Objective Measure of Listening Effort

June 8, 2017

Malcolm Slaney

Measuring Audio Quality

Mean Squared Error

- **Objective**
- **Not realistic (phase changes)**

Mean Opinion Score

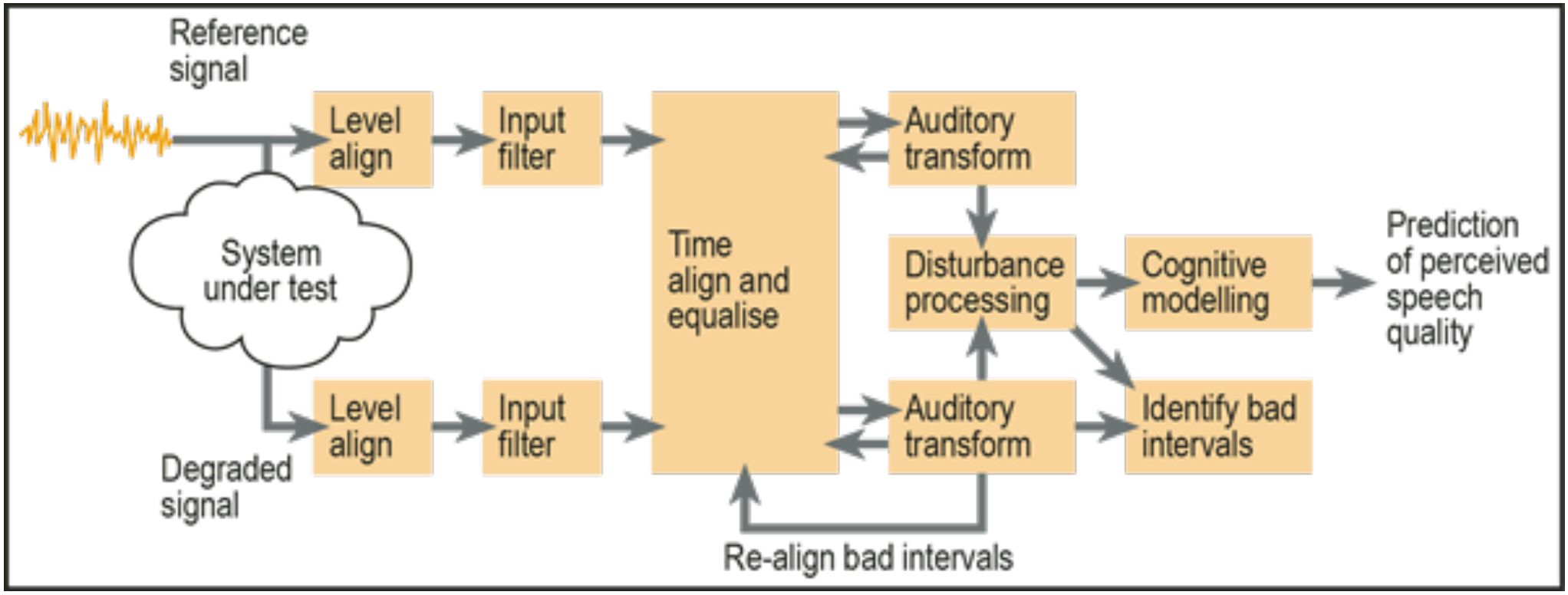
- **Uses human scorers**
- **Expensive**

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

PESQ

- **Models that approximate human listeners**

PESQ



Auditory Model (L2 in spike rates)

http://www.ap.com/wp-content/uploads/2016/05/PESQ_block_small.png

Observation

“Some listeners claim a subjective improvement from noise reduction,

yet, it has not been shown to improve speech intelligibility, often even making it worse.”

Why?

Measures of Listening Effort

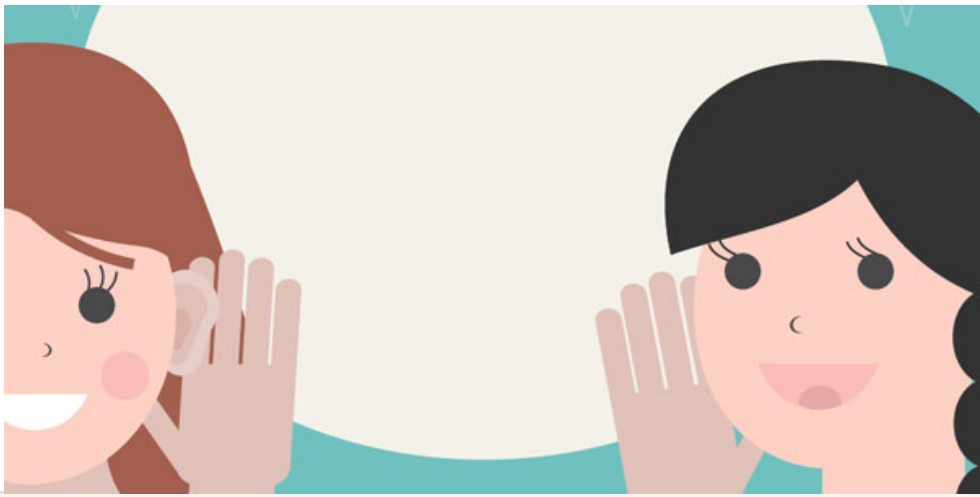
Self-reported Measures of Effort, Fatigue or Stress

Cognitive-Behavioral Measures

- Working memory ★
- Attention
- Speed of processing ★

Physiological Measures

- MEG and ERP (amplitude of P3a)
- Alpha power in EEG (higher power 8-13Hz)
- fMRI (frontal regions show higher BOLD)
- Pupil responses
  (peak size bigger under load)
- Cardiac responses
  (lower variability in rate)
- Skin conductance
  (increase indicates increasing demand)
- Hormonal Responses



http://jslhr.pubs.asha.org/article.aspx?articleid=2442495

Core Ideas

Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction

Anastasios Sarampalis
University of California at Berkeley

Sridhar Kalluri
Brent Edwards
Starkey Hearing Research Center, Berkeley, CA

Ervin Hafter
University of California at Berkeley

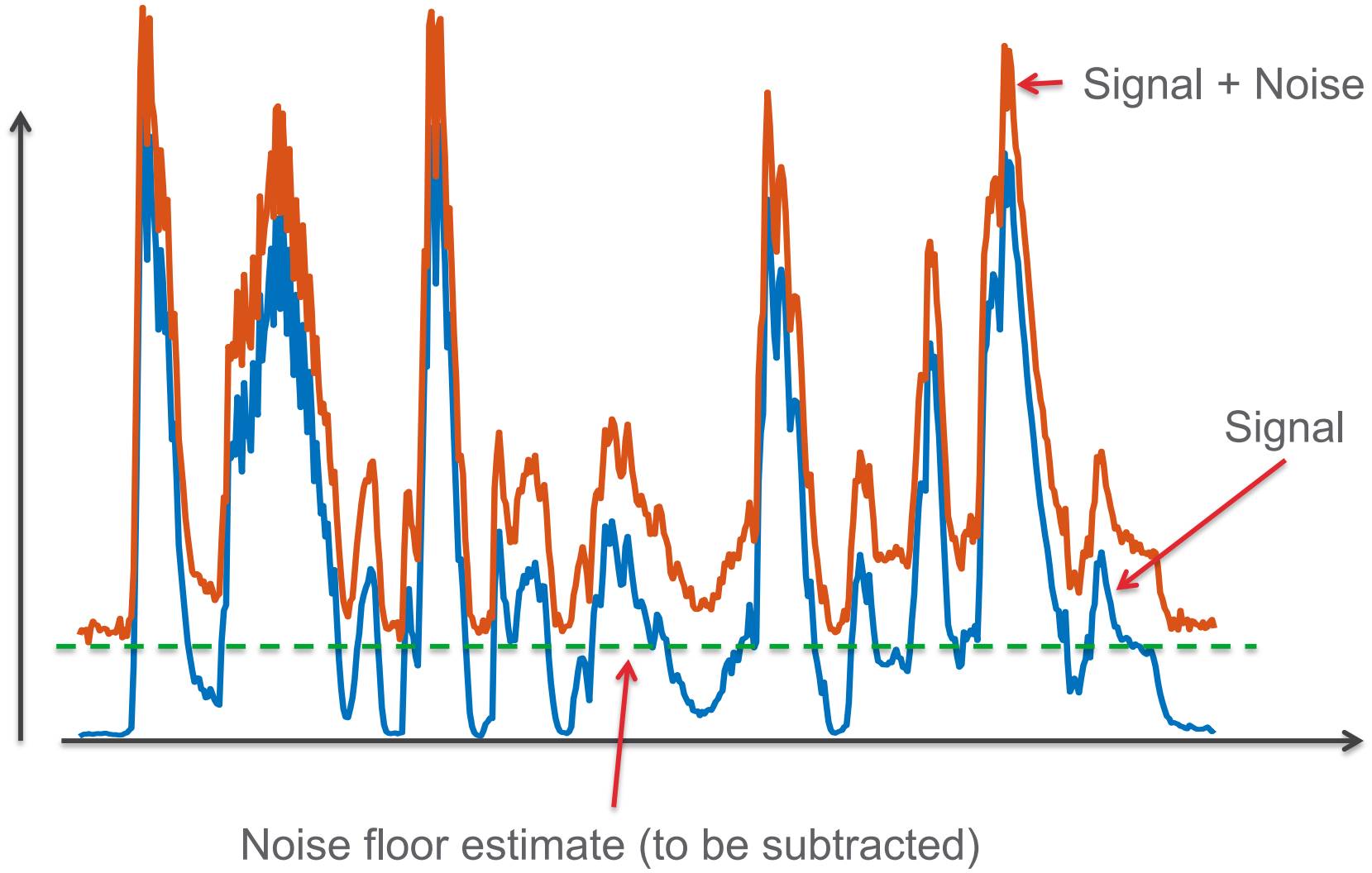
Purpose: This work is aimed at addressing a seeming contradiction related to the use of noise-reduction (NR) algorithms in hearing aids. The problem is that although some listeners claim a subjective improvement from NR, it has not been shown to improve speech intelligibility, often even making it worse.
Method: To address this, the hypothesis tested here is that the positive effects of NR might be to reduce cognitive effort directed toward speech reception, making it available for other tasks. Normal-hearing individuals participated in 2 dual-task experiments, in which 1 task was to report sentences or words in noise set to various signal-to-noise ratios. Secondary tasks involved either holding words in short-term memory or responding in a complex visual reaction-time task.
Results: At low values of signal-to-noise ratio, although NR had no positive effect on speech reception thresholds, it led to better performance on the word-memory task and quicker responses in visual reaction times.
Conclusions: Results from both dual tasks support the hypothesis that NR reduces listening effort and frees up cognitive resources for other tasks. Future hearing aid research should incorporate objective measurements of cognitive benefits.

Hearing-impaired (HI) listeners, despite understanding speech in quiet almost as well as normal-hearing (NH) listeners, have great difficulties when speech is presented in background noise (e.g., Plomp, 1994). This is true even when amplification is provided by means of a hearing aid such that the speech is within the range of audibility; this problem is a widely reported reason for hearing aid owners to stop using their devices (Kochkin, 2000). Furthermore, this difficulty becomes more pronounced as the degree of hearing loss increases (Killion, 1997).

Advances in digital hearing aid technology have allowed the widespread use of signal processing algorithms such as spectral feature enhancement, multiband compression, directional microphones, and noise reduction (NR), mainly with the aim of improving speech intelligibility, particularly in adverse listening conditions. The benefits, or lack thereof, of these algorithms on speech intelligibility are, understandably, well documented (e.g., Dillon & Lovegrove, 1993; Hickson, 1994; Levitt, Neuman, Mills, & Schwander, 1986; Ricketts, Lindley, & Henry, 2001). Objective measurements of benefits beyond those seen with speech tests, however, are not so prevalent. In particular, NR algorithms, which will be the main focus here, aim to counteract the effects of noise on speech perception and sound quality by improving the signal-to-noise ratio (SNR). These algorithms exist in many forms, but in general, they all work by adjusting

1230 *Journal of Speech, Language, and Hearing Research* • Vol. 52 • 1230–1240 • October 2009 • © American Speech-Language-Hearing Association
1092-4388/09/5205-1230

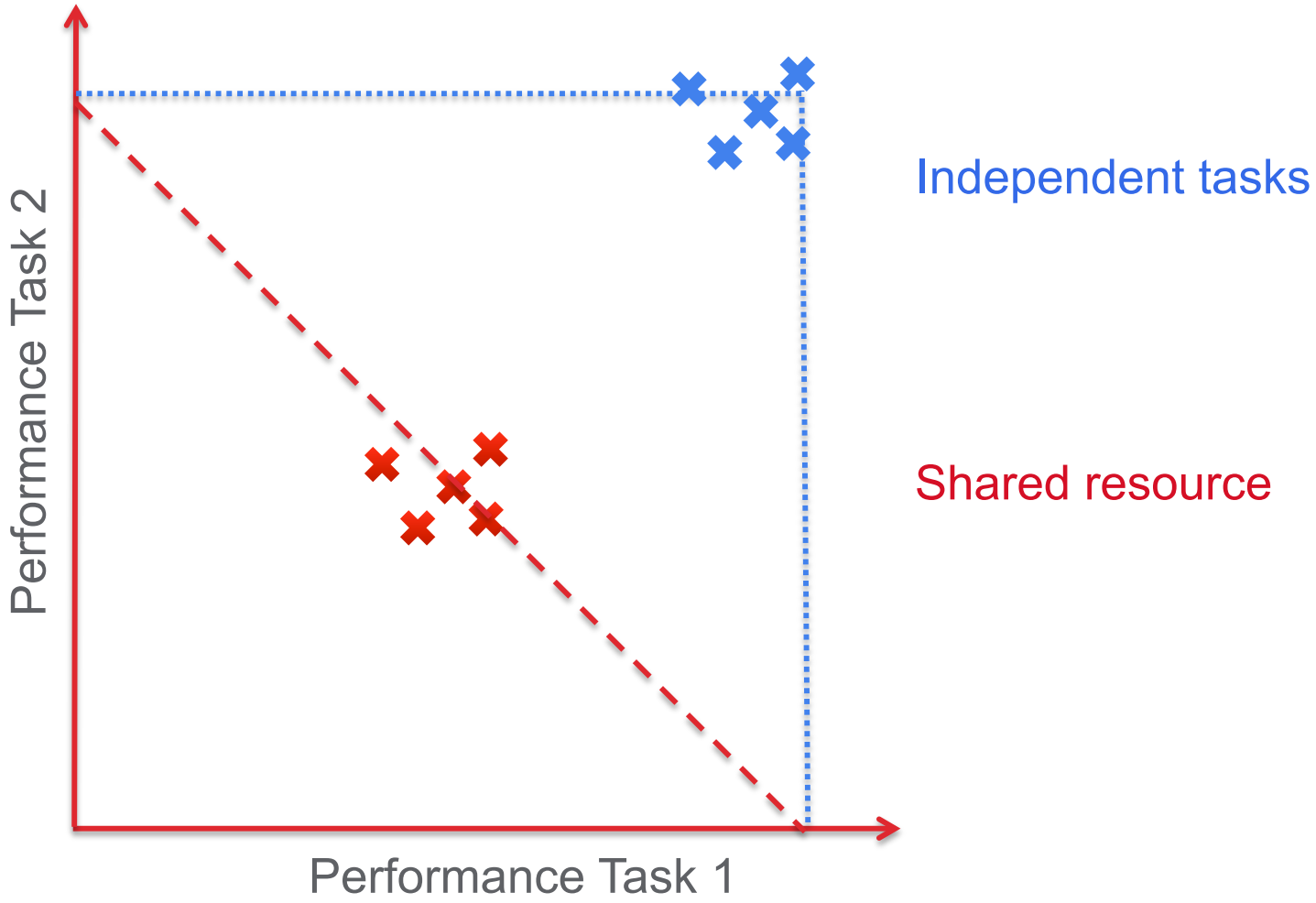
Ephraim and Malah Speech Enhancement



Issues: Hard to estimate floor, reconstruction errors

Dual Task

Evaluate performance with shared resource



Test #1 – Memory for low/high context words

Speech Perception in Noise (SPIN) sentences

- High context: “A chimpanzee is an ape.”
- Low context: “She might have discussed the ape.”

Processing

- +/- 2dB SNR (4 speaker babble)
- Ephraim-Malah NR algorithm

Primary task

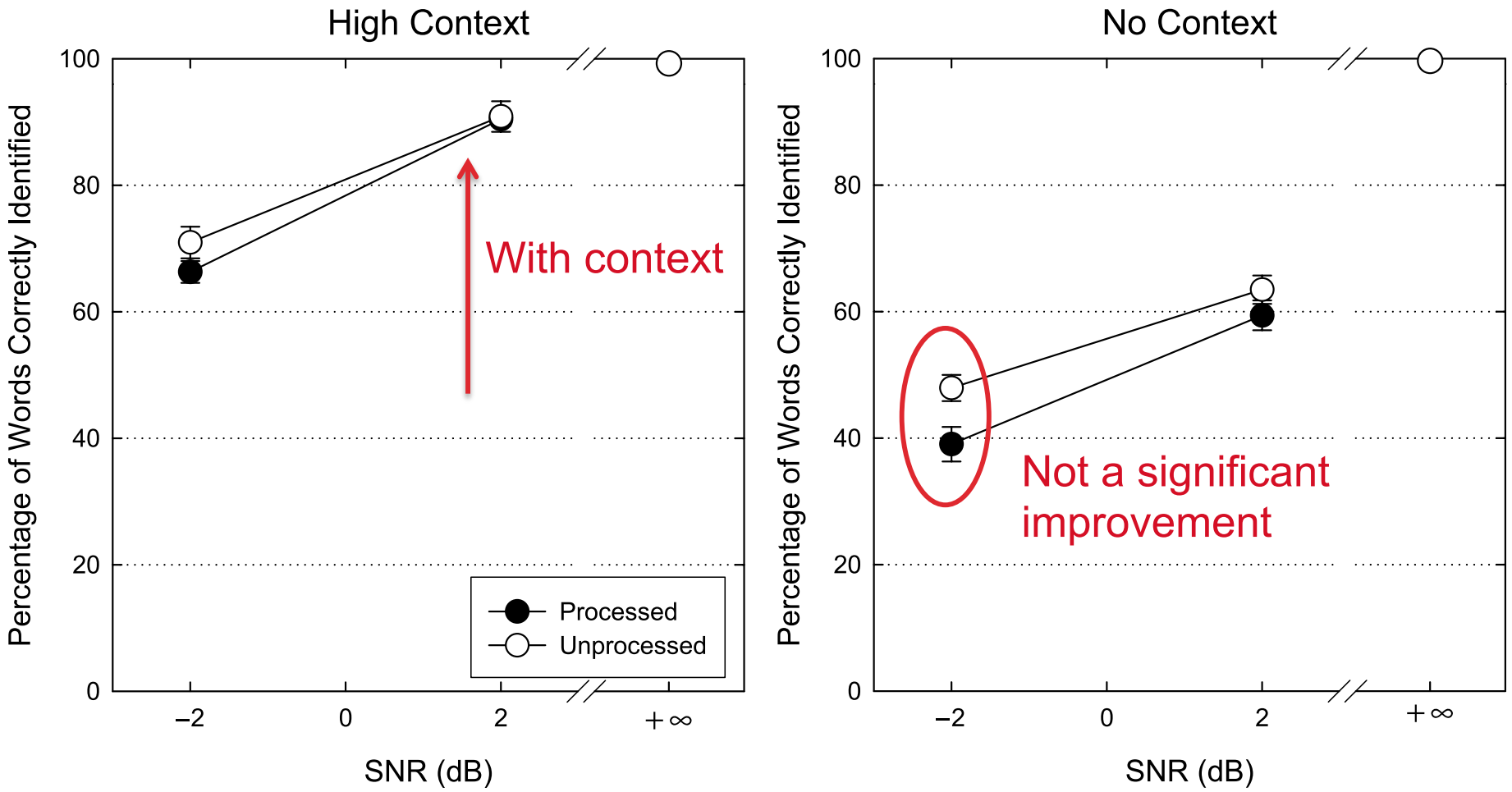
- Repeat last word of sentence

Secondary task

- Recall last 8 answers

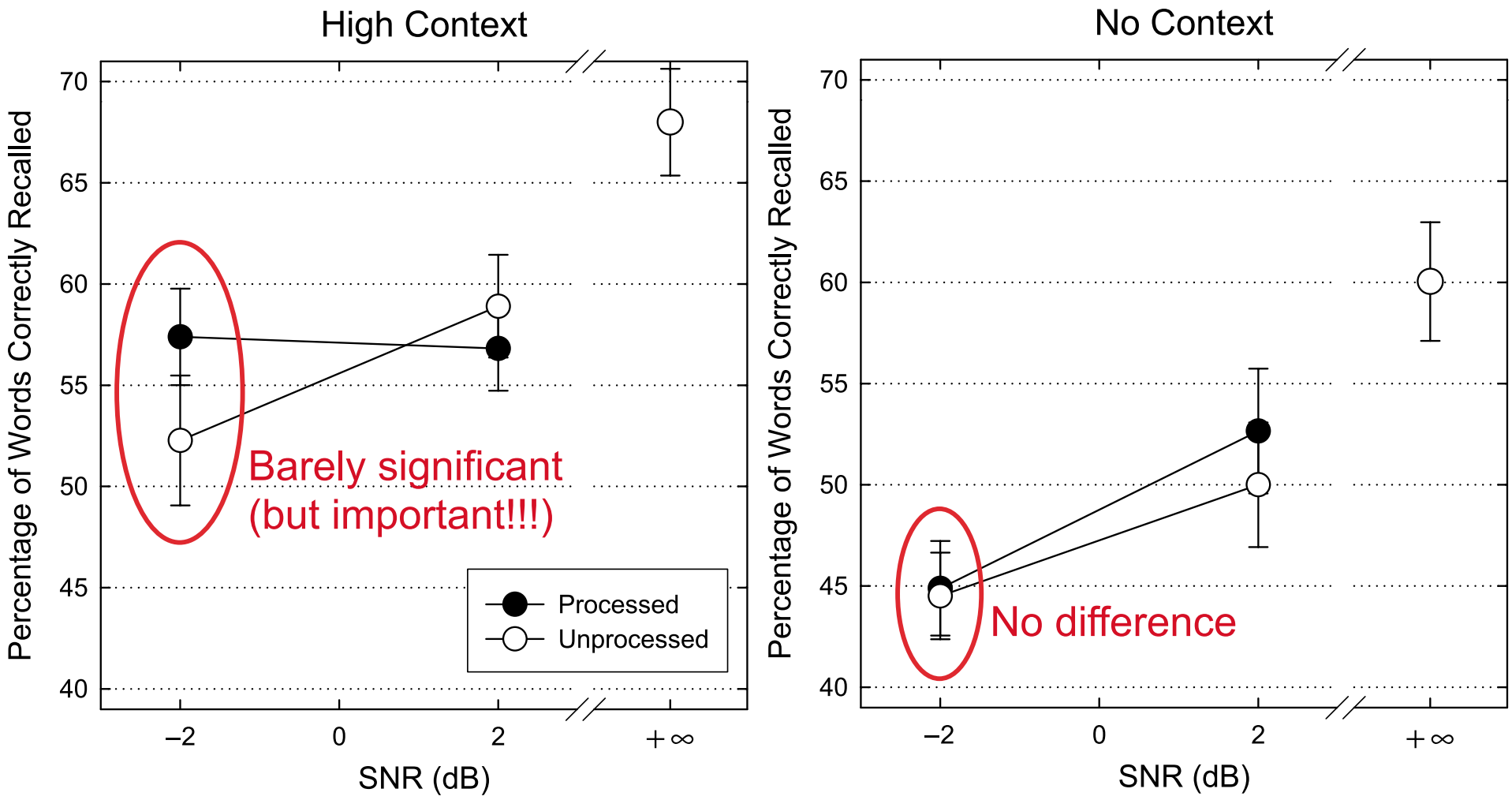
Test #1 – Primary Task - Intelligibility

Figure 1. Speech intelligibility as a function of signal-to-noise ratio (SNR), averaged across 25 listeners in Experiment 1. The left and right panels show performance for material having contextual information and for material lacking contextual information, respectively. Data with noise reduction (NR) processing are plotted with filled symbols, and those without NR processing are plotted with open symbols. The error bars denote 1 standard error of the mean (*SEM*).



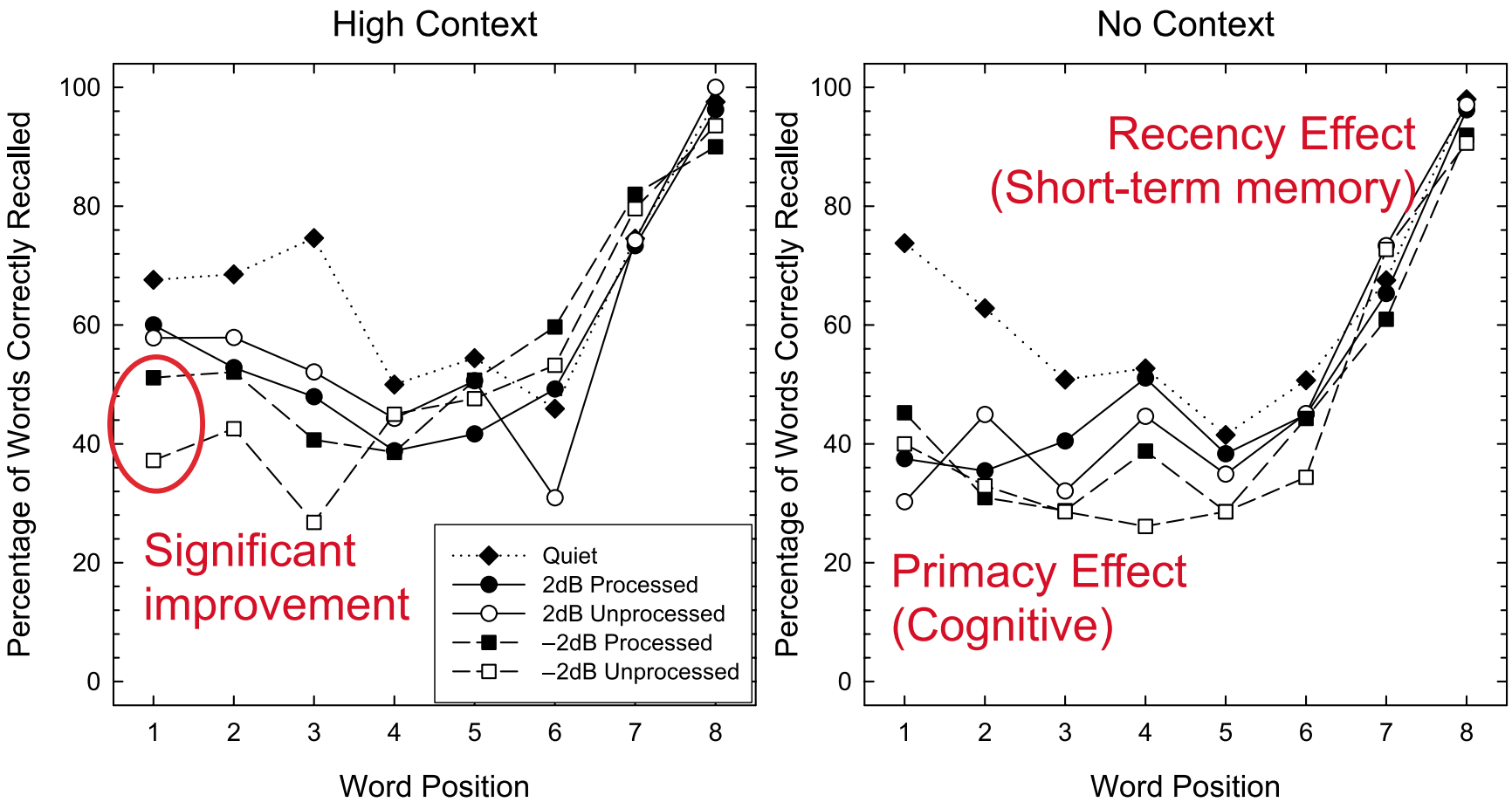
Test #1 – Secondary Task - Recall

Figure 2. See caption in Figure 1, but this figure illustrates free recall performance.



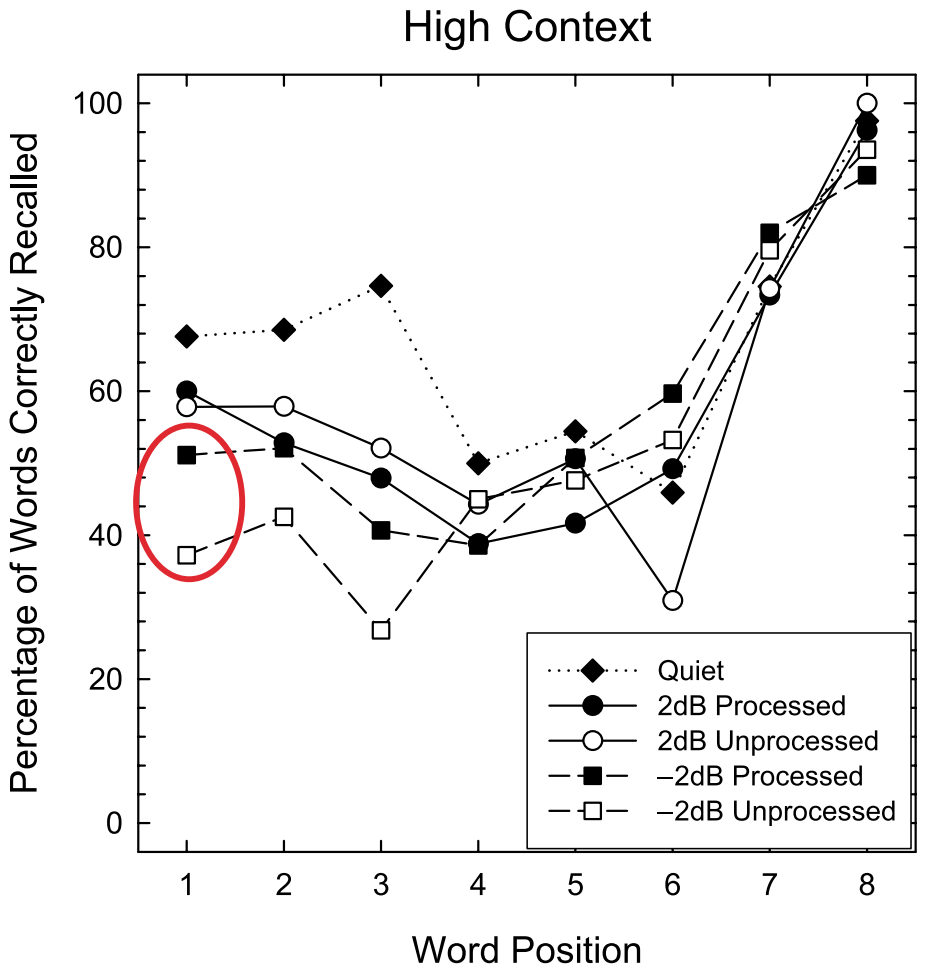
Test #1 – Word Recall (repeat latest, versus memory)

Figure 3. Free recall performance in Experiment 1, as a function of word position, averaged across 25 listeners. The left and right panels show performance for sentences with and without context, respectively. The parameter is presence of noise and NR processing.



Test #1 – Conclusions

“When context information was available, rehearsal was facilitated by providing NR processing, at least at the lowest SNR tested here.”



Test #2 – Words vs. Visual Response

IEEE sentences

- “The fruit peel was cut in thick slices.”

Processing

- -6, -2, +2 dB SNR (4-speaker babble)
- Ephraim-Malah NR algorithm

Primary task

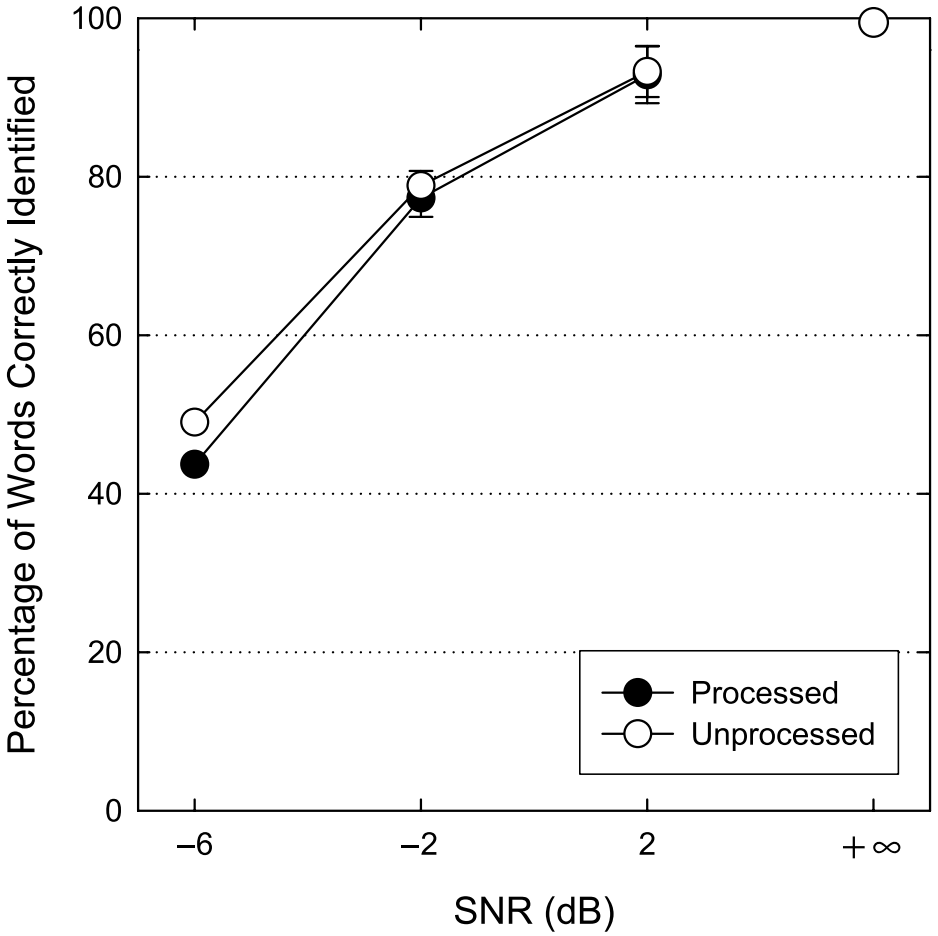
- Repeat entire sentence
- Score accuracy of repetition

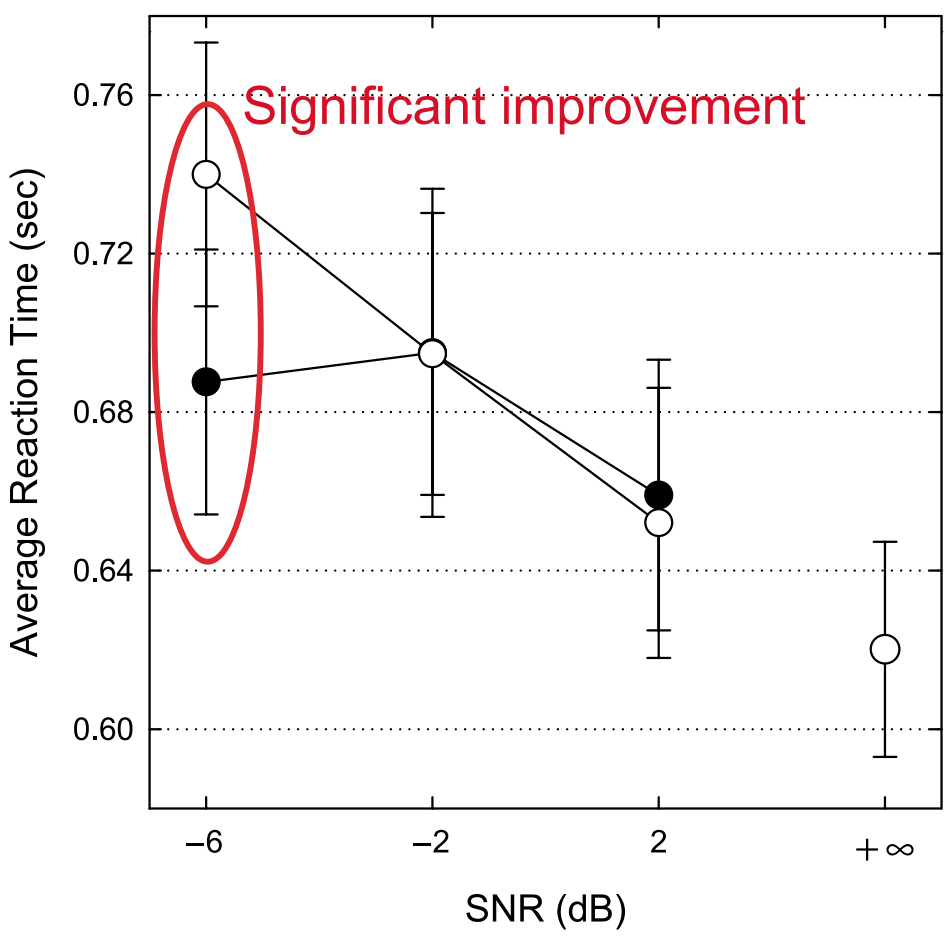
Secondary task

- Type digit appearing in one of two boxes.
- Uncorrelated appearance time
- Measure reaction time

Task 2 – Speech Intelligibility and Reaction Time

Figure 4. Mean speech intelligibility performance (left panel) and mean reaction times (right panel) as a function of SNR, averaged across 25 listeners in Experiment 2. The parameter is presence of NR processing. Error bars denote 1 *SEM*.

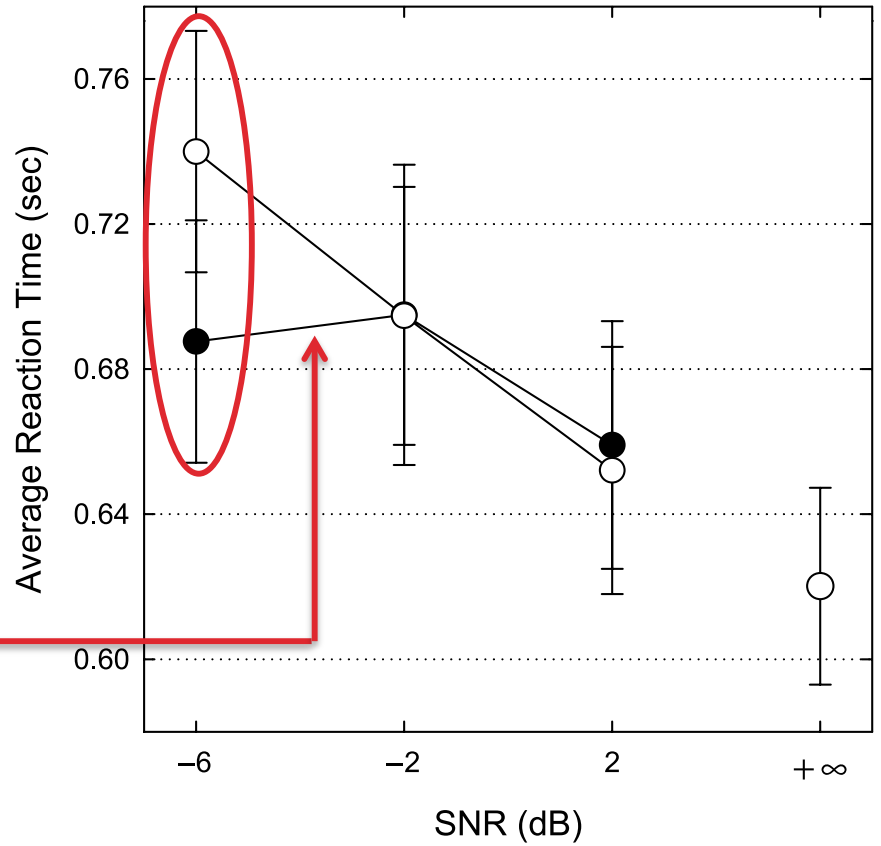




Test #2 – Conclusions

"This finding suggests that at this low SNR, use of an NR algorithm may free up cognitive resources that would otherwise be involved in extracting speech from noise, allowing them to be allocated to other simultaneous processing tasks."

Processing reduces reaction time!



Google

Other approaches

Memory Model

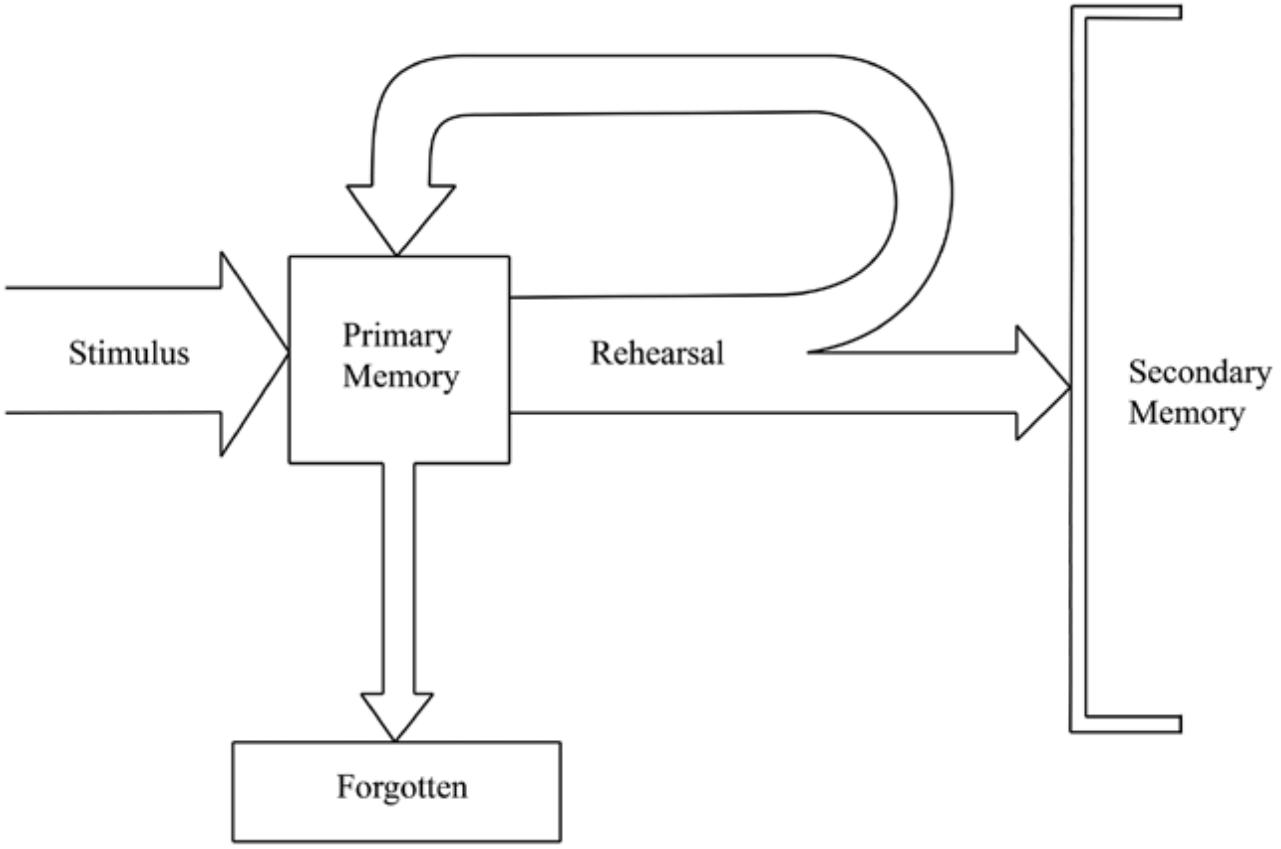


Figure 1. A simple flow diagram showing information in short-term ("primary") memory being transferred to long-term ("secondary") memory via rehearsal. If not successfully transferred, the material will be forgotten. (From N. C. Waugh & D.A. Norman (1965), *Psychological Review*, *72*, Figure 2, pg. 93. Reproduced with permission from the American Psychological Association.)

Physiological Factors

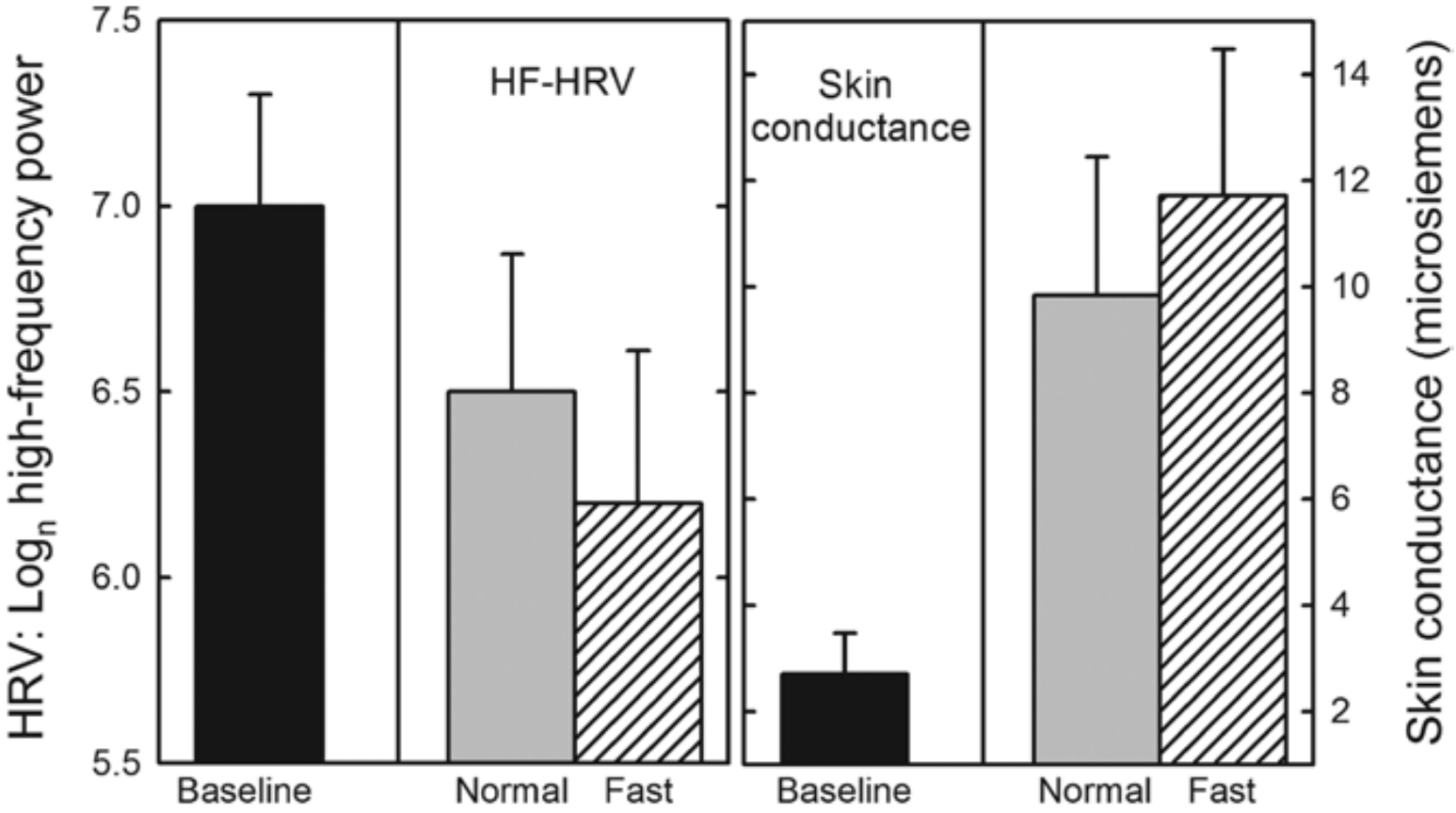


Fig. 2. Mean high-frequency heart rate variability (HF-HRV; left) and skin conductance (right) for baseline conditions and sentence repetition conditions for sentences presented at normal and fast speaking rates. The error bars denote ± 1 SE.

Age Differences

Younger listeners have more cognitive capacity?

- Similar word accuracy
- Different secondary task performance

Young had better performance with Visual

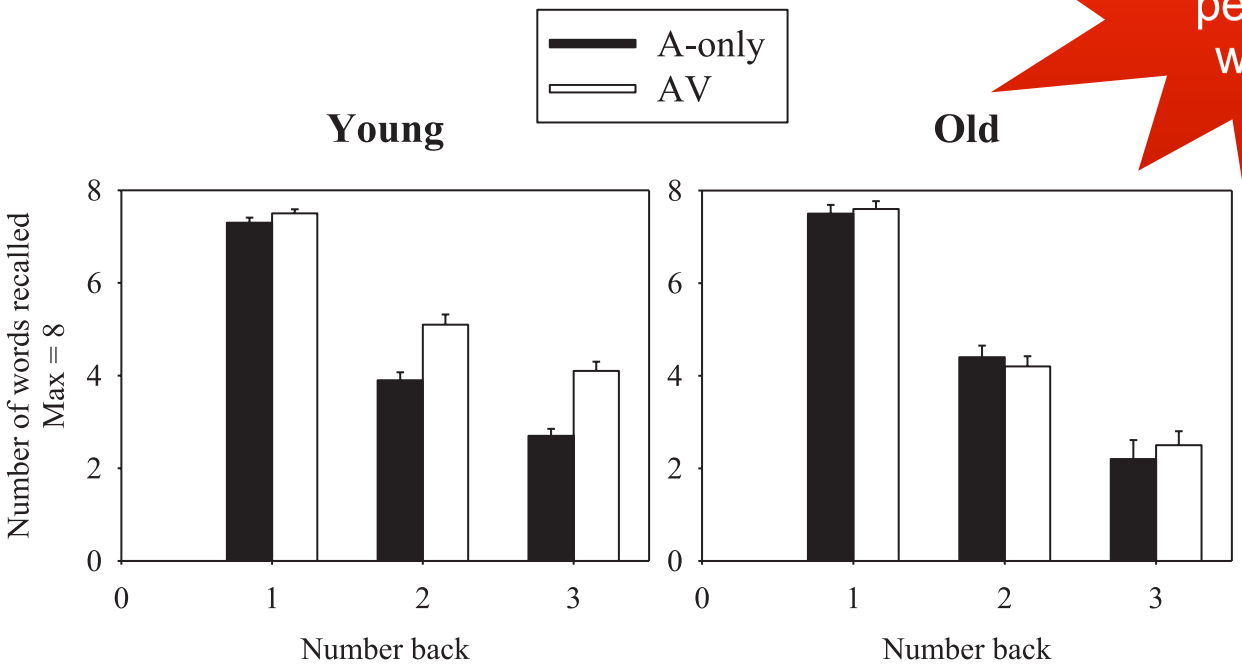


Fig. 2. Number of words recalled for younger (left panel) and older (right panel) adults in each of the three test positions (one-, two-back, and three-back). Error bars represent standard errors of the mean.

Ease of Language Understanding

Understanding Language

- Some is easy (implicit)
- Some is hard (effortful)

Auditory Stream Analysis

- Adds effort

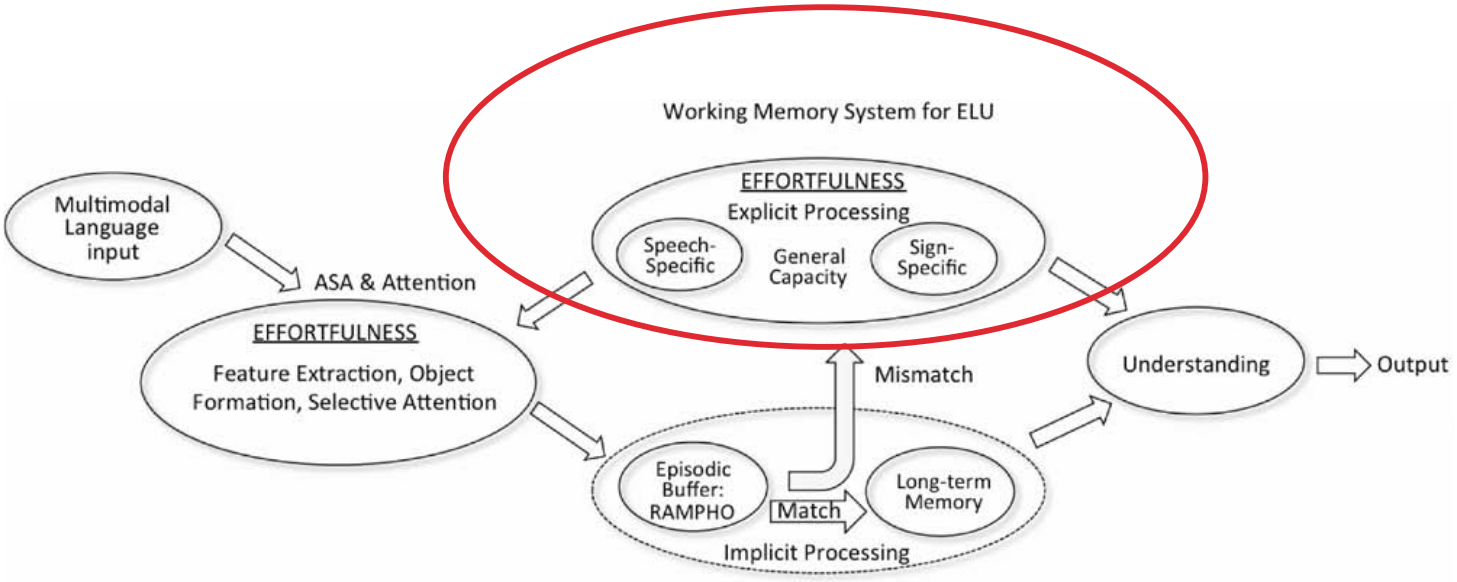
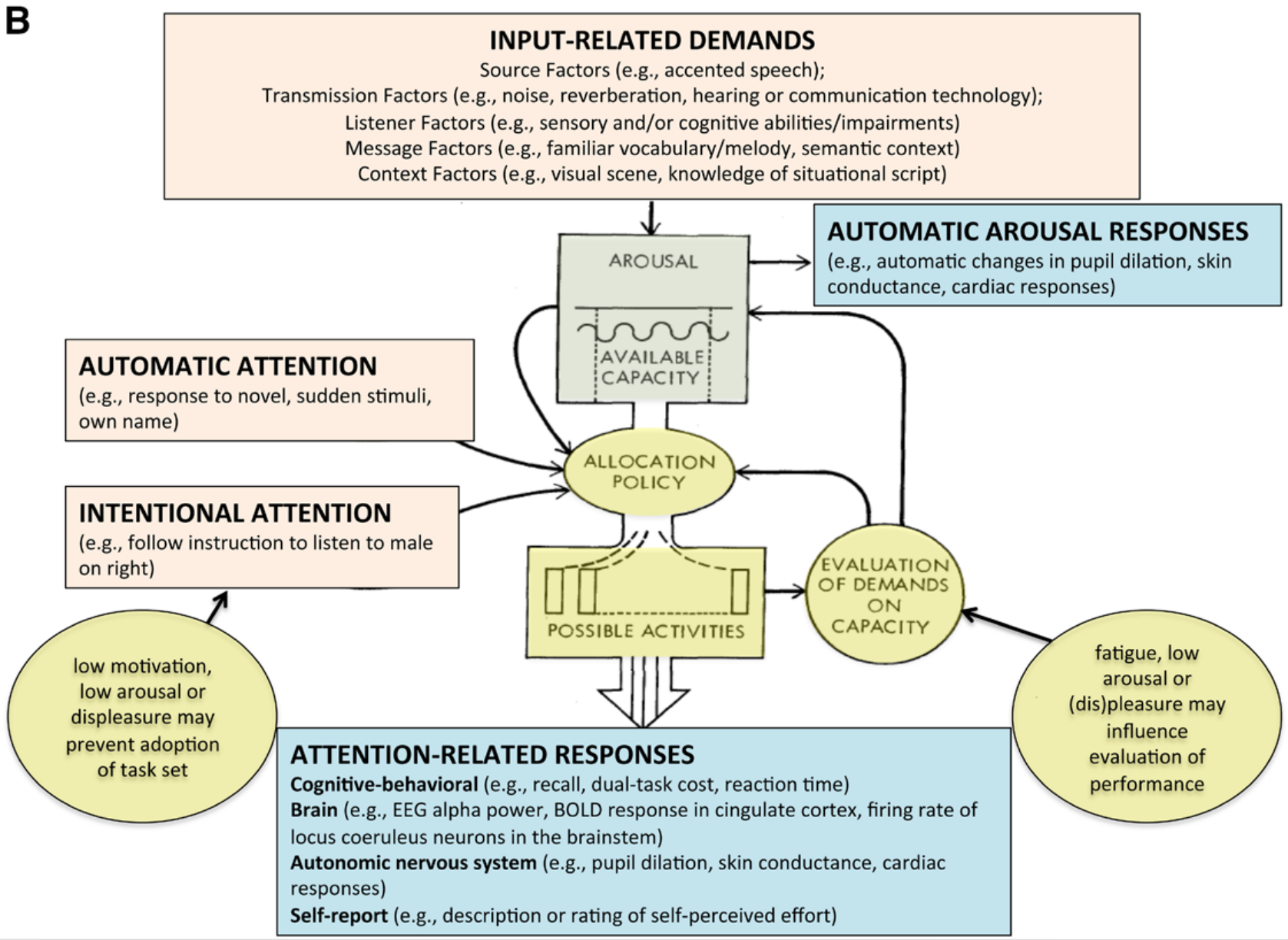


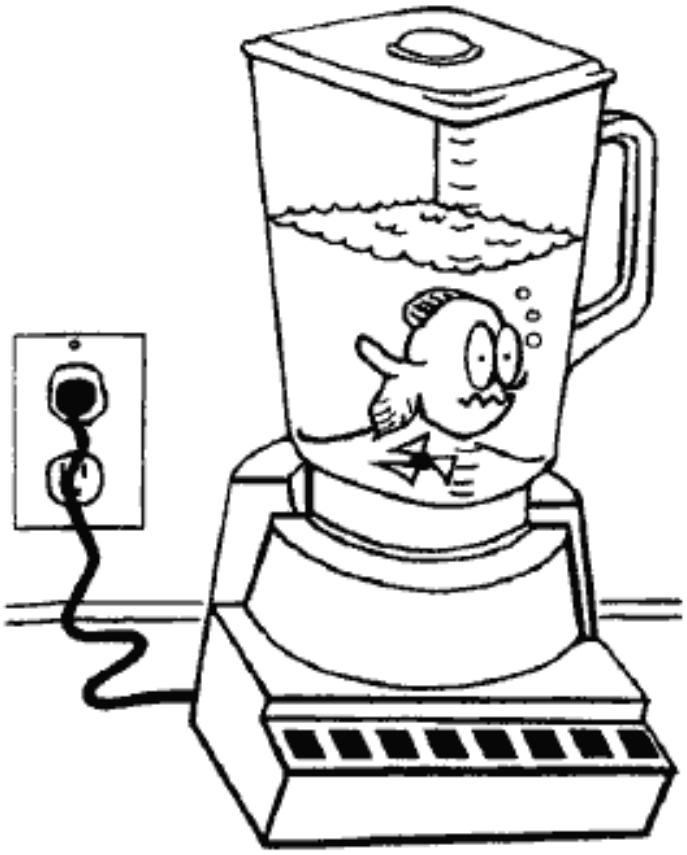
Fig. 2. A hybrid auditory scene analysis (ASA) and Ease of Language Understanding (ELU) model, based on the ELU model (Rönnberg et al. 2008). The new additions are the ASA & Attention module, and the arrow point to it from the Explicit Processing module. Cognitive effort is exerted within the ASA & Attention module and the Explicit Processing module.

Capacity Model of Attention

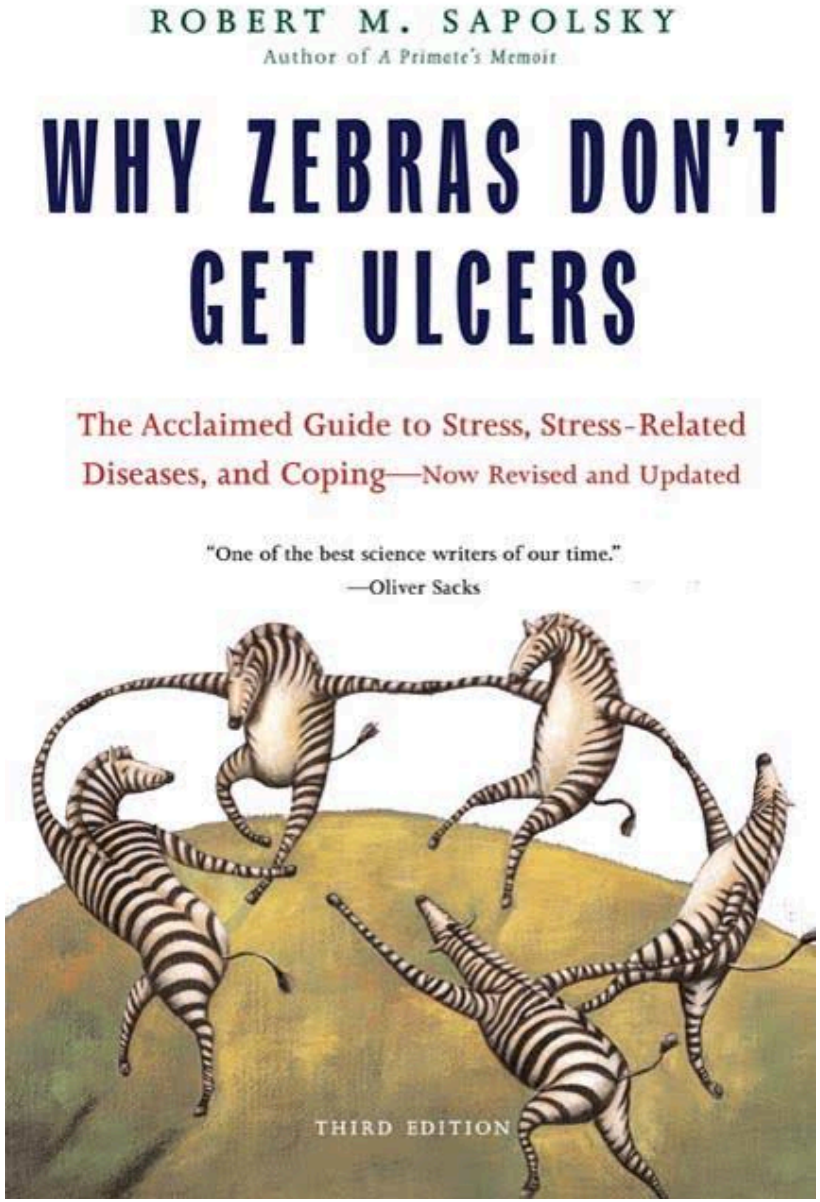


Pichora-Fuller, et al. Ear and Hearing, pp. 5S-27S, 2016.

Hormonal Indications



And you thought there was stress in *your* life !



http://www.wolfescape.com/Humour/NonMedThumbs/Stress-FishInBlender.gif

Conation

co·na·tion

/kōˈnāSH(ə)n/

noun PHILOSOPHY PSYCHOLOGY

the mental faculty of purpose, desire, or will to perform an action; volition.

Need for Aid

“Patients who are hard of hearing do not primarily seek hearing help because they have noticed poorer audibility of soft sounds, but instead they complain about an inability to function in complex everyday acoustical environments and demanding listening situations.”

“They complain of poorer environmental awareness, inability to distinguish different talkers in group conversations, and increased listening effort and fatigue from extended communication interactions.”

Edwards. Ear and Hearing, vol. 37, pp. 85S-91S, 2016.

Conclusions

“Furthermore, these results suggest that the benefit of using a digital NR algorithm is not in making speech more intelligible but, rather, in reducing the cognitive effort involved in the task. This can be seen as an improvement in performance in a simultaneous task.”

Summary

Mean-squared Errors

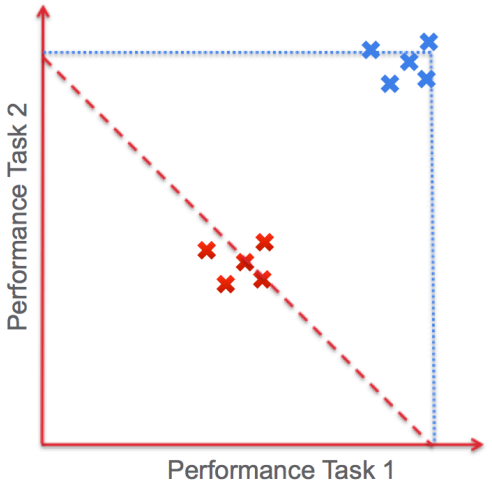
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Perceptual

- MOS
- Articulation Index
  - Fletcher->Allen

Cognitive

- Comprehension
  - Mach1(From Interval))
- Listening Effort



Thank You

Q&A