# Remembering prosody in discourse: Verbatim memory and regeneration

Jennifer Bellik & Tom Roberts
University of California, Santa Cruz
CUNY 2019, CU Boulder | March 30, 2019

# What was the title of this talk?

# Verbatim vs. gist memory

In memory tasks:

- Memory for **verbatim** form of expressions ('surface features') is good at first, but decays rapidly
- Memory for **gist** (broad semantic content) is retained much longer

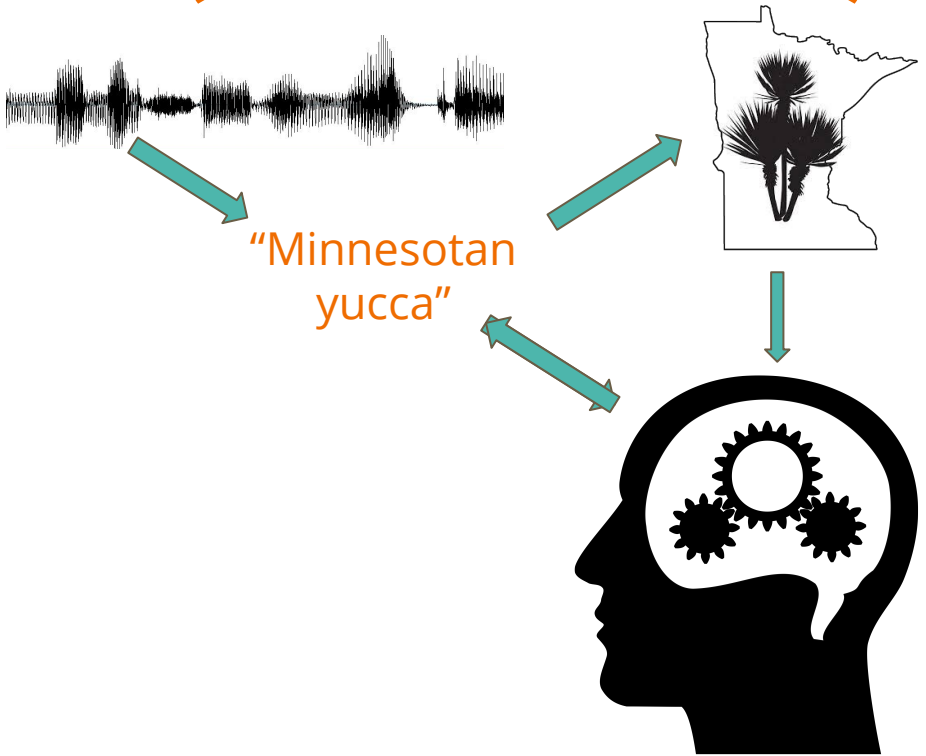
**Why is gist memory more durable?**

Sachs 1967, 1974, Jarvella 1971, Marslen-Wilson & Tyler 1976, *et seq*

# Regeneration Hypothesis (Potter & Lombardi)

Verbatim memory is illusory. People extract gist from verbatim expressions…

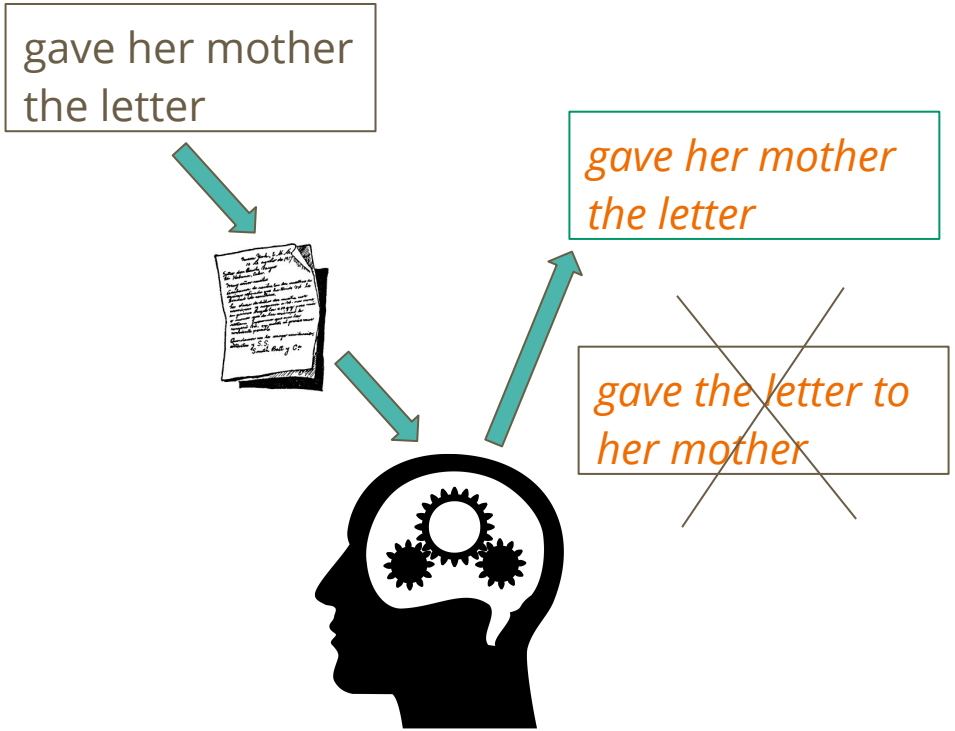
…then discard them, so only gist memory remains. Apparent memory for surface features in recall tasks is reconstructed from gist.




Bock & Brewer 1974, Johnson-Laird et al. 1974, Potter & Lombardi 1990, 1998, Holtgraves 2008

# Problems for Regeneration

- In P&L's recall tasks, participants reproduced surface syntax even when 'meaning-equivalent' alternatives were available
- P&L attribute this to syntactic priming, not verbatim memory.




Potter & Lombardi 1990, 1998, Lombardi & Potter 1992

# Problems for Regeneration

- Other research suggests a need for both verbatim and gist memory (e.g., Fuzzy Trace Theory)
- Previously encountered fine-grained *phonetic* detail impacts later processing (cf. Exemplar Theory)
- Many manipulations tested (e.g. ditransitive/double object alternation, active/passive alternation) plausibly differ in information structure

**Is there memory for truly surface features?**

**If so, what role does regeneration play in memory?**

**Fuzzy Trace Theory**: Reyna & Brainerd 1995
**Improving verbatim memory**: Johnson-Laird & Stevenson 1970, Anderson & Bower 1973, Gernsbacher 1985, Murphy & Shapiro 1994, Gurevich et al 2010
**Phonetic detail:** Sumner et al. 2014, Kimball & Cole 2016, Mitterer & Reinisch 2017, *i.a.*

# Research questions

How does memory for prosody work in **spoken** English discourse?

**Method:** Recognition tasks with auditory presentation

Is there verbatim memory for surface (non-contentful) prosody?

**Experiment 1:** Rhythm Rule

How does prosody impact gist memory?

**Experiments 2 & 3:** Regenerating questions from Focal Pitch Accent

# Experiment 1: Rhythm Rule

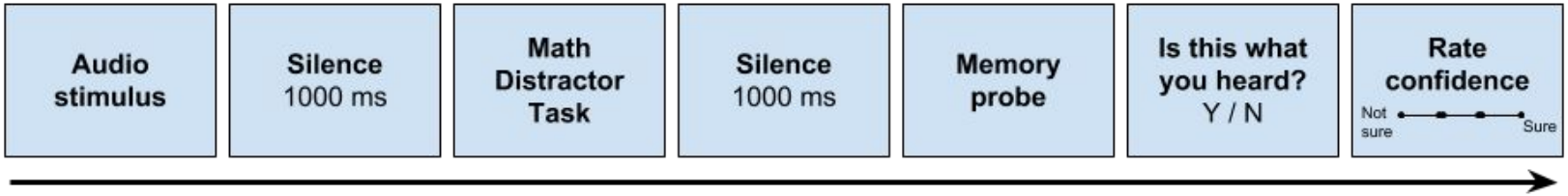
- Is there verbatim memory for features with no interpretive impact?
- Such a surface prosodic feature: Rhythm Rule (RR)
- RR: Move main stress leftward to maintain ‘strong’ beats at regular intervals. (***Optional***.)
- In an XP consisting of two words, the RR can be triggered if the second word has initial primary stress and the first has final or penultimate primary stress

*ThirTEEN MEN → THIRteen MEN*

*hypoDERmic NEEdle → HYpodermic NEEdle*

Liberman & Prince 1977, Hayes 1984, *et seq*

# Procedure: Recall Task




**Stimulus** Listen to the dialogue.

**Distractor task** 15 - 9 = ?

**Probe** Did you hear this?

**Confidence rating** How confident are you of your response? (1-4 scale)

# Exp. 1 Materials

Items (24) = 3-6 line dialogues between speakers A & B.

- Last line (probe) was spoken by B and ended with the RR target region
- RR target region = Adj+N pair, utterance final
- New vs. Old probes differed *only* in the last two words (spliced in)

Fillers (48):

- Counterbalanced position of probe in dialogue (2-4 sentences back)
- whether probe was spoken by A or B
- Manipulated both local prosody and sentence-level tunes, but new probes were still string-identical

# Exp. 1 Sample item

**Context**

B: How is your landscaping project going?
A: I'm having trouble finding good plants for this climate.
B: There are actually a lot of shrubs that can survive cold weather.
A: What's an example?

**Target**

B: Well, there's the { MINnesotan YUCca [+RR] / MinneSOtan YUCca [-RR] }

# Exp. 1 Method

2x2 design: [**±RR**] x [**Probe New/Old**]

Participants (n=36):

- UCSC undergraduates
- 3 excluded for not learning English before age 5
- 33 total analyzed

# Exp. 1 Predictions

If surface stress **is** stored verbatim in memory:

Participants will be sensitive to changes in application of the rhythm rule.

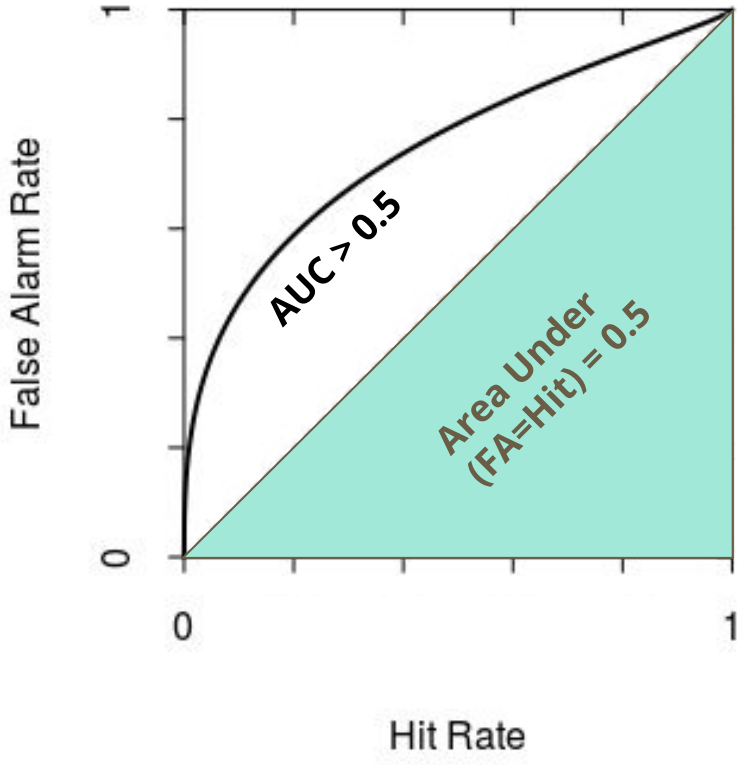
If surface stress **is not** stored verbatim in memory:

Participants will be insensitive to these changes.

If +/-RR application is in free variation:

No effect of ±RR.

# Analytical Method: Detection Theory




**AUC** = measure of sensitivity

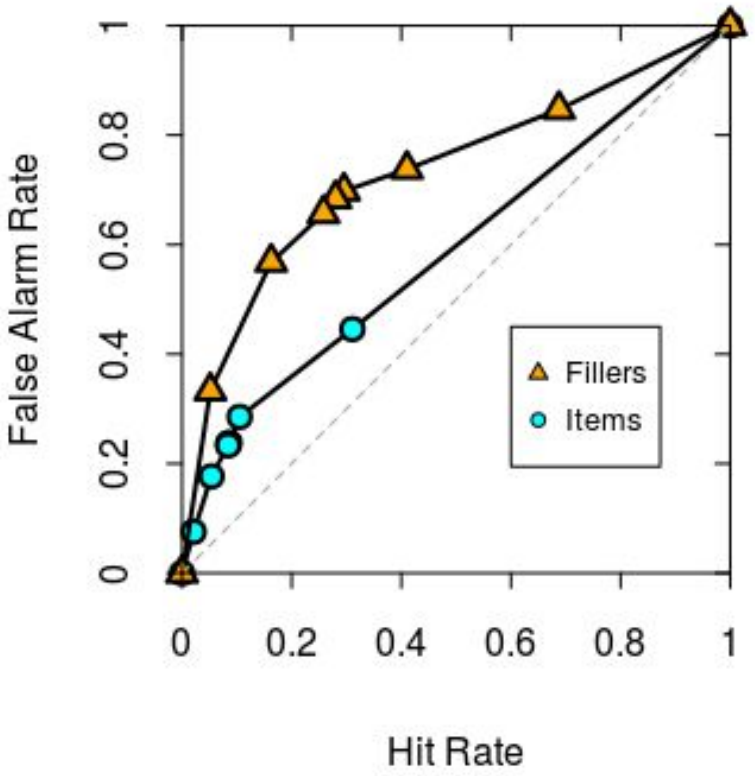
- Area under the Receiver Operating Characteristic curve
- ROC calculated using the Ratings Method
- AUC > 0.5 = sensitivity

***c*** = criterion = measure of response bias

- Negative *c* → 'yes' bias

Macmillan & Creelman 2005

# Exp. 1 Results: New vs. Old Probes




| | Experimental Items | Fillers |
|---|---|---|
| AUC (95% CI) | 0.590 (0.557-0.623) | 0.733 (0.708-0.758) |
| c | -1.05 | 0.1 |

**AUC > 0.5 : Participants can discriminate between new and old probes**

**c < 0: Bias to respond "yes, I heard this before"**

MacMillan & Creelman 2005, Robin et al. 2011

# Exp. 1 Discussion

- Although the deck was stacked against participants, they were significantly (albeit weakly) sensitive to changes in RR application.
- No effect of whether initial presentation was +RR or -RR: Sensitivity really reflects access to a verbatim representation, not regeneration

**Upshot**: Non-contentful surface prosody can be remembered, with some difficulty.

**Next question**: Can prosody also have an impact on gist memory?

# Experiments 2 & 3: Focal Pitch Accent

Prosody prototypically conveys “higher” discourse information.

A prosodic feature *with* interpretive impact: Focal Pitch Accent (FPA)

- FPA indicates that a constituent is interpreted under exhaustive focus.
- Felicitous answers must put FPA only on the constituent corresponding to the question’s wh-word. (ex. *Gertrude* in SA, *kale* in S2)

| | |
|---|---|
| SQ: Who massaged the kale? | OQ: What did Gertrude massage? |
| SA: **[$_F$GERTRUDE]** massaged the kale. | OA: Gertrude massaged the **[$_F$KALE]**. |

Selkirk, 1995

# Experiments 2 & 3: Focal Pitch Accent

- In case of a mismatch between answer FPA and question wh-word, will questions be reconstructed to match answer FPA?
- Experiment 2: Memory for interrogatives
- Experiment 3: Prosodic mismatch + regeneration

| | |
|---|---|
| Hear: | Q1: Who massaged the kale?<br>A2: Gertrude massaged the **[$_F$KALE]**. |
| Regenerate: | Q2: What did Gertrude massage? |

Selkirk 1995

# Experiment 2: Interrogatives vs. Declaratives

- Previous work focuses on “gist” memory for declarative sentences.
- But what is the “gist” of a question?
- Prevailing theoretical accounts of question meaning treat it as a set of propositions: i.e., a set of declarative denotations.
    - Translates to penalty for remembering interrogatives as opposed to declaratives?

Hamblin 1973, Karttunen 1977, *et seq*.

# Exp. 2 Method

Auditory recognition task, same procedure as Exp. 1

2x2 design: **[Interrogative / Declarative]** x **[Probe New/Old]**

Participants (n=55):

- UCSC undergraduates who learned English by age 5

# Exp. 2 Materials

Items (24) consisted of 5-8 line conversations between A & B

- Target = penultimate line; Interrogative or Declarative
- Last line = felicitous response; same in both conditions

Fillers (48) counterbalanced

- Position of the probe appeared in the dialogue (1, 3, or 5 sentences back)
- Whether probe was spoken by A or B

| | **Target + felicitous response** | **Old probe** | **New probe** |
|---|---|---|---|
| **Interrogative** | A: Who massaged the kale?<br>B: [Gertrude]$_F$ massaged the kale. | Who massaged the kale? | What did Gertrude massage? |
| **Declarative** | A: I suppose Linda massaged the kale.<br>B: [Gertrude]$_F$ massaged the kale. | I suppose Linda massaged the kale. | I suppose Gertrude massaged the mustard greens. |

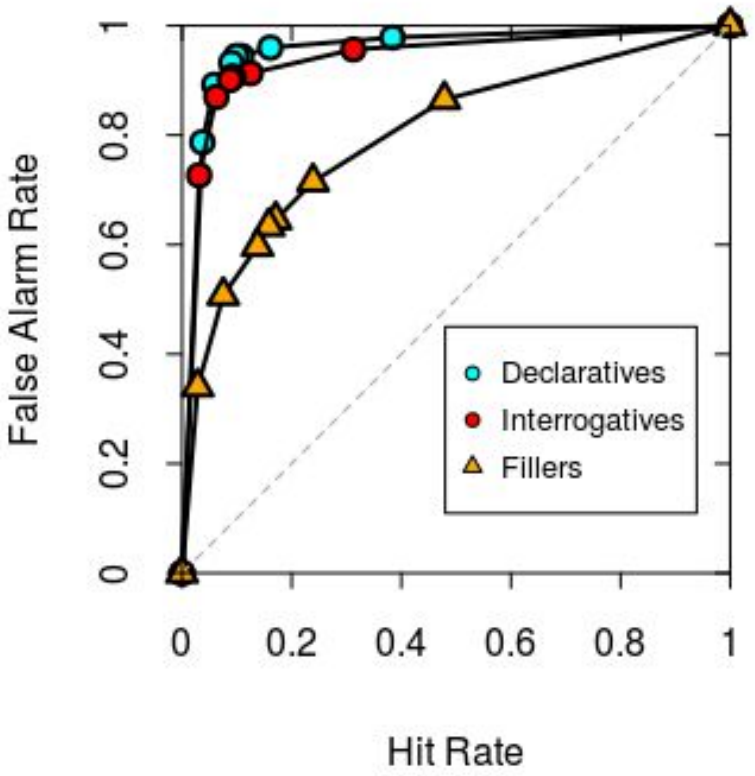
# Exp. 2 Predictions

If gist is related to **propositional content**…

> Changed interrogatives will be harder to remember than changed declaratives, because interrogative meaning is strictly more complex.

Both declarative and interrogative probe changes involve changes in lexical items, so there are strong cues to recognition of changes in both.

# Exp. 2 Results




| | Declarative | Interrogative | Fillers |
|---|---|---|---|
| AUC (95% CI) | 0.947 (0.937-0.970) | 0.940 (0.929-0.959) | 0.802 (0.785-0.818) |
| c | 0.16 | 0.00 | -0.33 |

**Recall of declaratives and interrogatives both extremely good and not significantly different.**

→ Established baseline for Exp. 3

# Experiment 3 (pilot): Regeneration of questions

**Experiment 1**: Stress is remembered independently of interpretative impact.

**Experiment 2**: Recognition memory for interrogatives is very good.

**Experiment 3**: Will questions be regenerated based on answer prosody?

Test case: Q-A focus mismatch.

| | |
|---|---|
| Q1: Who massaged the kale? | Q2: What did Gertrude massage? |
| A1: **[$_F$GERTRUDE]** massaged the kale. | A2: Gertrude massaged the **[$_F$KALE]**. |

# Exp. 3 Method

2x2 design: [**Felicitous/Infelicitous**] x [**Probe New/Old**]

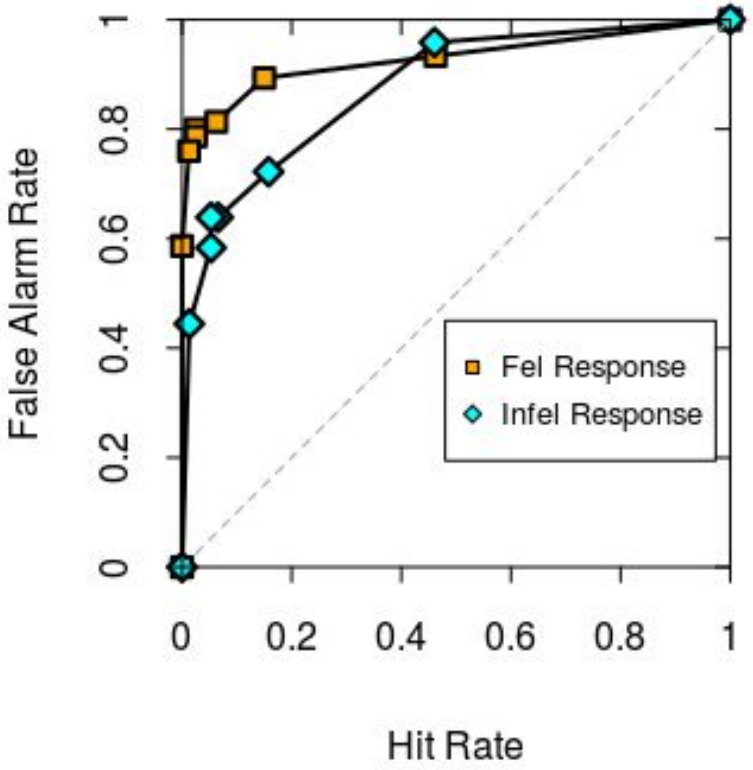
Participants (n=16): UCSC graduate and undergraduate students

Items (n=40):

- Question-Answer pairs produced by 2 speakers
  - Questions involve wh-extraction of the subject or object.
  - Answers put FPA on the subject or object.
- Presented via Latin square without fillers.

Procedure differed from Expt. 1 & 2 in using a distractor melody (no math problem).

# Exp. 3 Results (pilot) - Question Probes




| | Felicitous Response | Infelicitous Response |
|---|---|---|
| AUC (95% CI) | 0.928 (0.880-0.969) | 0.890 (0.836-0.938)) |
| c | -0.47 | -0.36 |

**Sensitivity to probe newness was better for questions followed by felicitous answers (though not by much).**

# Exp. 2 & 3 Discussion

- Exp 2 suggests that interrogatives are easily recognized when their follow-up response is felicitous.
  - Possible that lexical/syntactic changes are simply an extremely strong cue.
  - A recall task might show less of a ceiling effect.
- Exp 3 suggests that infelicitous prosody on a response may interfere with recall of a question.
- This suggests that prosodic information (at least FPA) may also be used to facilitate memory of prior discourse moves.

# Conclusion

Prosody appears to play a role in both verbatim and gist memory:

★ Surface (prosodic) features of language in spoken dialogues are stored in memory, not just gist + regeneration
★ However, incongruous answer prosody appears to be able to trigger erroneous recall of questions, suggesting that remembered prosody may impact gist memory.

Future work:

- Comparison of reconstruction of interrogatives vs. declaratives
- Effects on memory of highly interpretative semantic content, e.g. intonational tunes

# Thank you!

Special thanks to Steven Foley, Jed Pizarro-Guevara, Amanda Rysling, Matt Wagers, audiences at UCSC S/Lab and CAMP 2, and RAs Madeleine King and Koy Rugama.

Contact: {jbellik,rotom}@ucsc.edu

# References


- Anderson, J. R. & G. Bower. 1973. A propositional theory of recognition memory. *Memory and Cognition* 2(3): 406–412.
- Bates, E., M. Masling, & W. Kintsch. 1978. Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory, 4*(3): 187-197.
- Bock, K. & W. F. Brewer. 1974. Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology* 103(5): 837–843.
- Gernsbacher, M. A. 1985. Surface information loss in comprehension. *Cognitive Psychology* 17: 324–363.
- Grabe, E. & Warren, P. 1995. Stress Shift: Do Speakers Do It or Do Listeners Hear It? In B. Connell and A. Arvaniti (eds.), *Papers in Laboratory Phonology IV.* Cambridge University Press.
- Gurevich, O., M. A. Johnson & A. E. Goldberg. 2010. Incidental Verbatim Memory for Language. *Language and Cognition* 2(1): 45-78.
- Hayes, B. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15: 33-74.
- Holtgraves, T. 2008. Conversation, speech acts, and memory. *Memory and Cognition* 36(2): 361–374.
- Jarvella, R. J. 1971. Syntactic Processing of Connected Speech. *Journal of Verbal Learning & Verbal Behavior* 10(4): 409-416.
- Johnson-Laird, P., C. Robins & L. Velicogna. 1974. Memory for words. *Nature* 251: 704–705.
- Johnson-Laird, P. & R. Stevenson. 1970. Memory for syntax. *Nature* 227: 412.

# References


- Kimball, A. E. & J. Cole. 2014. Avoidance of Stress Clash in Perception of Conversational American English. *Speech Prosody 7*, Dublin.
- Kimball, A. & J. Cole. 2016. Pitch contour shape matters in memory. *Proceedings of Speech Prosody 8,* 1171-1175.
- Liberman, M. & A. Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2): 249-336.
- Lombardi, L. & M. C. Potter. 1992. The regeneration of syntax in short term memory. *Journal of Memory and Language* 31: 713–733.
- MacMillan, N. A. & C. D. Creelman. *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marslen-Wilson, W., & L. K. Tyler. 1976. Memory and levels of processing in a psycholinguistic context. *Journal of Experimental Psychology: Human Learning and Memory* 2(2): 112-119.
- Mitterer, H., & E. Reinisch. 2017. Surface forms trump underlying representations in functional generalizations in speech perception: The case of german devoiced stops. *Language, Cognition, and Neuroscience* 32: 1133-1147.
- Murphy, G. & A. M. Shapiro. 1994. Forgetting of verbatim information in discourse. *Memory and Cognition* 22: 85–94.
- Potter, M. C. & L. Lombardi. 1990. Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29: 633–654.

# References


- Potter, M. C. & L. Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38(3): 265–282.
- Reyna, V. F. & C. J. Brainerd. 1995. Fuzzy Trace Theory: Some Foundational Issues. *Learning and Individual Differences* 7(2): 145-162.
- Sachs, J. S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2(9): 437–443.
- Sachs, J. S. 1974. Memory in reading and listening to discourse. *Memory and Cognition* 2(1): 95-100.
- Sumner, M., S. K. Kim, E. King & K. B. McGowan. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology* 4:1015.
  Vogel, I., T. Bunnell & S. Hoskins. 1995. The Phonology and Phonetics of the Rhythm Rule. In B. Connell and A. Arvaniti (eds.), *Papers in Laboratory Phonology IV.* Cambridge University Press.

# Exp. 1 Item Creation

- For target sentences, two versions were recorded: one with stress shift and one without
- The authors selected whichever version sounded more natural (15/24 originally carried the stress shift) and spliced out the final two words for items with the other stress patterns
- Double-checked that splices did not sound obvious
- Anticipation of upcoming stress might affect prosody in item production
    - **BUT**: No significant  difference between % correct for items whose carrier phrase was from +RR or -RR recordings ($\chi^2(1) = 0.026$)

# Exp. 1 Response results

| **Stress shift in stimulus [+RR]** Target = “MINnesotan YUCca” | | |
|---|---|---|
| | Did you hear this before? | |
| | **“Yes”** | **“No”** |
| **Old Probe** “MINnesotan YUCca” | 197 (Hits) | 17 (Misses) |
| **New Probe** “MinneSOtan YUCca” | 157 (False Alarms) | 55 (Correct Rejections) |

| **No stress shift in stimulus [-RR]** Target = “MinneSOtan YUCca” | | |
|---|---|---|
| | Did you hear this before? | |
| | **“Yes”** | **“No”** |
| **Old Probe** “MinneSOtan YUCca” | 195 (Hits) | 19 (Misses) |
| **New Probe** “MINnesotan YUCca” | 167 (False Alarms) | 45 (Correct Rejections) |

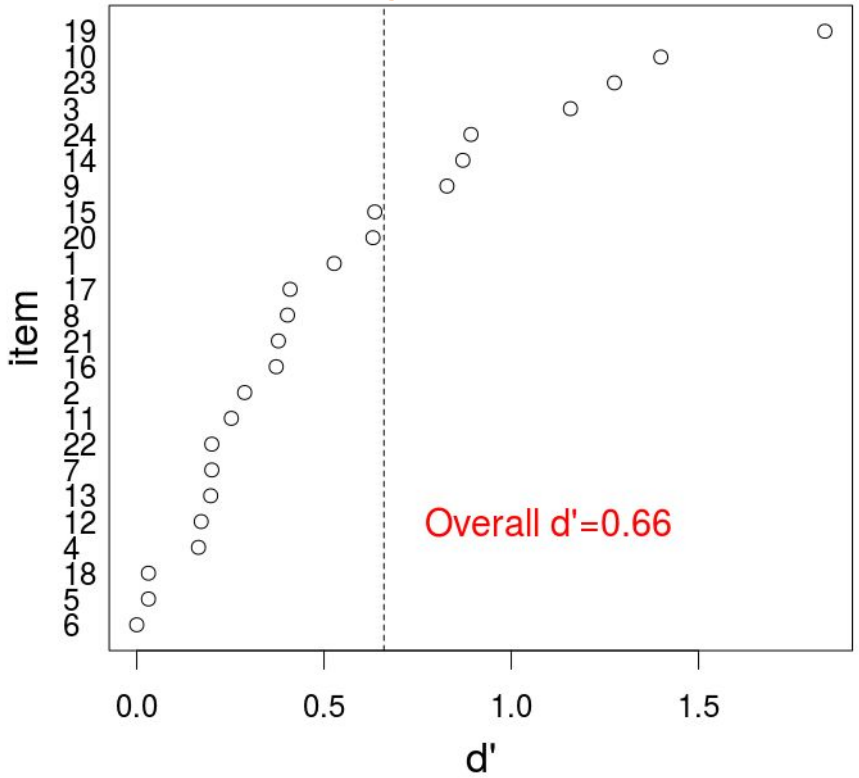
# Exp. 1 Confidence Results

- "Yes" bias:
  - Accuracy is much higher for Old items (92%) than New (24%).
  - "No" responses were more likely to be correct, but received lower confidence ratings

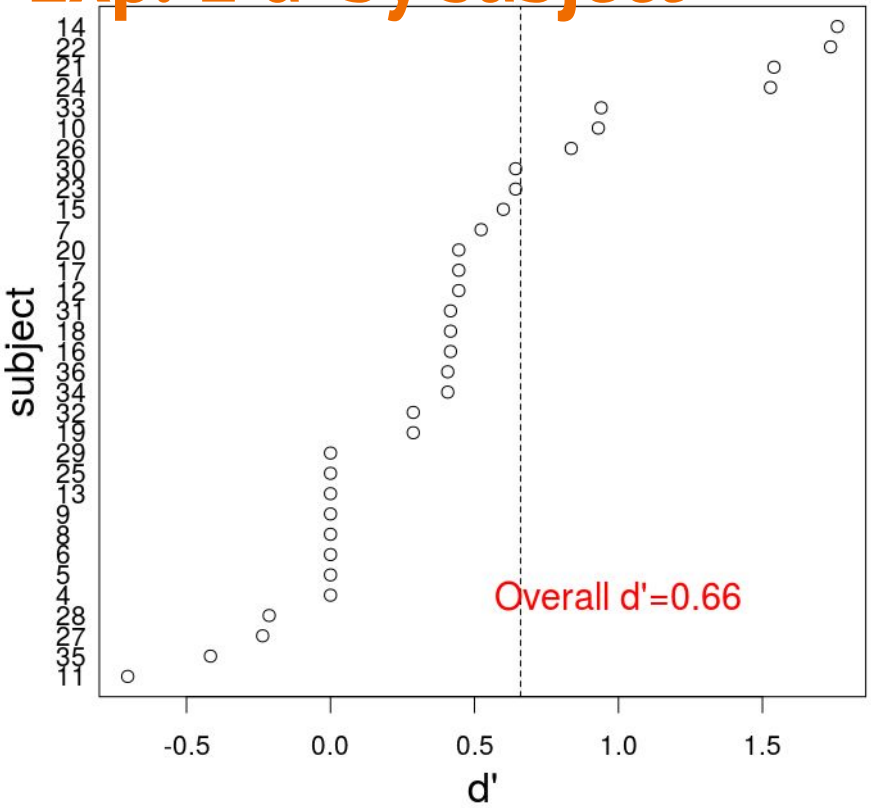
| | 1 (Not sure) | 2 | 3 | 4 (Very sure) |
|---|---|---|---|---|
| "Yes" | 2 | 28 | 156 | 530 |
| "No" | 1 | 37 | 57 | 41 |

- Overall proportion correct (57%) is not greater than chance (By items: $X^2(23)=17.9$, $p=0.76$; by subjects: $X^2(32) = 32.15$, $p=0.46$), but proportion correct doesn't capture sensitivity well with such a large bias.

# Exp. 1 d' by item



# Exp. 1 d’ by subject



# Why not just report chi-squared?

Not as useful a metric as AUC when there is bias

Chi-squared tests of proportion correct were not significant by subject or item for Exp. 1 ($p >> 0.05$), due to the huge yes-bias.

But AUC + c shows that there is significant sensitivity here.