An introduction to multisensory perception

Tobias Andersen

Cognitive Systems

DTU Compute

Multisensory perception

- The sensory modalities *complement* each other
  - Adds to the richness of the perceptual experience
  - We prefer a movie in color and surround sound to silent black and white
  - People who lose the sense of color often also lose their appetite

McGurk illusion: you hear one think but observe another one

- The sensory modalities contain redundant information
  - Sensory information should be integrated across sensory modalities
  - Estimate phonemes from the voice and face
  - Cats can hear and see the location of their prey

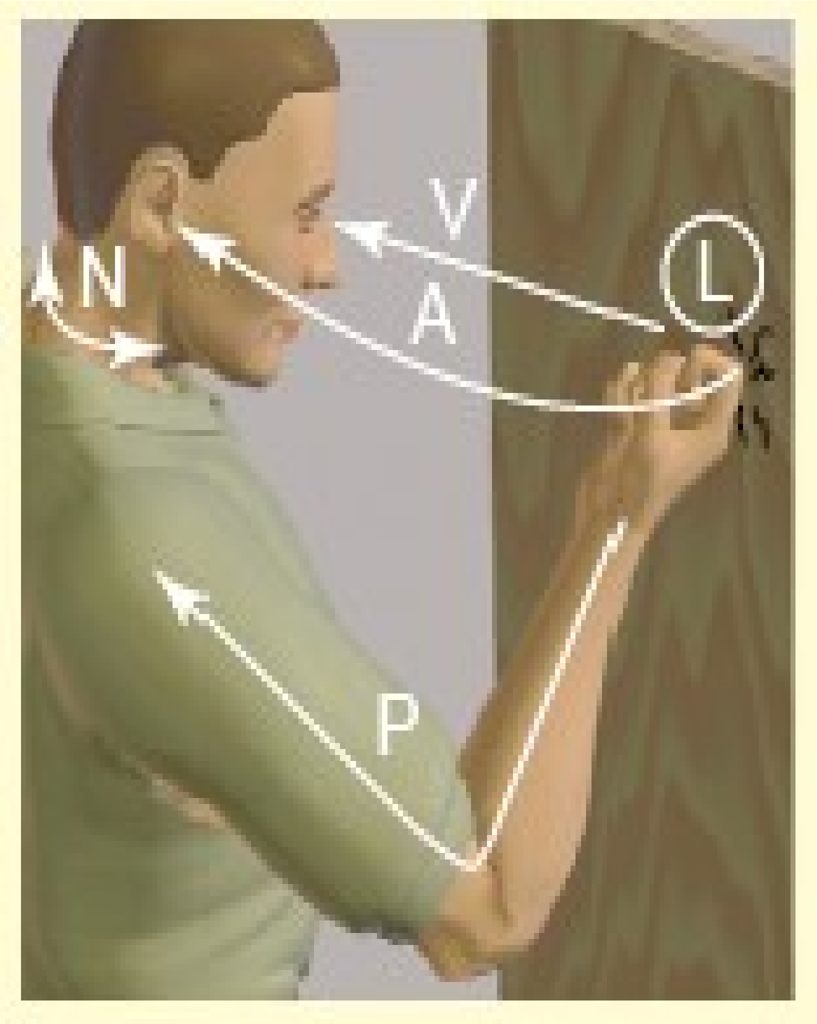
Audiovisual integration of speech

Demonstration of the McGurk illusion – find more examples on youtube

Ventriloquism

- We perceive the voice of the ventriloquist as coming from the dummy
- Often attributed to the ventriloquist being able to “project” his voice to a location
- Wrong!
- It is multisensory integration!

Multisensory integration of space



From Ernst and Bülthoff, TICS, 2004

Location (L) of the hand is estimated from

- Audition (A)
- Vision (V)
- Proprioception (P)

how torque your articulations and muscles are

But we need a single estimate!

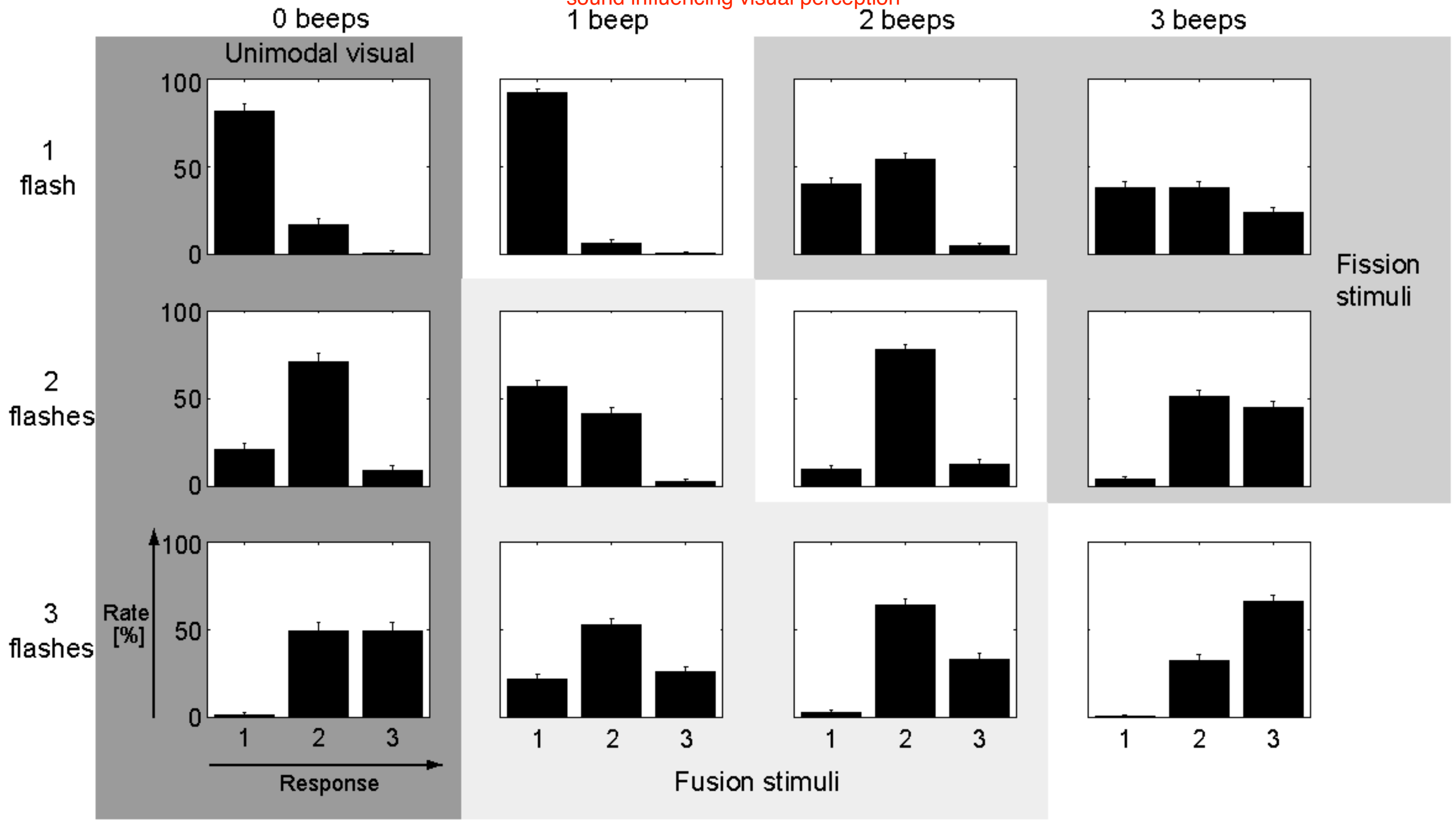
Multisensory Integration

- How do you integrate the information?
  - Should you weigh each source equally?

Multisensory integration - flashes and beeps

flash and two beeps,
then you ask to count the number of flashes

sound influencing visual perception

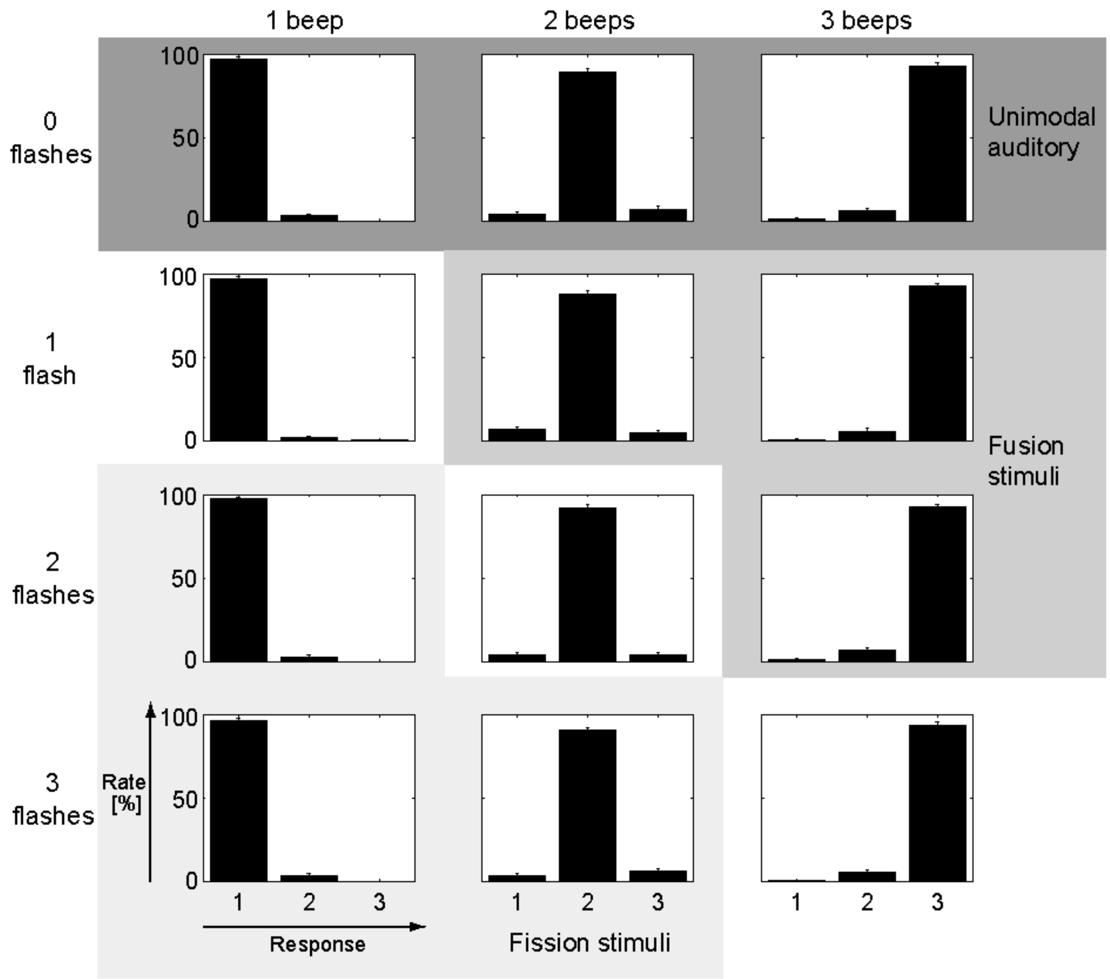


From Andersen, Tiippana & Sams, Cognitive Brain Research, 2004

Count the flashes; beeps at 80 dB

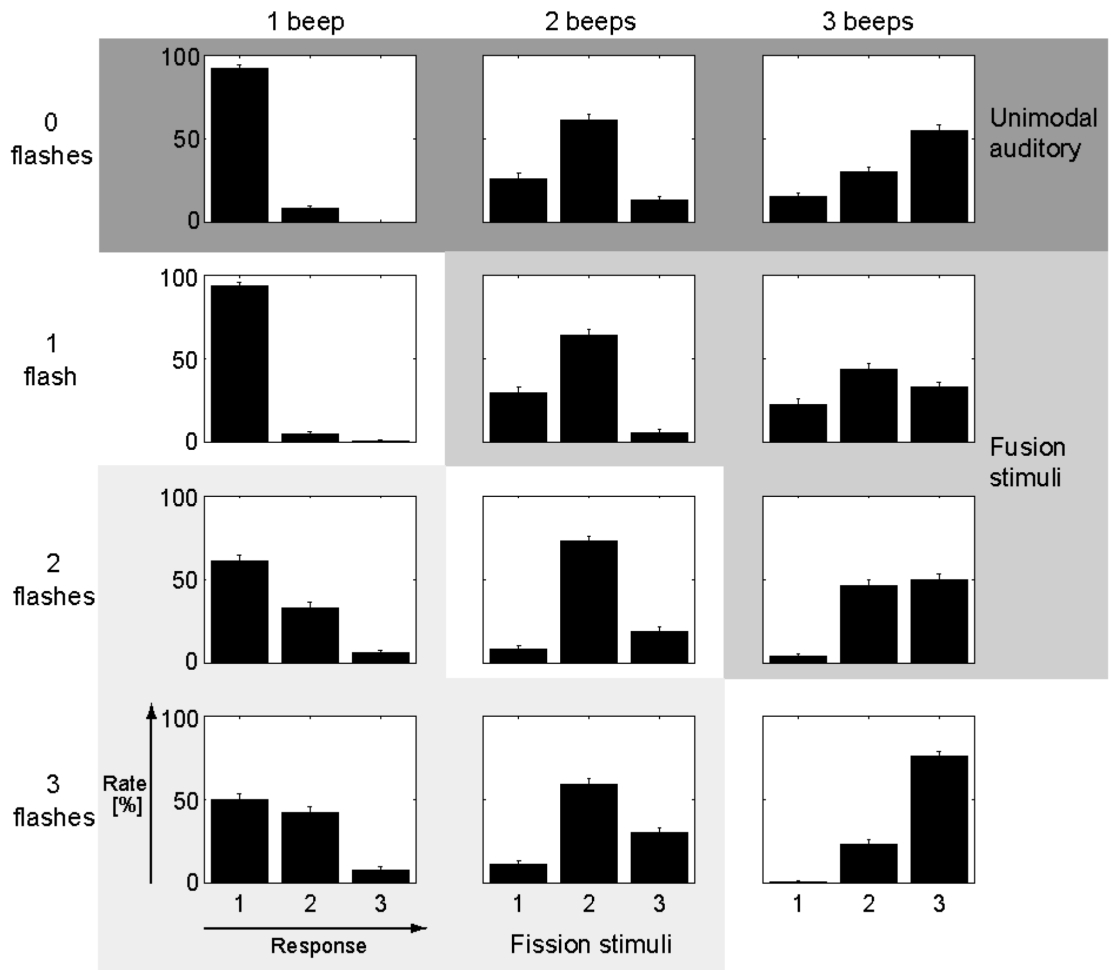
Multisensory integration - flashes and beeps

count the beeps instead



Count the beeps; beeps at 80 dB

Multisensory integration - flashes and beeps



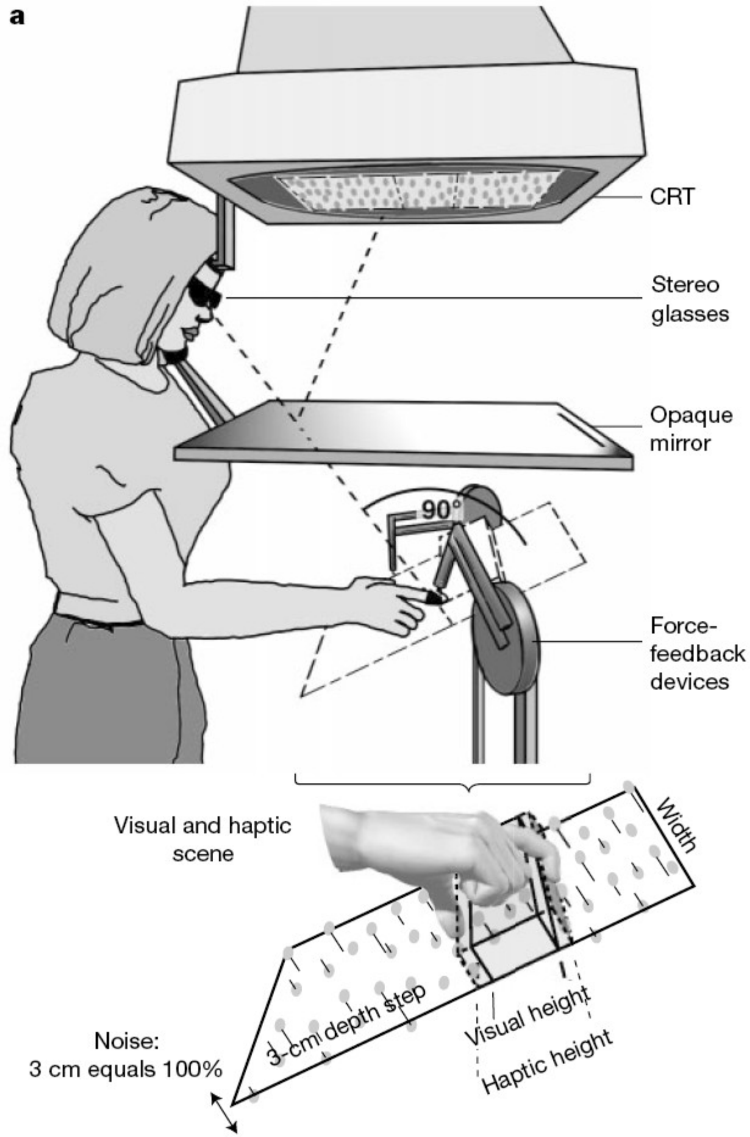
Count the beeps; beeps at 10 dB

Multisensory integration - flashes and beeps

- Governing principles
  - Information reliability
    - The strength of cross-modal influence depended on sound level
  - Modality appropriateness
    - The sound had to be at threshold to be influenced
    - The flashes was influenced also well above threshold
  - Directed attention
    - Possible to count either flashes or beeps

the reliability of the information depends so much on the weight we give when perceving

Multisensory integration - space



- Height can be estimated from
  - sight
  - proprioception
- Independent stimuli can be created with
  - Force feedback device
  - mirrored stereo display

From Ernst and Banks, Nature, 2002

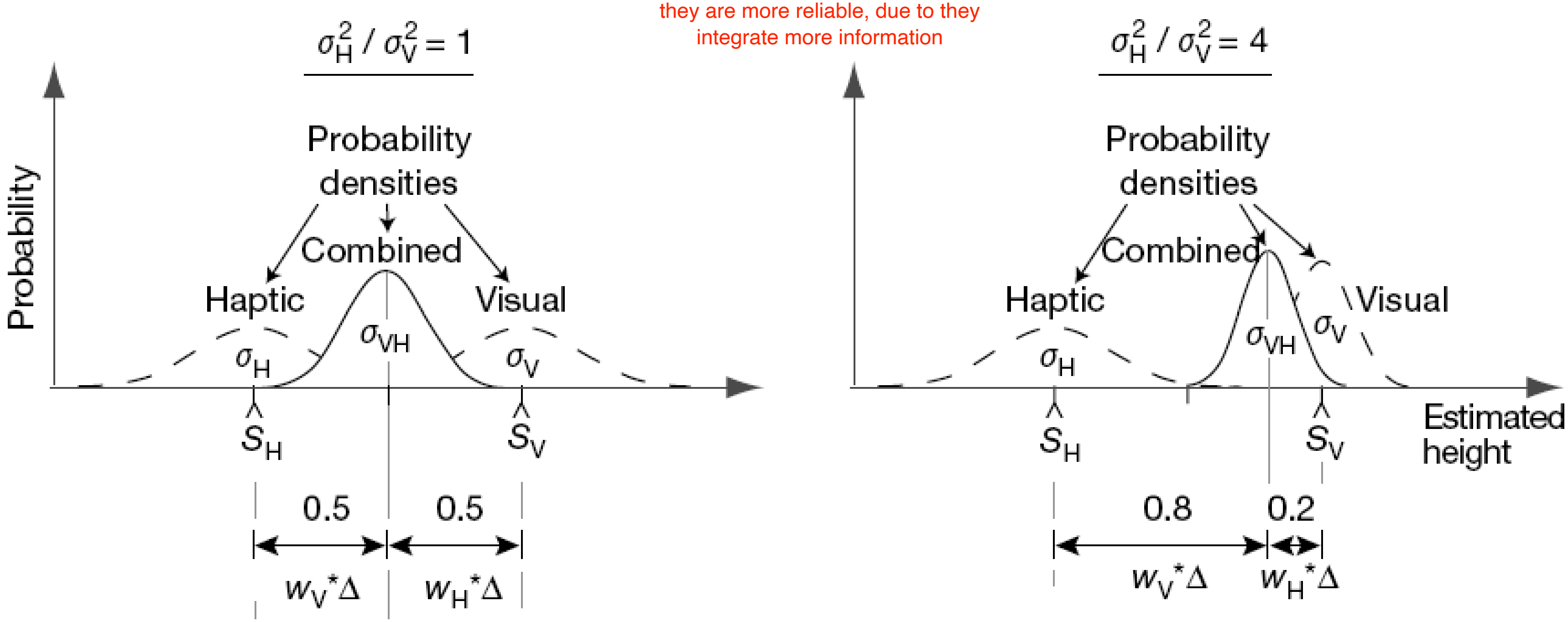
Maximum likelihood estimation model

- Perception is noisy
- Conditional independence P(S|A,V)=P(S|A)P(S|V)
  - nice and simple for continuous responses with Gaussian noise

unreliable: Gaussian much larger

when combined both perceptions, they are more reliable, due to they integrate more information

in this case vision is much more weighted, due to it is more reliable, narrow gaussian distribution



haptic and visual equally reliable

From Ernst and Banks, Nature, 2002

MLE for categorical responses (FLMP)

- We start with Bayes' rule…

$$P(r_i \mid S_A, S_V) = \frac{P(S_A, S_V \mid r_i)P(r_i)}{P(S_A, S_V)}$$

- and expand the numerator

$$P(r_i \mid S_A, S_V) = \frac{P(S_A, S_V \mid r_i)P(r_i)}{\sum_{j=1}^{N} P(S_A, S_V \mid r_j)P(r_j)}$$

denominator: normalization constant

- then we assume that the prior is flat

la probabilidad de la respuesta no depende del indice "i"

MLE for categorical responses (FLMP)

$$P(r_i \mid S_A, S_V) = \frac{P(S_A, S_V \mid r_i)}{\sum_{j=1}^{N} P(S_A, S_V \mid r_j)}$$

- Now we assume conditional independence of the auditory and visual stimulus, S_A and S_V given the response category r_i

$$P(r_i \mid S_A, S_V) = \frac{P(S_A \mid r_i) P(S_V \mid r_i)}{\sum_{j=1}^{N} P(S_A \mid r_j) P(S_V \mid r_j)}$$

MLE for categorical responses (FLMP)

- Then, we ’flip’ the likelihoods on the right hand side by inserting Bayes’ rule

$$P(r_i \mid S_A, S_V) = \frac{\frac{P(r_i|S_A)P(S_A)}{P(r_i)} \frac{P(r_i|S_V)P(S_V)}{P(r_i)}}{\sum_{j=1}^{N} \frac{P(r_j|S_A)P(S_A)}{P(r_j)} \frac{P(r_j|S_V)P(S_V)}{P(r_j)}}$$

- This expression needs to be reduced by dividing out $P(S_A), P(S_V)$ and $P(r_i) = P(r_j)$

MLE for categorical responses (FLMP)

- Finally we arrive at the FLMP

$$P(r_i \mid S_A, S_V) = \frac{P(r_i \mid S_A)P(r_i \mid S_V)}{\sum_{j=1}^{N} P(r_j \mid S_A)P(r_j \mid S_V)}$$

- The response probabilities for audiovisual stimuli is the normalized product of the response probabilities for auditory and visual stimuli

Homework project

- Overall goal: reproduce (some of the) findings in the paper by Andersen (JASA, 2015)

create a function that takes 4 free parameters (ca,cv,sigmaa,simgav) and outputs the model response probabilities
muA=1:5
~uA=ua-ca

Homework project

discrete responses, to map this responses to probabilities
1) insert a criteria into my model: continuum between ba and da “linear” (middle step is where observers would be 50/50)
2) therefore we can fit a psychometric function to our data

- Speech perception experiment
  - Auditory stimuli
    - Synthetic speech on a 5-point ba-da continuum
  - Visual stimuli
    - Synthetic speech on a 5-point ba-da continuum
  - Audiovisual stimuli
    - The 25 combinations of the 5 auditory and 5 visual stimuli

Homework project

- Data
  - The number (response count) of d-responses out of 24 trials from 5 subjects
- 1st problem
  - Fit the FLMP to the data