Visual World Paradigm

An Eye-Tracking Technique to Study the Real Time Processing of Spoken Language

Likan Zhan

2018-11-07

zhanlikan@blcu.edu.cn

Table of Contents

Introduction

The Visual World Paradigm

(Salverda & Tanenhaus, 2017)

The Visual World Paradigm

- The **visual world paradigm** (VWP) is a family of experimental methods for studying real-time language processing in language comprehension and production that can be used with participants of all ages and most special populations.

(Salverda & Tanenhaus, 2017)

The Visual World Paradigm

- The **visual world paradigm** (VWP) is a family of experimental methods for studying real-time language processing in language comprehension and production that can be used with participants of all ages and most special populations.
- Participants' eye movements to objects in a visual workspace or pictures in a display are monitored as they listen to, or produce, spoken language that is about the contents of the visual world.

(Salverda & Tanenhaus, 2017)

The Visual World Paradigm

- The **visual world paradigm** (VWP) is a family of experimental methods for studying real-time language processing in language comprehension and production that can be used with participants of all ages and most special populations.
- Participants' eye movements to objects in a visual workspace or pictures in a display are monitored as they listen to, or produce, spoken language that is about the contents of the visual world.
- Eye-movements in the VWP provide a sensitive, time-locked response measure that can be used to investigate a wide range of psycholinguistic questions on topics running the gamut from speech perception to interactive conversation in collaborative task-oriented dialogue.

(Salverda & Tanenhaus, 2017)

The Linking Hypothesis

(Salverda & Tanenhaus, 2017)

The Linking Hypothesis

- As visual attention shifts to an object in the workspace, as a consequence of planning or comprehending an utterance, there is a high probability that a saccadic eye movement will rapidly follow to bring the attended area into foveal vision.

(Salverda & Tanenhaus, 2017)

The Linking Hypothesis

- As visual attention shifts to an object in the workspace, as a consequence of planning or comprehending an utterance, there is a high probability that a saccadic eye movement will rapidly follow to bring the attended area into foveal vision.
- Where a participant is looking, and in particular when and to where saccadic eye movements are launched in relationship to the speech, can provide insights into real-time language processing.

(Salverda & Tanenhaus, 2017)

A Brief History

COGNITIVE PSYCHOLOGY 6, 84–107 (1974)

The Control of Eye Fixation by the Meaning of Spoken Language

A New Methodology for the Real-Time Investigation of Speech Perception, Memory, and Language Processing

ROGER M. COOPER[1,2]
Stanford University

A Brief History

antiserum with
the primary an-
ed for 30 min at
ated in the sec-
hours at 22°C.
identical to the
cyanine (Cy3)–
oglobulin G (no.
rch Labs, West
ation of 1:600)
the tissue sec-
(pH 7.4, 22°C),
d coverslipped
-phenylenedia-

within the cell,
-600 Confocal
A) and with an
for epifluores-

Matus, S. P. Hunt, *Eur. J. Neurosci.* **3**, 551 (1991).
19. P. W. Mantyh, unpublished observations.

30 September 1994; accepted 2 March 1995

Integration of Visual and Linguistic Information in Spoken Language Comprehension

Michael K. Tanenhaus,* Michael J. Spivey-Knowlton,
Kathleen M. Eberhard, Julie C. Sedivy

Psycholinguists have commonly assumed that as a spoken linguistic message unfolds over time, it is initially structured by a syntactic processing module that is encapsulated from information provided by other perceptual and cognitive systems. To test the effects of relevant visual context on the rapid mental processes that accompany spoken language

A Brief History

JOURNAL OF MEMORY AND LANGUAGE **38,** 419–439 (1998)
ARTICLE NO. ML972558

Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models

Paul D. Allopenna, James S. Magnuson, and Michael K. Tanenhaus

University of Rochester

A Brief History

ELSEVIER

Cognition 73 (1999) 89–134

COGNITION

www.elsevier.com/locate/cognit

The kindergarten-path effect: studying on-line sentence processing in young children

John C. Trueswell*, Irina Sekerina, Nicole M. Hill, Marian L. Logrip

University of Pennsylvania, Philadelphia, PA, USA

Received 18 August 1998; received in revised form 29 January 1999; accepted 1 May 1999

A Brief History

ELSEVIER

Cognition 66 (1998) B25–B33

COGNITION

Brief article

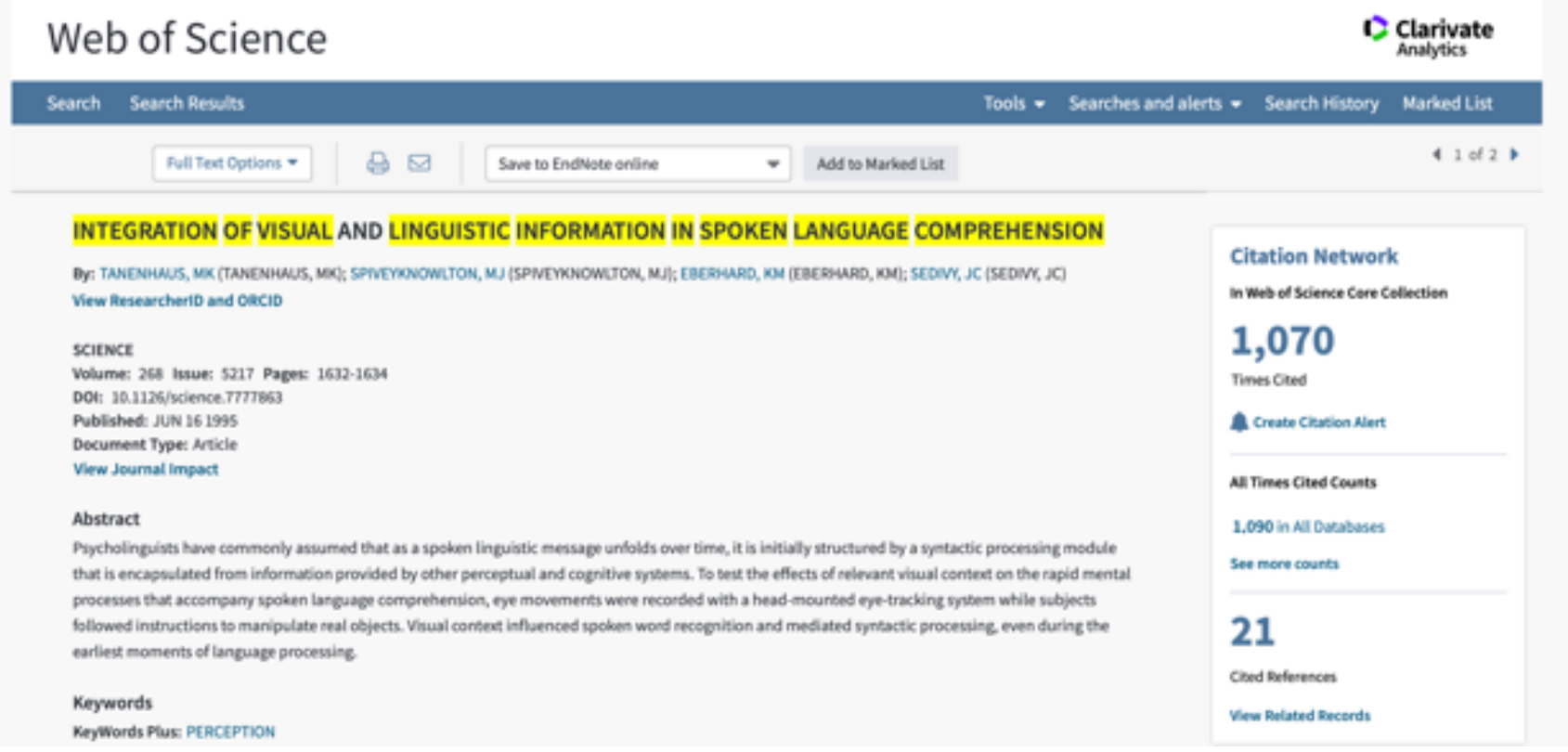
Viewing and naming objects: eye movements during noun phrase production

Antje S. Meyer*, Astrid M. Sleiderink, Willem J.M. Levelt

Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands

Received 25 September 1997; accepted 5 March 1998

A Brief History



Common Variations

Apparatus

(Zhan, 2018b)

Apparatus

- The simplest, least expensive, and most portable system is just a normal video camera, which records an image of the participant's eyes.

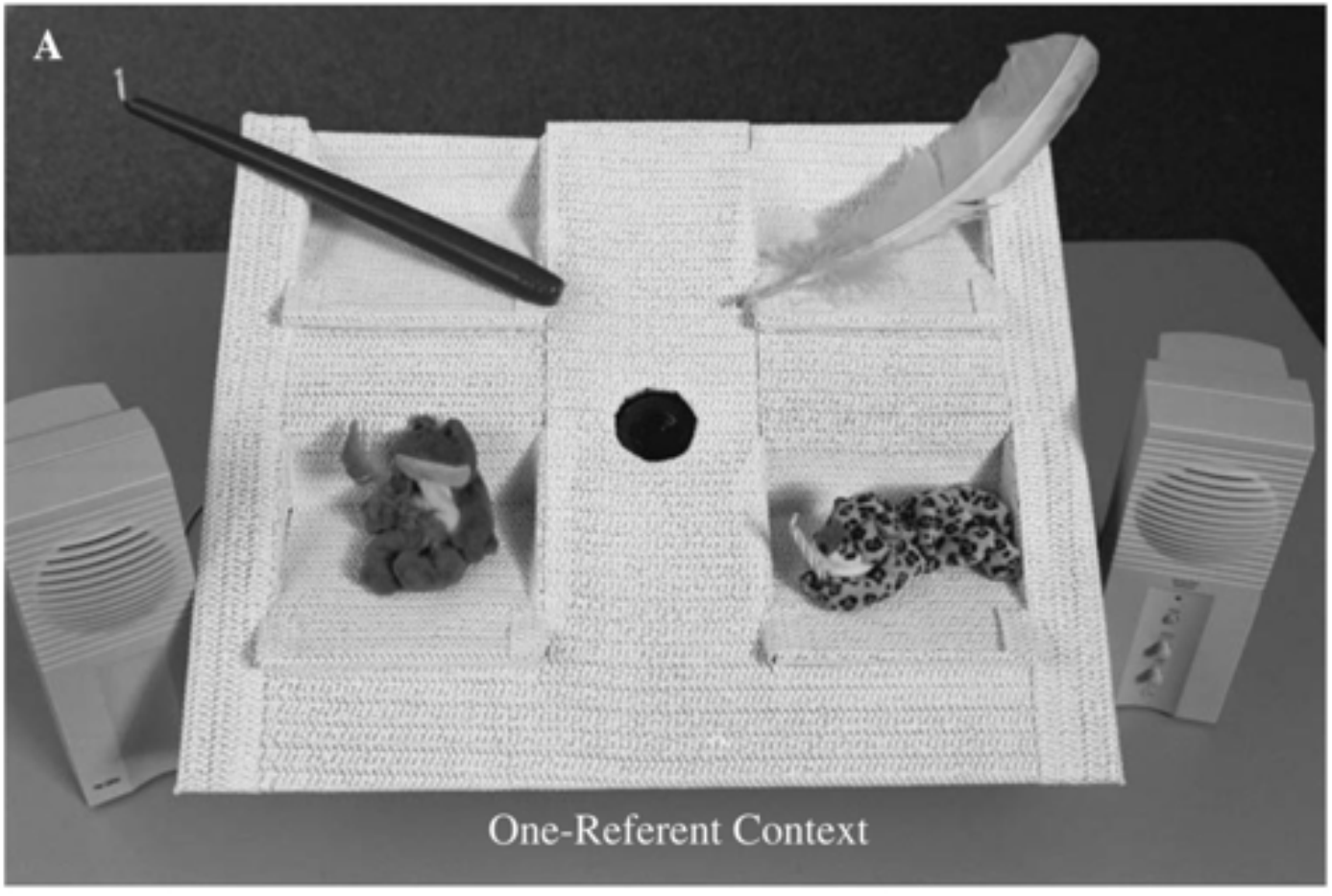
(Zhan, 2018b)

Apparatus

- The simplest, least expensive, and most portable system is just a normal video camera, which records an image of the participant's eyes.
- A contemporary commercial eye tracking system normally uses optical sensors measuring the orientation of the eye in its orbit.

(Zhan, 2018b)

Apparatus



(Snedeker & Trueswell, 2004)

Apparatus


TM4
(EyeTech DS)
VT2
(EyeTech DS)
VT2 Mini
(EyeTech DS)
EyeFollower
(LC Technologies)
EyeGaze 600
(LC Technologies)
S2
(Mirametrix)
RED 250/500
(Sensomotoric Instruments)
RED-m
EyeLink 1000
(SR Research)
EyeLink II
(SR Research)
faceLAB 5
(Seeing Machines)
T60/T120
(Tobii Technology)
TX300
(Tobii Technology)
X1
(Tobii Technology)
X2-30/X2-60
(Tobii Technology)
X60/120
(Tobii Technology)
SMI Eye Tracking
Glasses (Sensomotoric
Instruments)
ViewPoint EyeFrame Scene
Camera (Arringto Research)

Visual World

(Zhan, 2018b)

Visual World

- A visual display is normally a screening display depicting an array of pictures.

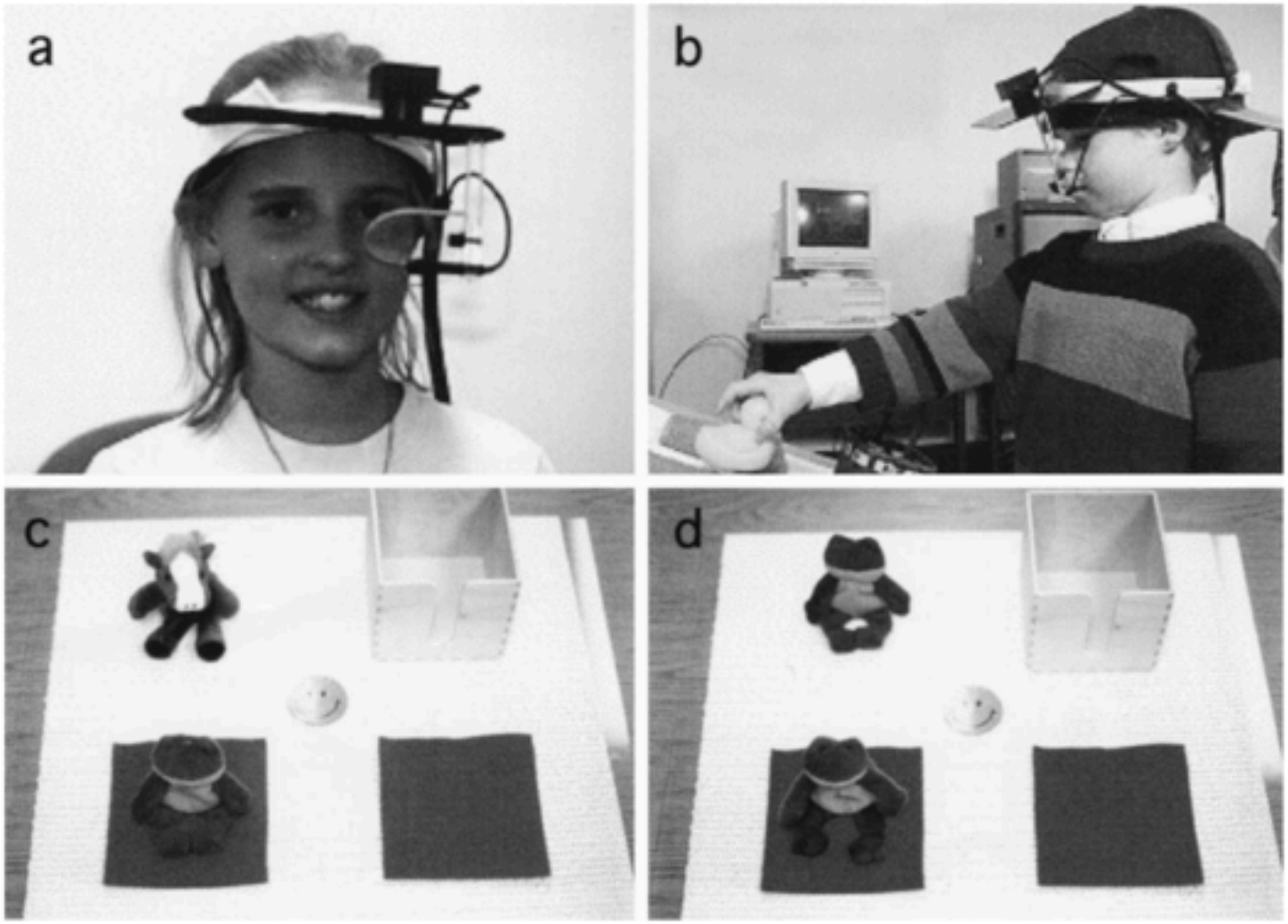
(Zhan, 2018b)

Visual World

- A visual display is normally a screening display depicting an array of pictures.
- It can also be a screening display depicting an array of printed words, a schematic scene, or a real world scene containing real objects.

(Zhan, 2018b)

Visual World



(Trueswell, Sekerina, Hill, & Logrip, 1999)

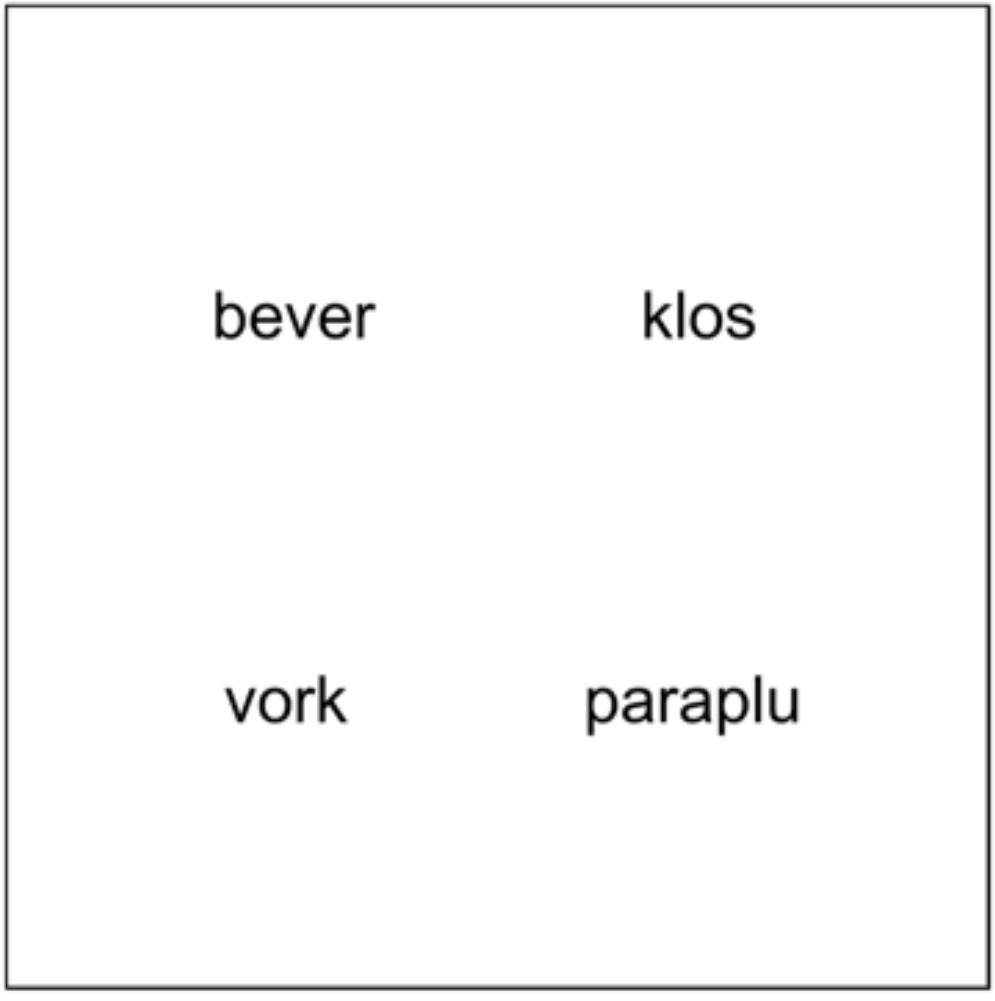
Visual World

(Huettig & McQueen, 2007)

Visual World

(Zhan, Zhou, & Crain, 2018)

Visual World

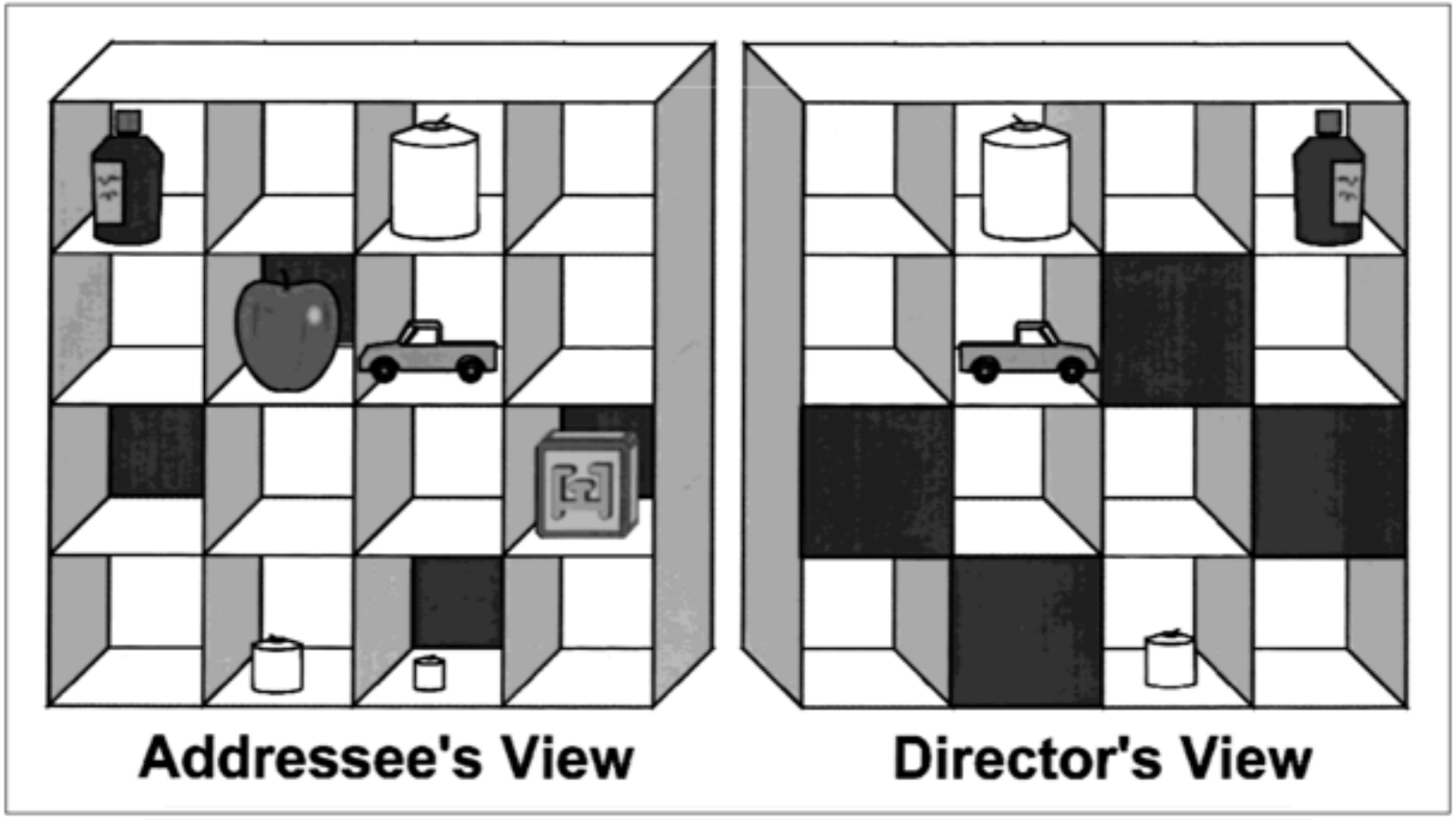


(Huettig & McQueen, 2007)

Visual World

(Altmann & Kamide, 2007)

Visual World



(Keysar, Barr, Balin, & Brauner, 2000)

Spoken Language

(Salverda & Tanenhaus, 2017)

Spoken Language

- The language can differ along any number of dimensions, from manipulations of fine-grained acoustic phonetic features (duration, VOT, formant structure, fundamental frequency, etc.) to properties of words (syntactic category, semantic features, frequency of occurrence, etc.) to linguistic structure (syntactic structure, information structure, semantic and pragmatic properties such as implicating and questioning, etc.).

(Salverda & Tanenhaus, 2017)

Spoken Language

- The language can differ along any number of dimensions, from manipulations of fine-grained acoustic phonetic features (duration, VOT, formant structure, fundamental frequency, etc.) to properties of words (syntactic category, semantic features, frequency of occurrence, etc.) to linguistic structure (syntactic structure, information structure, semantic and pragmatic properties such as implicating and questioning, etc.).
- The language often comes from a disembodied voice, which provides a narrative (e.g., *The doctor will hand the scalpel to the nurse*) or an instruction (e.g., *Put the large candle above the fork*).

(Salverda & Tanenhaus, 2017)

Spoken Language - Our Researches

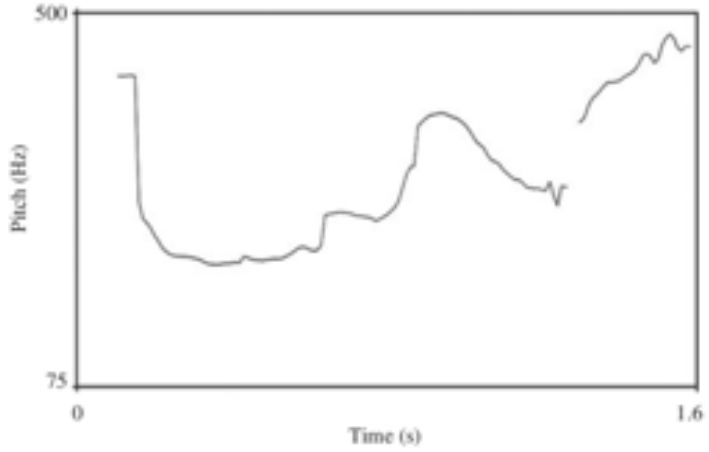
Spoken Language - Our Researches

- The spoken language can differ in their verbs (Zhou et al., 2018), their phonological stresses (Zhou, Su, et al., 2012), their sentential prosodies (Zhou, Crain, & Zhan, 2012), their aspect markers (Zhou et al., 2014), and their epistemic modals (Moscati et al., 2017).

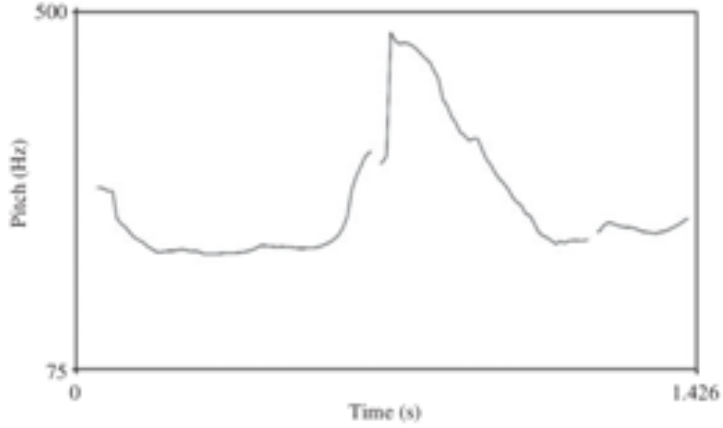
Spoken Language - Our Researches

- The spoken language can differ in their verbs (Zhou et al., 2018), their phonological stresses (Zhou, Su, et al., 2012), their sentential prosodies (Zhou, Crain, & Zhan, 2012), their aspect markers (Zhou et al., 2014), and their epistemic modals (Moscati et al., 2017).
- The spoken language can also be semantically complex statements that differ in their logical structures, such as concessives and biconditionals (Zhan et al., 2015), conditionals (Zhan et al., 2018), and disjunctions (Zhan, 2018a).

Spoken Language - Our Researches



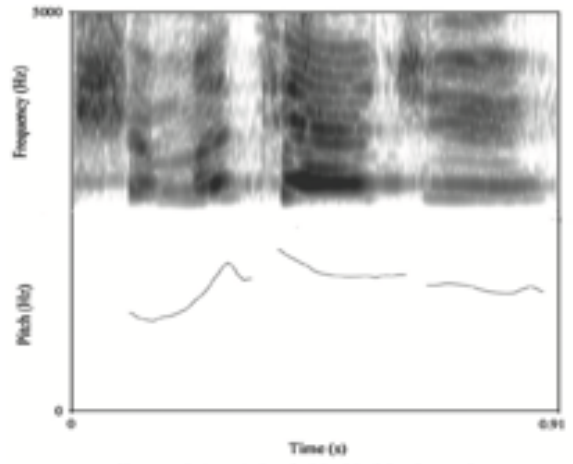
Xiaoming de NAOZHONG 'Xiaoming's CLOCK'



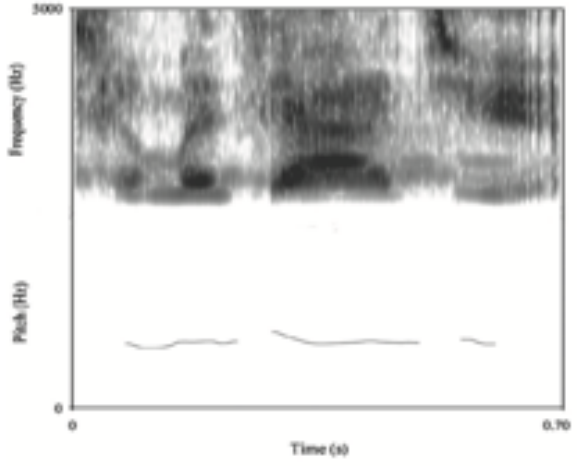
XIAOMING de naozhong 'XIAOMING's clock'

(Zhou, Su, et al., 2012)

Spoken Language - Our Researches



Shenme shuiguo 'what fruit' with rising intonation



Shenme shuiguo 'what fruit' with level intonation

(Zhou, Crain, & Zhan, 2012)

Spoken Language - Our Researches

(7) a. Laonainai zhong-le yi-duo xiaohua.
old lady plant-PERF one-CL flower
'The old lady has planted a flower.'
b. Laonainai zhong-zhe yi-duo xiaohua.
old lady plant-DUR one-CL flower
'The old lady is planting a flower.'

(Zhou et al., 2014)

Spoken Language - Our Researches

(3) a monkey *might* be in the orange box
(4) a monkey *must* be in the orange box

(Moscati et al., 2017)

Spoken Language - Our Researches

a).**And**

小明的	箱子里	有	一只	奶牛	和	一只	公鸡
Xiaoming de	xiang zi li	you	yi zhi	nai niu	he	yi zhi	gong ji
Xiaoming's	box in	have	one-CL	cow	and	one-CL	rooster

Xiaoming's box contains a cow and a rooster.

b).**But**

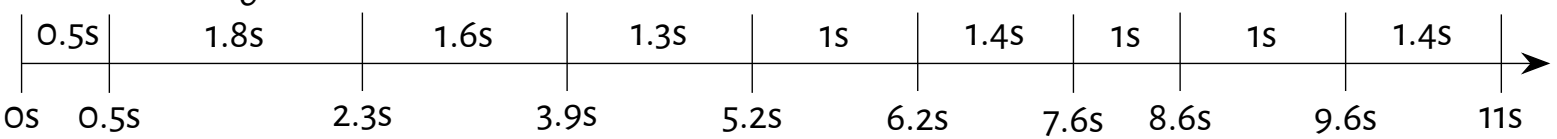
小明的	箱子里	有	一只	奶牛	但	没有	公鸡
Xiaoming de	xiangzi li	you	yi zhi	nai niu	dan	meiyou	gong ji
Xiaoming's	box in	have	one-CL	cow	but	not	rooster

Xiaoming's box contains a cow but not a rooster.

c).**Or**

小明的	箱子里	有	一只	奶牛	或	一只	公鸡
Xiaoming de	xiang zi li	you	yi zhi	nainiu	huo	youzhi	gongji
Xiaoming's	box in	have	one-CL	cow	or	one-CL	rooster

Xiaoming's box contains a cow or a rooster.



(Zhan, 2018b)

Behavioral Task

Behavioral Task

- In *Task* or *action based studies*, participants interact with real-world objects or, more typically, interact with pictures in a screen based workspace to perform a motor task, typically clicking and dragging pictures to follow explicit instructions (*Put the clown above the star*), clicking on a picture when its name is mentioned, or manipulating real objects (e.g., *Pick up the apple. Now put it in the box*).

Behavioral Task

- In *Task* or *action based studies*, participants interact with real-world objects or, more typically, interact with pictures in a screen based workspace to perform a motor task, typically clicking and dragging pictures to follow explicit instructions (*Put the clown above the star*), clicking on a picture when its name is mentioned, or manipulating real objects (e.g., *Pick up the apple. Now put it in the box*).
- *Look and listen studies* (Altmann & Kamide, 1999, 2007) do not require participants to perform an explicit task other than to look at the computer screen.

Behavioral Task

- In *Task* or *action based studies*, participants interact with real-world objects or, more typically, interact with pictures in a screen based workspace to perform a motor task, typically clicking and dragging pictures to follow explicit instructions (*Put the clown above the star*), clicking on a picture when its name is mentioned, or manipulating real objects (e.g., *Pick up the apple. Now put it in the box*).
- *Look and listen studies* (Altmann & Kamide, 1999, 2007) do not require participants to perform an explicit task other than to look at the computer screen.
- Participants are asked to determine whether or not the auditory utterance applies to the visual display (Zhan et al., 2018), or to choose the correct image in the visual display the spoken utterance is talking about (Zhan, 2018a).

Participants

(Zhan, 2018b)

Participants

- The visual world paradigm can be used in a wide of populations, including those who cannot read and/or who cannot overtly give their behavioral responses,

(Zhan, 2018b)

Participants

- The visual world paradigm can be used in a wide of populations, including those who cannot read and/or who cannot overtly give their behavioral responses,
- The eligible participants include preliterate children, elderly adults, and patients, such as who with aphasics or with ASD.

(Zhan, 2018b)

General Considerations

Speech and Spoken Language

(Salverda & Tanenhaus, 2017)

Speech and Spoken Language

- Speech is a temporal, rapidly changing signal. Acoustic cues are transient, and there are no acoustic signatures that correspond to linguistic categories.

(Salverda & Tanenhaus, 2017)

Speech and Spoken Language

- Speech is a temporal, rapidly changing signal. Acoustic cues are transient, and there are no acoustic signatures that correspond to linguistic categories.
- Relevant cues to a category, or even a phonetic feature such as voicing, are determined by multiple cues, many of which arrive asynchronously and are impacted by both high and low level linguistic subsystems.

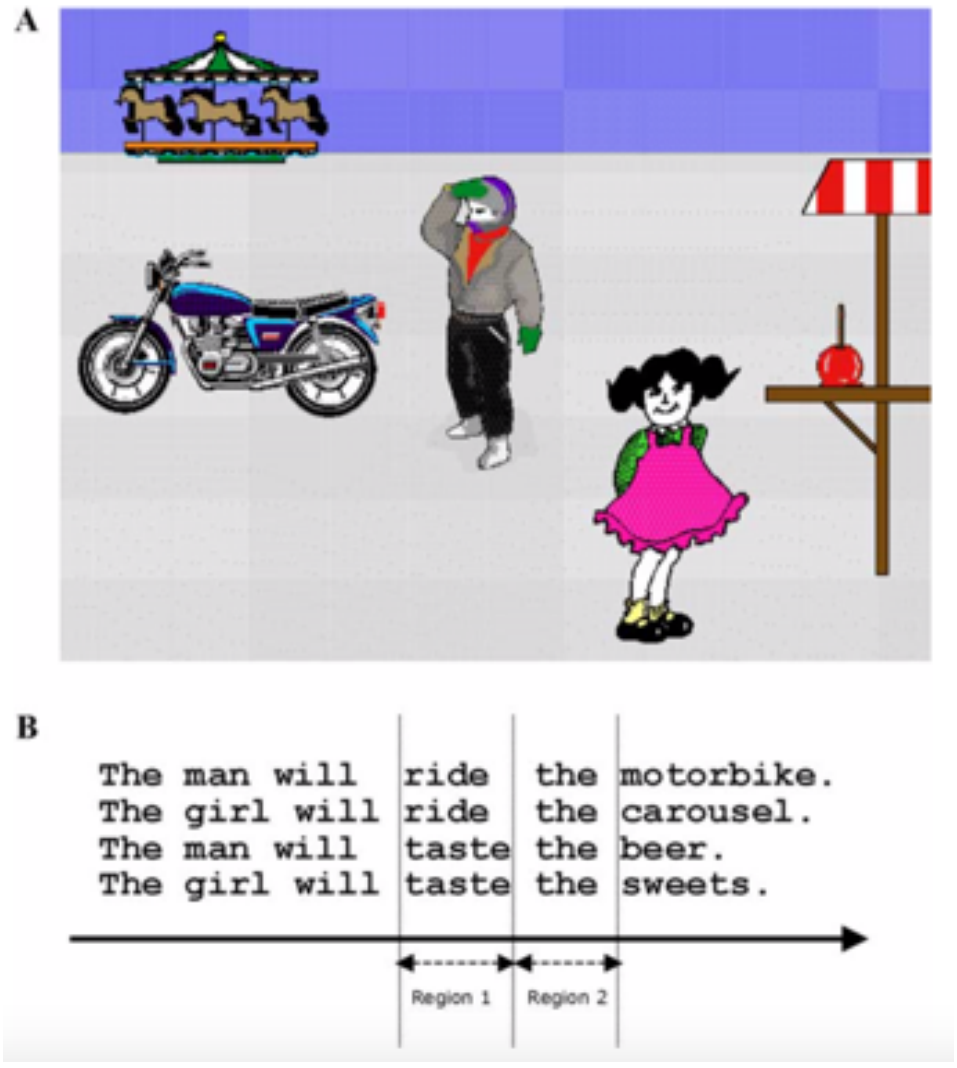
(Salverda & Tanenhaus, 2017)

Speech and Spoken Language

- Speech is a temporal, rapidly changing signal. Acoustic cues are transient, and there are no acoustic signatures that correspond to linguistic categories.
- Relevant cues to a category, or even a phonetic feature such as voicing, are determined by multiple cues, many of which arrive asynchronously and are impacted by both high and low level linguistic subsystems.
- Linking eye movements to relevant linguistic information in the speech signal is therefore critically dependent on having some understanding of where, when, and why information in the speech signal provides information about linguistic structure.

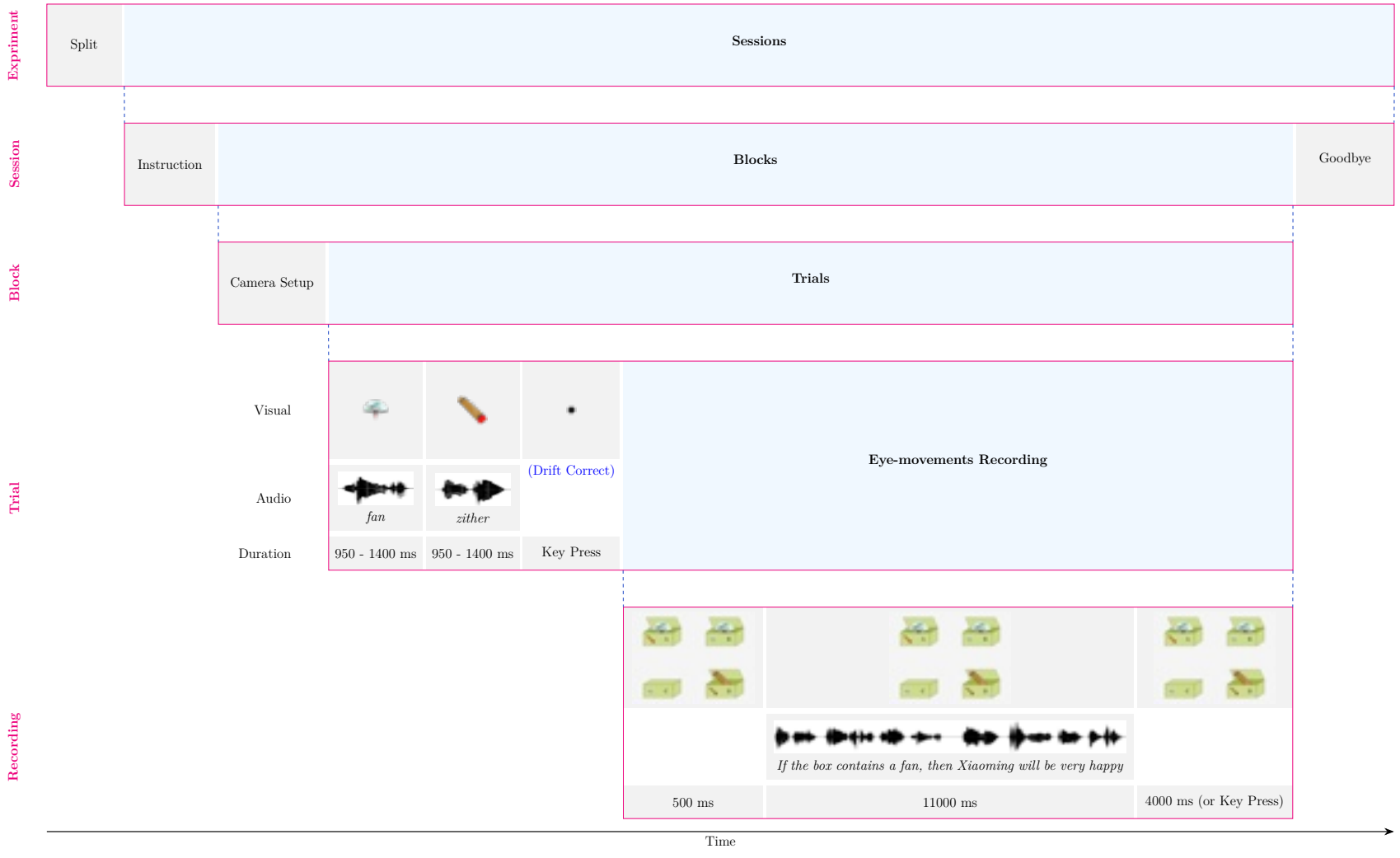
(Salverda & Tanenhaus, 2017)

Eye Movements in Natural Tasks



(Kamide, Scheepers, & Altmann, 2003)

Eye Movements in Natural Tasks



(Zhan, 2018b)

Disadvantages, Limitations, and Concerns

(Zhan, 2018b)

Disadvantages, Limitations, and Concerns

- Participants' interpretation of the spoken language is deduced from their eye movements on the visual world, not from the actual interpretation of the language stimuli per se.

(Zhan, 2018b)

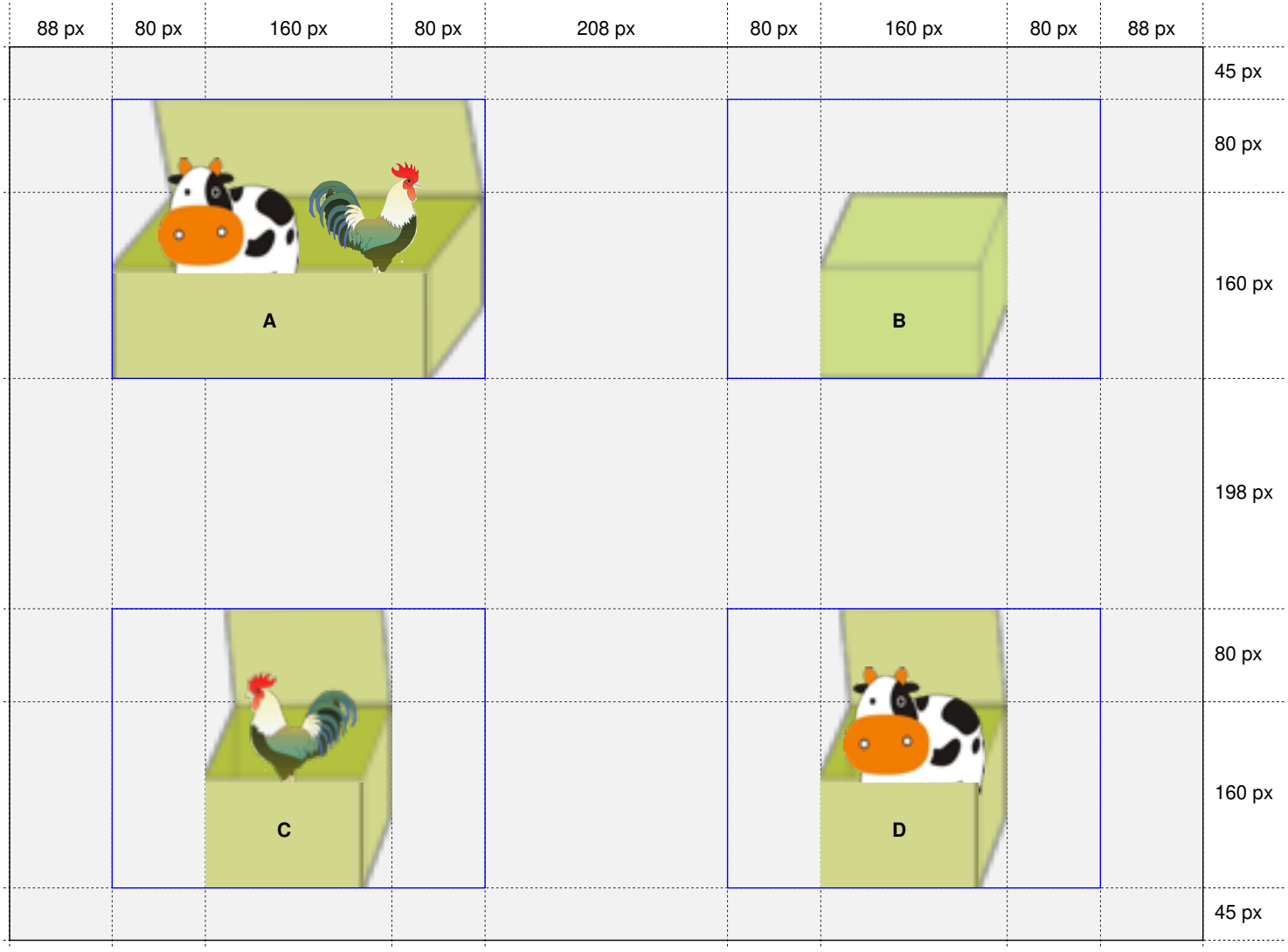
Disadvantages, Limitations, and Concerns

- Participants' interpretation of the spoken language is deduced from their eye movements on the visual world, not from the actual interpretation of the language stimuli per se.
- The visual world paradigm used is normally more restricted than the actual visual world, with a limited set of pictured referents and a limited set of potential actions.

(Zhan, 2018b)

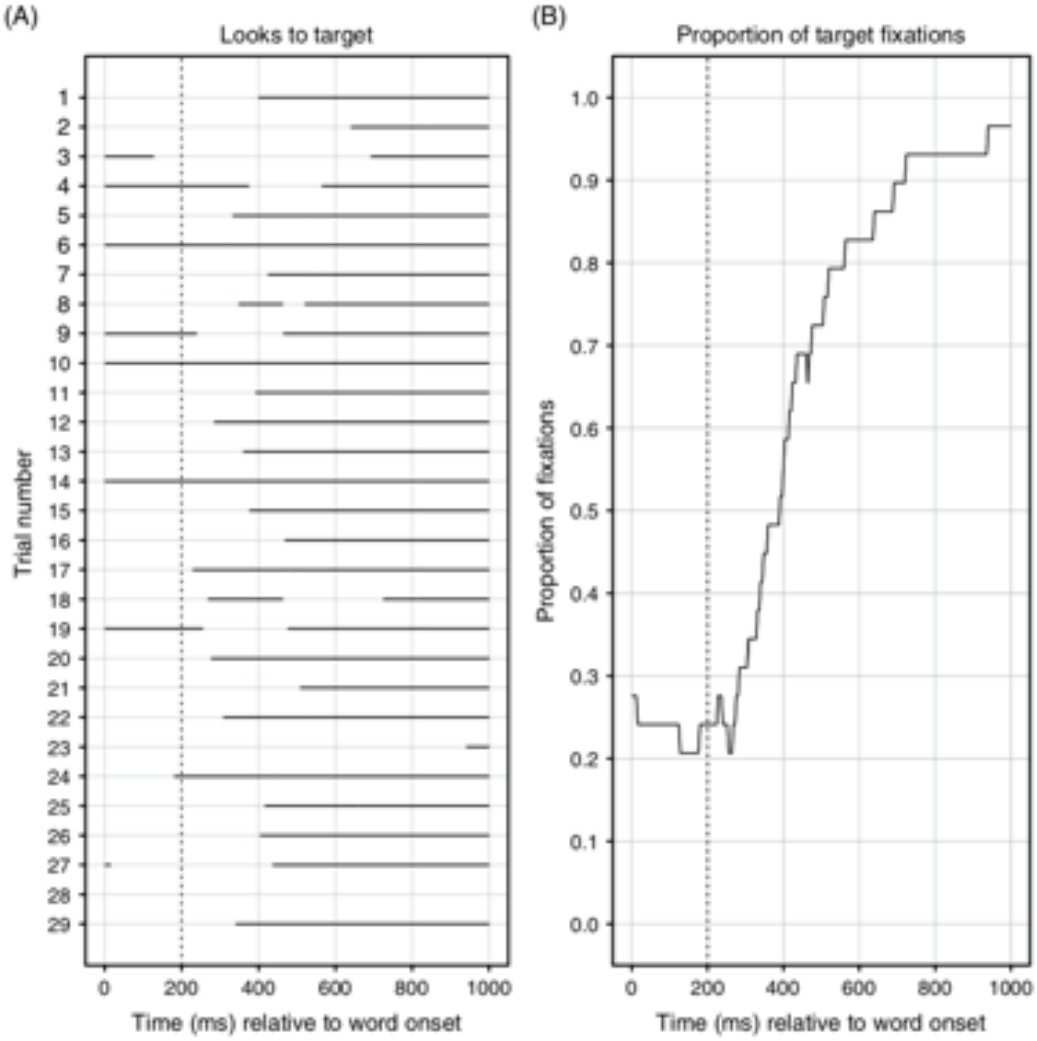
Data Analysis

Regions Of Interest



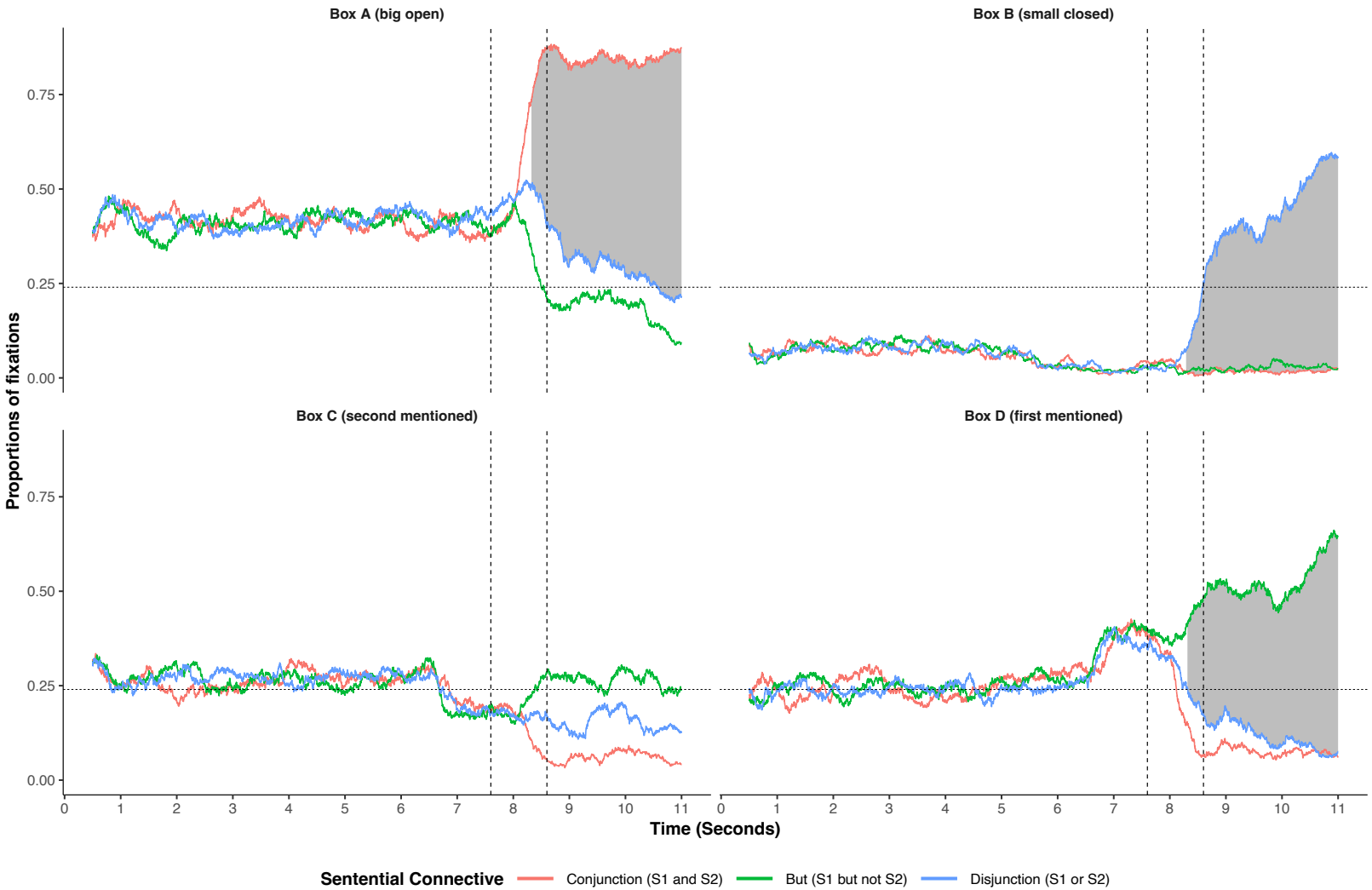
(Zhan, 2018a, 2018b)

Proportion of Fixations



(Salverda & Tanenhaus, 2017)

Data Visualization



(Zhan, 2018a, 2018b)

Questions Could Be Answered

(Zhan, 2018b)

Questions Could Be Answered

- On the coarse-grain level, are participants' eye movements in the visual world affected by different auditory linguistic input?

(Zhan, 2018b)

Questions Could Be Answered

- On the coarse-grain level, are participants' eye movements in the visual world affected by different auditory linguistic input?
- If there is an effect, what is the trajectory of the effect over the course of the trial? Is it a linear effect or high-order effect? and

(Zhan, 2018b)

Questions Could Be Answered

- On the coarse-grain level, are participants' eye movements in the visual world affected by different auditory linguistic input?
- If there is an effect, what is the trajectory of the effect over the course of the trial? Is it a linear effect or high-order effect? and
- If there is an effect, then on the fine-grain level, when is the earliest temporal point where such an effect emerges and how long does this effect last?

(Zhan, 2018b)

Statistical Analyses

(Zhan, 2018b)

Statistical Analyses

- The response variable, i.e., proportions of fixations, is both below and above bounded (between 0 and 1), which will follow a multinomial distribution rather than a normal distribution.

(Zhan, 2018b)

Statistical Analyses

- The response variable, i.e., proportions of fixations, is both below and above bounded (between 0 and 1), which will follow a multinomial distribution rather than a normal distribution.
- To explore the changing trajectory of the observed effect, a variable denoting the time-series has to be added into the model.

(Zhan, 2018b)

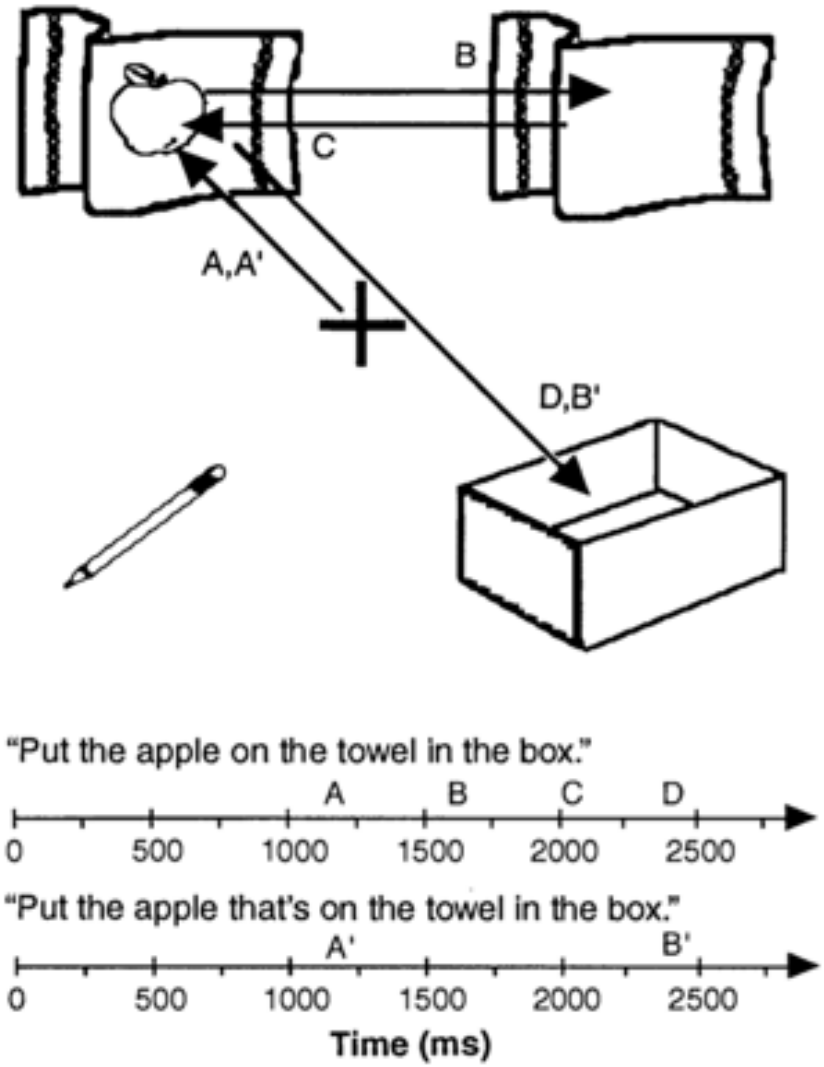
Statistical Analyses

- The response variable, i.e., proportions of fixations, is both below and above bounded (between 0 and 1), which will follow a multinomial distribution rather than a normal distribution.
- To explore the changing trajectory of the observed effect, a variable denoting the time-series has to be added into the model.
- When a statistical analysis is repeatedly applied to each time bin of the periods of interest, the familywise error induced from these multiple comparisons should be tackled.

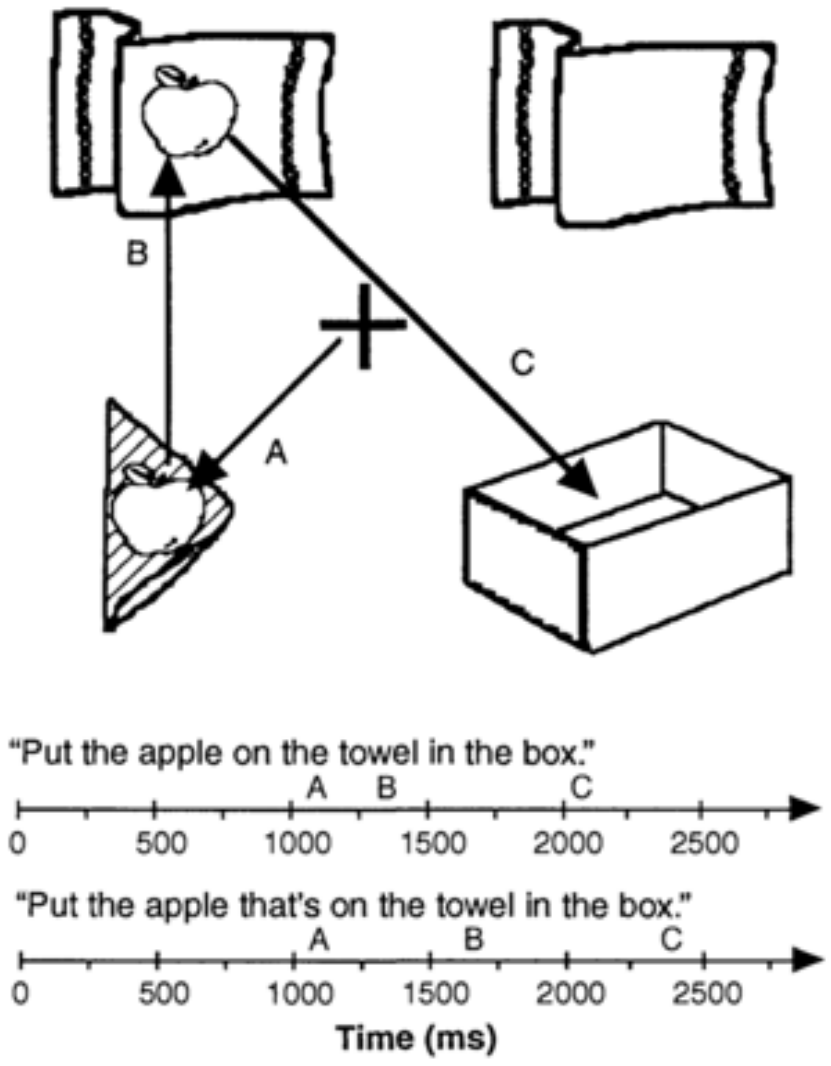
(Zhan, 2018b)

Example Studies

Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995)



Tanenhaus et al. (1995)



Tanenhaus et al. (1995)

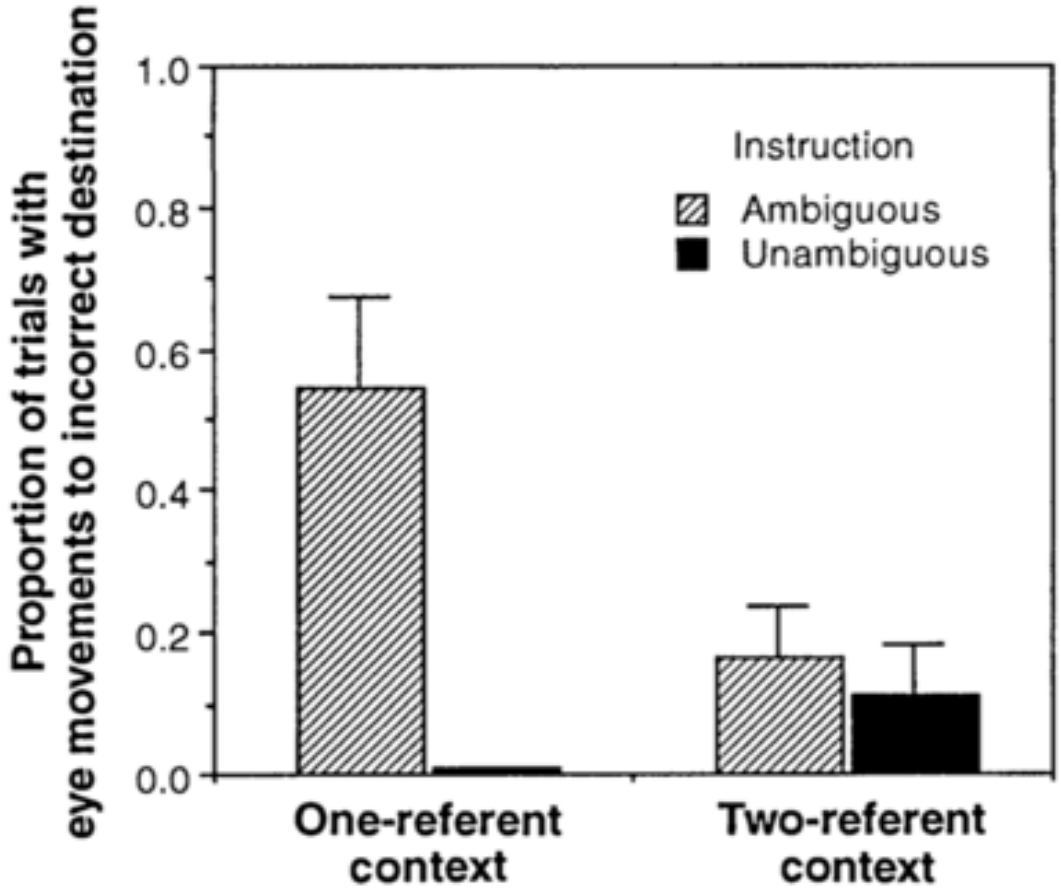
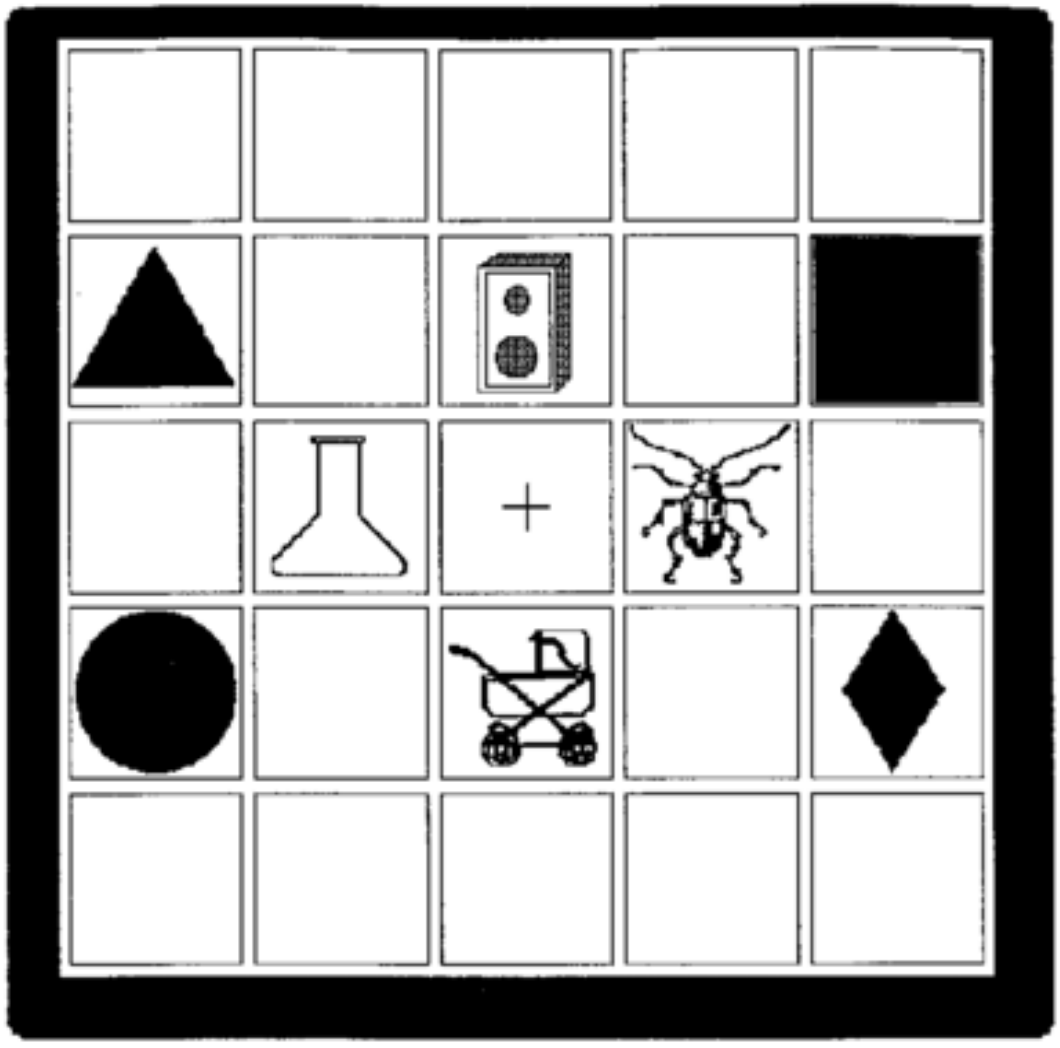
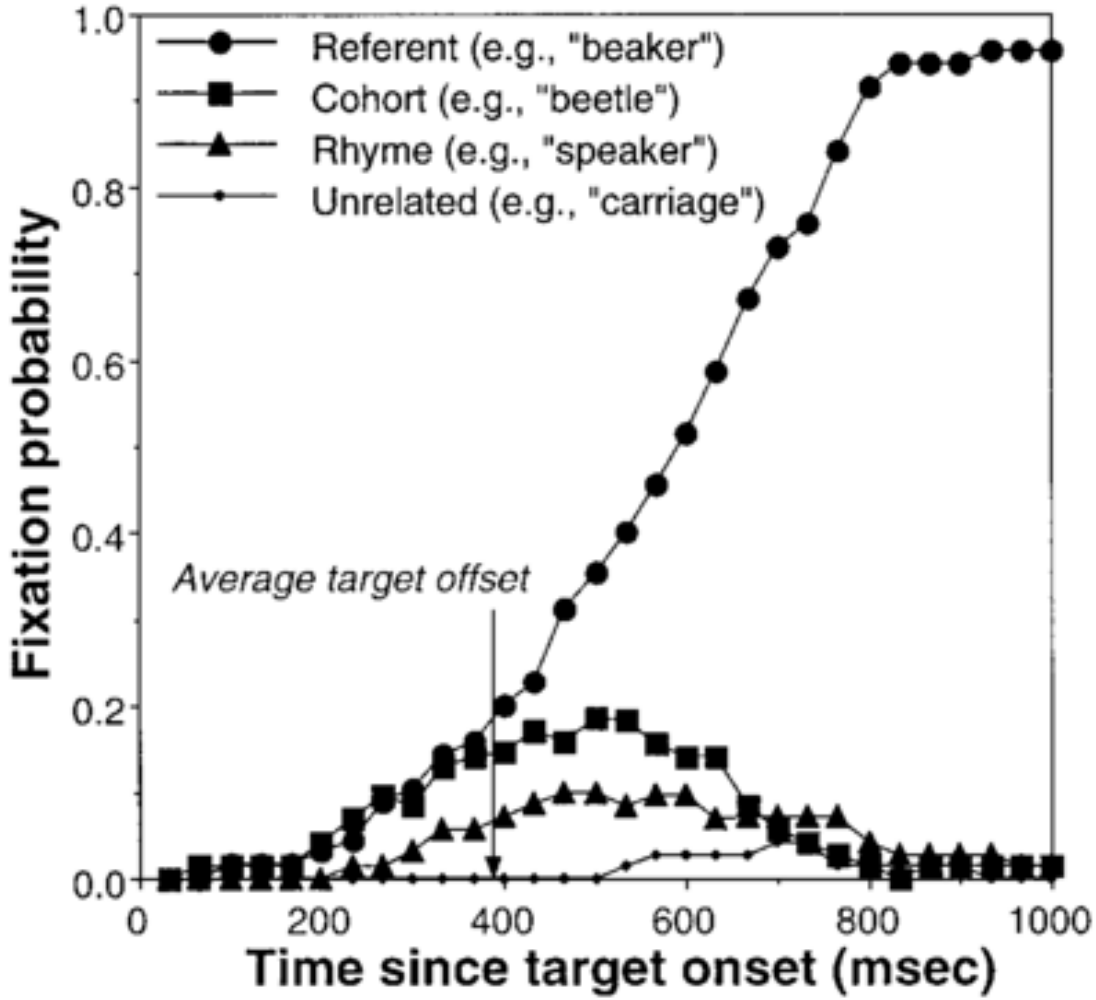


Fig. 3. Proportion of trials in which participants looked at the incorrect destination.

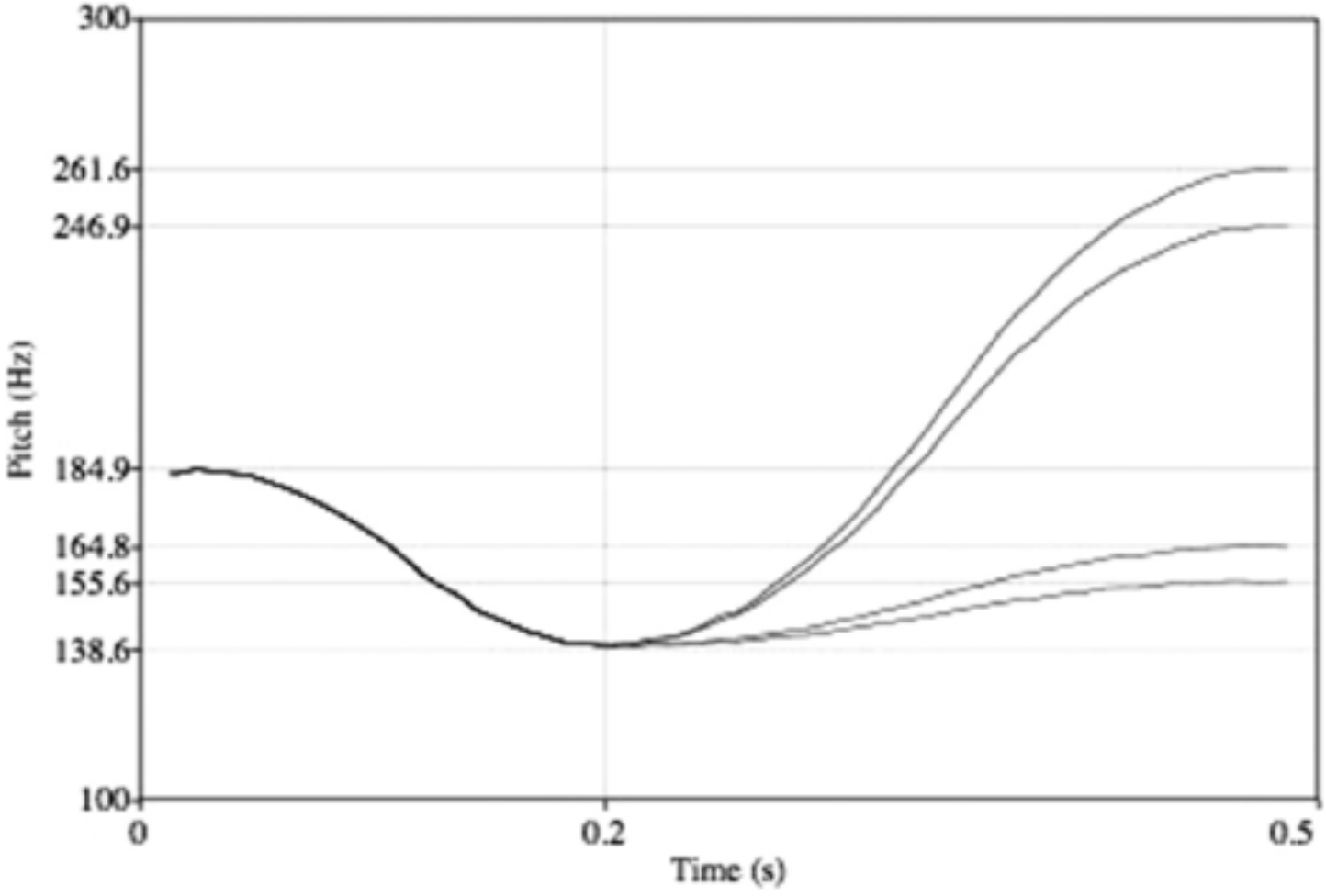
Allopenna, Magnuson, and Tanenhaus (1998)



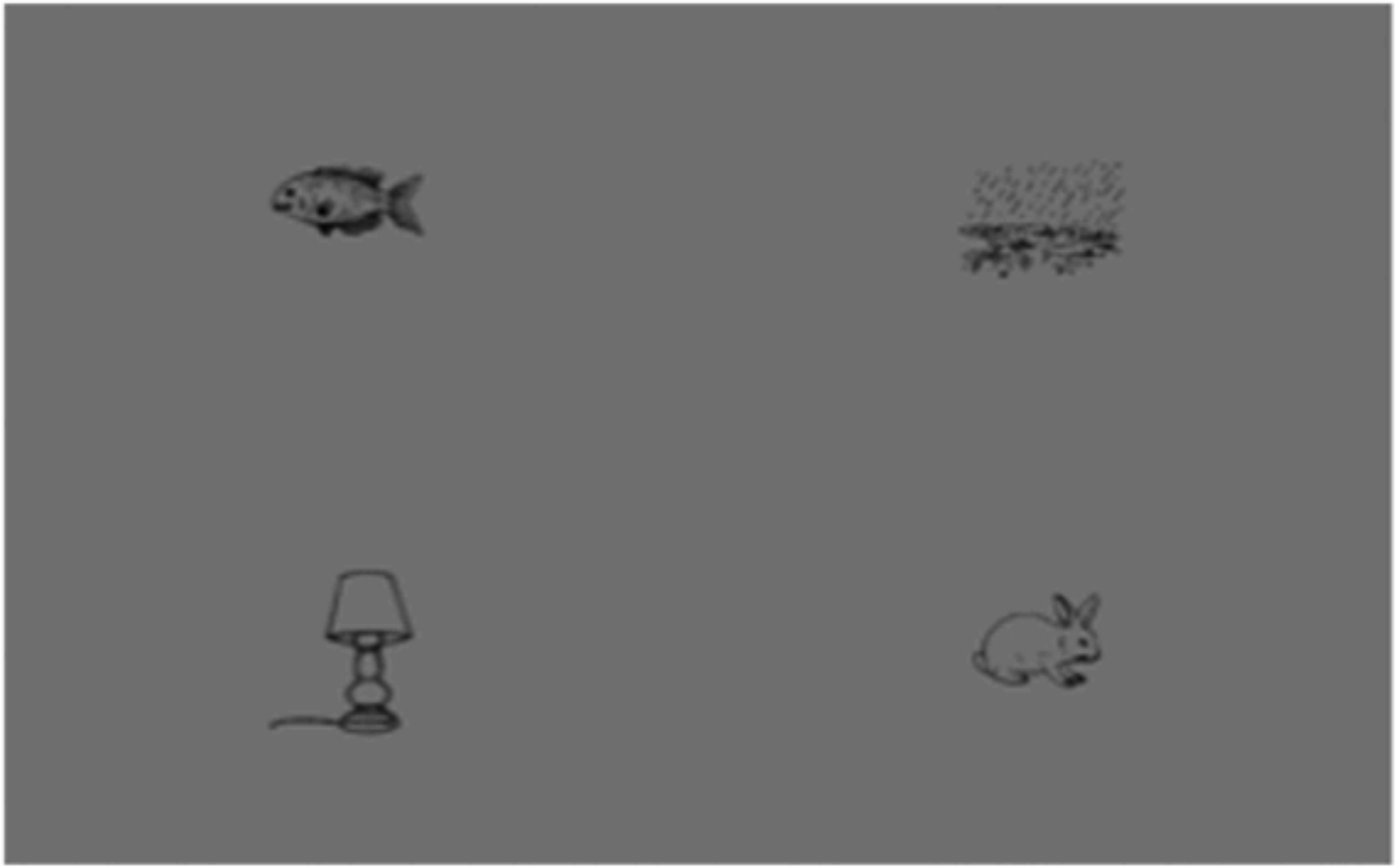
Allopenna et al. (1998)



Shen, Deutsch, and Rayner (2013)

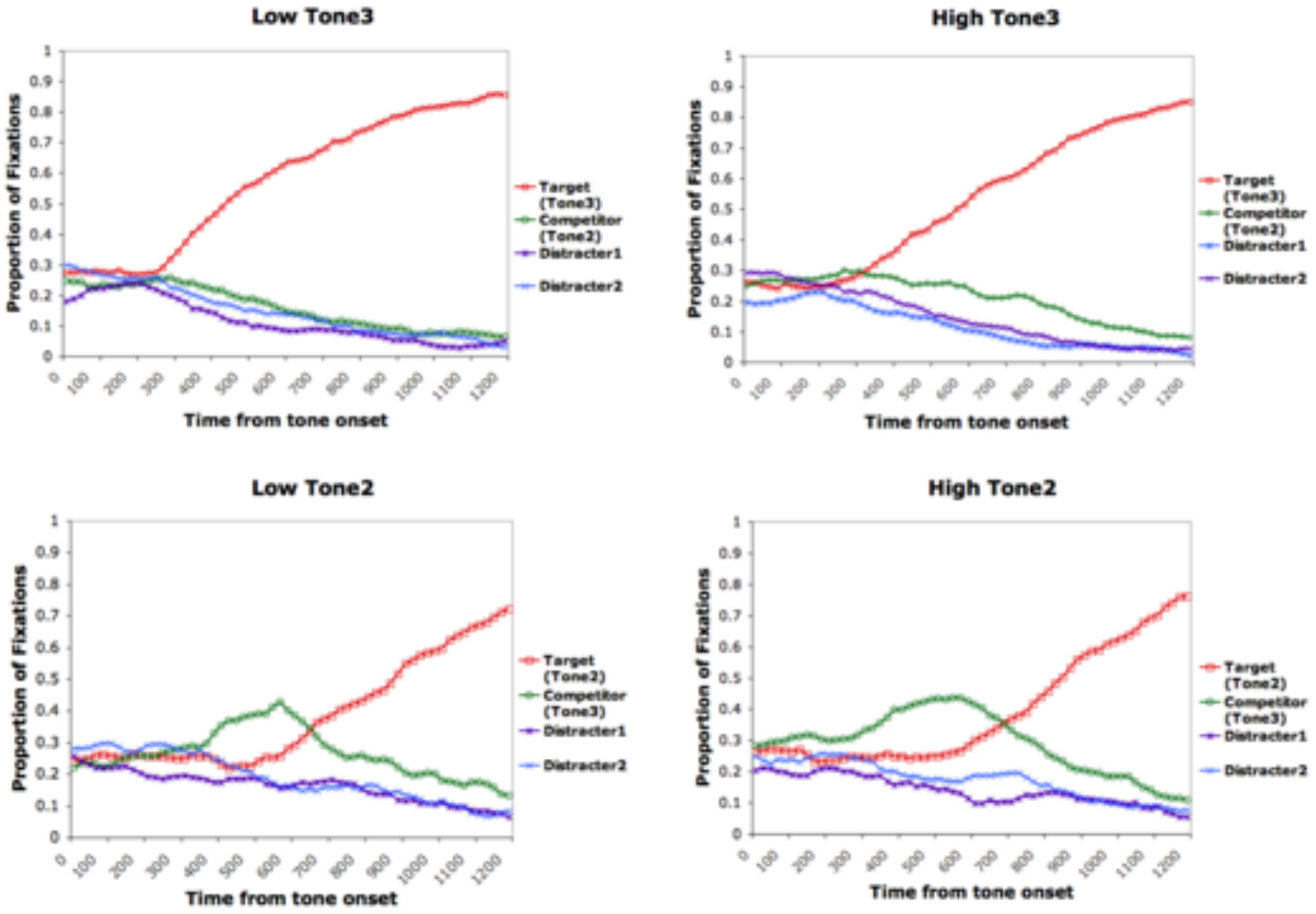


Shen et al. (2013)

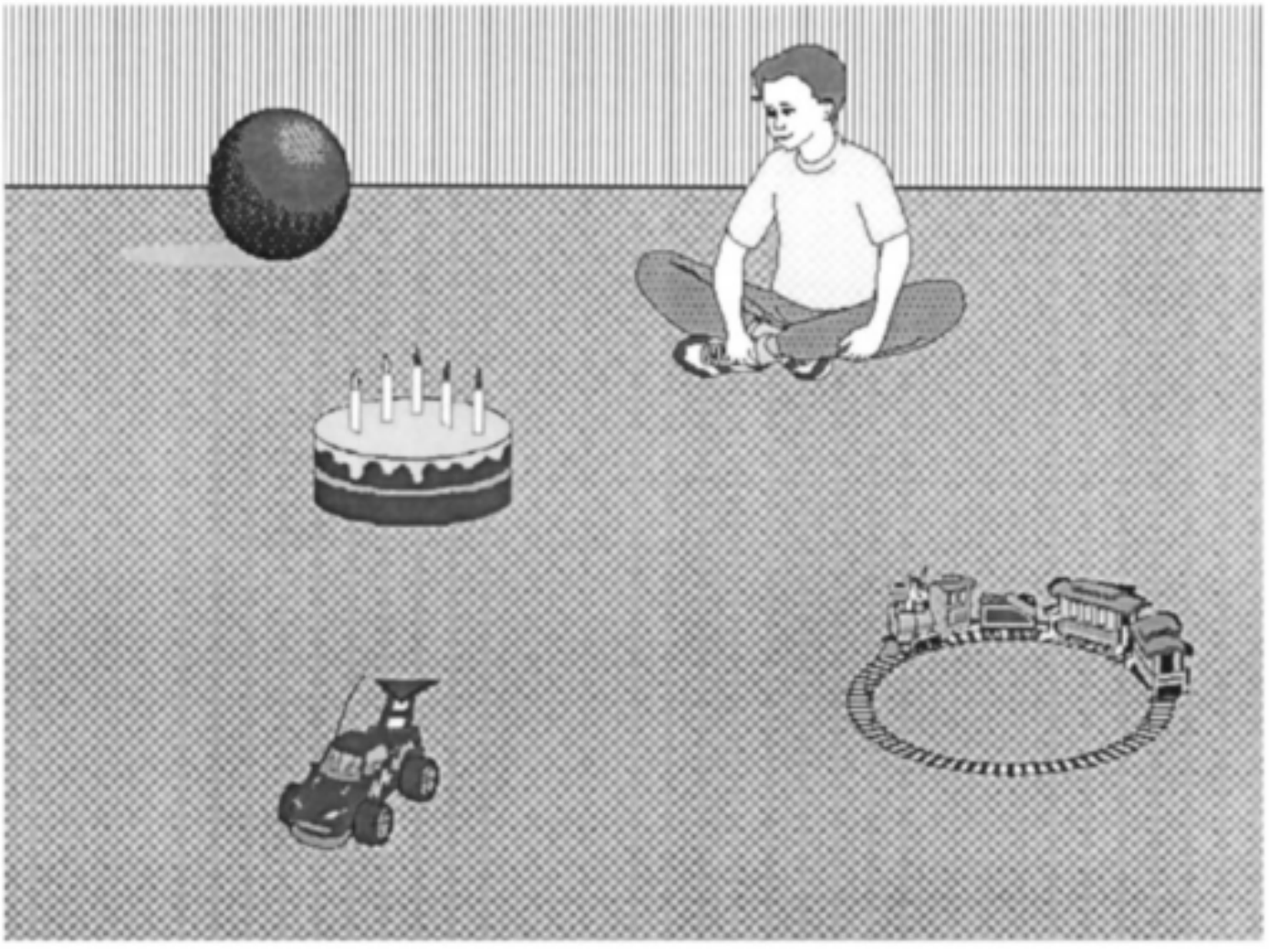


Shen et al. (2013)

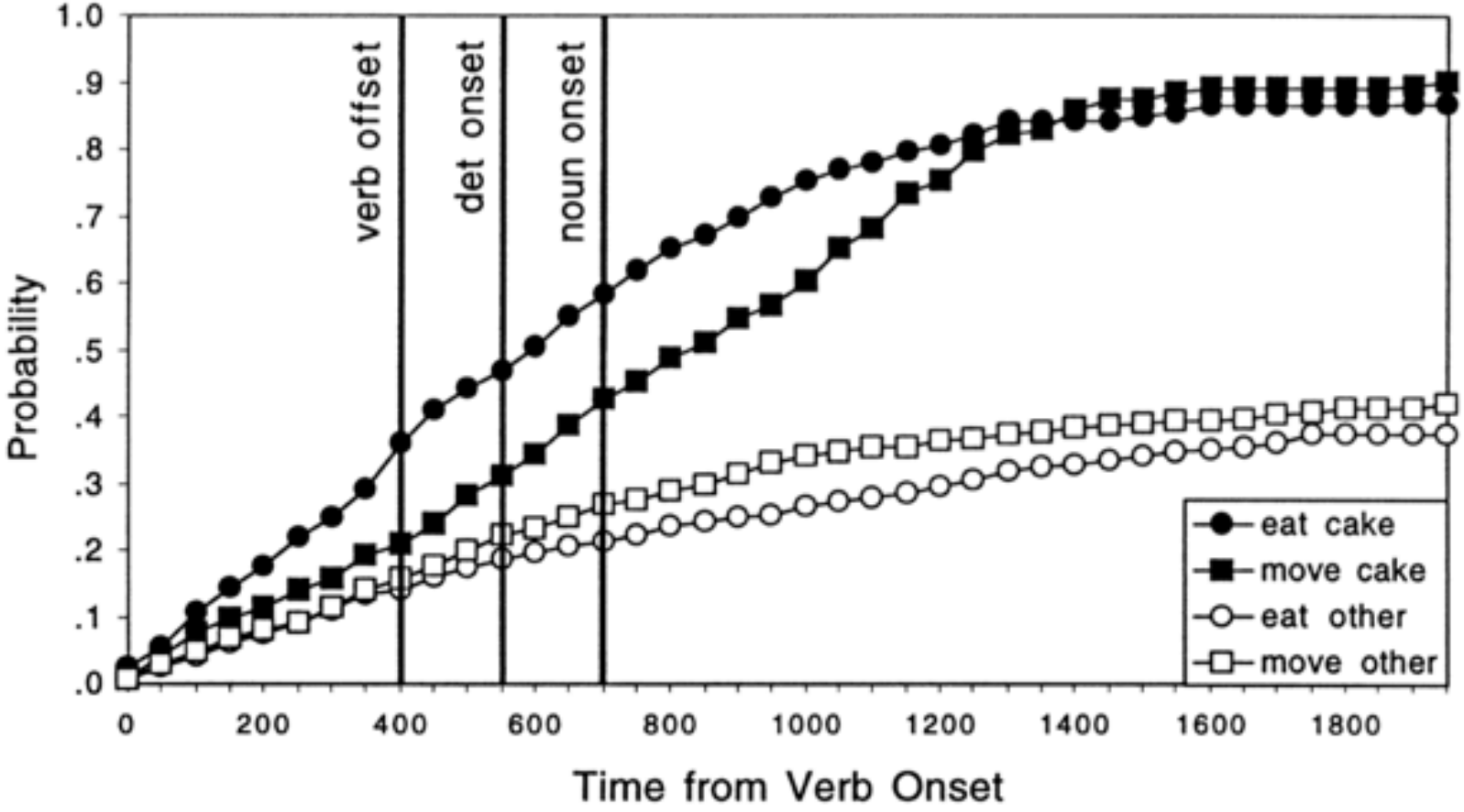
Shen et al. (2013)



Altmann and Kamide (1999)

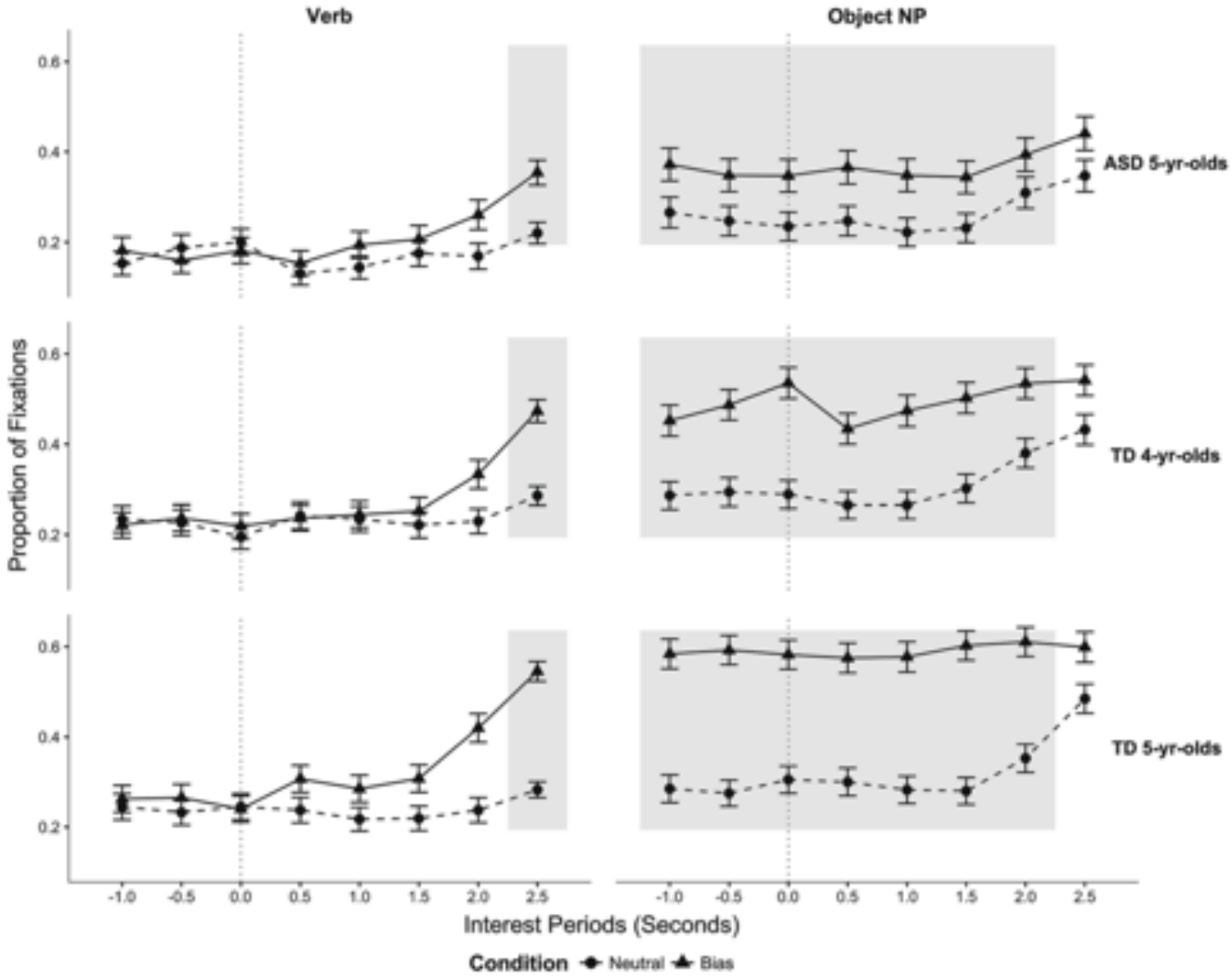


Altmann and Kamide (1999)

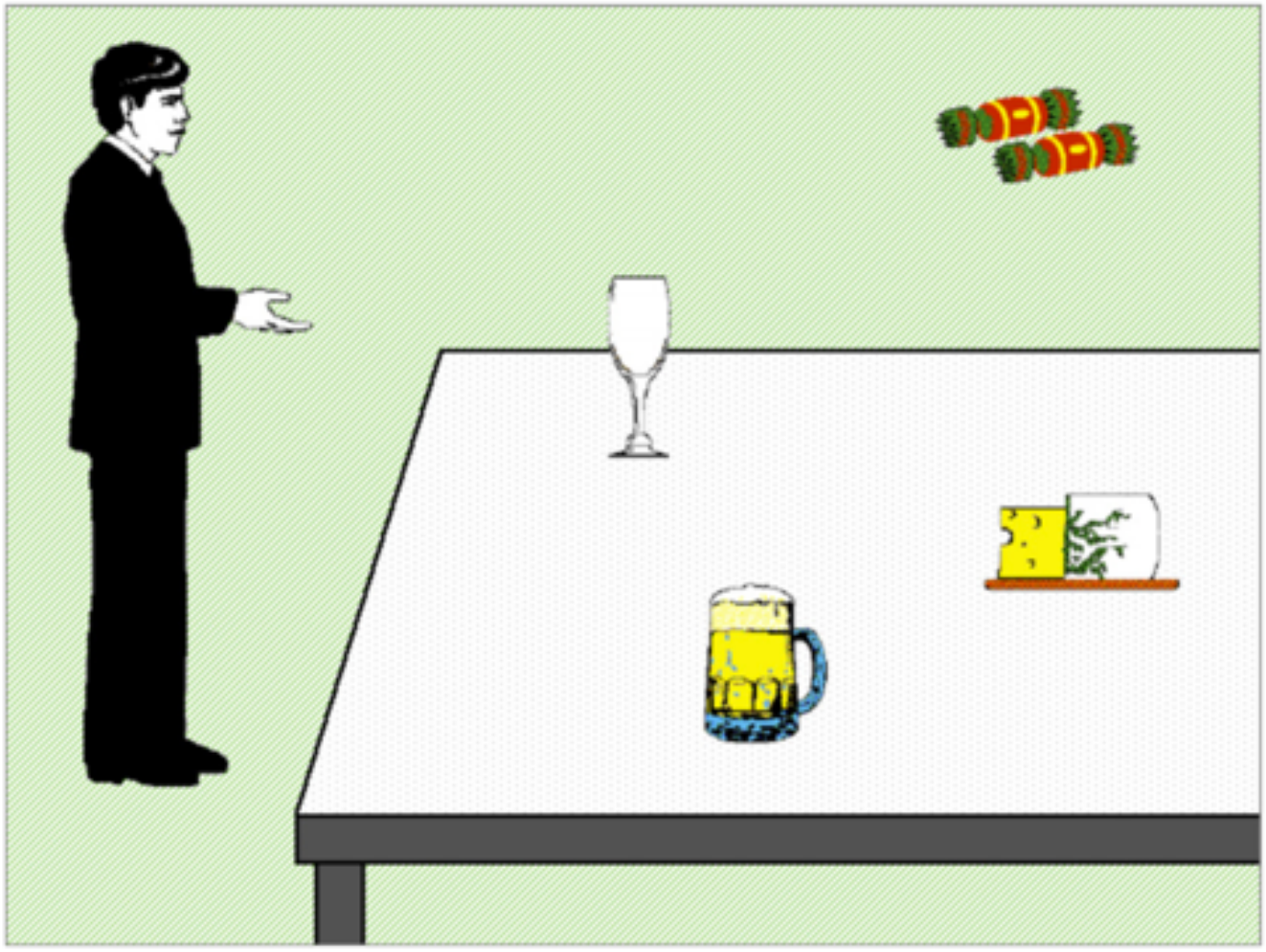


Zhou et al. (2018)

Zhou et al. (2018)



Altmann and Kamide (2007)

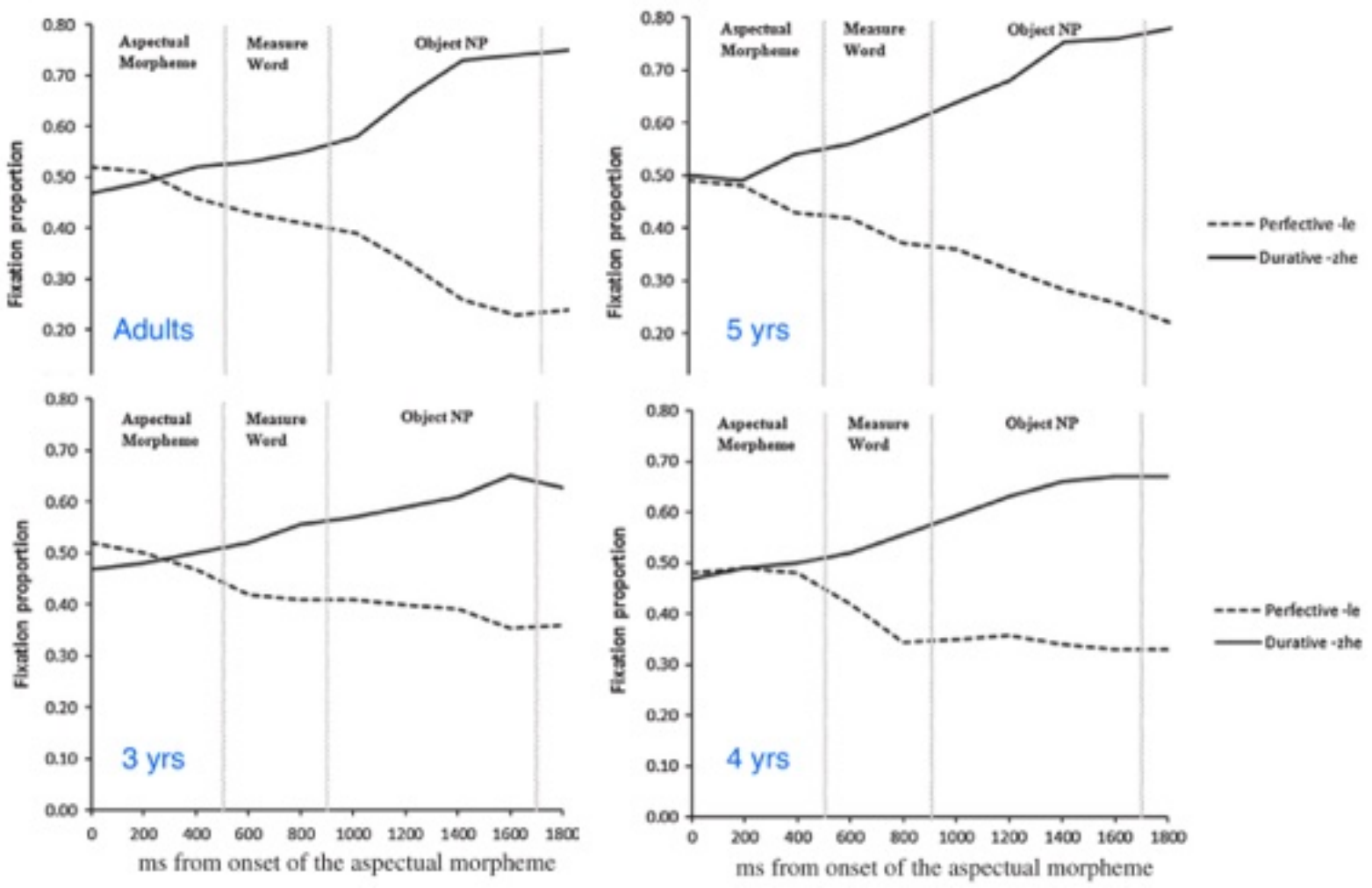


Zhou et al. (2014)

Completed Event Area

Ongoing Event Area

Zhou et al. (2014)



References

References i

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439. doi: https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264. doi: 10.1016/s0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518. doi: 10.1016/j.jml.2006.12.004

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460-482. doi: 10.1016/j.jml.2007.02.001

References ii

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*(1), 37-55. doi: 10.1023/a:1021933015362

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32-38. doi: 10.1111/1467-9280.00211

Moscati, V., Zhan, L., & Zhou, P. (2017). Children's on-line processing of epistemic modals. *Journal of Child Language*, *44*(5), 1025-1040. doi: 10.1017/s0305000916000313

Salverda, A. P., & Tanenhaus, M. K. (2017). The visual world paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide.* Hoboken, NJ: Wiley.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of mandarin tones 2 and 3: Evidence from eye movements. *Journal of the Acoustical Society of America*, *133*(5), 3016-3029. doi: 10.1121/1.4795775

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238-299. doi: 10.1016/j.cogpsych.2004.03.001

References iii

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634. doi: 10.1126/science.7777863

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134. doi: 10.1016/s0010-0277(99)00032-3

Zhan, L. (2018a). Scalar and ignorance inferences are both computed immediately upon encountering the sentential connective: The online processing of sentences with disjunction using the visual world paradigm. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.00061

Zhan, L. (2018b). Using eye movements recorded in the visual world paradigm to explore the online processing of spoken language. *Journal of Visualized Experiments*, *140*, e58086. doi: 10.3791/58086

Zhan, L., Crain, S., & Zhou, P. (2015). The online processing of only if and even if conditional statements: Implications for mental models. *Journal of Cognitive Psychology*, *27*(3), 367-379. doi: 10.1080/20445911.2015.1016527

References iv

Zhan, L., Zhou, P., & Crain, S. (2018). Using the visual-world paradigm to explore the meaning of conditionals in natural language. *Language, Cognition and Neuroscience*, *33*(8), 1049-1062. doi: 10.1080/23273798.2018.1448935

Zhou, P., Crain, S., & Zhan, L. (2012). Sometimes children are as good as adults: The pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language*, *67*(1), 149-164. doi: 10.1016/j.jml.2012.03.005

Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, *133*(1), 262-276. doi: 10.1016/j.cognition.2014.06.018

Zhou, P., Su, Y., Crain, S., Gao, L. Q., & Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: a view from mandarin chinese. *Journal of Child Language*, *39*(4), 687-730. doi: 10.1017/s0305000911000249

Zhou, P., Zhan, L., & Ma, H. (2018). Predictive language processing in preschool children with autism spectrum disorder: An eye-tracking study. *Journal of Psycholinguistic Research*. doi: 10.1007/s10936-018-9612-5