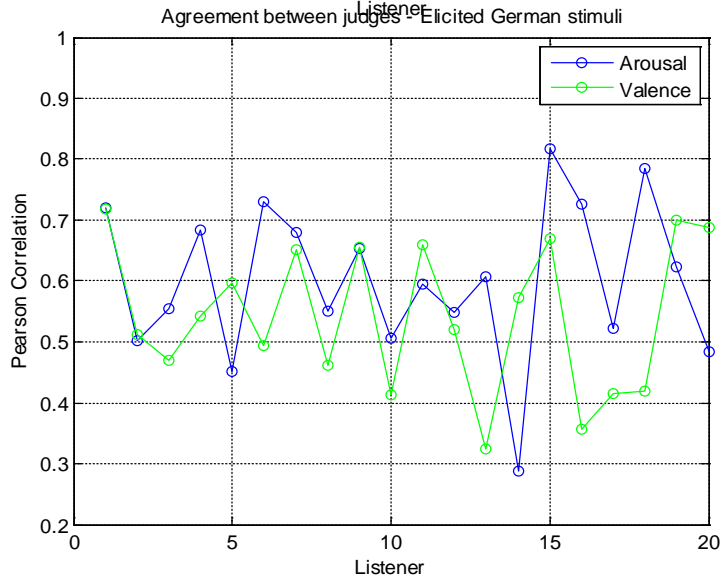
משוחקים



טבעיים  
עברית



מעוררים  
גרמנית



# דיון- הסכמה בין שופטים (פירסון):

- ערכיות < עוררות
  - למעט בגירויים המעוררים בגרמנית
- נבדקות 2 ו10 בעלות מתאמים נמוכים
  - למעט בגירויים המעוררים בגרמנית
- עברית משוחקים- הסכמה פחותה לגבי רמות העוררות



# דיון-הסכמה בין שופטים:

- בהשערה השנייה, שיערנו כי תמצא היררכיה בהסכמה בין שופטים בין סוגי הגירויים:
  - משוחקים ← טבעיים+מעוררים
- בתוצאות: קיימת היררכיה בסדר שונה
  - טבעיים ← מעוררים ← משוחקים
    - הגירויים המשוחקים מכילים מילים קצרות
    - איכות השחקנים
    - תומכים בממצאי המחקר push&pull effects מאפייני

(Scherer 1986)

# סיכום ומסקנות:

- את חמשת הרגשות שנבדקו ניתן:
  - להכליל לקטגורית רגש נפרדת
  - לדרג על שני צירים במרחב
- שמחה מובחנת משאר הרגשות - ערכיות חיובית
- ממצאי מחקר זה שונים ממחקרים קודמים אשר מיקמו את הרגשות על הצירים באופן שרירותי ולא אמפירי



# ביקורת והמלצות:

- במחקרנו ארסנל הגירויים מצומצם
- נבחרו 5 רגשות בסיסיים
- לתוצאות הוכנסו רק גירויים בעלי הסכמה של 8 שיפוטים ומעלה
- כדאי להמשיך ולחקור את הנושא, ע"מ להוסיף אור על הממצאים
  - ארנסל גירויים שונה
  - קטגוריות רגש נוספות
  - אוכלוסיית מחקר אחרת