Models and Modes of Audiovisual integration

Andersen, Tobias

Publication date: 2015

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Models and Modes of Audiovisual integration

Tobias Andersen
Technical University of Denmark

Cognitive Systems

DTU Compute
Department of Applied Mathematics and Computer Science
Outline

• Categorical audiovisual perception
  – What’s so special?
    • Categorical, non-linear changes
      – The McGurk effect
      – Flashes and beeps
McGurk

McGurk and MacDonald, Nature, 1976
Illusory flashes and beeps

Illusory flashes and beeps

Andersen, Tiippana & Sams, Cognitive Brain Research, 2004
Illusory flashes and beeps

Andersen, Tiippana & Sams, Cognitive Brain Research, 2004
Illusory flashes and beeps

Andersen, Tiippana & Sams, Cognitive Brain Research, 2004
Illusory 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
Maximum Likelihood Estimation (MLE)

- 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
Multisensory integration

• Maximum likelihood rule nice and simple for Gaussian noise

\[
P(S|H,V) = \frac{P(S|H)P(S|V)}{\int P(S'|H)P(S|V) dS'}
\]

From Ernst and Banks, Nature, 2002
Early MLE - Classification

Probability density

$x_{12}$

$x_{23}$

Internal representation
Early MLE - Classification

\[ P(R = 1 | S) = \frac{P(x < x_{12} | S)}{P(x_{12} < x < x_{23} | S)} \]

\[ P(R = 2 | S) = \frac{P(x_{12} < x < x_{23} | S)}{P(x_{23} < x | S)} \]
Late MLE (a.k.a. FLMP)

\[ P(R_i \mid A, V) = \frac{P(R_i \mid A) \times P(R_i \mid V)}{\sum_{j=1}^{N} P(R_j \mid A) \times P(R_j \mid V)} \]

- Late integration (occurs after categorization)
- Only parameters: Unimodal response probabilities
- Generally good fits
Early vs. Late MLE

• Applied to illusory flashes and beeps
  – Early MLI generally has fewer free parameters
  – Early MLI fits our data better
  – Early MLI parameterizes reliability
    • a more parsimonious model
  – Early MLI orders responses / stimuli
    • 1 flash < 2 flashes < 3 flashes

Andersen, Tiippana & Sams, 2005
The UCSC corpus

Massaro (1998)
Early MLE applied to the UCSC corpus

Andersen (forthcoming), JASA
Early MLE applied to the UCSC corpus

Andersen (forthcoming), JASA
Early MLE applied to the UCSC corpus

Andersen (forthcoming), JASA
Early MLE applied to the UCSC corpus

• Linear spacing constraint
  – Reflect the experimental design
  – Reduces model complexity (10 -> 4 free parameters)
  – Allows Early MLE

Andersen (forthcoming), JASA
Results

![Graph showing RMSE for FLMP, Late MLE, and Early MLE with free parameters of 10, 4, and 4 respectively.]

Andersen (forthcoming), JASA
Results

Andersen (forthcoming), JASA
Fits by stimulus

Andersen (forthcoming), JASA
Results by subject

\[ N = 82 \]

\[ p < .01 \]

Andersen (forthcoming), JASA
Other models

• Free weight model
  – Separates spacing from variance
  – 1 additional free parameter
  – Better fit – worse prediction

• Equal weight model
  – with a logistic noise distribution it is equivalent to late MLE
  – No improvement in fit / prediction

Andersen (forthcoming), JASA
Audiovisual speech perception

Andersen & Winther, in preparation
Audiovisual speech perception

Andersen & Winther, in preparation
Audiovisual speech perception

Andersen & Winther, in preparation
Audiovisual speech perception

No Visual

No Auditory

Auditory /eke/

Auditory /epe/

Auditory /ete/

Andersen & Winther, in preparation
Audiovisual speech perception

Andersen & Winther, in preparation
The continuous internal representation

Auditory /T/  Audiovisual  Visual /P/

T  PT  PT  P

Andersen & Winther, in preparation
The continuous internal representation

Auditory /T/  Audiovisual  Visual /P/

Andersen & Winther, in preparation
The continuous internal representation

Summerfield, Phonetica, 1979
Andersen & Winther, in preparation
The continuous internal representation

Cyclical model without integration
35 parameters
RMSE = 0.004

Summerfield, Phonetica, 1979
Andersen & Winther, in preparation
The continuous internal representation

Cyclical model without integration
35 parameters
RMSE = 0.004

Early MLE
13 parameters
RMSE = 0.02

Summerfield, Phonetica, 1979
Andersen & Winther, in preparation
The continuous internal representation

Cyclical model without integration
35 parameters
RMSE = 0.004

Early MLE
13 parameters
RMSE = 0.02

Late MLE / FLMP
30 parameters
RMSE = 0.01

Summerfield, Phonetica, 1979
Andersen & Winther, in preparation
Cross-validation

• Leave one-out cross-validation
  – Late MLE / FLMP: poor results
  – Early MLE: Less poor results (but still poor)

• Why?
  – Non-linearity (not just number of free parameters)
  – Model fits very sensitive to small changes in parameter values
  – Assumes that the internal representation is unrealistically precise

Andersen & Winther, in preparation
Early MLE - Classification

\[
P(R = 1 \mid S) = \frac{P(x < x_{12} \mid S)}{P(x_{12} < x < x_{23} \mid S)}
\]

\[
P(R = 2 \mid S) = \frac{P(x_{12} < x < x_{23} \mid S)}{P(x_{23} < x \mid S)}
\]

Andersen & Winther, in preparation
regularization

• Don’t like something about your model?
• Optimize it away!

• I don’t like
  – Too high internal precision
    • Unrealistic
    • Makes models too flexible
    • Kills predictive power

• So, I add a penalizing term to the error when fitting

Andersen & Winther, in preparation
regularization

• Early MLE - Continuous representation
  • The critical parameter is the width, \( \sigma \), of the distributions
  • Apply a Gaussian prior on \( 1/\sigma \) centered at zero (flat distribution)
  • Penalizes for high precision

• Late MLE / FLMP
  • The parameters are the unimodal response probabilities
    - Apply a uniform symmetric Dirichlet prior
    \[
    P(P(R)) = \frac{1}{B(\alpha)} \prod_{r=1}^{N} P(R_r)^{\alpha-1}
    \]
    - Penalize for negative log prior w/o the normalization term, \( B(\alpha) \)
    \[
    -\log(P(P(R))) = -(\alpha - 1) \sum_{r=1}^{N} \log(P(R_r))
    \]
  - When the concentration parameter, \( \alpha = 1 \), the distribution is flat
    • Regularization penalizes peaked distributions
    • Peaked distributions are unstable

  Andersen & Winther, in preparation
regularization

FLMP           MLE           Hybrid

Regularized cross-validation RMSE

Best regularized cross-validation RMSE

RMSE

Multinomial expectation value

Regularization weight index (lower means less regularization)

Andersen & Winther, in preparation
How good is Early MLE with regularization?

Andersen & Winther, in preparation
How good is Early MLE with regularization?

Andersen & Winther, in preparation
How good is Late MLE / FLMP with regularization?

Andersen & Winther, in preparation
How good is Late MLE / FLMP with regularization?

Andersen & Winther, in preparation
How good is Late MLE / FLMP with regularization?

Andersen & Winther, in preparation
Conclusion

• Leave-one (stimulus/condition) out cross-validation is a great way to test models
  – Gives a good estimate of the right kind of generalization error

• Models of audiovisual speech perception benefits from
  – An underlying continuous parameter
  – Regularization

• The computational mechanism of integration is still unknown
  – Current results favor Early MLE
  – The Hybrid model performs almost as well
  – Weighted models make more sense

Andersen & Winther, in preparation
Modes of perception
Modes of perception
Sine-wave speech
Sine Wave Speech

• Created by placing time-varying sine wave tones at the three lowest formants of the speech signal

• Naïve observers do not recognize sine wave speech as speech

• Informed observers can understand the phonetic content
Sine Wave Speech - Stimuli

From Tuomainen, Andersen, Tiippana and Sams, Cognition, 2005
Sine Wave Speech - Paradigm

1. Training in non-speech mode (SWS)
2. Testing in non-speech mode (SWS)
3. Testing natural speech
4. Training in speech mode (SWS)
5. Testing in speech mode (SWS)

From Tuomainen, Andersen, Tiippana and Sams, Cognition, 2005
Sine Wave Speech - Results

From Tuomainen, Andersen, Tiippana and Sams, Cognition, 2005
Sine Wave Speech - Conclusion

• Strong audiovisual integration of sine wave speech and the talking face

• But! Only when observers are in speech mode

• Demonstrates strong top-down influence on audiovisual integration of speech

From Tuomainen, Andersen, Tiippana and Sams, Cognition, 2005
Audiovisual detection advantage

- The AV detection advantage (Grant & Seitz, JASA, 2000)
  - Acoustic speech detection threshold lowered by congruent visual speech
  - AV gain sizes reported between 1.6 and 2.7 dB, depending on method
  - Not just a response bias
    - 2 AFC w/ adaptive staircase procedure – visual information identical in the 2 alternatives

- Is it speech specific?

AV detection - results

The AV detection advantage occurs also for SWS

No difference in AV detection advantage between nonspeech and speech conditions

Identification - results

Identification, percent correct

EEG – mismatch negativity MMN

Stekelenburg & Vroomen (2012), Neuropsychologia
EEG – N1 and P2

Baart, Stekelenburg & Vroomen (2014), Neuropsychologia
Margaret Thatcher
Margaret Thatcher
The McThatcher MMN

Eskelund, MacDonald & Andersen (2015), Neuropsychologia
Modes of perception

• Phonetic audiovisual integration varies for very similar stimuli
  – Sine-wave speech
  – McThatcher effect

• Audiovisual integration is a multi-stage process
  – Speech mode in the McGurk illusion and the detection advantage

• Phonetic audiovisual integration is reflected in the MMN and the P2
  – But not the N1
Thanks for listening

Any ??
Audiovisual SDT
Audiovisual SDT

• Audiovisual integration in signal detection
  – Sound can enhance visual sensitivity
  – Frasinetti et al., 2003

• Integration of magnitude in weak signals
  – Cat chasing mouse in the dusk
  – Involves the Superior Colliculus
  – Directs attention to the location of a change
    • Stein et al.

• Loudness increase perceived brightness
Paradigm

Andersen & Mamassian, Vision Research, 08
Perceptual effects

• Sound carries no information
• Bias free paradigm
• Two stimulus attributes may integrate audiovisually:
  – Transients
  – Sustained loudness and brightness

Andersen & Mamassian, Vision Research, 08
Attention

• Directional effects
  – If louder makes brighter, then a luminance decrease should be more difficult to detect when the sound becomes louder

• Additional task
  – Identify the luminance change as an increase or decrease

• Attentional effects
  – Exogenous attentional cueing
  – Reduction of temporal uncertainty

Andersen & Mamassian, Vision Research, 08
Paradigm

Lumin: ↑
Sound: —
Lumin: ↑
Sound: —
Lumin: ↓
Sound: —
Lumin: ↓
Sound: ↑
Lumin: ↓
Sound: ↓

Andersen & Mamassian, Vision Research, 08
Predictions
Loudness/brightness interaction

Andersen & Mamassian, Vision Research, 08
Predictions
Attention and Uncertainty

Andersen & Mamassian, Vision Research, 08
Predictions
Transient interactions

Andersen & Mamassian, Vision Research, 08
Results

Andersen & Mamassian, Vision Research, 08
Transient hypothesis

• A true perceptual integration of rapid transients in the intensity of auditory and visual signals
• In excellent agreement with physiological studies of the Superior Colliculus
• These studies predict a temporal window of integration of 100 ms
• This can be tested by varying the audiovisual SOA
  – Should eliminate uncertainty reduction
Predictions
Transient interactions
Predictions

Attentional cueing
Results

![Graph showing results with sound lag in ms and percentage change in hit rate.](image)
Conclusions

• Sound intensity increase visual sensitivity
  – when lagging with 75 ms but not when lagging 150 ms
    • Cannot be due to exogenous attention
  – When stimulus asynchrony varies randomly
    • Cannot be due to reduction of uncertainty

• In good agreement with response properties of SC neurons
Summary

• Categorical audiovisual perception
  – Special: Strong, non-linear effects
    • Tricky to model!
    • Needs regularization
  – Not so special
    • Information reliability
    • Modality appropriateness
    • Continuous quantitative models apply
      – When adding a response boundary
      – Provides predictive power when regularized
    – McGurk Depends on top-down effects (Speech mode)
    – Multi-dimensional (multi-faceted)
Summary

- Audiovisual integration in signal detection
  - Based on transients
    - Not on intensity
  - Separable from attentional cueing
    - And reduction of temporal uncertainty