ECVP 2007

The Attention Cascade Model and Attentional Blink in RSVP

Dr. Shui-I Shih
University of Southampton

This set of slides is prepared for the convenience of ECVP reviewers. Most animated graphical or diagramic illustrations, which one would expect to accompany the oral presentation, are replaced with written descriptions.
Search in Rapid Serial Visual Presentation and Attentional Blinks

T1 occurs

Target Onset Asynchrony

%Correct Response

Attentional Blink

P(T1)

P(T2|T1)
Cognitive Accounts of the AB

- Inhibition model (Raymond, Shapiro, & Arnell, 1992)
- Interference model (Shapiro, Raymond, & Arnell, 1994)
- Two-stage model (Chun & Potter, 1995)
- Two-stage competition model (Potter, Staub, & O’Connor, 2002)
- Central interference model (Jolicoeur, 1999)
- Hypothesis of attentional dwell time (Duncan, Ward, & Shapiro, 1994)
- Hypothesis of temporal loss of control (Di Lollo, Kawahara, Ghorashi, Enns, 2005)
Neural Networks

- Simultaneous type, serial token model of temporal attention and working memory (Bowman & Wyble, 2007)
- Connectionist model of the AB (Chartier, Cousineau, & Charbonneau, 2004)
- Neuronal workspace model of conscious access (Dehaene, Sergent, & Changeux, 2003)
- Neurodynamic model of the AB (Fragopanagagos, Kockelkoren, & Taylor, 2005)
The Attention Cascade Model

- A mathematical model
- A simple mathematical function (e.g., rectangular, exponential, or Gaussian) to characterize each processing stage at a macro level (e.g., the net outcome of an ensemble of neurons)
- Much fewer parameters than neural networks (6 free parameters)
Five Common Assumptions of AB Models

see the next slide for descriptions
Descriptions of the Five Common Assumptions

1. Each RSVP stimulus activates its LTM traces and forms a preliminary representation (PR)
2. A PR is unreportable unless it has been transferred and consolidated in working memory
3. The transfer is initiated by any stimulus (e.g., T1 and T2) that sufficiently matches the target templates (i.e., with high top-down salience)
4. The stimulus trailing a potential target (i.e., T1+1 and T2+1) may enter working memory if the presentation rate is fast
5. The AB effect arises from limitations in working memory
Attention Cascade Model

See subsequent slides for descriptions
Descriptions of the Attention Cascade Model

• It shares the five common assumptions.

• A rectangular function of SOA wide is assumed for each preliminary representation (PR)

• A rectangular function of width \( w \) is assumed for the attention window (that transfers PRs to working memory); \( w \) depends on presentation rate, task demand, etc.

• Two routes to trigger the attention window:
  – Controlled: a top-down (but not bottom-up) salient stimulus \( \rightarrow \) sensory processor \( \rightarrow \) LTM \( \rightarrow \) peripheral buffer (which holds PRs) \( \rightarrow \) attention control mechanism (ACM) [4 processing stages]
  – Automatic: a bottom-up salient stimulus \( \rightarrow \) sensory processor \( \rightarrow \) ACM [2 processing stages]

• The triggering time is thus a random variable of a sum of 4 (controlled) or 2 (automatic) exponential functions (see next slide)
Triggering Time Distribution of the Attention Window

- **Sensory Processor**
- **LTM**
- **Peripheral Buffer**
- **Attention Control Mechanism**

**Controlled**

- Sensory Processor: *
- LTM: *
- Peripheral Buffer: *
- Attention Control Mechanism: =

**Automatic**

- Sensory Processor: *
- LTM: *
- Peripheral Buffer: *
- Attention Control Mechanism: =
• The area of the preliminary representation (PR) function of a stimulus that overlaps with the attention window (AW) gives the strength value $s$ that the stimulus is initially held in the working memory buffer.
  – If the strength is greater than a response threshold, then the stimulus is passed to the response buffer ready for output.
  – Otherwise, it requires further processing in the consolidation processor (CP).

• If the CP is engaged, the stimulus has to queue and its strength decays exponentially with the rate parameter $\lambda$ contingent on its initial strength $s$ so that the higher the $s$ is the slower the decay.
The consolidation processor (CP) with limited capacity $C$, once available, inputs all stimuli in the working memory buffer.

Stimulus strengths are changed along the successive operations in the CP.

- The CP weighs each stimulus according to its top-down salience (e.g., target x 1, distractor x 0).
- If the sum of weighted strengths $\sum q_i > C$, each $q_i$ is weighted by $C/\sum q_i$ so that all items are weakened just enough that the limit $C$ is not breached.
- During processing, each strength grows as a $CDF$ of an exponential distribution with the rate parameter contingent on the weighted/weaken strength $r_i$ so that the higher the $r_i$ is the more rapidly it grows.
• After $\pi$ ms of consolidation, a stimulus is correctly identified if its resulting strength is greater than that of the internal noise at that moment; otherwise a guess will be made.
  – Because the internal noise reflects the trial to trial variation, the current application assumes constant processing duration $\pi$.

• A Gaussian distribution with a mean $\mu_n$ and a standard deviation $\sigma_n$ is assumed for the internal noise.
### Summary of the Model Parameters

<table>
<thead>
<tr>
<th></th>
<th>Assumed</th>
<th>Estimated</th>
<th>Derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary Rep (PR)</td>
<td>[SOA]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Window (AW)</td>
<td>[ ]</td>
<td>( w )</td>
<td></td>
</tr>
<tr>
<td>Pre-AW stage</td>
<td>2 or 4 stages</td>
<td>( \beta )</td>
<td>TTD of AW</td>
</tr>
<tr>
<td>Initial strength, ( s_i )</td>
<td>PR-AW overlap</td>
<td></td>
<td>( 0 \leq s_i \leq 1 )</td>
</tr>
<tr>
<td>Consolidation duration</td>
<td></td>
<td>( \pi )</td>
<td></td>
</tr>
<tr>
<td>Queuing time</td>
<td></td>
<td></td>
<td>( \text{Max}(0, \pi - TOA) )</td>
</tr>
<tr>
<td>Decay rate</td>
<td>Exponential pdf</td>
<td></td>
<td>( 1 - s_i )</td>
</tr>
<tr>
<td>Strength after decay, ( q_i )</td>
<td></td>
<td></td>
<td>( 0 \leq q_i \leq 1 )</td>
</tr>
<tr>
<td>Resource</td>
<td>( C )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned resource, ( r_i )</td>
<td></td>
<td></td>
<td>( r_i = q_i \cdot \text{Min}(1, \ C/\sum q_i) )</td>
</tr>
<tr>
<td>Growth rate</td>
<td>Exponential CDF</td>
<td></td>
<td>( r_i )</td>
</tr>
<tr>
<td>Mean of internal noise</td>
<td>Gaussian</td>
<td>( \mu )</td>
<td></td>
</tr>
<tr>
<td>SD of internal noise</td>
<td>Gaussian</td>
<td>( \sigma )</td>
<td></td>
</tr>
</tbody>
</table>
The Attention Cascade Model
Accounts for Shih and Reeves (2007)

• Two experiments
  – Stimuli: photometrically equiluminant red, green, and yellow characters
  – Single RSVP at 10 Hz (SOA = 100 ms)
  – T1: odd digit (3, 5, 7, 9)
  – T2: even digit (2, 4, 6, 8)
  – Distractors: captial letters
  – Yielded 150 data points (84 and 66 from Exps 1 and 2 respectively)

• The 6 parameters are estimated to account for the 150 data points simultaneously.

• The next slide depicts the two experiments. It is followed by 3 slides that present the data and model predictions.
• Experiment 1: T1 Salience x T2 Salience x TOA
  – Yellow distractors, yellow or red T1, yellow or green T2
  – TOA: target onset asynchrony, 100 – 800 ms
• Experiment 2: Task Relevance of Salience x S2-T2 Lag x TOA
  – Two salient items (S1 and S2) in each RSVP
  – T2 was always non-salient
  – Task relevance of saliency (the two levels were run in separate blocks)
    • T1 = S1 OR Non-salient T1 and S1 occurred several items before T1
    – S2-T2 Lag: S2 as T2-1 OR T2+1

<table>
<thead>
<tr>
<th>Examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exp 1</strong></td>
</tr>
<tr>
<td>N-N-100</td>
</tr>
<tr>
<td>N-S-200</td>
</tr>
<tr>
<td>S-N-300</td>
</tr>
<tr>
<td>S-S-600</td>
</tr>
<tr>
<td><strong>Exp 2</strong></td>
</tr>
<tr>
<td>R-100-T2+1</td>
</tr>
<tr>
<td>I-300-T2-1</td>
</tr>
</tbody>
</table>
Experiment 1, Shih and Reeves (2007)

Data

SN: Salient T1, Non-salient T2
SS: Salient T1, Salient T2
NS: Non-salient T1, Salient T2
NN: Non-salient T1, Non-salient T2

Model

TOA (ms)

Mean P(T1)

Mean P(T2)

Mean P(T2|T1)

Model

TOA (ms)
Relevant Salience Condition, Experiment 2, Shih and Reeves (2007)

**Data**

- **Graph a**: Mean $P(T_1)$ vs. TOA (ms)
- **Graph b**: Mean $P(T_2)$ vs. TOA (ms)
- **Graph c**: Mean $P(T_2|T_1)$ vs. TOA (ms)

**Model**

- **Graph d**: Model $P(T_1)$
- **Graph e**: Model $P(T_2)$
- **Graph f**: Model $P(T_2|T_1)$
Irrelevant Salience Condition, Experiment 2, Shih and Reeves (2007)

**Data**

- Mean $P(T1)$
- Mean $P(T2)$
- Mean $P(T2|T1)$

**Model**

- Mean $P(T1)$
- Mean $P(T2)$
- Mean $P(T2|T1)$

*Irrelevant Salience Condition, Experiment 2, Shih and Reeves (2007)*
The Attention Cascade Model
Accounts for Potter, Staub, and O’Connor (2002)

- SOA = 53 ms
- T1 and T2: words
- Distractors: #### or %%%%

<table>
<thead>
<tr>
<th></th>
<th>Exp 1</th>
<th>Exp 2</th>
<th>Exp 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSVP</td>
<td>Single</td>
<td>Unsynchronized dual</td>
<td></td>
</tr>
<tr>
<td>TOA (ms)</td>
<td>53, 107, 213</td>
<td>0, 40, 107, 213</td>
<td>0, 13, 27, 40</td>
</tr>
</tbody>
</table>

- Exp 2&3 vs. 1: differ in acuity, number of items entering working memory, noise, etc.
- These differences are absorbed into the internal noise $N(\mu, \sigma)$
- 22 data points
Potter, Staub, and O'Connor (2002)

Data

Model

Accuracy, Exp. 1

Accuracy, Exp. 2

Accuracy, Exp. 3

TOA (ms)
A comparison:

Simultaneous type, serial token model of temporal attention and working memory
(Bowman & Wyble, 2007, *Psychological Review*)
Another comparison:

Connectionist model of the AB (Chartier, Cousineau, & Charbonneau, 2004)

Figure 9: Examples of individual data from 4 participants (top row) and from the simulations (bottom row).
# Summary of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Shih and Reeves Exps 1 and 2</th>
<th>Potter et al. Exp 1 2 and 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (ms)</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>$w$ (ms)</td>
<td>148</td>
<td>80</td>
</tr>
<tr>
<td>$C$</td>
<td>103</td>
<td>38</td>
</tr>
<tr>
<td>$\pi$ (ms)</td>
<td>525</td>
<td>530</td>
</tr>
<tr>
<td>$\mu_n$ (ms)</td>
<td>-3</td>
<td>-4</td>
</tr>
<tr>
<td>$\sigma_n$ (ms)</td>
<td>63</td>
<td>38</td>
</tr>
<tr>
<td>Mean $R^2$</td>
<td>0.87</td>
<td>0.72 (0.79)</td>
</tr>
</tbody>
</table>

\[
\frac{w}{SOA} : \frac{148}{100} \approx \frac{80}{53} \approx 1.5
\]

\[
\frac{C}{SOA} : \text{the number of item the available resources can process}
\]

\[
\frac{103}{100} \sim 1 \text{ letter}
\]

\[
\frac{38}{53} \sim 0.7 \text{ word}
\]
Conclusions

• The attention cascade model provides reasonable, computational accounts for the results of Potter, Staub, and O'Connor (2002) and Shih and Reeves (2007).
• The model parameters have reasonable values.