The role of surprise in theory testing
A case study from music cognition

Henkjan Honing

ILLC | CSCA
University of Amsterdam

www.hum.uva.nl/mcg
Thanks to

- Institute for Logic, Language & Computation (ILLC)
- Cognitive Science Center Amsterdam (CSCA)
- Royal Netherlands Academy of Arts and Sciences (KNAW)
Outline

- Computational modeling of music cognition
- Case study
  - How to select among alternative models?
  - What makes a model surprising?
- Discussion
Computational modeling

- Music
- Stimuli
- Machine
- Knowledge
- Algorithm
- Behavior

Computational modeling

Mind
Stimuli
Machine

Mental Process
Algorithm

Behavior

exhibits
models
agrees


Friday, 25 March 2011
Computational modeling

- Mind
- Stimuli
- Machine

Mental Process → Behavior
- exhibits
- models
- validates

Algorithm → Behavior
- exhibits
- exhibits


Friday, 25 March 2011
Model selection methods

A. Measure of goodness-of-fit (e.g. Rodgers & Rowe, 2002)
B. Measure of simplicity (e.g. Pitt, Myung, & Zhang, 2002)
C. Measure of surprise (cf. Popper, 1963)

D. ‘Ground-truth’ verification (e.g. Rijsbergen, 1979)
E. ...
Case study

- Computational models of expressive timing in music performance (e.g., Sundberg & Verillo, 1980; Kronman & Sundberg, 1987; Longuet-Higgins & Lisle, 1989; Feldman, Epstein & Richards, 1992; Todd, 1992; Epstein, 1994; Todd, 1995; Friberg & Sundberg, 1999; Large & Palmer, 2002)

- Modeling the Final Ritard: Typical slowing down at the end of a music performance (e.g., Hudson, 1996; Clarke, 1999; Gabrielsson, 1999)
Two computational approaches

- **Kinematic approach** (*K model*)
  Predicts shape of expressive timing patterns and how they conform to the laws of physical motion (*commonality*)

- **Perception-based approach** (*P model*)
  Predicts the amount of expressive freedom a performer has in the interpretation of a rhythmic fragment before being ‘misinterpreted’ as an altogether different rhythm (*diversity*)
\[ v(t) = u + at \quad (1) \]
\[ v(x) = \left( u^2 + 2ax \right)^{1/2} \quad (2) \]
\[ v(x) = \left[ 1 + (w^q - 1)x \right]^{1/q} \quad (3) \]
Mechanical version of $K$ model
Mechanical version of K model

Constant Braking Force Model of the Final Ritard

www.hum.uva.nl/mmm/fr
Mechanical version of $K$ model
P model

Two components:

- Model of perceived regularity (*tempo tracker*)
  (Large & Jones, 1999; Toiviainen, 1999)

- Model of rhythmic categorization (*quantizer*)
  (Longuet-Higgins, 1987; Desain & Honing, 1989)
Effect of rhythm and tempo on predictions

Honing (2005)
A. Measure of good fit

**Table 2. Results for Datasets F&S99 and S&V80 with (+) and without (−) the Last IOI.**

<table>
<thead>
<tr>
<th>Set</th>
<th>Last</th>
<th>Mean r²</th>
<th>SD</th>
<th>Mean r²</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F&amp;S99</td>
<td>+</td>
<td>.98</td>
<td>.01</td>
<td>.90</td>
<td>.07</td>
<td>*</td>
</tr>
<tr>
<td>F&amp;S99</td>
<td>−</td>
<td>.89</td>
<td>.08</td>
<td>.97</td>
<td>.04</td>
<td>*</td>
</tr>
<tr>
<td>S&amp;V80</td>
<td>+</td>
<td>.95</td>
<td>.05</td>
<td>.91</td>
<td>.07</td>
<td>n.s.</td>
</tr>
<tr>
<td>S&amp;V80</td>
<td>−</td>
<td>.86</td>
<td>.09</td>
<td>.97</td>
<td>.05</td>
<td>**</td>
</tr>
</tbody>
</table>

* p < .01. ** p < .001.
Conclusion of method A

- Measures of GOF only assess fit
- GOF is *not* able to distinguish between variations in the data caused by noise and those that the model was designed to capture
- Even if one model would have a significant better fit we could not select that model over the other
B. Measure of flexibility

‘Response area’: Range of possible predictions

K model

P model

Honing (2005)
B. Measure of flexibility

Effect of note density and rhythmic structure

P model

Honing (2005)
Conclusion of method B

- Both models making roughly similar fits to the data
- K model simpler than the P model
- However, the P model show less flexibility, and should hence be preferred
Conclusion of method B

- Both models making roughly *similar fits* to the data
- K model *simpler* than the P model
- However, the P model show less *flexibility*, and should hence be preferred

- Still, we can wonder how ‘surprising’ all this is in the context of the phenomenon modeled
C. Element of ‘surprise’
C. Element of ‘surprise’
Towards a *measure of surprise*

- ‘Confirmations [of a theory] should count only if they are the result of risky predictions; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory — an event which would have refuted the theory.’ (Popper, 1963:47)
Towards a measure of surprise

- Correct prediction of an unlikely event is more surprising than the correct prediction of something that was expected anyway.

- Prefer the model that:
  - Minimizes the intersection of $H_{\text{predicted}}$ with respect to $H_{\text{plausible}}$
  - While preferring the $H_{\text{predicted}}$ that is least smooth
  - ...

Honing & Romeijn (in prep.)
Towards a *measure of surprise*

Prefer the model that:

1. Fits the empirical data well (*best fit*)
2. Makes limited range predictions (*least flexible*)
3. Makes unexpected predictions (*most surprising*)
Meeting a friend in the corridor, Wittgenstein said:

“Tell me, why do people always say it was natural for men to assume that the sun went round the earth rather than that the earth was rotating?”

His friend said: “Well, obviously, because it just looks as if the sun is going round the earth.”

To which the philosopher replied, “Well, what would it have looked like if it had looked as if the earth was rotating?”
Meeting a friend in the corridor, Wittgenstein said: “Tell me, why do people always say it was natural for men to assume that the sun went round the earth rather than that the earth was rotating?”

His friend said: “Well, obviously, because it just looks as if the sun is going round the earth.”

To which the philosopher replied, “Well, what would it have looked like if it had looked as if the earth was rotating?”
References


