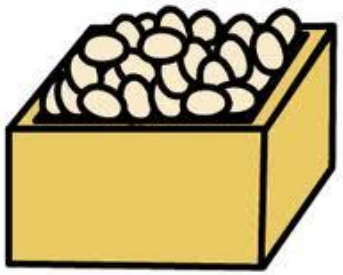
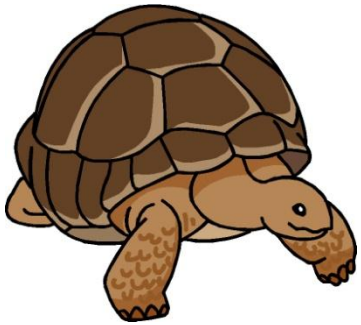
Sharing the beginning vs. the end: Spoken word recognition in the visual world paradigm in Japanese

Hideko Teruya
&
Vsevolod Kapatsinski
University of Oregon

Visual World Paradigm

*m**ame***
(rhyme)





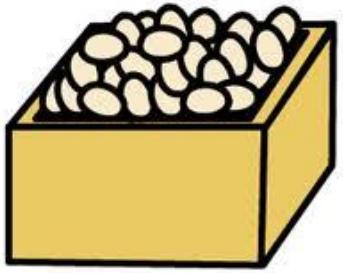
kame
(target)

tora
(unrelated)

***k**uchi*
(cohort)

Visual World Paradigm

*m**ame***
(rhyme)



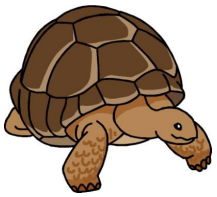
kame
(target)

tora
(unrelated)

***k**uchi*
(cohort)

General Question

- How do people process words in spoken word recognition?
  - Incremental models (McClelland & Elman, 1986)



/ k a m e /

  - Possible-word constraint (Norris et al., 1997)

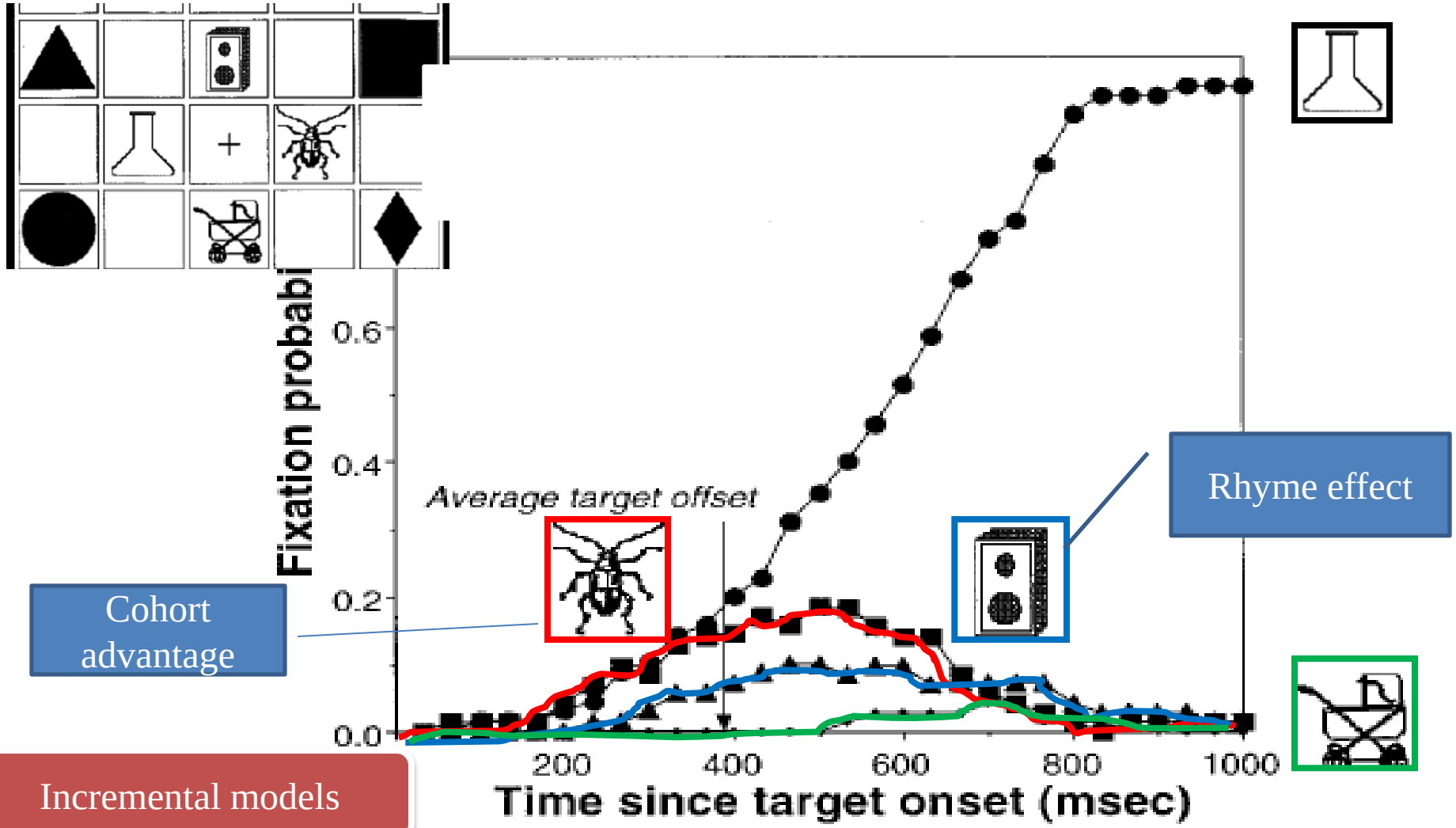
/ k a m e /

How much info do we need to move the eyes?

- Our hypothesis:
  - Processing is incremental
    - We continuously update word probabilities as acousix input comes in
    - Words that match the input enough get activated
      - Even if not “neighbors” of the target according to one-phoneme addition-deletion-substitution rule
  - Eye movements require a certain minimum amount of information
    - Only move once the lexical search space has been constrained enough
    - One consonant may not be enough (especially if it’s a common one)

Cohort & rhyme effect (Allopenna, et al., 1998)

"Pick up the beaker"



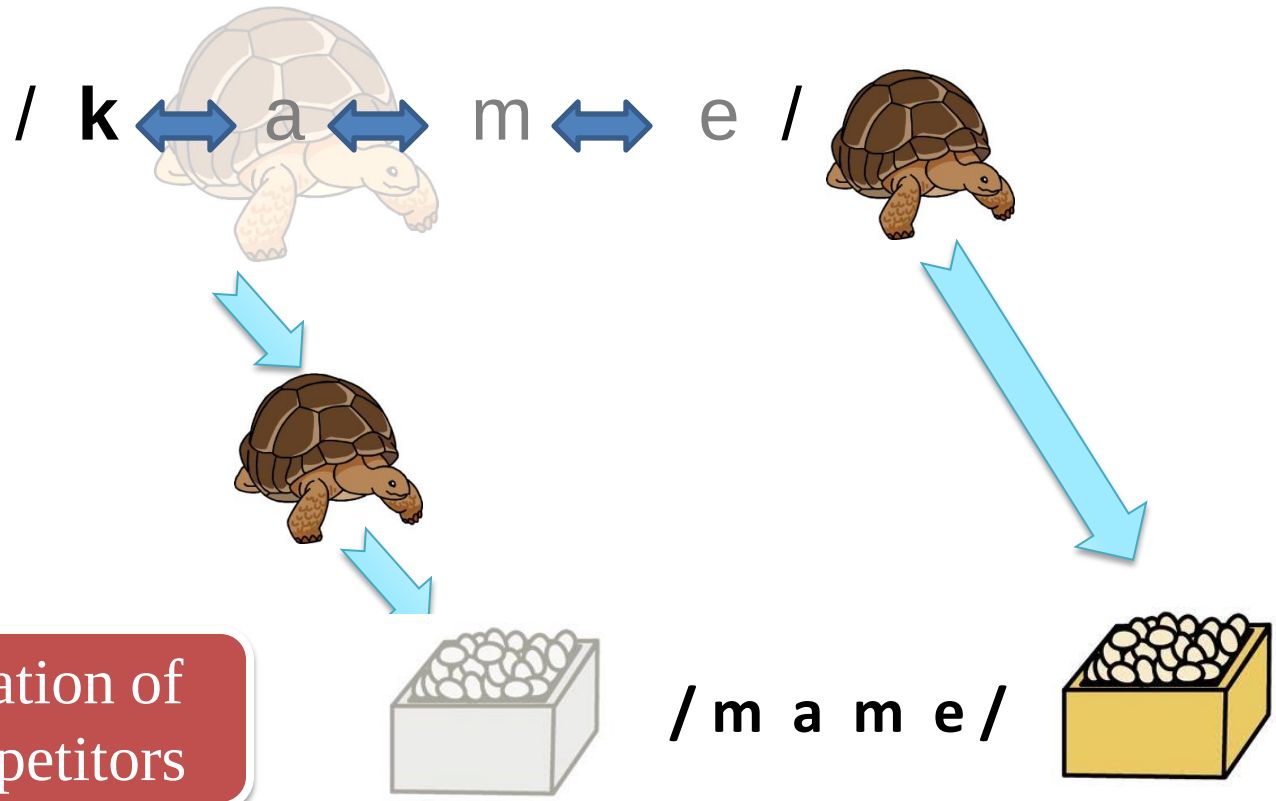
Incremental models

Coarticulatory effect

(Dahan et al., 2001)

- Phonetic cues

  → Guess the end from the beginning



Early activation of rhyme competitors

Purpose of our study

To investigate:

- how the Japanese CVCV words are processed in spoken word recognition

  Most of the previous studies examined English or Dutch closed syllables

- the influence of the amount of overlap between the competitors and the target

  None (?) manipulated the amount of overlap between the competitors & the target

Our study

Participant

- 10 native Japanese speakers

Language

- Japanese

Materials

- Audio words --- Japanese CVCV words
- Pictures --- color pictures (a set of 4 picture)
- 16 experimental sets & 16 filler sets

Our study (Contd.)

Target Conditions

1) Sharing the first syllable (OO_ _ vs. _ _OO)

- **tsume**---**tsu**ki---sa**me**---koppu
  'fingernails' 'moon' 'shark' 'glass'

(Filler)

koppu--**tsume**---**tsu**ki---sa**me**

2) Sharing the first segment (O_ _ _ vs. _OOO)

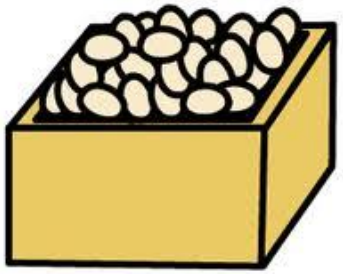
- **kame**---**k**uchi---m**ame**--tora
  'turtle' 'mouth' 'bean' 'tiger'

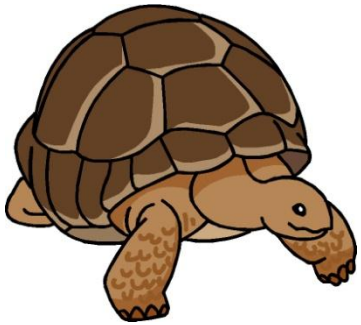
(Filler)

tora --**kame**---**k**uchi---m**ame**

Example Stimulus (O_ _ _ vs. _ OOO)

mame
(rhyme)





kame
(target)

tora
(unrelated)

kuchi
(cohort)

Analysis

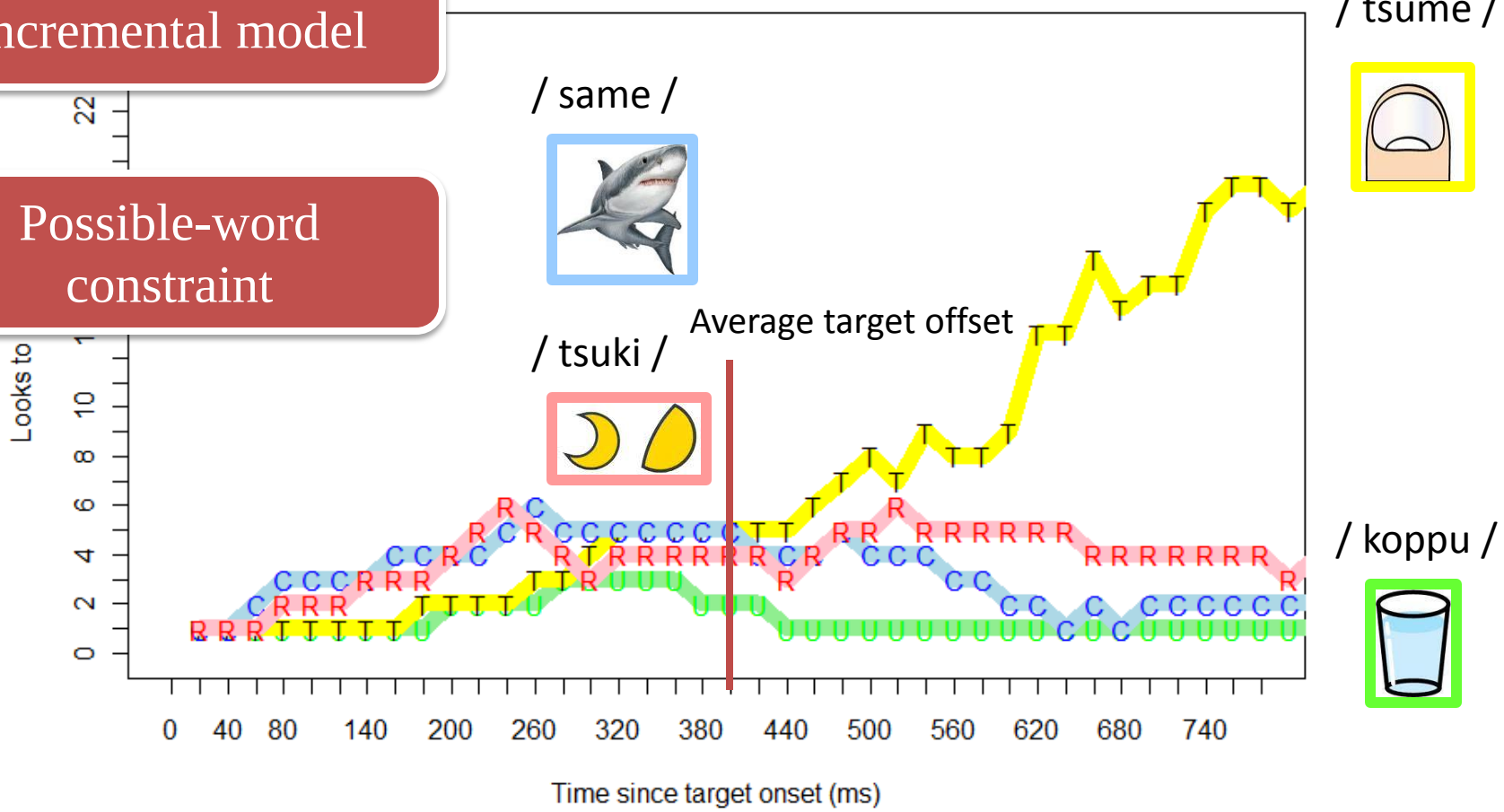
- Measurements
  - →where the participants fixated in every millisecond
  - →Averaged within-subjects in 20 ms increments
- A growth curve analysis
  - → (linear mixed effects model with time, time^2 and time^3 and type of competitor or amount of overlap as fixed effects and subject as random effect) (as suggested by Magnuson et al. 2008 or Mirman et al. 2008)

Results (Condition 1)

Sharing the first syllable (OO_ _ vs. _ _OO)

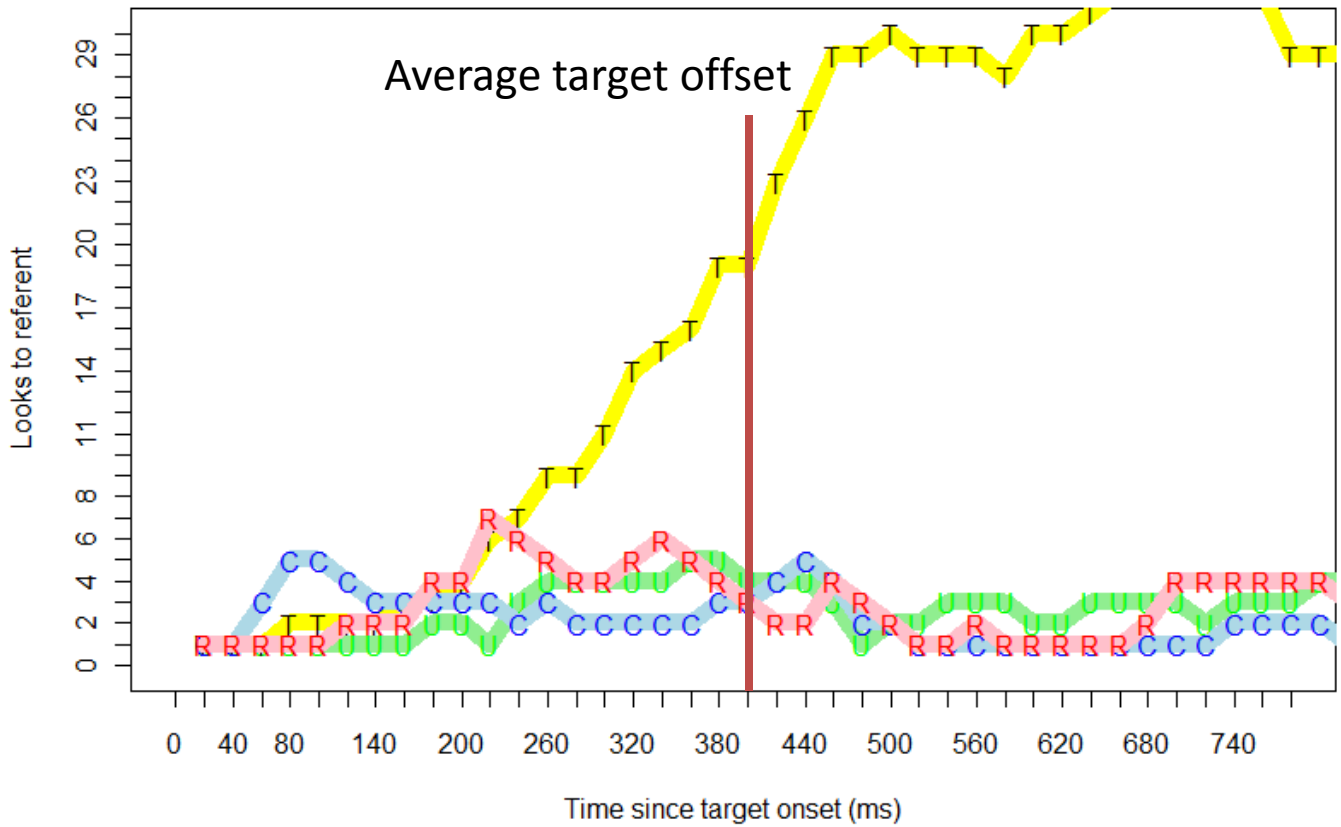
Incremental model

Possible-word constraint



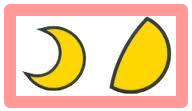
Results (Condition 1)

Sharing the first syllable (OO_ _ vs. _ _OO)



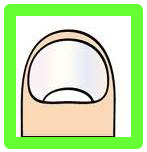
/ koppu /

/ tsuki /



/ same /

/ tsume /

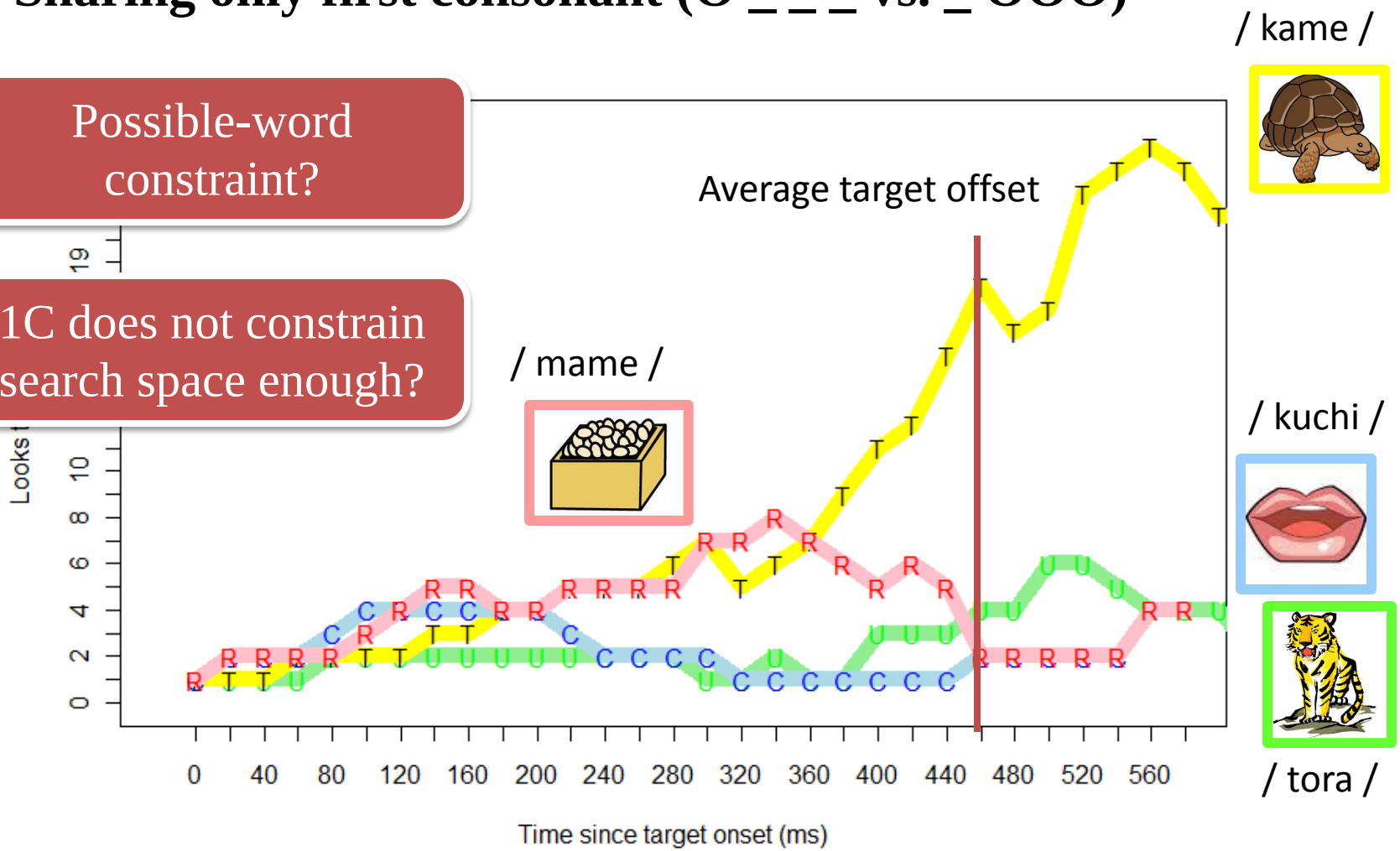


Results (Condition 2)

Sharing only first consonant (O _ _ _ vs. _ OOO)

Possible-word constraint?

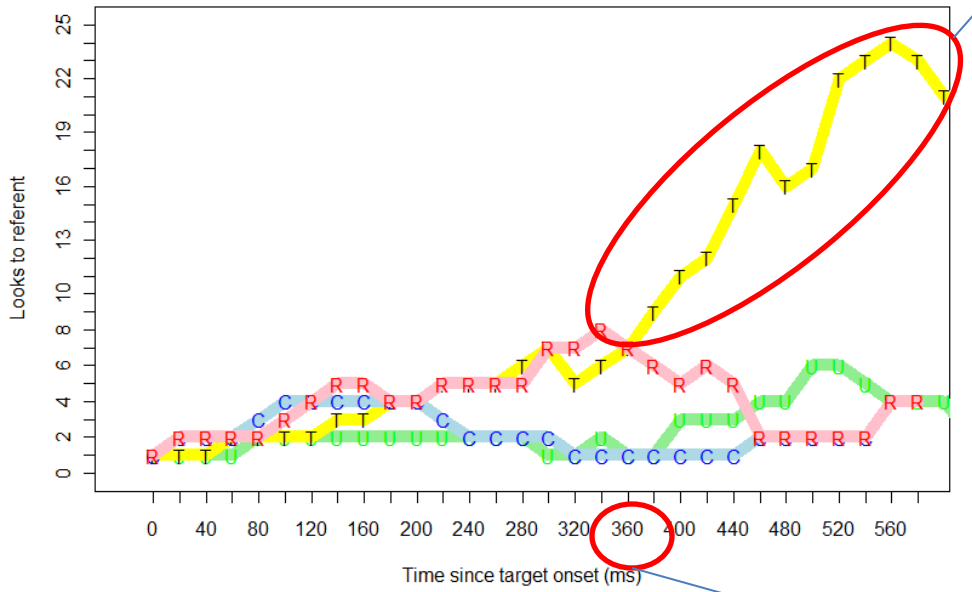
1C does not constrain search space enough?

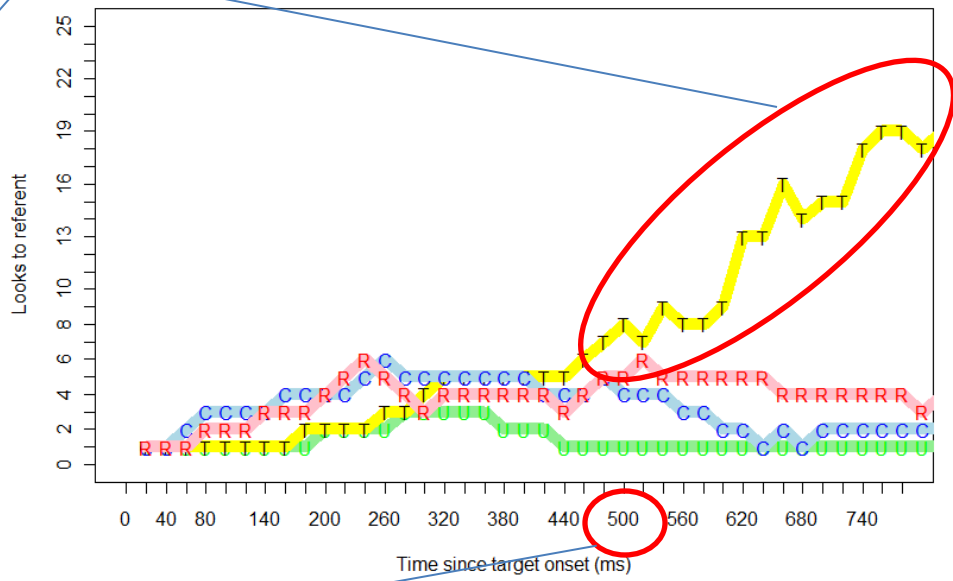


Results (Comparison)

Shared only first consonant **Shared the first syllable**

Looks to target are marginally reduced





Looks to target are marginally delayed

General discussion

Language specificity

- It is unclear whether the result would also hold in English, as previous studies did not manipulate degree of overlap

Cohort effect

- No effect of one consonant sharing may be derived from amount of information, not length of sharing
- Need enough confidence that X could be the referent to saccade to X
- The effect of syllable sharing is consistent with the possible-word constraint

General discussion (Contd.)

Strong and early rhyme effect

- Pitch information could be a cue to trigger rhyme competitor
- Coarticulation could also be a cue, suggesting the incremental models of word processing
- Japanese are in favor of rhyme information??

Conclusions

- Words are processed incrementally but the acoustics-so-far need to constrain the search space enough to drive a saccade.
  - Eye movements are promiscuous (Tanenhaus et al. 2000) but not completely...
- The results demonstrated a great degree of influence of rhyme overlap. This may be due to anticipatory coarticulation.

Conclusions (Contd.)

- Degree of overlap between the referents and the target had an effect on spoken word recognition which is problematic for one-phoneme deletion-addition-substitution neighborhood definition (Kapatsinski 2007, Felty & Buchwald 2008)
- Visual world studies (along with gating) can be used to empirically determine which words compete for recognition, moving us towards a grounded definition of “neighbor”.

Future work

- More participants
- Controlling pitch accent within a set of picture words
- More conditions needed (i.e., OOO_ vs. _ _ _O) in order to fully cross degree of overlap and location of mismatch
- Informativeness differences controlling for overlap duration, e.g., #k... vs. #r... (Kapatsinski 2006)
- Consider the influence of acoustic prominence of initial consonant
- English

Selected references

Allopenna, P. D, Magnuson, J. S. & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory & Language, 38, 419-439.*

Baayen, R. H. (2008). Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge University Press.

Dahan, D, Magnuson, J. S., Tanenhaus, M. K. & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16 (5/6), 507–534.*

Felty, R. & Buchwald, A. (2008). Neighbors as competitors: Lexical analyses of spoken word recognition errors. Poster presented at the 155th Meeting of the Acoustical Society of America, Paris, France, June 29 – July 4, 2008.

Kapatsinski, V. (2006). Having something common in common is not the same as sharing something special: Evidence from sound similarity judgments. LSA Annual Meeting, Albuquerque, NM.

Kapatsinski, V. (2007). Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech Research Lab Progress Report #27*, 133-152. Indiana University, Speech Research Lab.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25, 71-102.*

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18, 1-86.*

Mirman, D., Dixon, J. A. & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory & Language, 59, 475-494.*

Norris, D., McQueen, J. M., Cutler, A. & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34, 191-243.*

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557-580.

Thank you!

Hideko Teruya (**teruya**@uoregon.edu)

Vsevolod Kapatsinski (**vkapatsi**@uoregon.edu)

University of Oregon