



HAL
open science

Bayesian modelling of visual attention in word recognition: simulating optimal viewing position

Thierry Phénix, Sylviane Valdois, Julien Diard

► To cite this version:

Thierry Phénix, Sylviane Valdois, Julien Diard. Bayesian modelling of visual attention in word recognition: simulating optimal viewing position. 19th Conference of the European Society for Cognitive Psychology (ESCOP), 2015, Paphos, Cyprus. hal-02004341

HAL Id: hal-02004341

<https://hal.science/hal-02004341v1>

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction

- Word Recognition** is the cornerstone of reading. It is a dynamic process that emerges from the interaction between low-level visual processing of input letters and the activation of memorized orthographic knowledge.
- Visual Attention (VA)** is critical to process multiple elements simultaneously. VA capacity constrains VA span (max number of items identified in parallel), thus the number of letters that can be simultaneously processed when reading. This capacity is highly limited (4-5 items) [1,2].



Current word recognition models do not include visual attention as a key mechanism.
How to model the role of Visual Attention in Word Recognition?

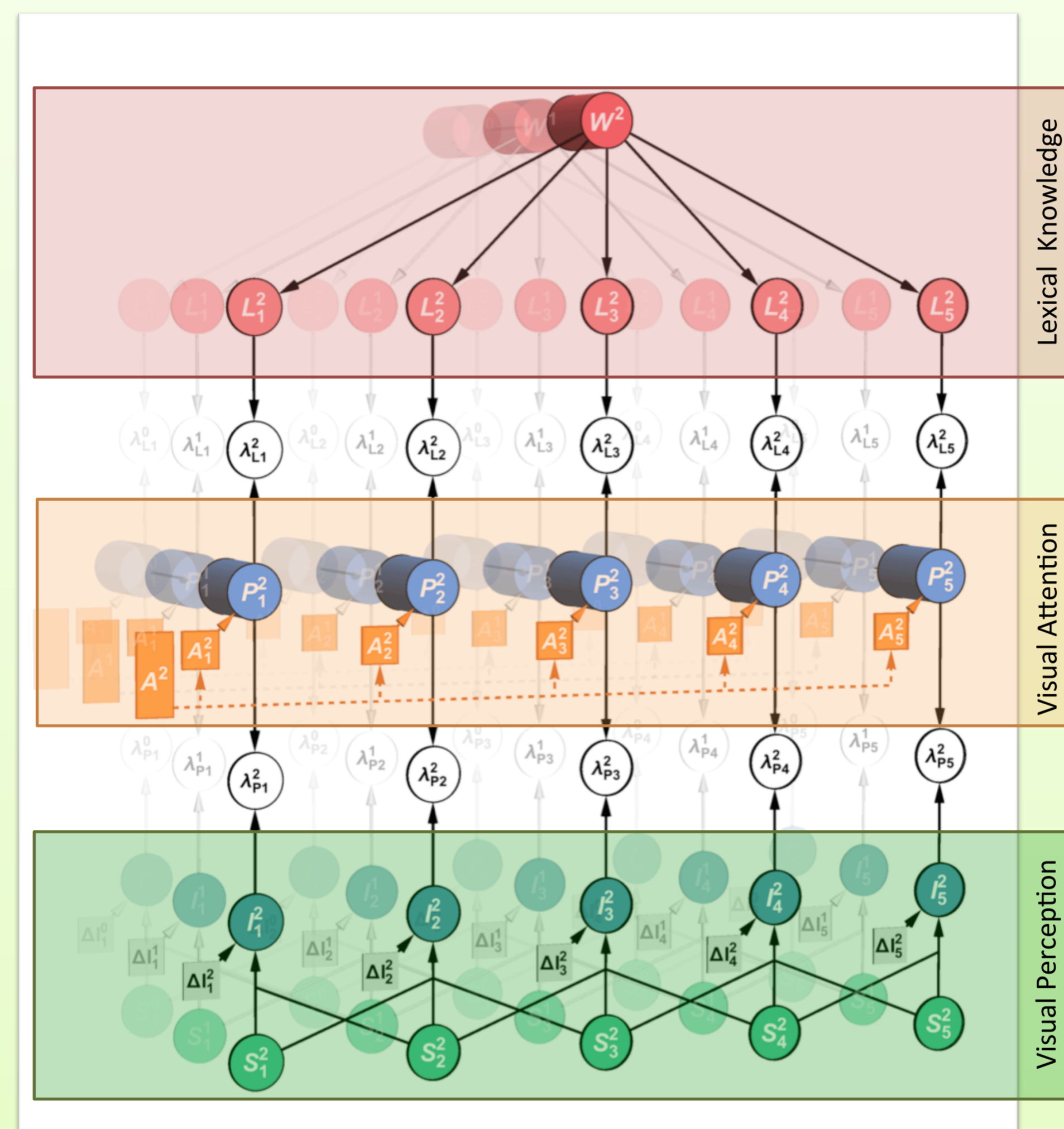


- BRAID** (Bayesian word Recognition with Attention, Interference and Dynamics). **BRAID** is a probabilistic word recognition model that incorporates control of **attention** resources, lateral **interference** between visual inputs (crowding effect), and temporal **dynamics** of information processing, in addition to bottom-up letter identification and top-down orthographic knowledge.
- OVP Effect** (Optimal Viewing Position): word recognition performance varies as a function of fixation location within the word. Can **BRAID** simulate the OVP function?

BRAID Bayesian Reading with Attention, Interference and Dynamic

Variables

- W^t Internal word representation at time t
- L_n^t Letter of internal word at position n and time t
- λ_n^t Coherence variable between P and L at position n and time t
- P_n^t Dynamic percept at position n and time t
- A^t Attention over percepts at time t
- C_n^t Attention control of percept n at time t
- λ_n^t Coherence variable between I and P at position n and time t
- I_n^t Interferences at position n and time t
- ΔI_n^t Interference weights at position n and time t
- S_n^t Letter at position n in the stimulus at time t

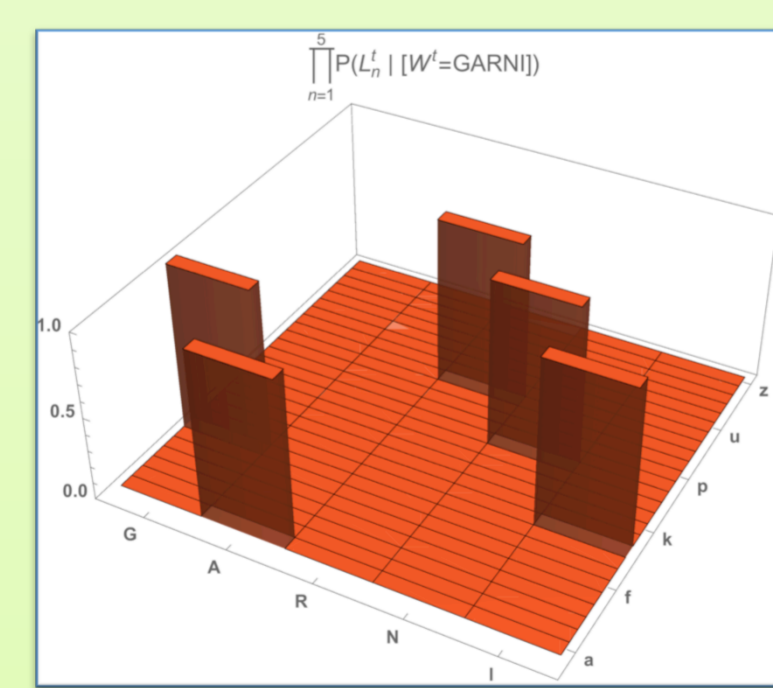


Mathematical formulation of the joint distribution of the model

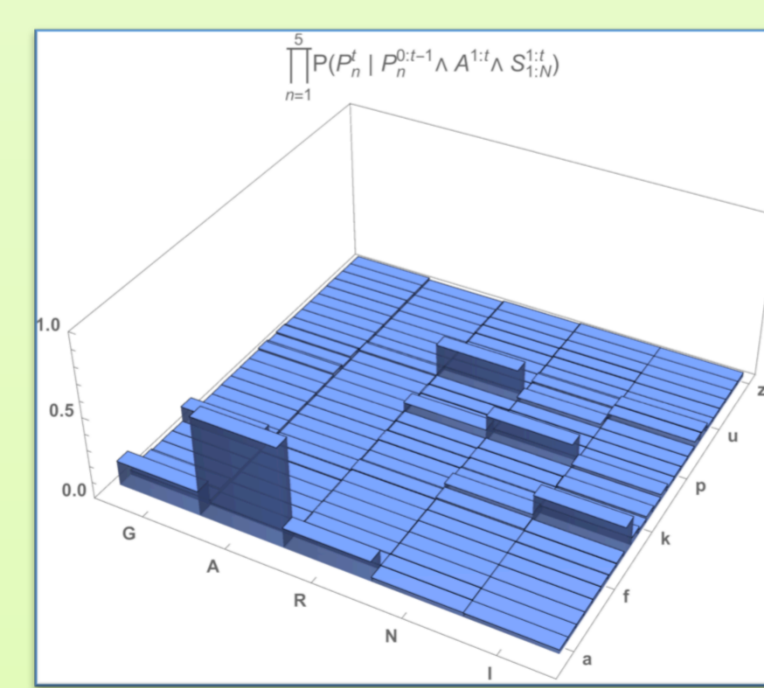
$$P(W^{0:T} L_{1:N}^{1:T} \lambda_{1:N}^{1:T} A^{1:T} C_{1:N}^{1:T} P_{1:N}^{0:T} \lambda_{1:N}^{1:T} \Delta I_{1:N}^{1:T} I_{1:N}^{1:T} S_{1:N}^{1:T}) = \prod_{n=1}^N \left[P(W^n | W^{n-1}) \prod_{t=1}^T \left[P(L_n^t | W^n) P(\lambda_n^t | L_n^t P_n^t) \right] \right] \prod_{n=1}^N \left[P(A^t) \prod_{n=1}^N \left[P(C_n^t | A^t) P(P_n^t | P_n^{t-1} C_n^t) P(\lambda_n^t | P_n^t I_n^t) \right] \right] \prod_{n=1}^N \left[P(S_n^t) P(\Delta I_n^t) P(I_n^t | S_{1:N}^t \Delta I_n^t) \right]$$

Distribution definitions

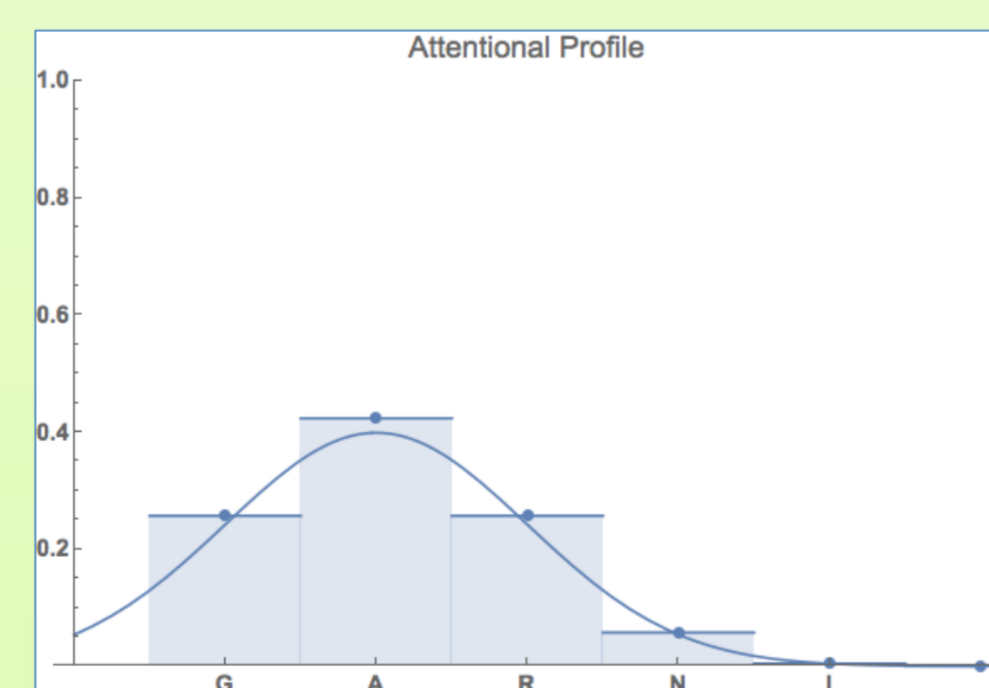
- $P(W^0)$ Represents the resting state of the internal word activation. Typically, it is based on word frequency.
- $P(W^t | W^{t-1})$ Represents transition of internal word activation from time $t-1$ to t . We implement a memory decay to the resting state.
- $P(L_n^t | W^t)$ Represents orthographic knowledge.
- $P(\lambda_n^t | L_n^t P_n^t)$ Guarantees L_n^t and P_n^t have the same value during the computation.
- $P(P_n^t | P_n^{t-1} C_n^t)$ Represents transition of percept n from time $t-1$ to t . Decay here is controlled by attention $P(C_n^t | A^t)$. Allocating attentional resources blocks decay.
- $P(A^t)$ Represents the distribution of attention over percepts. We use a Gaussian distribution centered on the viewing position. Its standard deviation models Visual Attention Span.
- $P(\lambda_n^t | P_n^t I_n^t)$ Guarantees P_n^t and I_n^t have the same value during the computation.
- $P(I_n^t | S_{1:N}^t \Delta I_n^t)$ Represent the crowding effect. We implement only direct neighbor interferences. The result is weighted by $P(\Delta I_n^t)$.
- $P(S_n^t)$ Represents prior knowledge about sensory stimulus. Typically, we use a uniform distribution.



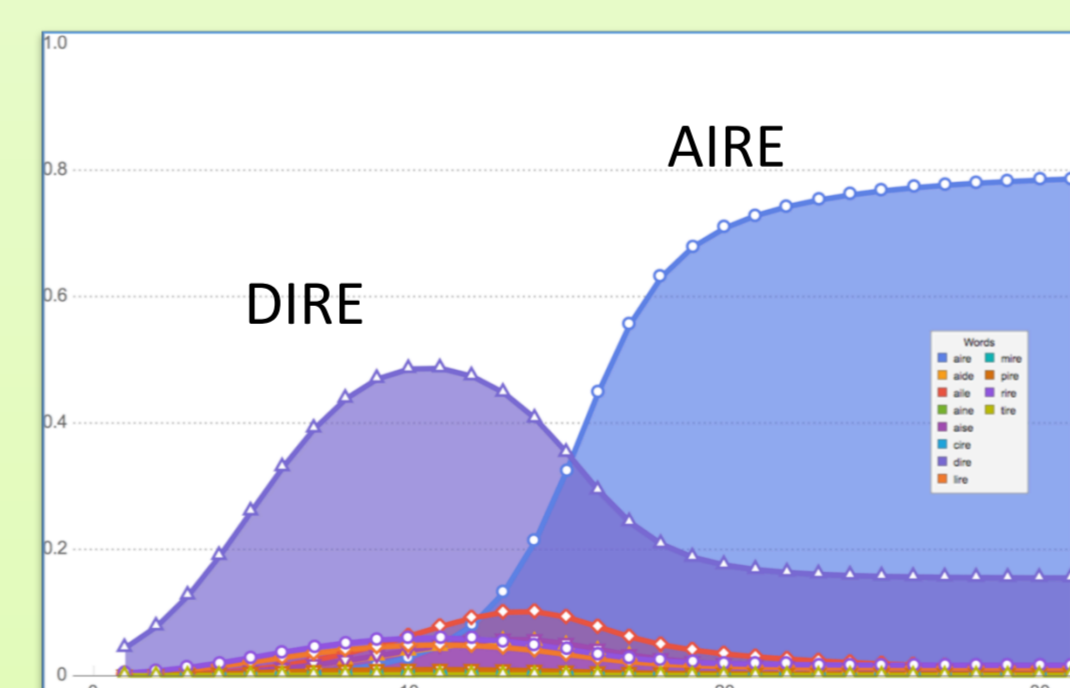
Internal representation for the word "GARNI".



Percept build on the stimulus "GARNI" after 10 iterations.



Attentional profile with a mean = 2 and standard deviation = 1.



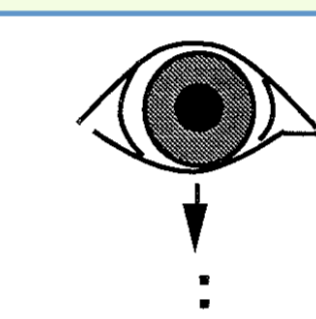
Recognition of stimulus "AIRE". Massive effect of frequency due to competition with "DIRE".

Experiment: simulating the OVP Effect

Behavioral Experiment (Montant, Nazir & Poncet 1998) [4]

Word recognition task with variable viewing position:

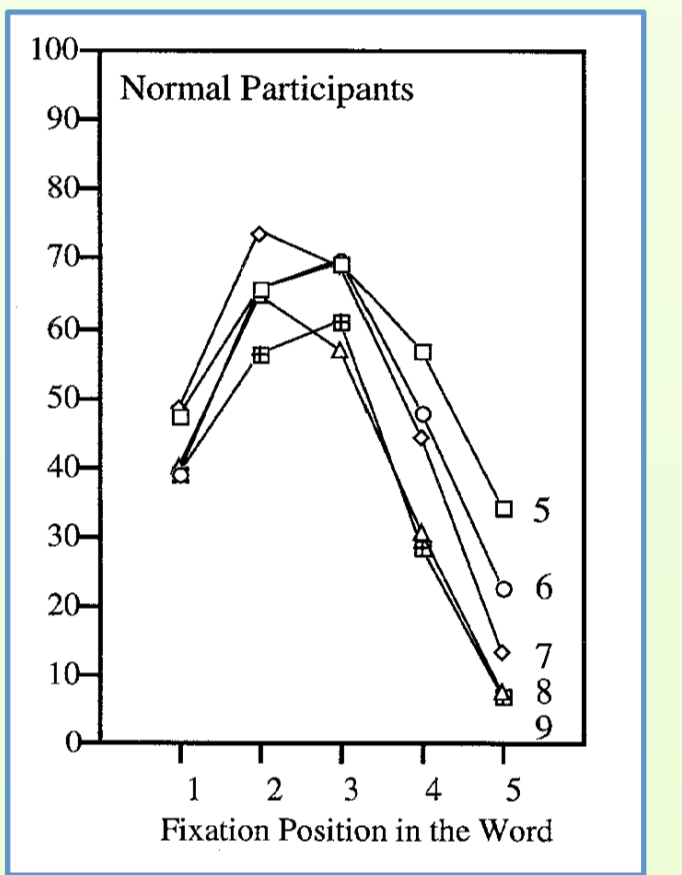
After looking at a fixation point for 1 s, a word is presented to the participant for a short duration, avoiding any eye movement, and shifted in function of desired viewing position.



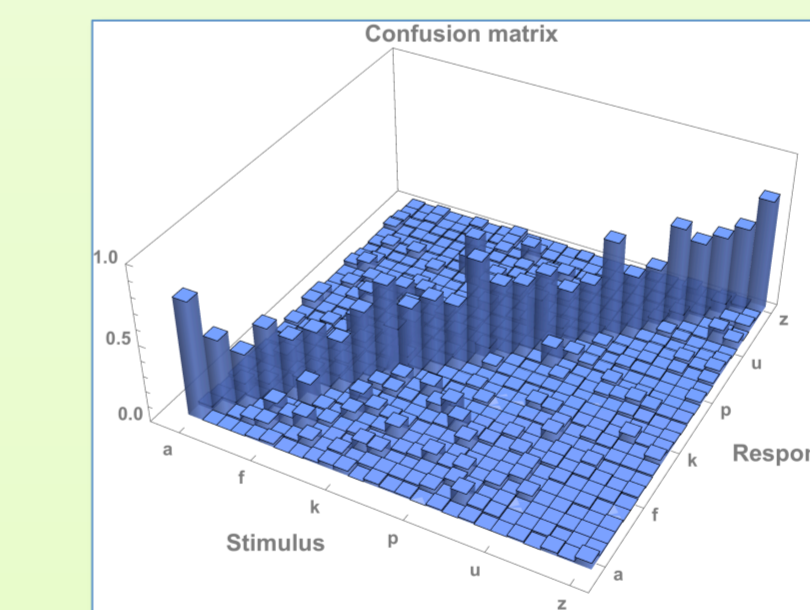
Word length factor: Word length varies between 5 to 9 letters, with a total of 250 words (50 per length).

Viewing Position factor: Word was divided into 5 equally wide zones. The center of each zone was the initial fixation position.

- Viewing position effect was significant for all lengths ($p < .001$).
- Performance was optimal when words were fixated slightly left of their center ($p < .001$).
- Performance was highly sensitive to word frequency ($t(248) = 5.27, P < .001$).

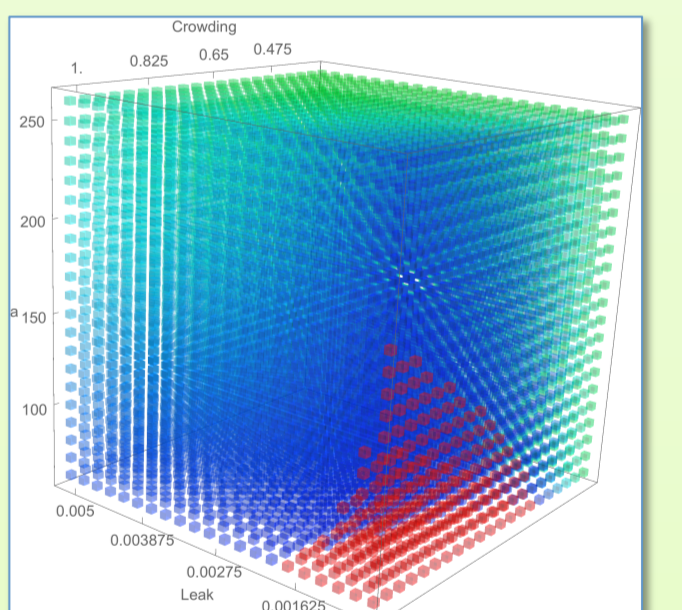


Computational simulation



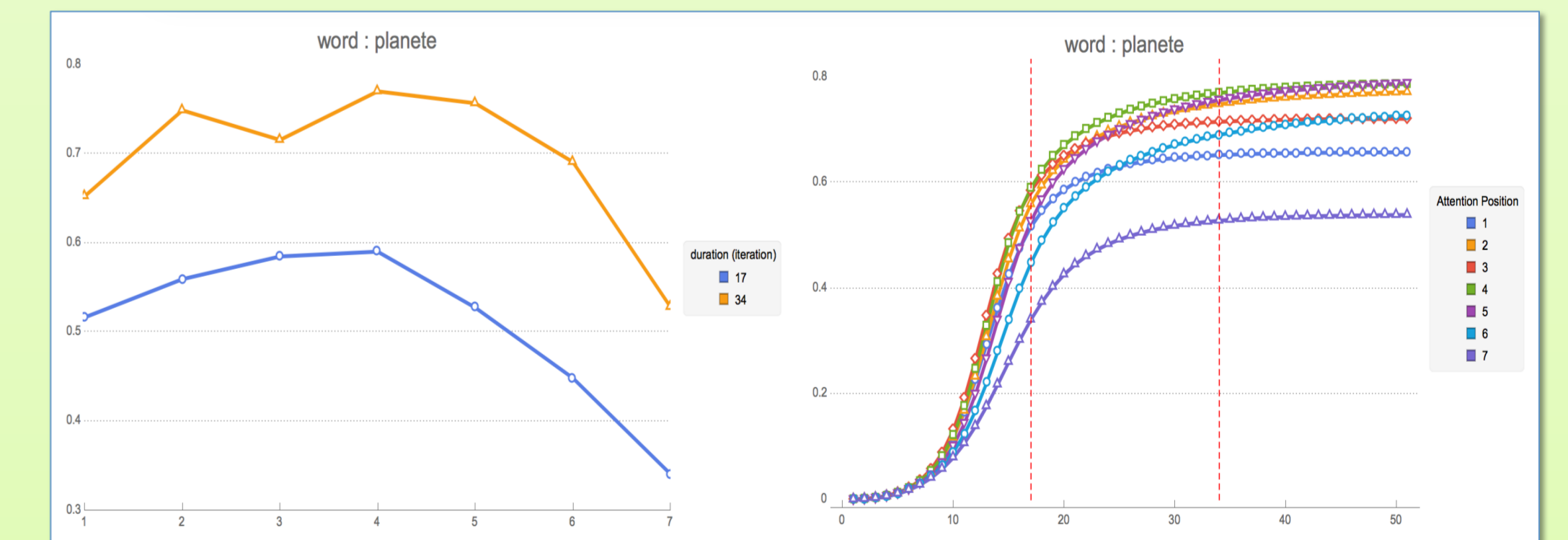
Confusion Matrix (on the left): Letter identification is based on a confusion matrix from Townsend [5]. But this matrix is the result of around 50 ms of exposition. We modified the matrix using a Laplace succession law to simulate the results at shorter duration.

Parameters evaluation (on the right): Each cube represents the level of identification in function of decay (leak), crowding and Laplace succession law parameters. Red corresponds to at least 95% identification.



Preliminary results

- On the left, percentage of recognition as a function of viewing position (OVP curve) for the word "PLANETE".
- On the right, dynamics of recognition for each viewing position. The OVP curve should vary depending on the duration we consider.
- A more systematic study is underway.



Conclusion

BRAID is the first word recognition model designed with structured probabilistic modeling [3]. Letter identification is a temporally dynamic process, building up a percept distribution by accumulation of sensory evidence. It relies on a letter confusion matrix [5] and a temporal decay parameter. The weighted fusion of letter neighbors' distributions allows flexibility in letter position coding and further accounts for crowding. Attention modulates letter processing and enhances letter identification under attentional focus. Acquired orthographic knowledge (lexical database of 36,000 words) serves as a top-down influence during letter identification.

- BRAID** integrates noisy position coding, which allows recognizing a word even if some of its letters are transposed, or ineligible. The same mechanism accounts for **crowding** effects [6] (outer letters are better recognized within strings).
- BRAID** integrates word **frequency** as a prior. In dynamic point of view, frequency can be seen as a resting state of the word distribution: it is the starting point of the dynamic process of recognition, but also the value to which the distribution will go if we remove the input stimulus.
- BRAID** integrates an **Attentional** component computed as a distribution over the percepts that controls for the amount of attention allocated to every single percept. In the dynamic process, attention **prevents decay of information** on letter identity. Typically, the distribution is a Normal. Its mean corresponds to the **viewing position** and its standard deviation **delimits the VA span**.

- We demonstrate the potential of **BRAID** to **simulate the OVP effect**. It is the first word recognition model able to do so.
- Why can **BRAID** simulate OVP effects? Because **BRAID** is the first word recognition model that incorporates **attention as a key component**. Attention allows focusing on different letter positions within the letter string.
- BRAID** provides new insights on OVP effects, by **simulating the temporal dynamics** of word recognition. **BRAID** predicts variations of OVP effects depending on the dynamics of the system, with potentially different effects at different temporal location.

[1] Bosse, Tainturier & Valdois (2007) : Developmental dyslexia: the visual attention span deficit hypothesis. Cognition 104 (2) 198-230
[2] Lobier, Dubois & Valdois (2013) : The role of visual processing speed in reading speed development. PLoS ONE 8 (4)
[3] Bessière, Mazer, Ahuactzin & Mekhnacha (2013) : Bayesian Programming. CRC Press, Boca Raton, Florida

[4] Nazir & Poncet (1998) : Pure Alexia and the Viewing Position Effect in Printed Words. Cognitive Neuropsychology 15 (1/2) 93-140
[5] Townsend (1971) : Alphabetic confusion : A test of models for individuals. Perception & Psychophysics 9 (6) 449-454
[6] Pelli, Tillman, Freeman, Su, Berger & Majaj (2007) : Crowding and eccentricity determine reading rate. Journal of Vision 7 (2) 20-36