Visual World Paradigm

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

ZHAN, Likan
Beijing Language and Culture University

2023-11-26

https://likan.org
zhan@likan.org

Table of Contents

Table of Contents

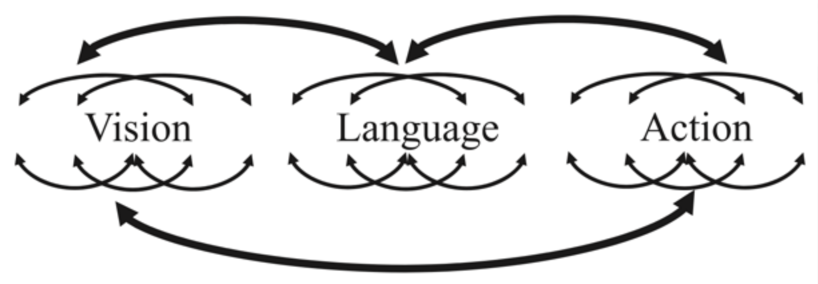
The Definition

Language in Interactive Frameworks

Language in Interactive Frameworks

- Cognitive processes are better understood not as computations that take place solely inside the brain, but instead as emergent properties resulting from the interaction of the brain with the body and the environment. (Spivey, 2023, P.230)

Language in Interactive Frameworks



(Spivey, 2023)

The Visual World Paradigm

The Visual World Paradigm

- The visual world paradigm has allowed spoken language processing studies to be grounded, revealing a system that rapidly integrates multiple constraints, including effects of the information in the visual context and task goals. (Wei & Tanenhaus, 2023)

The Visual World Paradigm

(Salverda & Tanenhaus, 2017)

The Visual World Paradigm

- In the **visual world paradigm** (VWP), 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

- In the **visual world paradigm** (VWP), 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.
- It 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

(Salverda & Tanenhaus, 2017)

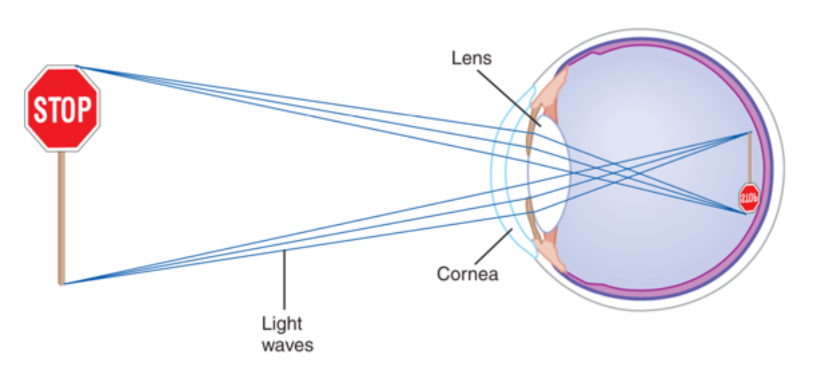
The Visual World Paradigm

- 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)

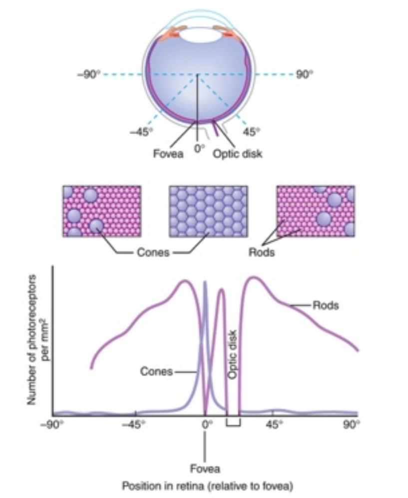
The Linking Hypothesis

The Linking Hypothesis



(Stanfield, 2013, PP.269-284)

The Linking Hypothesis



(Stanfield, 2013, PP.269-284)

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

Visual World: 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

(Cooper, 1974)

Visual World: A Brief History

北京语言大学

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

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

Visual World: 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

(Allopenna, Magnuson, & Tanenhaus, 1998)

Visual World: 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

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

Visual World: 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

(Meyer, Sleiderink, & Levelt, 1998)

Visual World: A Brief History

Seeing is believing: testing an explicit linking assumption for visual world eye-tracking in psycholinguistics

Judith Degen (jdegen@stanford.edu)

Leyla Kursat (lkursat@stanford.edu)

Daisy Leigh (ddleigh@stanford.edu)
Department of Linguistics, 450 Jane Stanford Way
Stanford, CA 94305 USA

Abstract

Experimental investigation is fundamental to theory-building in cognitive science, but its value depends on the *linking assumptions* made by researchers about the mapping between empirical measurements and theoretical constructs. We argue that sufficient clarity and justification are often lacking for linking assumptions made in *visual world eye-tracking*, a widely used experimental method in psycholinguistic research. We test what we term the *Referential Belief* linking assumption: that the proportion of looks to a referent in a time window reflects participants' degree of belief that the referent is the intended target in that time window. We do so by comparing eye-tracking data against explicit beliefs collected in an incremental decision task (Exp. 1), which replicates a scalar implicature processing study (Exp. 3 of Sun & Breheny, 2020). In Exp. 2, we replicate Sun and Breheny (2020) in a web-based eye-tracking paradigm using `WebGazer.js`. The results provide support for the Referential Belief link and cautious optimism for the prospect of conducting web-based eye-tracking. We discuss limitations on both fronts.

Keywords: psycholinguistics; experimental pragmatics; scalar implicature; linking functions; visual world; eye-tracking

coarse-grained temporal measures like response times from button presses. Notable VWP findings that could not have been obtained with more coarse-grained measures include the diverse insights that visual context is rapidly integrated into syntactic structure assignment (Tanenhaus et al., 1995), that words are processed incrementally and listeners maintain uncertainty about past input (Allopenna et al., 1998; Clayards et al., 2008), and that listeners anticipate upcoming linguistic material based on selectional restrictions and rapid pragmatic reasoning (Altmann & Kamide, 1999; Sedivy et al., 1999).

These notable successes notwithstanding, we still have a poor understanding of how to link observed eye movements to the underlying mental processes that generate them (Salverda & Tanenhaus, 2017; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Allopenna et al., 1998; Magnuson, 2019). The problem of interpretability is compounded by the fact that the VWP is used for vastly different tasks (for an overview, see Huettig, Rommers, & Meyer, 2011). Consider the difference between active referential tasks, in which participants' goal is to identify and select the speaker's intended

(Degen, Kursat, & Leigh, 2021)

Visual World: A Brief History

Integration of visual and linguistic information in spoken language comprehension.

By Tanenhaus, M K; Spivey-Knowlton, M J; Eberhard, K M; Sedivy, J C

View Web of Science ResearcherID and ORCID (provided by Clarivate)

Source Science (New York, N.Y.)

Volume: 268 Issue: 5217 Page: 1632-4

DOI: 10.1126/science.7777863

Published 1995-Jun-16

Indexed 1995-06-16

Document Type Journal Article; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, Non-P.H.S.; Research Support, U.S. Gov't, P.H.S.

Abstract 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 comprehension, eye movements were recorded with a head-mounted eye-tracking system while subjects followed instructions to manipulate real objects. Visual context influenced spoken word recognition and mediated syntactic processing, even during the earliest moments of language processing.

Total citations: Up to *2023-11-26*

Table of Contents

Apparatus

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.

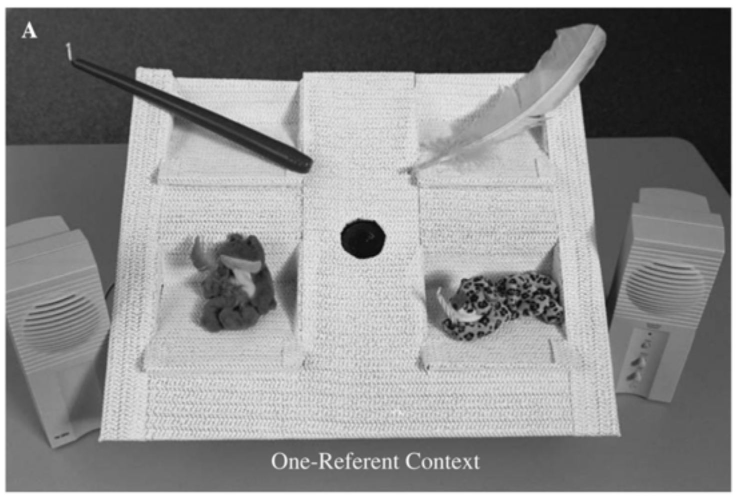
(Zhan, 2018b)

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

Visual World

(Zhan, 2018b)

Visual World

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

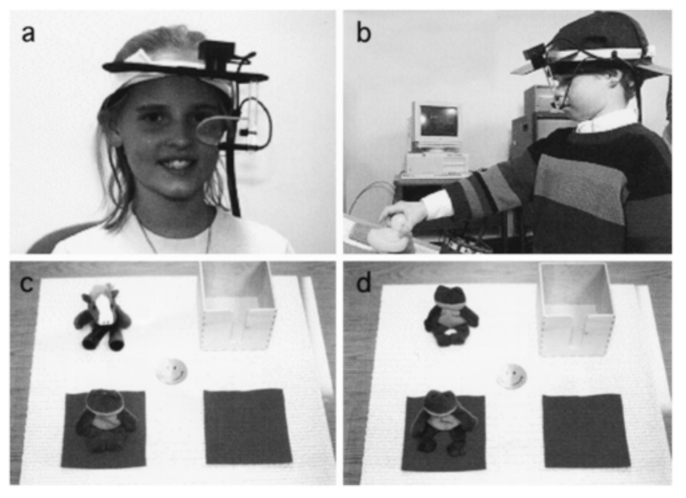
(Zhan, 2018b)

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: In the Literature

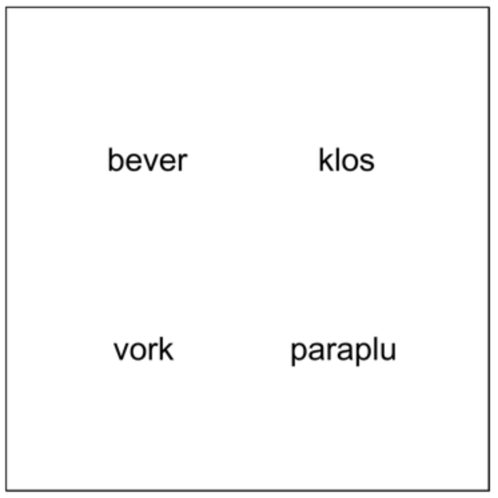


(Trueswell et al., 1999)

Visual World: In the Literature

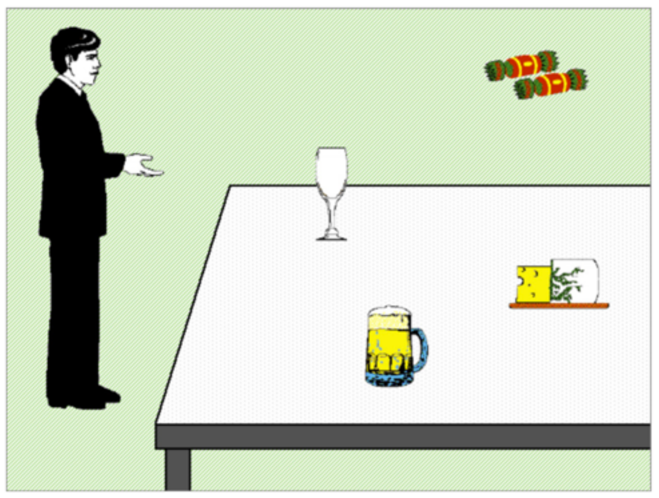
(Huettig & McQueen, 2007)

Visual World: In the Literature



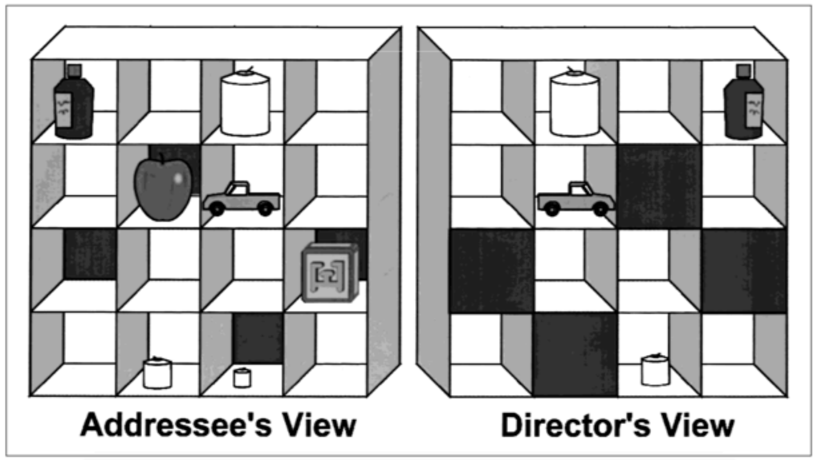
(Huettig & McQueen, 2007)

Visual World: In the Literature



(Altmann & Kamide, 2007)

Visual World: In the Literature

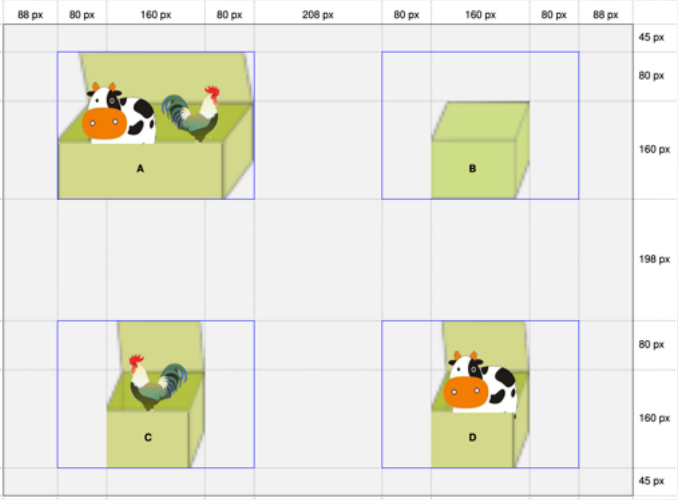


(Keysar, Barr, Balin, & Brauner, 2000)

Visual World: Our Studies

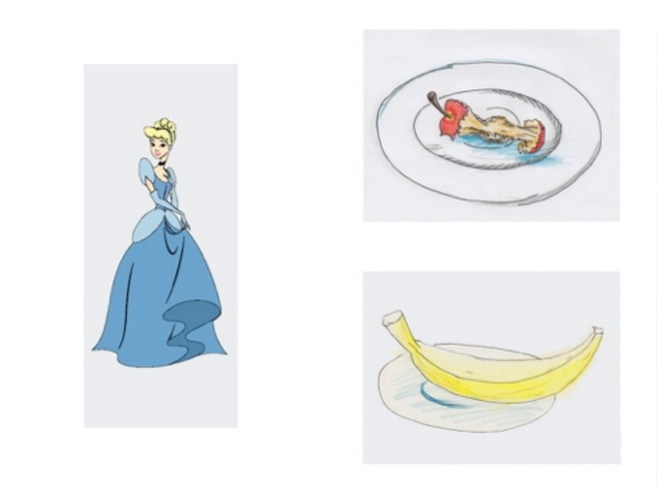
(Zhan, Crain, & Zhou, 2015)

Visual World: Our Studies



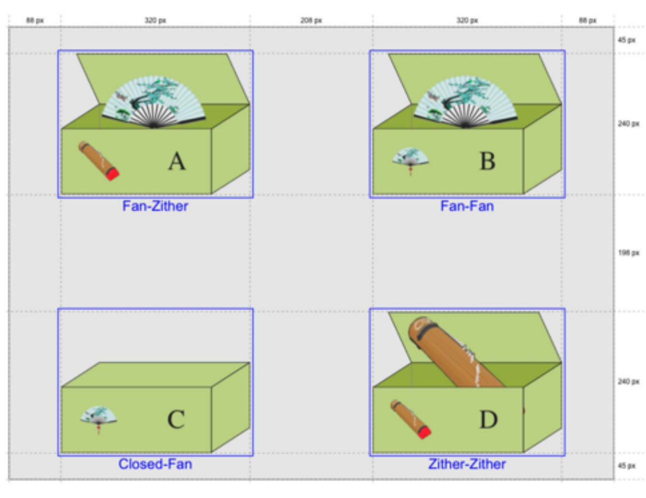
(Zhan, 2018a, 2018b)

Visual World: Our Studies



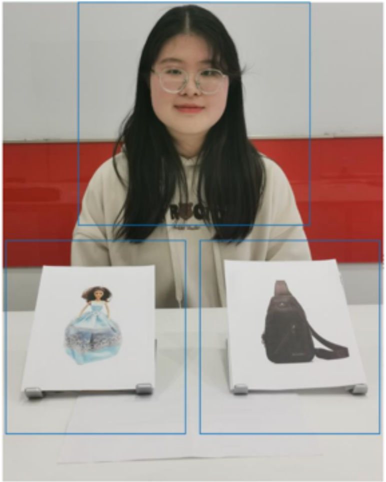
(Zhan, Zhou, & Crain, 2018)

Visual World: Our Studies

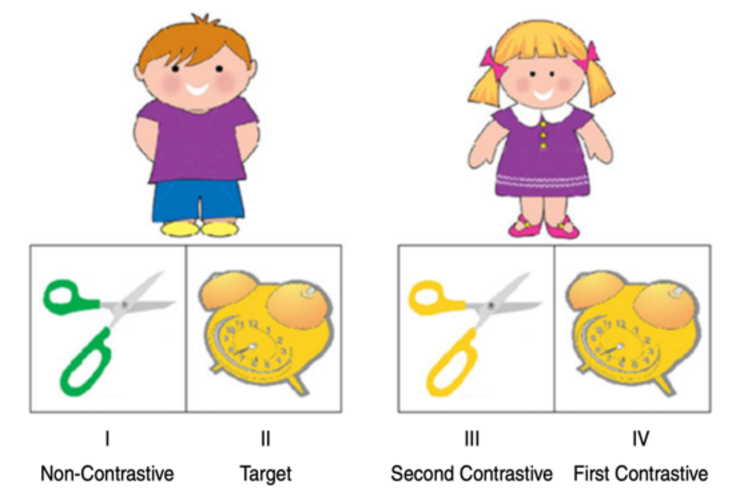


(Zhan & Zhou, 2023)

Visual World: Our Studies

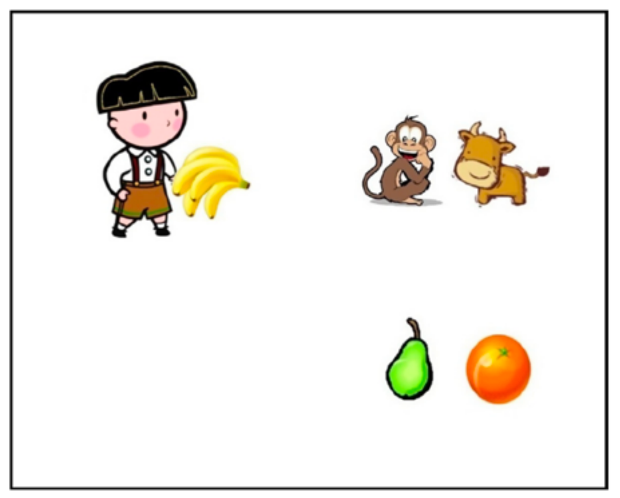


Visual World: Our Studies



(Zhou, Su, Crain, Gao, & Zhan, 2012)

Visual World: Our Studies

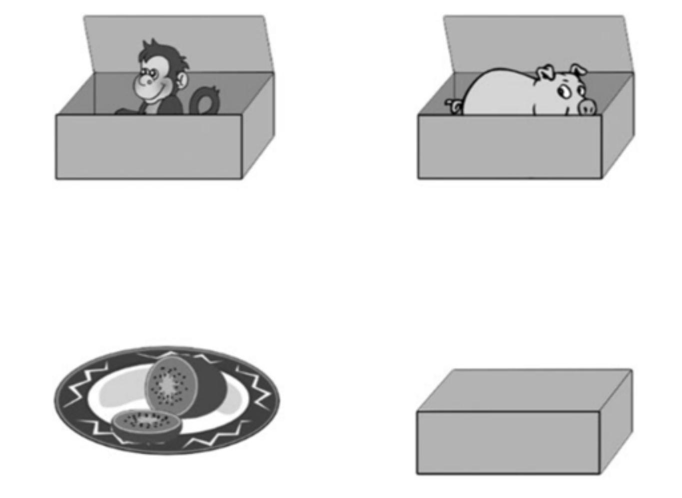


(Zhou, Crain, & Zhan, 2012; Zhou, Ma, & Zhan, 2019)

Visual World: Our Studies

(Zhou, Crain, & Zhan, 2014)

Visual World: Our Studies

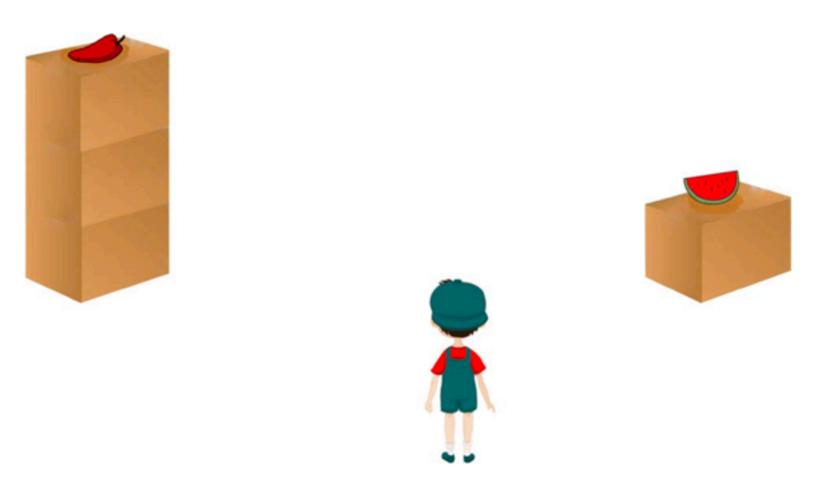


(Moscati, Zhan, & Zhou, 2017)

Visual World: Our Studies

(Zhou, Ma, Zhan, & Ma, 2018)

Visual World: Our Studies



(Zhou, Zhan, & Ma, 2019b)

Visual World: Our Studies

(Zhou, Zhan, & Ma, 2019a)

Visual World: Our Studies

(Zhou, Shi, & Zhan, 2021)

Spoken Language

Spoken Language: In the Literature

The language can differ along any number of dimensions, from

(Salverda & Tanenhaus, 2017)

Spoken Language: In the Literature

The language can differ along any number of dimensions, from

- Manipulations of fine-grained acoustic phonetic features (duration, voice onset time, formant structure, fundamental frequency, etc.) to

(Salverda & Tanenhaus, 2017)

Spoken Language: In the Literature

The language can differ along any number of dimensions, from

- Manipulations of fine-grained acoustic phonetic features (duration, voice onset time, formant structure, fundamental frequency, etc.) to
- Properties of words (syntactic category, semantic features, frequency of occurrence, etc.) to

(Salverda & Tanenhaus, 2017)

Spoken Language: In the Literature

The language can differ along any number of dimensions, from

- Manipulations of fine-grained acoustic phonetic features (duration, voice onset time, 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: In the Literature

(Salverda & Tanenhaus, 2017)

Spoken Language: In the Literature

- 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 Studies

Spoken Language: Our Studies

- The spoken language can differ in their verbs (Zhou, Zhan, & Ma, 2019a), syntactic structure (Zhou et al., 2018, 2021), their phonological stresses (Zhou, Su, et al., 2012), their sentential prosodies (Zhou, Crain, & Zhan, 2012; Zhou, Ma, & Zhan, 2019), their aspect markers (Zhou et al., 2014), their epistemic modals (Moscati et al., 2017), and their pronouns (Xie & Zhan, 2023).

Spoken Language: Our Studies

- The spoken language can differ in their verbs (Zhou, Zhan, & Ma, 2019a), syntactic structure (Zhou et al., 2018, 2021), their phonological stresses (Zhou, Su, et al., 2012), their sentential prosodies (Zhou, Crain, & Zhan, 2012; Zhou, Ma, & Zhan, 2019), their aspect markers (Zhou et al., 2014), their epistemic modals (Moscati et al., 2017), and their pronouns (Xie & Zhan, 2023).
- 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; Zhan & Zhou, 2023), and disjunctions (Zhan, 2018a, 2018b).

Spoken Language: Our Studies

(Zhan et al., 2015)

Spoken Language: Our Studies

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.

0.5s	1.8s	1.6s	1.3s	1s	1.4s	1s	1s	1.4s	
0s	0.5s	2.3s	3.9s	5.2s	6.2s	7.6s	8.6s	9.6s	11s

(Zhan, 2018a, 2018b)

Spoken Language: Our Studies

a). **And**

你看	公主	吃掉了	那个	苹果	/*香蕉	之后	她	就会受到惩罚
nikan	gongzhu	chidiaole	nage	pingguo	/*xiangjiao	zhihou	ta	jiu hui shoudao chengfa
look	princess	eat-Asp	that	apple	/*banana	then	she	will be punished

*Look, the princess eats that apple/*banana, then she will be punished.*

b). **Because**

因为	公主	吃掉了	那个	苹果	/*香蕉	所以	她	就会受到惩罚
yinwei	gongzhu	chidiaole	nage	pingguo	/*xiangjiao	suoyi	ta	jiu hui shoudao chengfa
because	princess	eat-Asp	that	apple	/*banana	therefore	she	will be punished

*Because the princess eats that apple/*banana, therefore she will be punished.*

c). **If**

如果	公主	吃掉了	那个	*苹果	/香蕉	那么	她	就会受到惩罚
ruguo	gongzhu	chidiaole	nage	*pingguo	/xiangjiao	name	ta	jiu hui shoudao chengfa
if	princess	eat-Asp	that	*apple	/banana	then	she	will be punished

*If the princess eats that *apple/banana, then she will be punished.*

Length:	0.95s	0.95s	1.2s	0.9s	1.1s	(1.1s)	0.9s	0.4s	1.2s
Onset:	0s	0.95s	1.9s	3.1s	4.0s	4.0s	5.1s	6.0s	6.4s	7.6s

(Zhan et al., 2018)

Spoken Language: Our Studies

Because

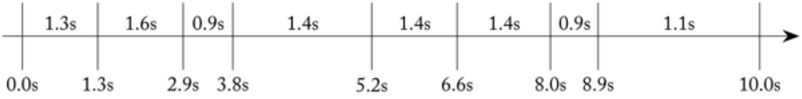
因为	箱子里	是	扇子/古筝	所以	小明	很	高兴/* 伤心
yinwei	xiangzi li	shi	shanzi/guzheng	suoyi	Xiaoming	hen	gaoxing/*shangxin
because	box in	is	fan/zither	therefore	Xiaoming	very	happy/*sad

*Because the box contains a fan/zither, therefore John is very happy/*sad.*

If

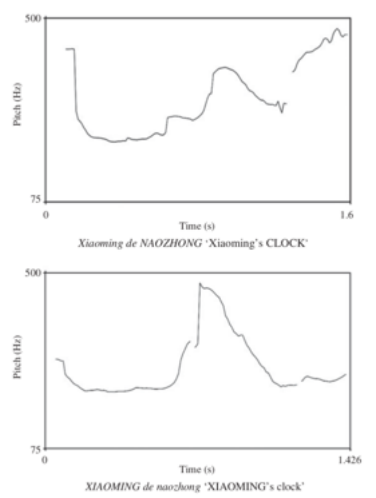
如果	箱子里	是	扇子/古筝	那么	小明	就	高兴/伤心
Ruguo	xiangzi li	shi	shanzi/guzheng	name	Xiaoming	jiu	gaoxing/shangxin
If	box in	is	fan/zither	then	John	will	happy/sad

If the box contains a fan/zither, then John will be very happy/sad.



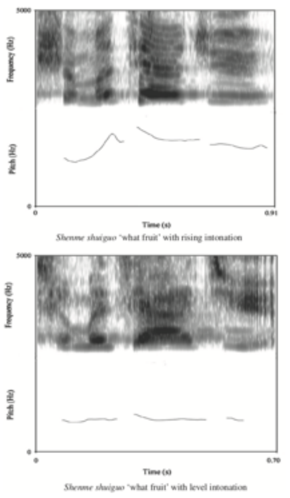
(Zhan & Zhou, 2023)

Spoken Language: Our Studies



(Zhou, Su, et al., 2012)

Spoken Language: Our Studies



(Zhou, Crain, & Zhan, 2012; Zhou, Ma, & Zhan, 2019)

Spoken Language: Our Studies

(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 Studies

Scenario	Modal	Examples
Undetermined	might	(1) *a monkey might be in the orange box*
	must	(2) *a monkey must be in the orange box*
Determined	must	(3) *a monkey must be in the orange box*
	might	(4) *a pear might be in the orange box*

(Moscati et al., 2017)

Spoken Language: Our Studies

a. **BA** shizi qingqingdi bao-le qilai.
BA lion gently hold up
Meaning: Someone gently holds the lion.

b. **BEI** shizi qingqingdi bao-le qilai.
BEI lion gently hold up
Meaning: Someone is gently held by the lion.

(Zhou et al., 2018)

Spoken Language: Our Studies

Spoken Language: Our Studies

(1) a.	Kangkang	yao qu	**chi**	di-shang-de	dangao.
	Kangkang	will go	eat	floor-top	cake
	‘Kangkang is going to eat the cake on the floor.’				
b.	Kangkang	yao qu	**zhao**	di-shang-de	dangao.
	Kangkang	will go	find	floor-top	cake
	‘Kangkang is going to find the cake on the floor.’				

(Zhou, Zhan, & Ma, 2019a)

Spoken Language: Our Studies

(8)	Xiaomao	yaoqu	ti	xiaogou	DE	piqiu
	cat	will	kick	dog	DE	ball

“The cat is going to kick the dog’s ball.”

(Zhou et al., 2021)

Behavioral Task

Behavioral Task

Behavioral Task

- *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

- *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).

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*).

Participants

Participants

(Spivey, 2023)

Participants

- Eye movements provide a relatively unconscious measure of overt attention without interrupting the task with a metacognitive report (such as a lexical decision task) or a concurrent motor task (such as button-pressing).

(Spivey, 2023)

Participants

- Eye movements provide a relatively unconscious measure of overt attention without interrupting the task with a metacognitive report (such as a lexical decision task) or a concurrent motor task (such as button-pressing).
- Participants just carry out the instructions as naturally as possible, unaware that the precise timing and locations of their eye movements are giving us all the data we need.

(Spivey, 2023)

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)

Table of Contents

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

- 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)

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

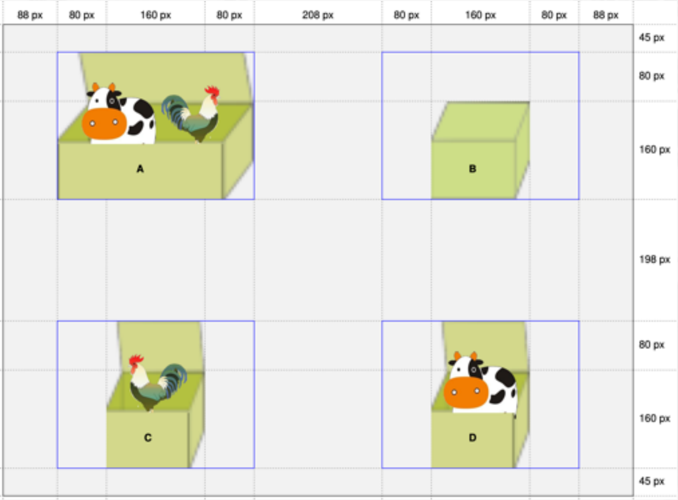
- Participants' interpretation of the spoken language is deduced from their eye movements on the visual world, not from the actual interpretation of the language stimuli per se.
- The visual world paradigm used is normally more restricted than the actual visual world, with a limited set of pictured referents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

(Zhan, 2018b)

Table of Contents

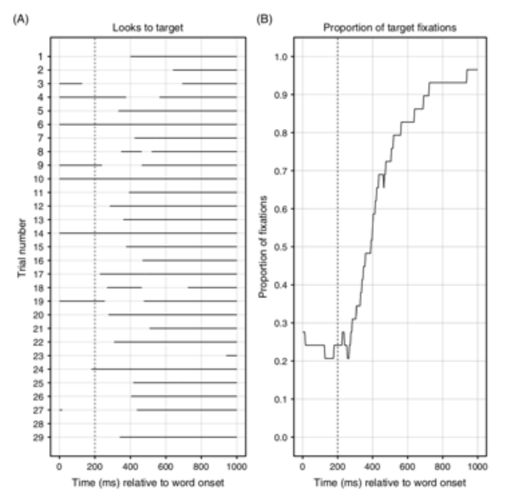
Descriptive Analysis

Regions of Interest



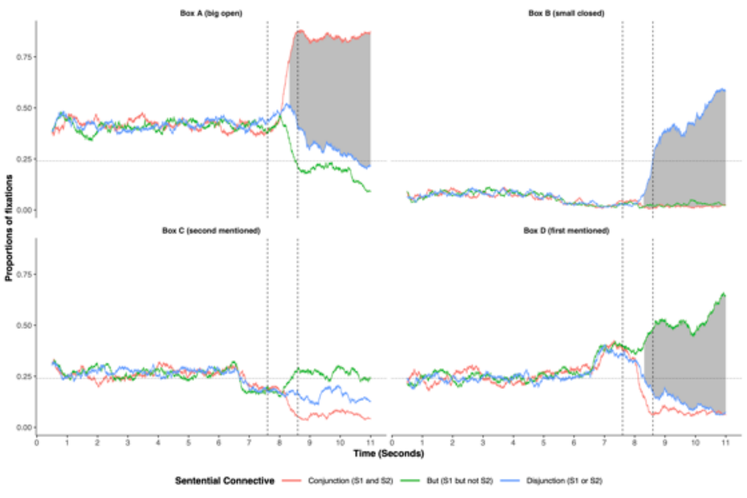
(Zhan, 2018a, 2018b)

Proportion of Fixations



(Salverda & Tanenhaus, 2017)

Data Visualization



(Zhan, 2018a, 2018b)

Inferential Analysis

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 binomial 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 binomial 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 binomial 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)

Statistical Analyses

- T-test, ANOVA
- LME: Linear Mixed-Effects Model
- GCA: Growth Curve Analysis
- CPA: Cluster-based Permutation Analysis
- BDOTS: Bootstrapped Difference of Time Series
- GAMM: Generalised Additive Mixed Modelling
- DPA: Divergence Point Analysis

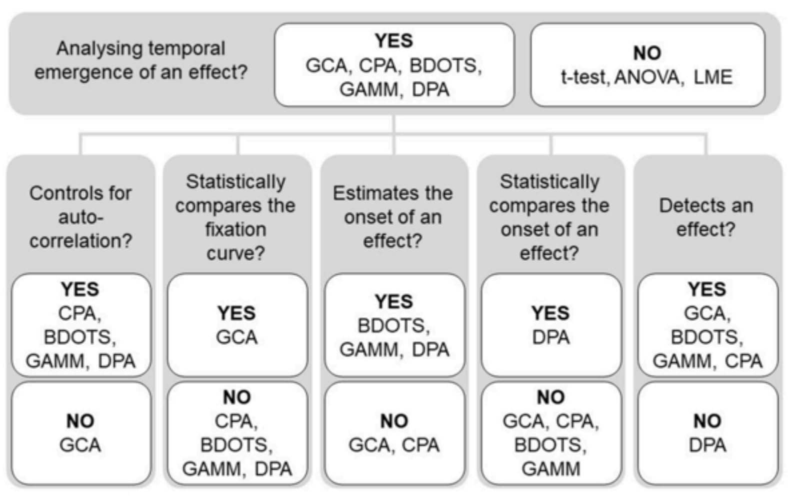
(Ito & Knoeferle, 2023)

Statistical Analyses

Analysis	Advantages	Disadvantages
GCA	▪ Can test differences in fixation curve across multiple conditions/groups	▪ Does not control for autocorrelation ▪ Including many higher-order polynomials may increase the chance of a false positive (Huang & Snedeker, 2020)
CPA	▪ Can test a difference in fixation proportion between two conditions/groups	▪ Cannot estimate or statistically compare the onset/offset of an effect ▪ May have reduced power to detect second (and later) clusters with smaller effects
BDOTS	▪ Can estimate when a fixation proportion difference between two conditions/groups occurred ▪ Can model typical fixation curves for target (mentioned) objects and competitor objects	▪ Requires a large number of data points (trials) to fit a good curve to the data ▪ Cannot statistically compare the onset/offset of an effect (e.g., test whether the onset was earlier in one condition than in another)
GAMM	▪ Can estimate when a fixation proportion difference between two conditions/groups occurred ▪ Can model linear and non-linear curves	▪ Does not control for autocorrelation for binomially coded data ▪ Cannot statistically compare the onset/offset of an effect (e.g., test whether the onset was earlier in one condition than in another)
DPA	▪ Can test a difference in the onset of an effect between two conditions/groups ▪ Can compute Bayes factors	▪ Does not detect (but assumes) an effect ▪ Cannot estimate multiple divergence points

(Ito & Knoeferle, 2023)

Statistical Analyses



(Ito & Knoeferle, 2023)

Table of Contents

Three Dicretions

(Wei & Tanenhaus, 2023)

Three Dicretions

- Coregister with brain imaging signals (EEG and fMRI)

(Wei & Tanenhaus, 2023)

Three Dicretions

- Coregister with brain imaging signals (EEG and fMRI)
- Three-dimensional virtual reality (VR)

(Wei & Tanenhaus, 2023)

Three Dicretions

- Coregister with brain imaging signals (EEG and fMRI)
- Three-dimensional virtual reality (VR)
- Virtual eye-tracking experiment via webcam

(Wei & Tanenhaus, 2023)

Apparatus: fMRI

(Frey, Nau, & Doeller, 2021)

Apparatus: Online

https://www.pcibex.net

Apparatus: Online

GORILLA

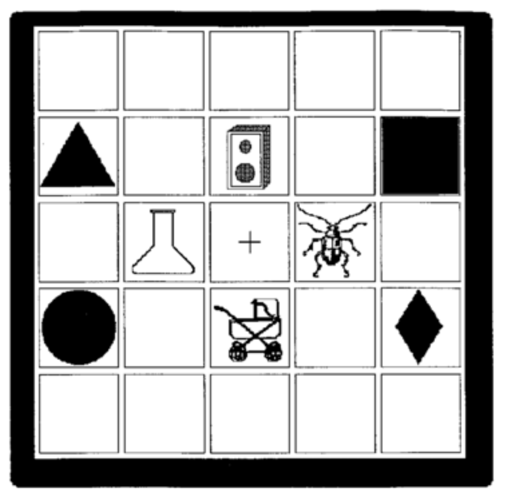
Behavioural Science in the Cloud

www.gorilla.sc

https://gorilla.sc

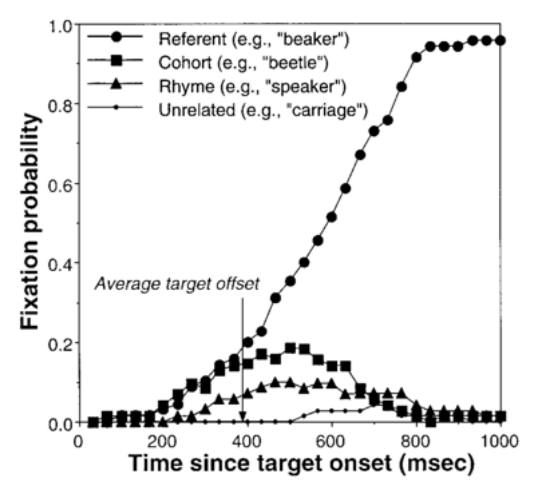
Table of Contents

Spoken Word Recognition



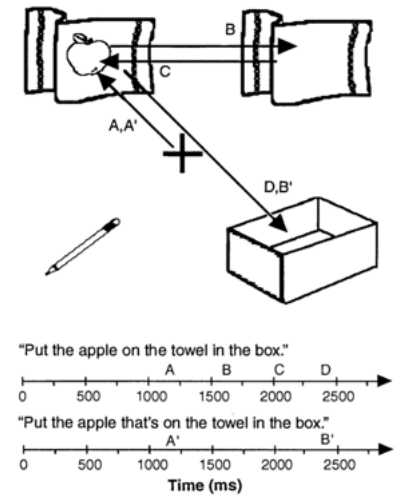
(Allopenna et al., 1998)

Spoken Word Recognition



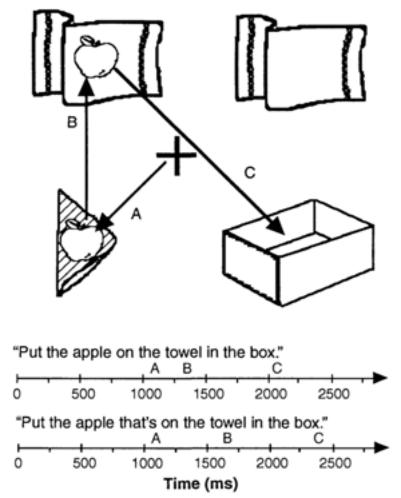
(Allopenna et al., 1998)

Syntactic Parsing



(Tanenhaus et al., 1995)

Syntactic Parsing



(Tanenhaus et al., 1995)

Syntactic Parsing

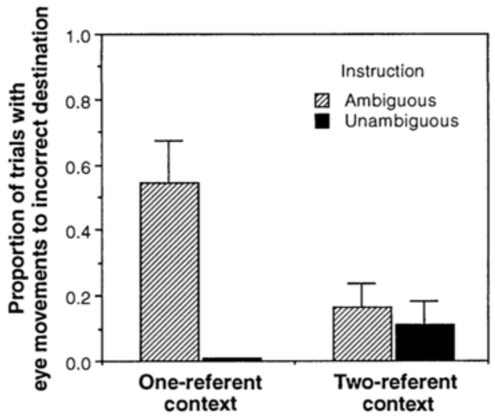
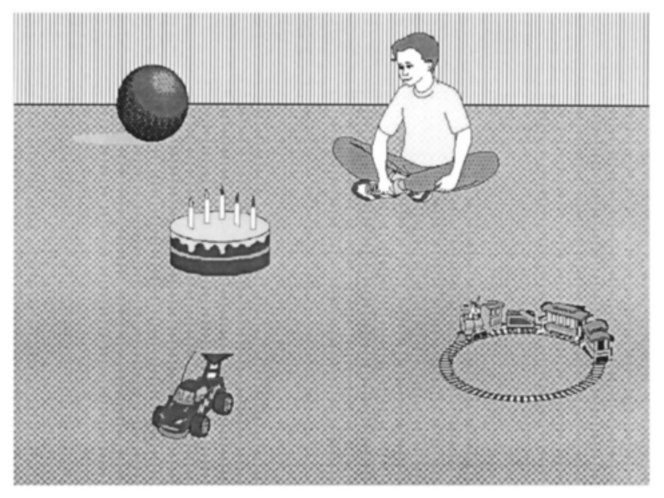


Fig. 3. Proportion of trials in which participants looked at the incorrect destination.

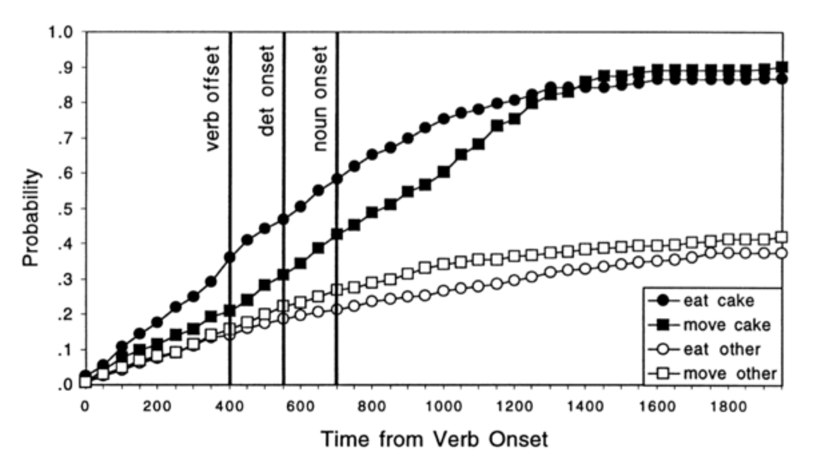
(Tanenhaus et al., 1995)

Semantic Integration: Verb



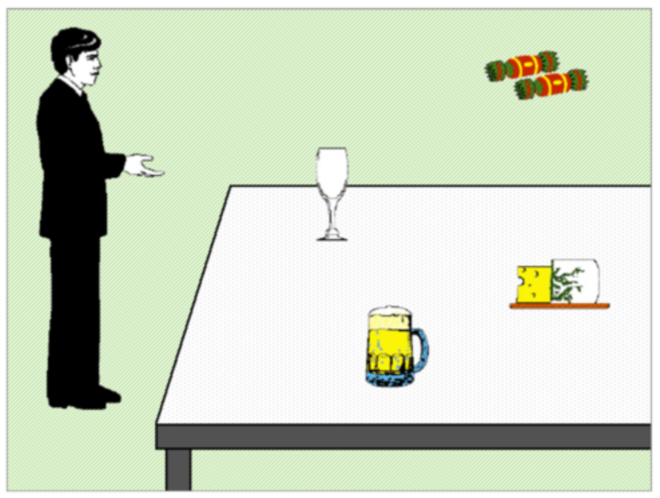
(Altmann & Kamide, 1999)

Semantic Integration: Verb



(Altmann & Kamide, 1999)

Semantic Integration: Tense Marker

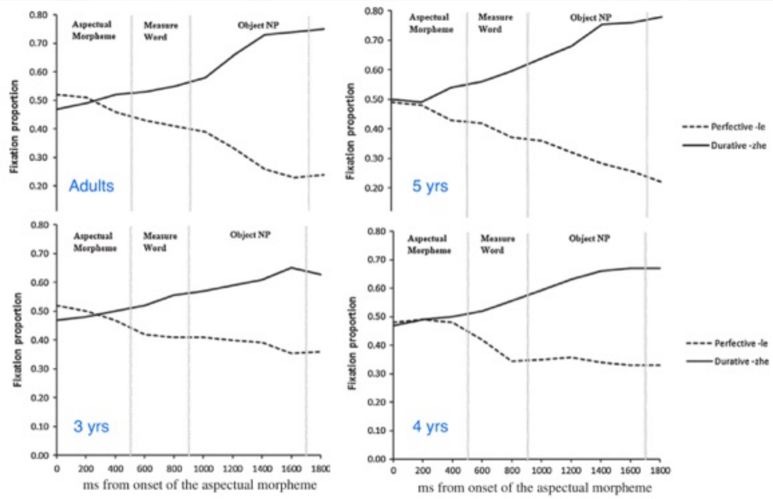


(Altmann & Kamide, 2007)

Semantic Integration: Aspect Marker

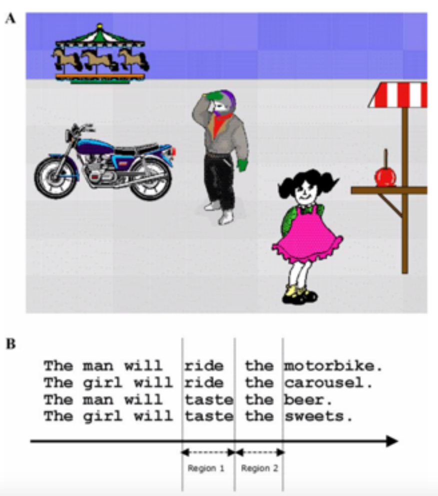
(Zhou et al., 2014)

Semantic Integration: Aspect Marker



(Zhou et al., 2014)

Semantic Integration: Subject



(Kamide, Scheepers, & Altmann, 2003)

Pragmatics and Communication

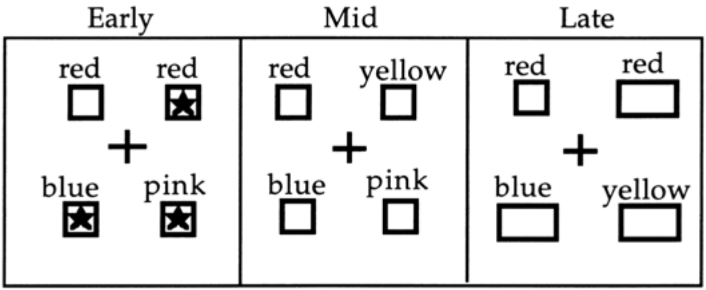
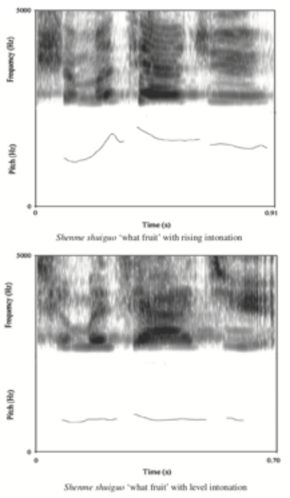


Fig. 1. Example displays from conditions manipulating the point at which a spoken instruction becomes unambiguous with respect to its referent. The accompanying instruction to this example was 'Touch the plain red square'.

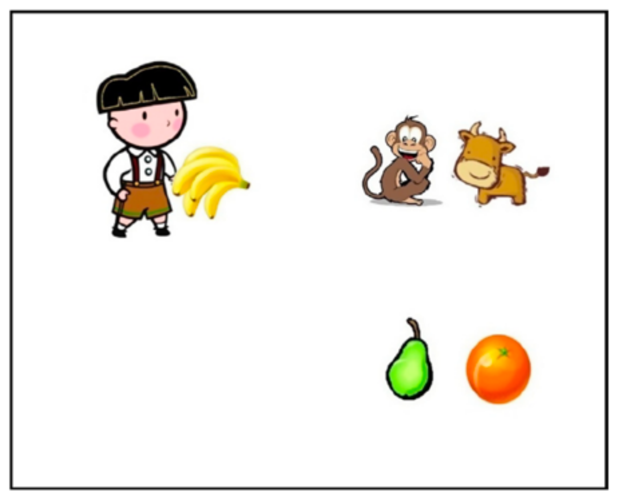
(Sedivy, Tanenhaus, Chambers, & Carlson, 1999)

Sentential Prosody: Question vs Statement



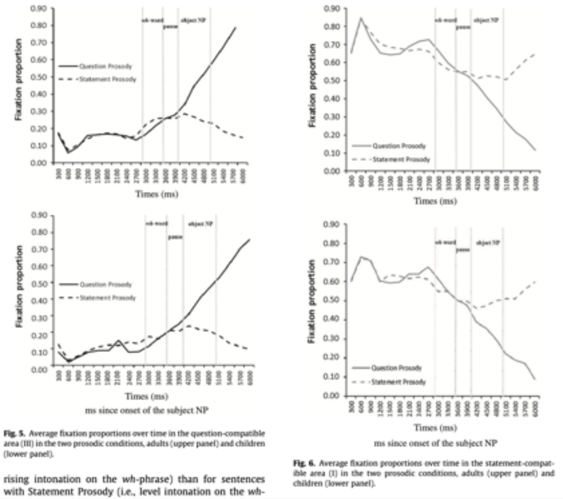
(Zhou, Crain, & Zhan, 2012)

Sentential Prosody: Question vs Statement



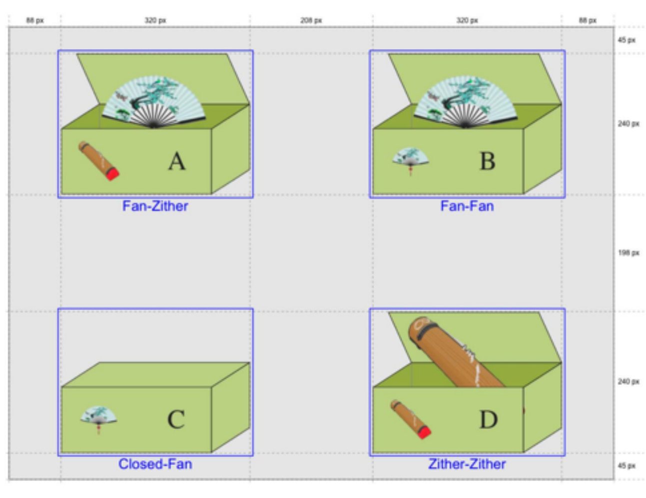
(Zhou, Crain, & Zhan, 2012)

Sentential Prosody: Question vs Statement



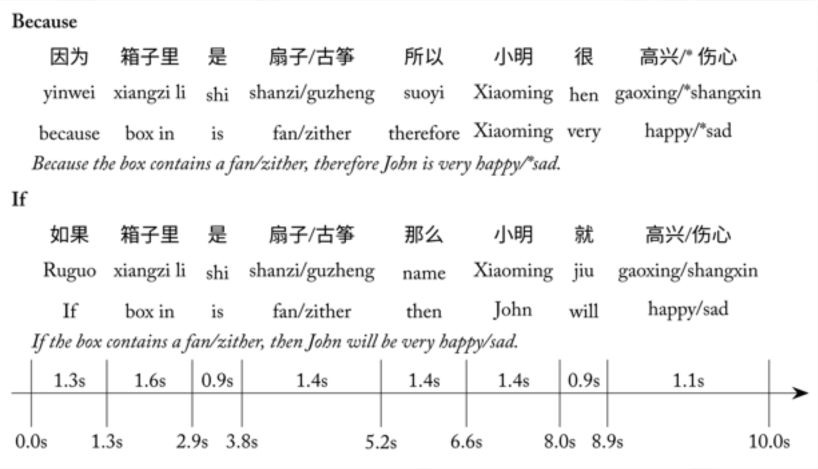
(Zhou, Crain, & Zhan, 2012)

Discouse Processing: If vs Because



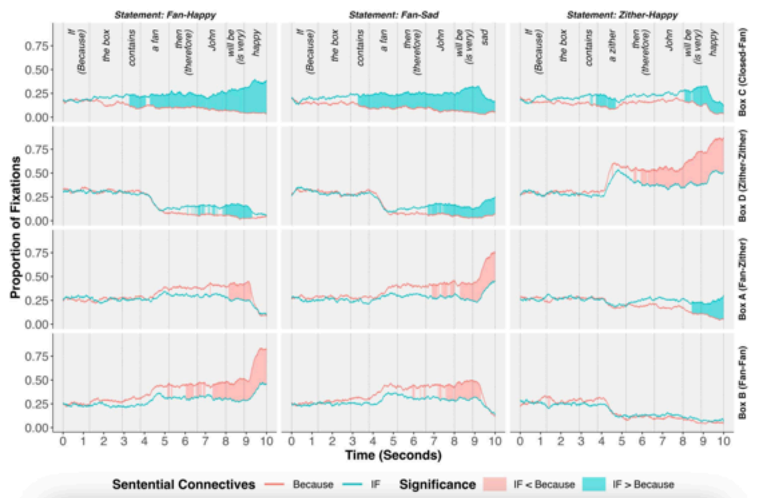
(Zhan & Zhou, 2023)

Discouse Processing: If vs Because



(Zhan & Zhou, 2023)

Discouse Processing: If vs Because



(Zhan & Zhou, 2023)

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 Article]. *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 [Journal Article]. *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 Article]. *Journal of Memory and Language*, *57*(4), 502-518. doi: 10.1016/j.jml.2006.12.004

Cooper, R. M. (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 [Journal Article]. *Cognitive Psychology*, *6*(1), 84-107. doi: 10.1016/0010-0285(74)90005-x

Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: testing an explicit linking assumption for visual world eye-tracking in psycholinguistics [Journal Article]. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

Frey, M., Nau, M., & Doeller, C. F. (2021). Magnetic resonance-based eye tracking using deep neural networks [Journal Article]. *Nature Neuroscience*, *24*(12), 1772-1779. doi: 10.1038/s41593-021-00947-w

References ii

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

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods [Journal Article]. *Behavior Research Methods*, *55*(7), 3461-3493. doi: 10.3758/s13428-022-01969-3

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 Article]. *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 [Journal Article]. *Psychological Science*, *11*(1), 32-38. doi: 10.1111/1467-9280.00211

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production [Journal Article]. *Cognition*, *66*(2), B25-B33. doi: 10.1016/s0010-0277(98)00009-2

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

References iii

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

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation [Journal Article]. *Cognition*, *71*(2), 109-147. doi: 10.1016/s0010-0277(99)00025-6

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 [Journal Article]. *Cognitive Psychology*, *49*(3), 238-299. doi: 10.1016/j.cogpsych.2004.03.001

Spivey, M. J. (2023). Cognitive science progresses toward interactive frameworks [Journal Article]. *Topics in Cognitive Science*, *15*(2), 219-254. doi: 10.1111/tops.12645

Stanfield, C. L. (2013). *Principles of human physiology* (5th ed.) [Book]. Person Education.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension [Journal Article]. *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 [Journal Article]. *Cognition*, *73*(2), 89-134. doi: 10.1016/s0010-0277(99)00032-3

References iv

Wei, Y., & Tanenhaus, M. K. (2023). Analysing spoken language comprehension with eye tracking [Book Section]. In S. Zufferey & P. Gygax (Eds.), *The routledge handbook of experimental linguistics* (chap. 17). Routledge.

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 [Journal Article]. *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 Article]. *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 Article]. *Journal of Cognitive Psychology*, *27*(3), 367-379. doi: 10.1080/20445911.2015.1016527

Zhan, L., & Zhou, P. (2023). The online processing of hypothetical events: A visual world eye-tracking study on conditionals and causal statements [Journal Article]. *Experimental Psychology*, *70*(2), 108-117. doi: 10.1027/1618-3169/a000579

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

References v

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 Article]. *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 [Journal Article]. *Cognition*, *133*(1), 262-276. doi: 10.1016/j.cognition.2014.06.018

Zhou, P., Ma, W., & Zhan, L. (2019). A deficit in using prosodic cues to understand communicative intentions by children with autism spectrum disorders: An eye-tracking study [Journal Article]. *First Language*, *40*(1), 41-63. doi: 10.1177/0142723719885270

Zhou, P., Ma, W., Zhan, L., & Ma, H. (2018). Using the visual world paradigm to study sentence comprehension in mandarin-speaking children with autism [Journal Article]. *Journal of Visualized Experiments*(140). doi: 10.3791/58452

Zhou, P., Shi, J., & Zhan, L. (2021). Real-time comprehension of garden-path constructions by preschoolers: A mandarin perspective [Journal Article]. *Applied Psycholinguistics*, *42*(1), 181-205. doi: 10.1017/s0142716420000697

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 Article]. *Journal of Child Language*, *39*(4), 687-730. doi: 10.1017/s0305000911000249

References vi

Zhou, P., Zhan, L., & Ma, H. (2019a). Predictive language processing in preschool children with autism spectrum disorder: An eye-tracking study [Journal Article]. *Journal of Psycholinguistic Research*, *48*(2), 431-452. doi: 10.1007/s10936-018-9612-5

Zhou, P., Zhan, L., & Ma, H. (2019b). Understanding others' minds: Social inference in preschool children with autism spectrum disorder [Journal Article]. *Journal of Autism and Developmental Disorders*, *49*(11), 4523-4534. doi: 10.1007/s10803-019-04167-x