Measuring the impact of signal enhancement on the quality of noisy speech

Mark Huckvale

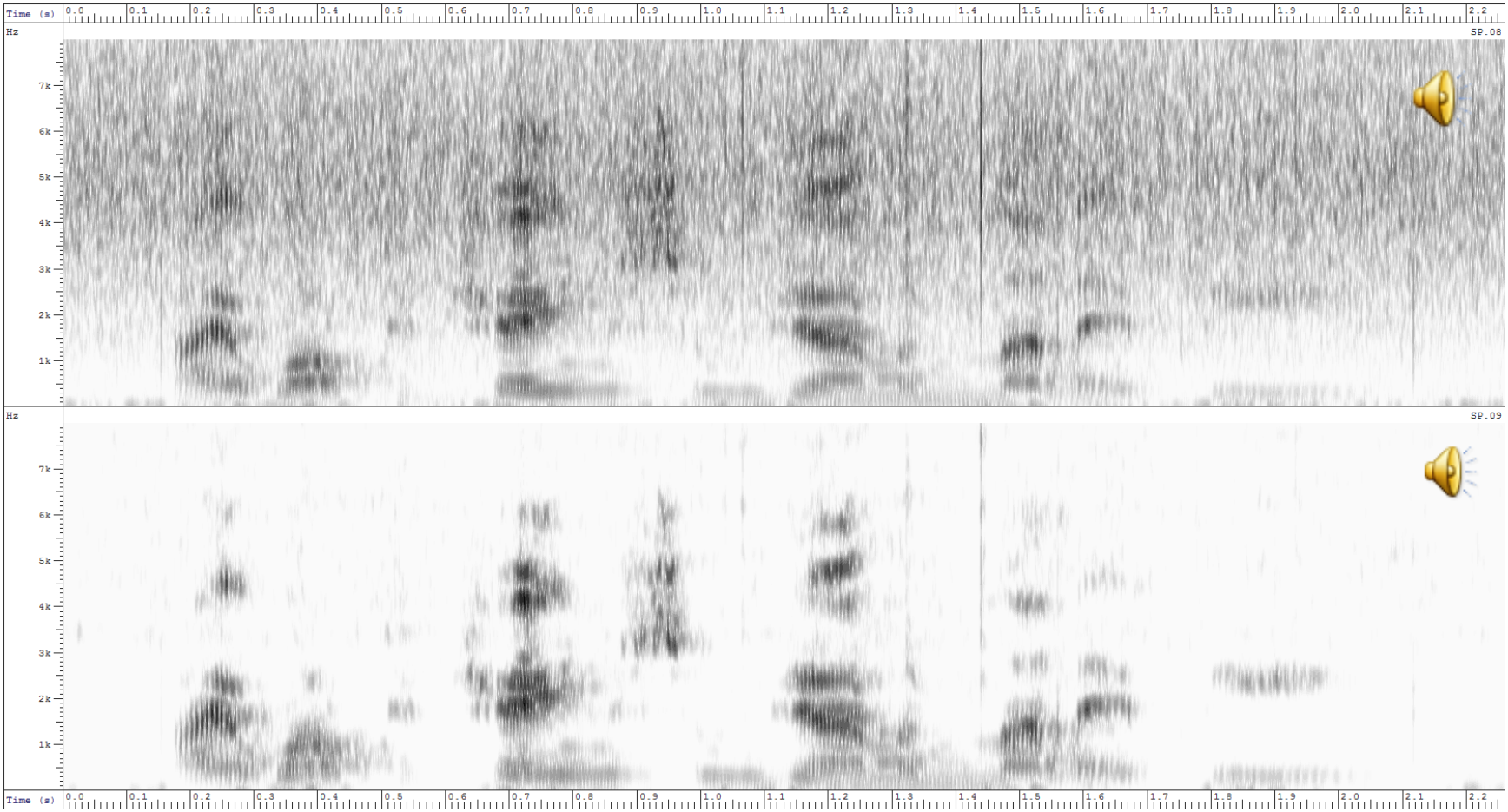
Centre for Law-Enforcement Audio Research

Speech, Hearing and Phonetic Sciences

University College London

www.clear-labs.com

Noise Reduction Example

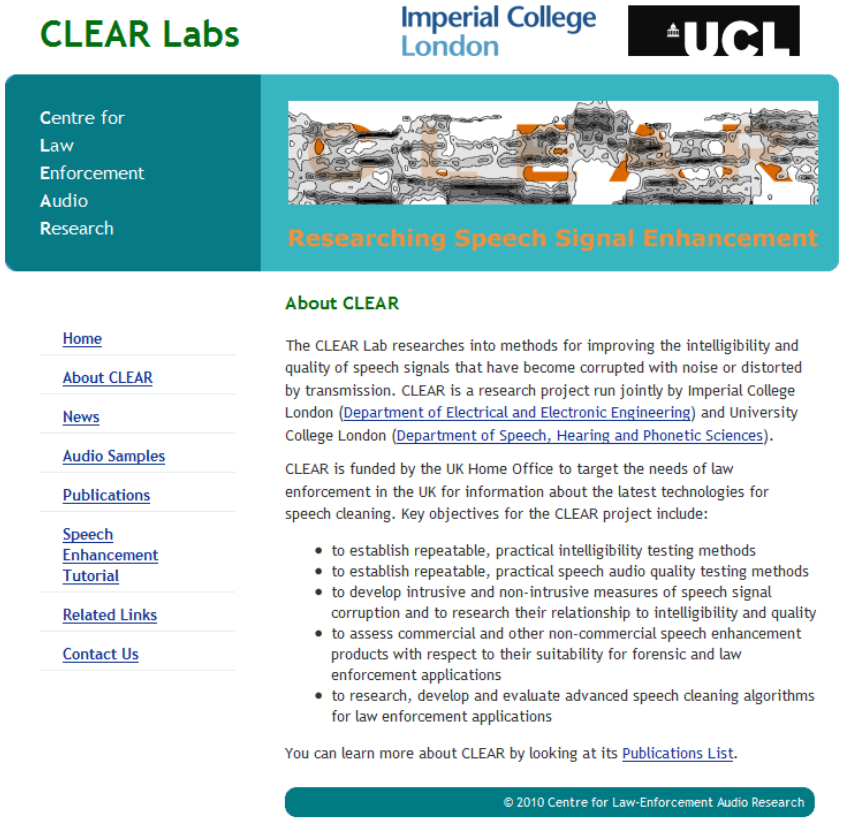


Overview

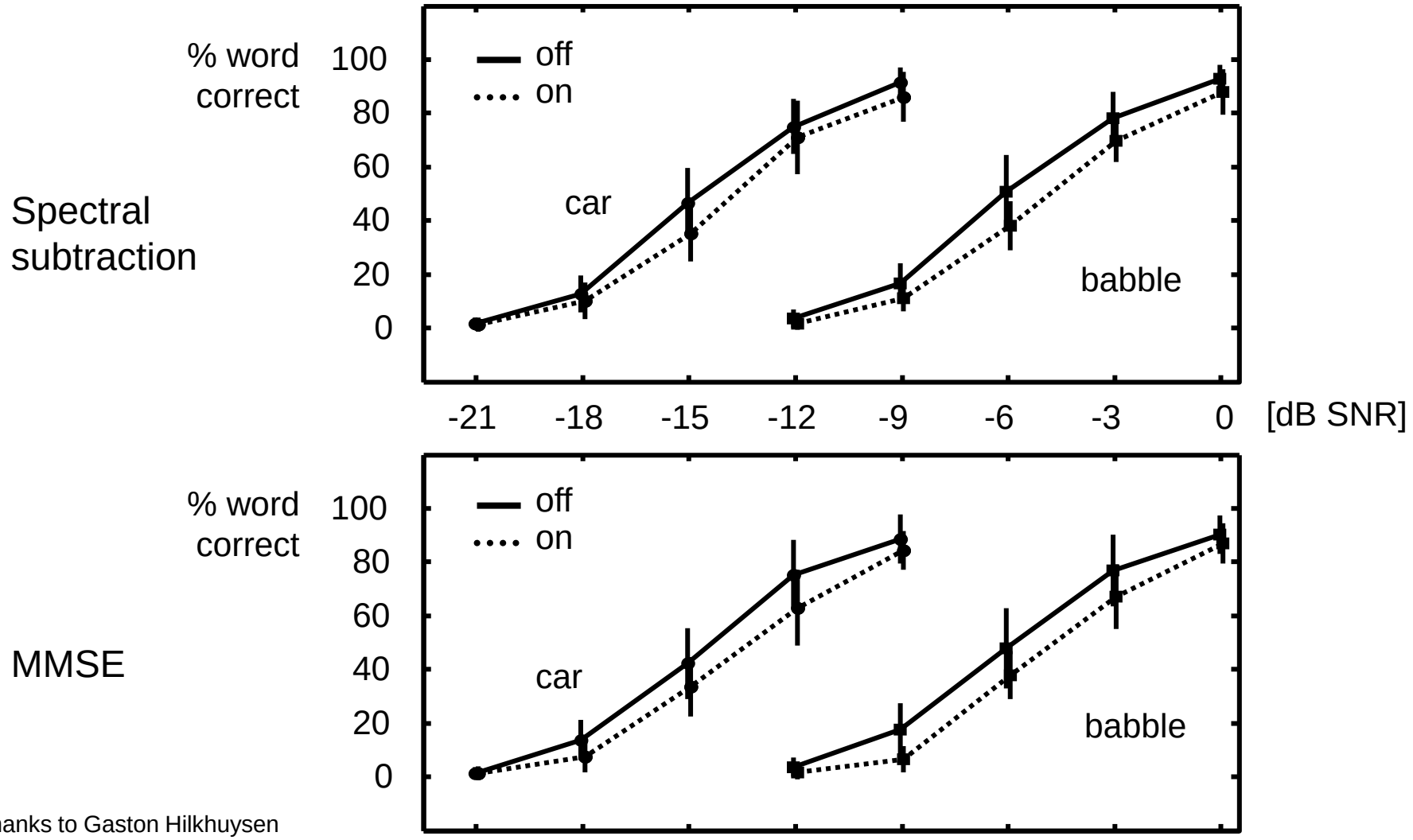
- The CLEAR project
- The Quality and Intelligibility puzzle
- The Typometer
- The Proofometer
- What have we learned?

The CLEAR Project

- Centre for Law-Enforcement Audio Research
  - Funded by UK Home Office
  - Joint Imperial College / University College London
  - Establish reliable techniques for testing the quality and intelligibility of speech signals after enhancement
  - Develop predictive models of quality and intelligibility
  - Evaluate commercial products for speech enhancement
  - Research new enhancement techniques



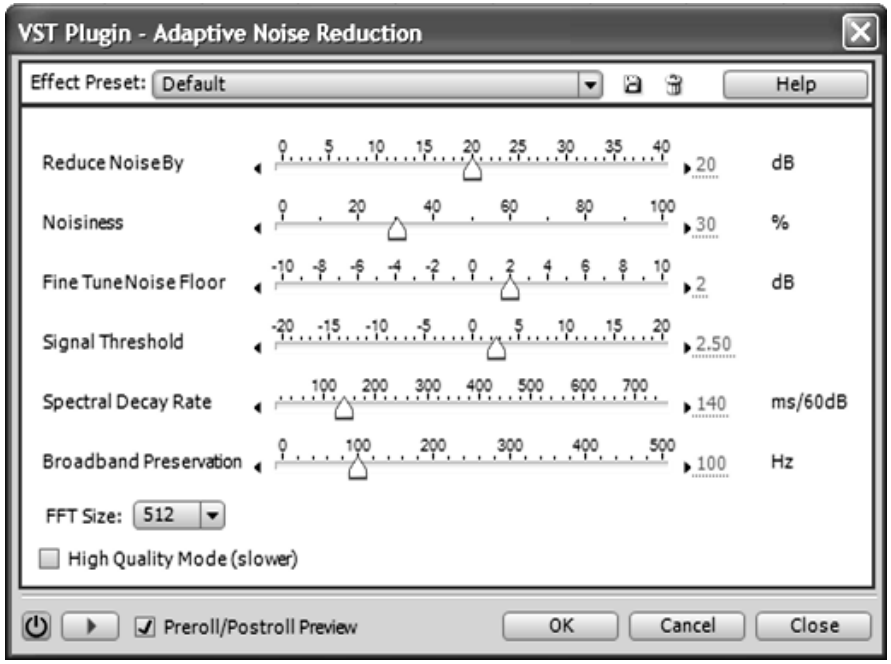
Intelligibility testing



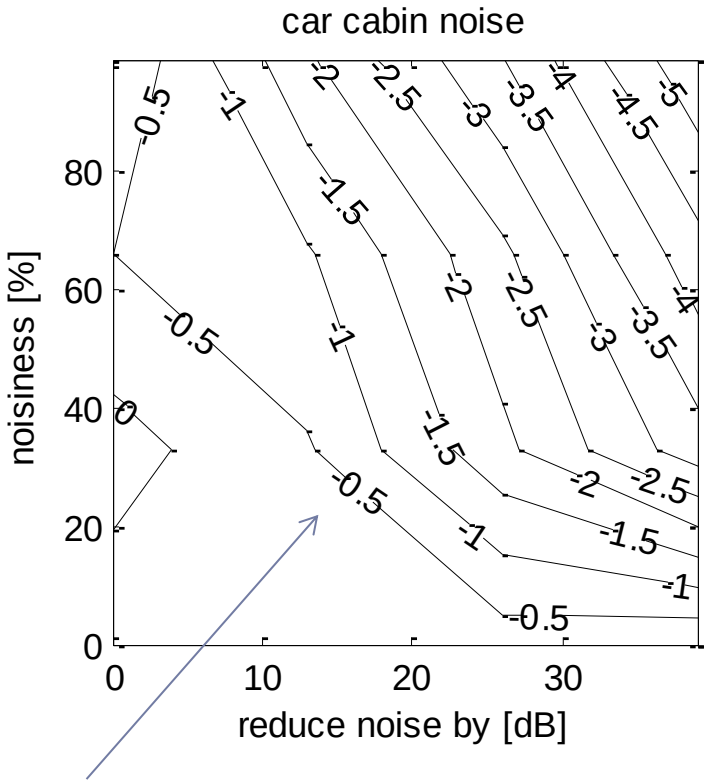
Thanks to Gaston Hilkhuysen

Optimising use of commercial system

Adobe Audition Noise Reduction



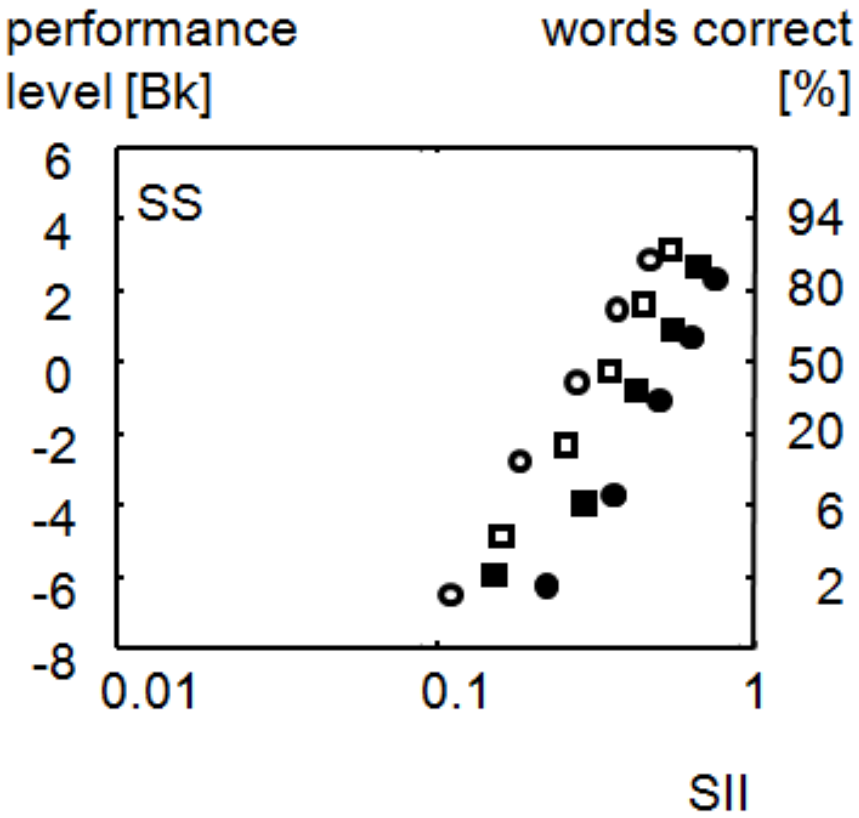
Change in intelligibility

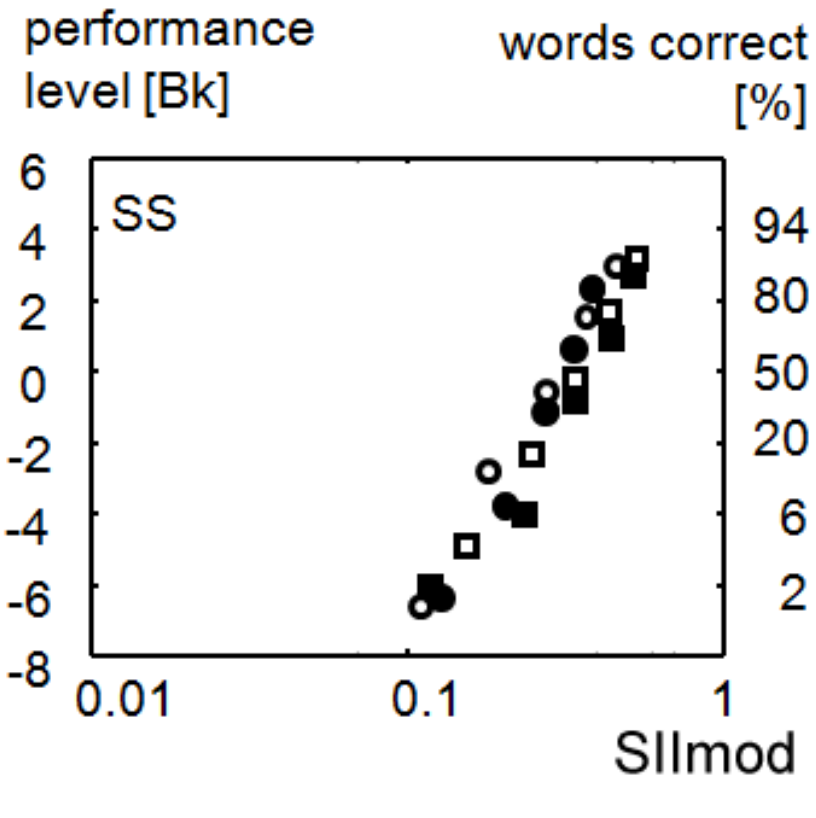


Change measured in log odds ratio

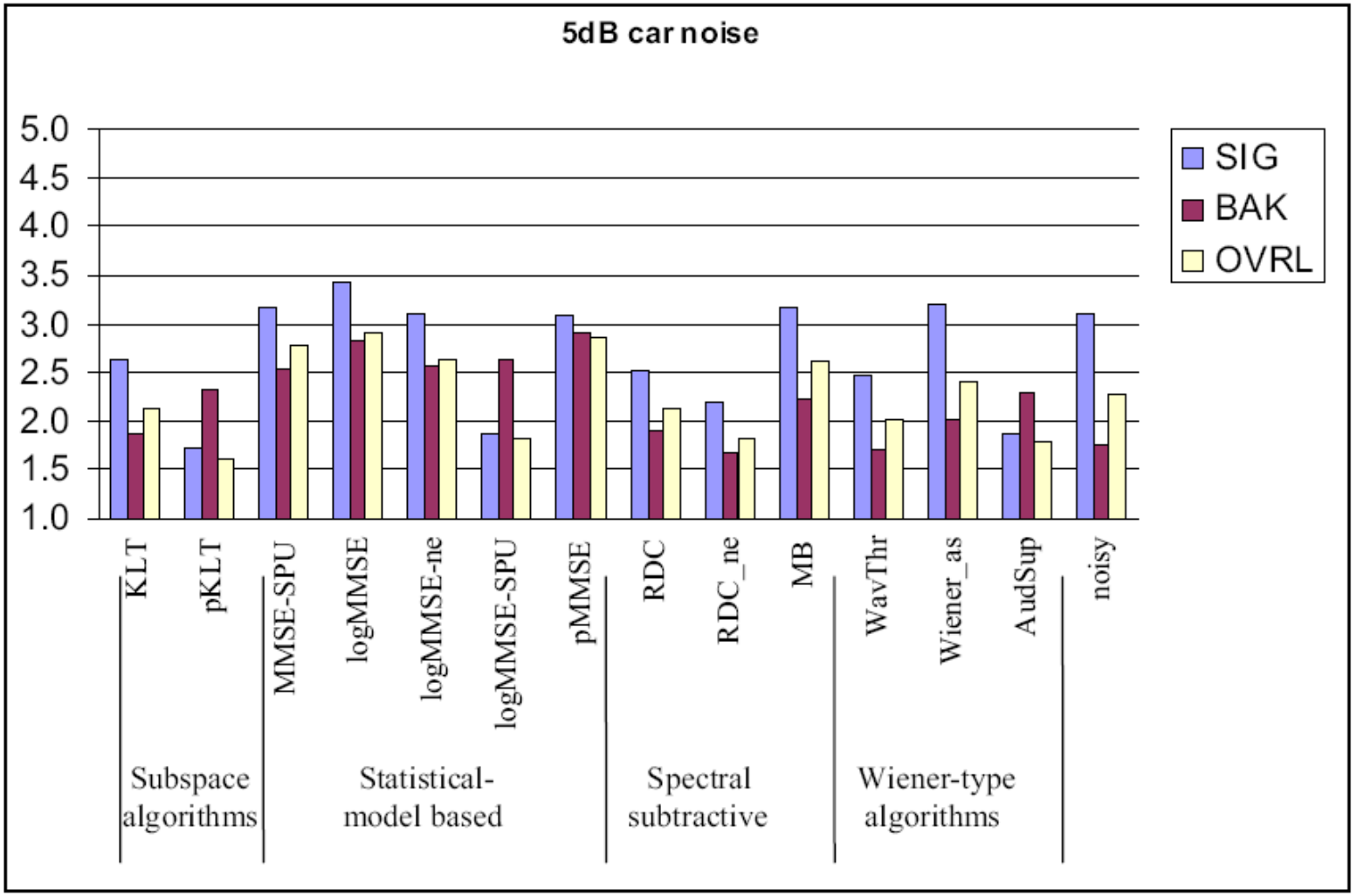
Thanks to Gaston Hilkhuysen

Modelling and Prediction





Effect of Enhancement on Quality



The Quality and Intelligibility Puzzle

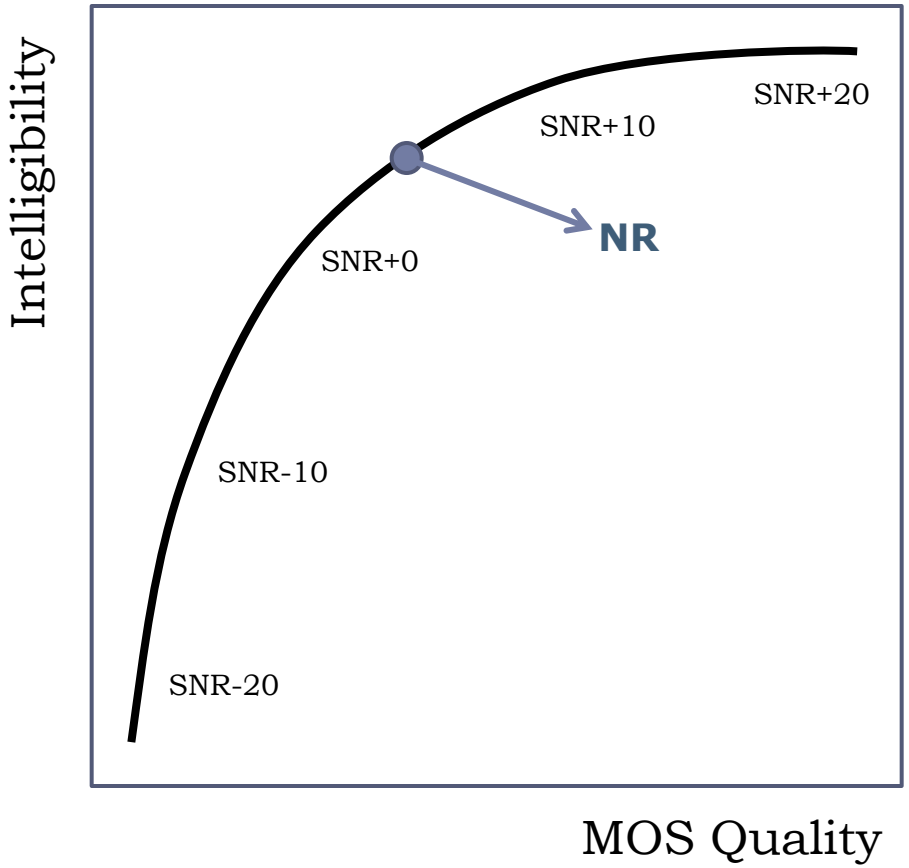
Signal Quality

✓ Noise reduction can lead to an increase in perceived **signal quality** (mean opinion scale)

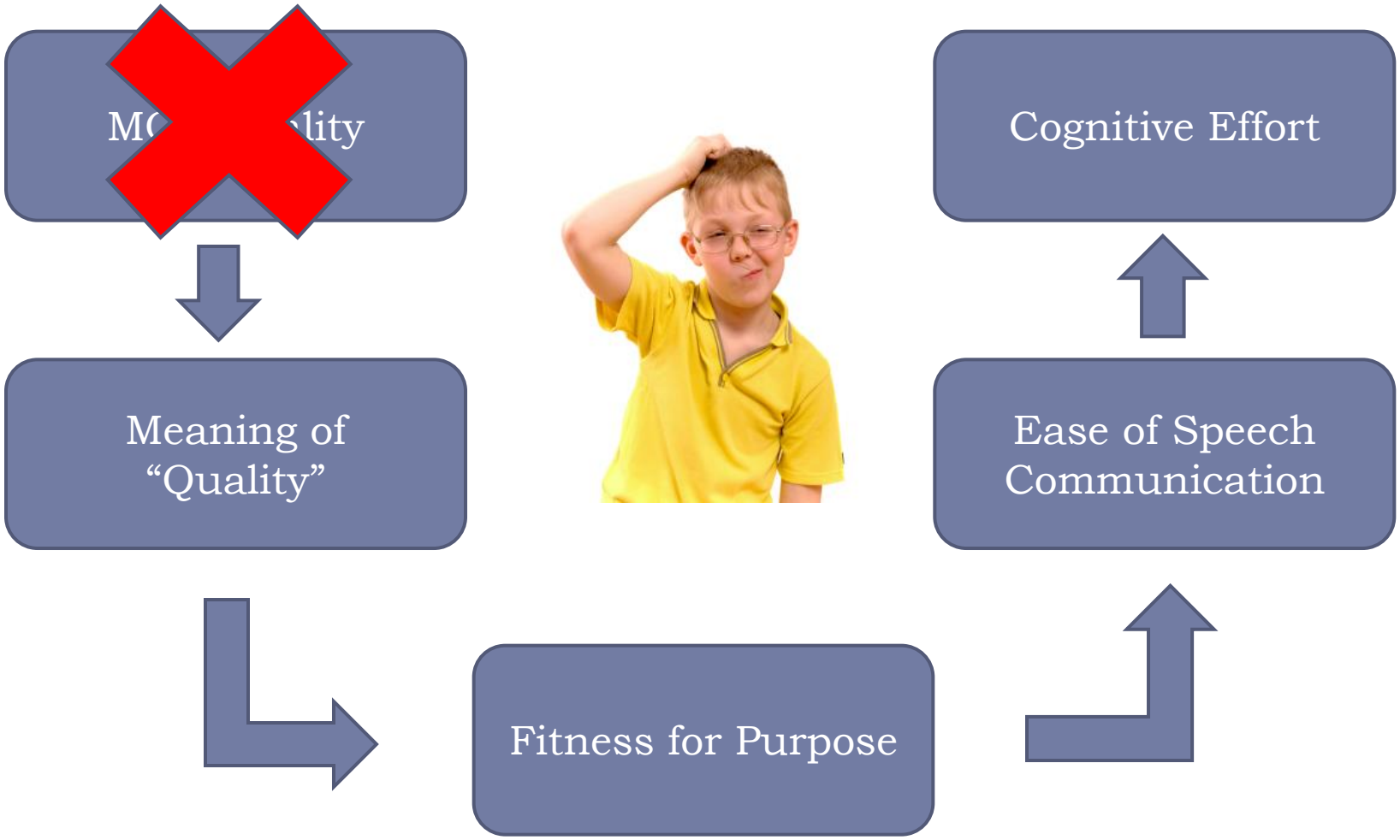
Intelligibility

✗ Noise reduction has little effect or a detrimental effect on **speech intelligibility** (% words correct)

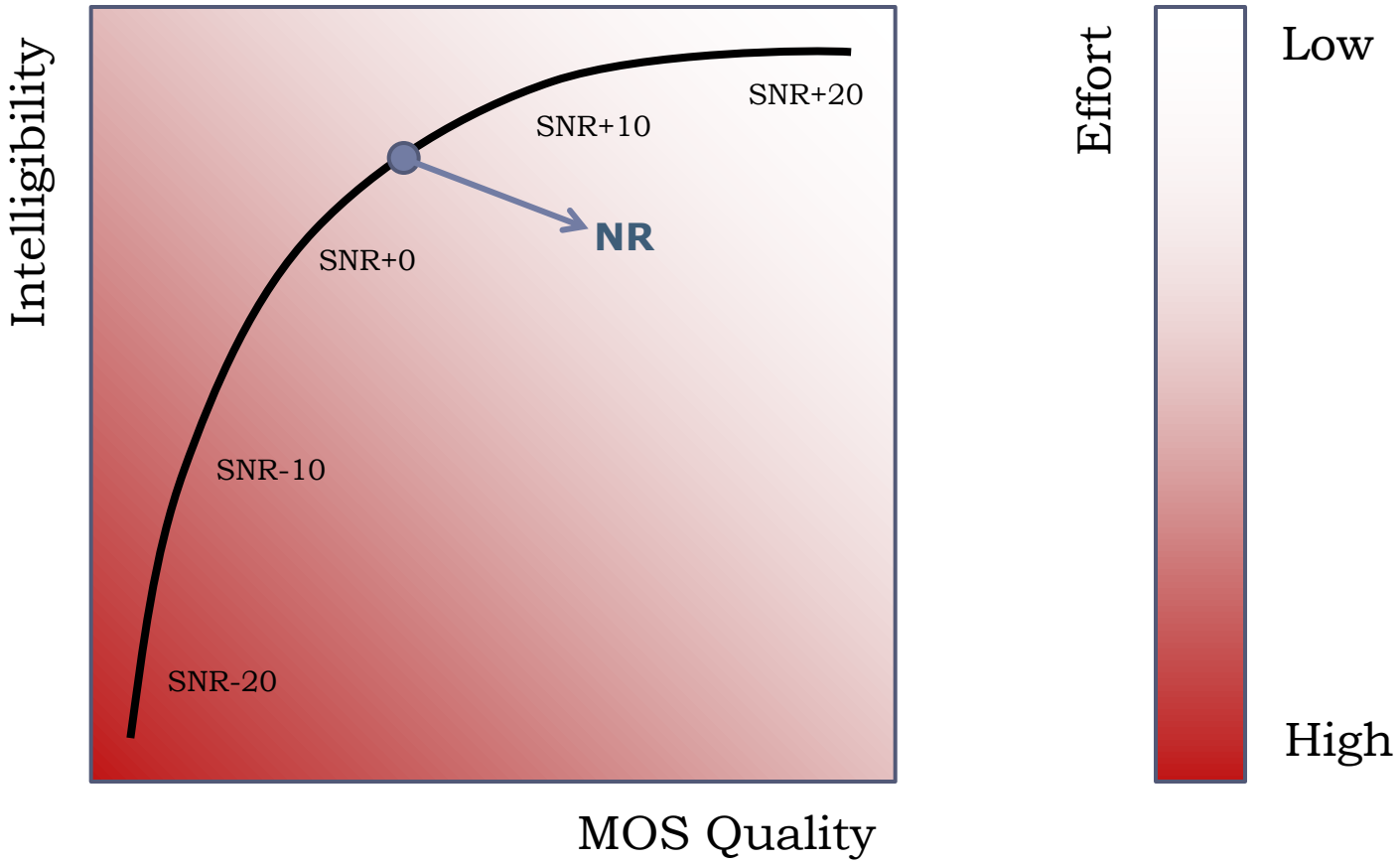
Impact of Noise Reduction



Rebasing Measurements of Quality

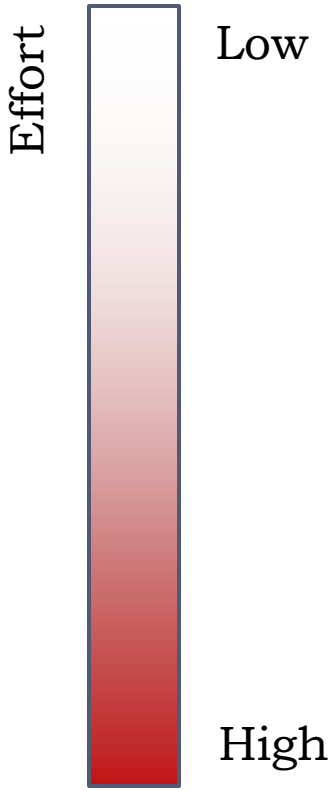


Possible Impact of NR on "Effort"



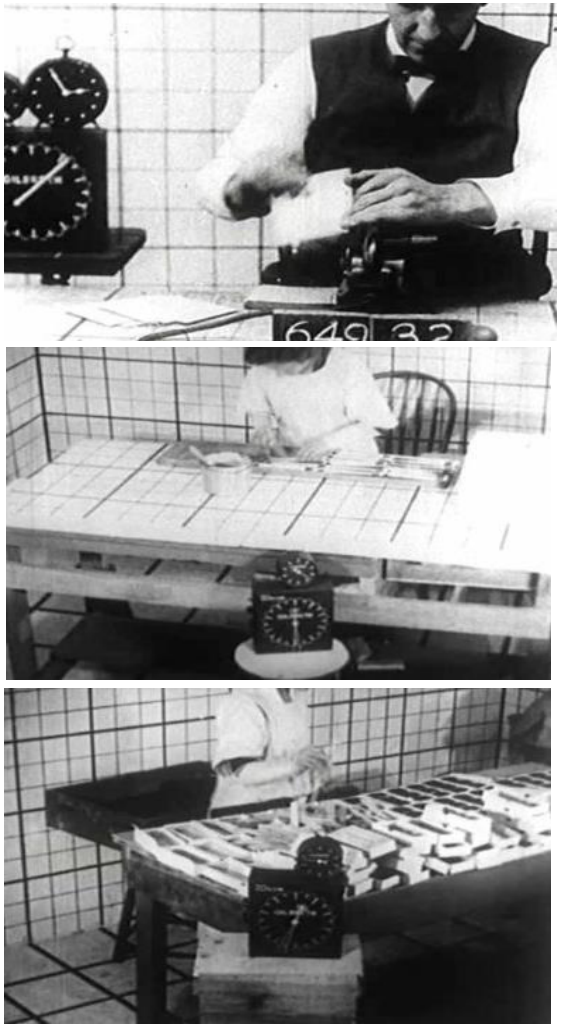
Requirements of a test of "effort"

- be **based on objective measurements**, that is, measurements of human performance not human opinion,
- **use a sufficiently complex task** to shift the psychometric function of intelligibility so that subjects make errors even for otherwise highly intelligible signals,
- **include measures of reaction time** or other physiologically-based signals to add a dimension of measurement directly related to cognitive effort,
- be **based on a speech task relevant to the situation** in which the communication systems is used.

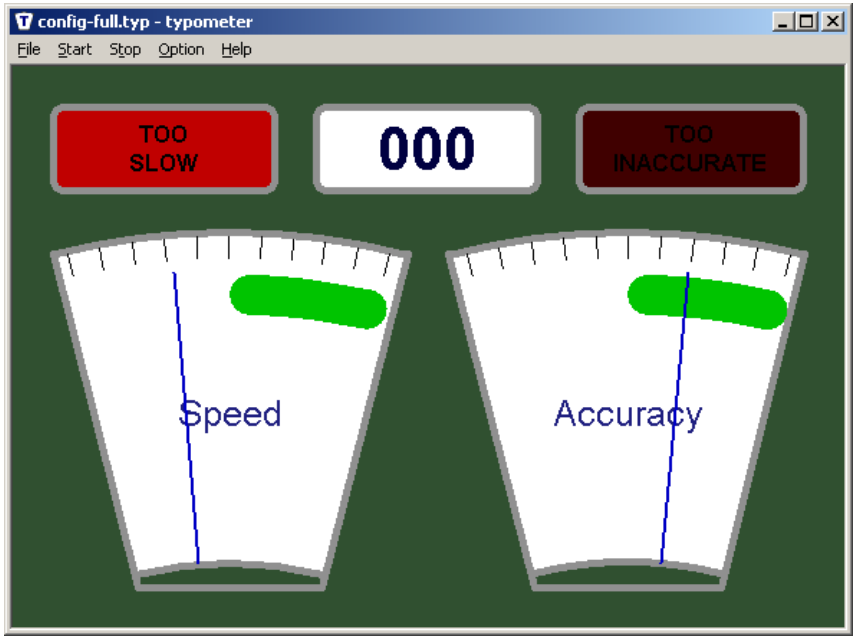


Measuring effort using lab tasks

- Word recall
  - Sarampalis at al (2009)
- Letter & Digit recall
  - Durin at al (2008)
- Digit reaction time
  - Huckvale & Leak (2009)
- Audio proof-reading
  - Huckvale & Frasi (2010)
- Other tasks?
  - Lexical decision task
  - Comprehension tests

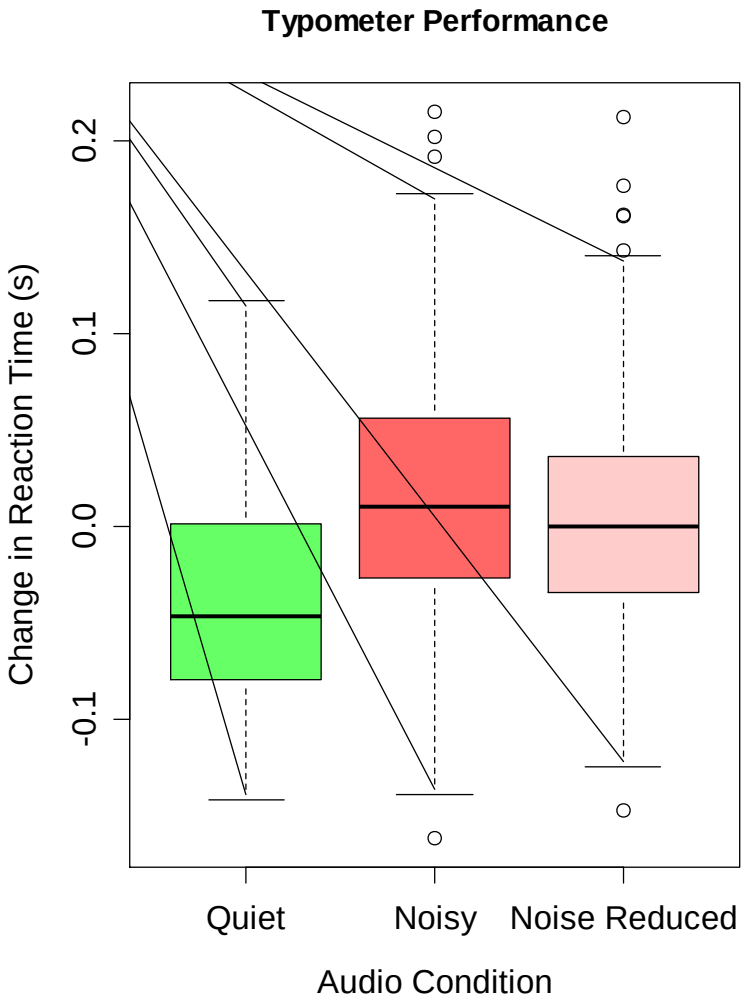


Typometer - Design



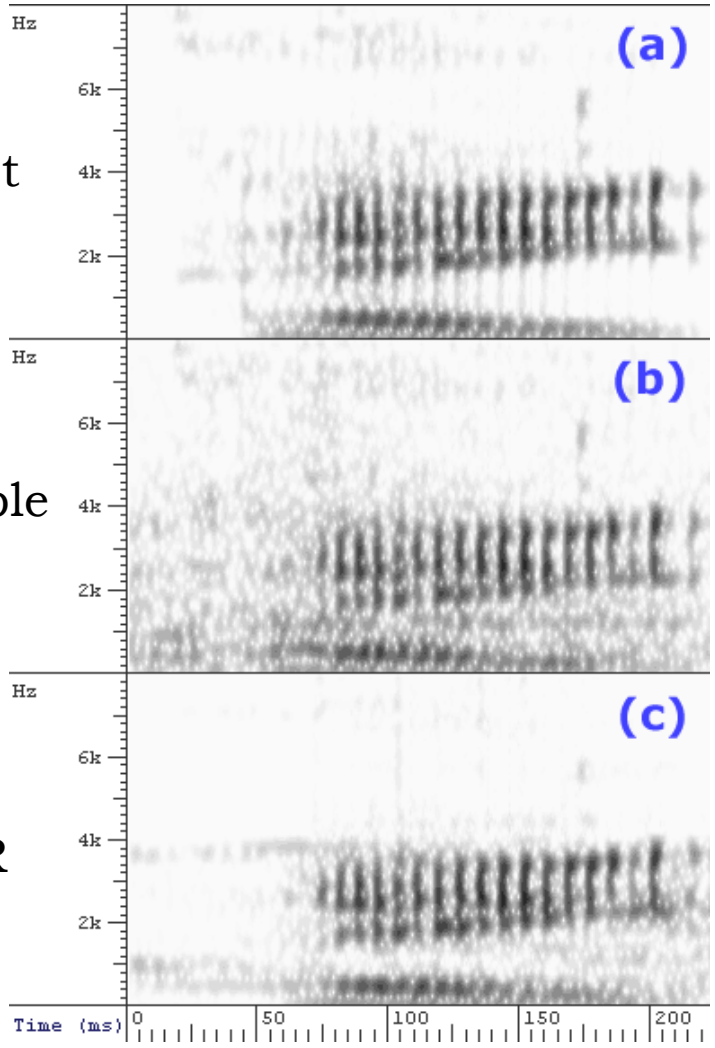
- Measures reaction time to spoken digits
- Subjects encouraged to be fast and accurate
- Measure mean reaction time in Quiet, Noisy, and Noise-reduced conditions

Typometer - Results



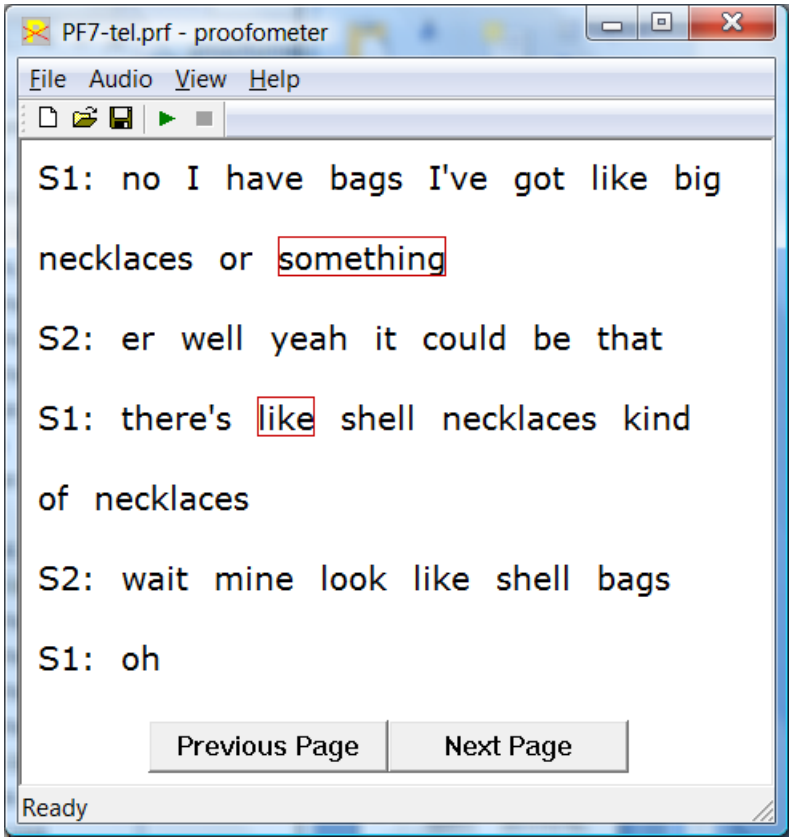
- Noise and noise-reduction did not affect task accuracy
- Reaction time increased in the presence of noise
- Reaction time did not significantly decrease again after noise reduction

Typometer – Auditory Effects



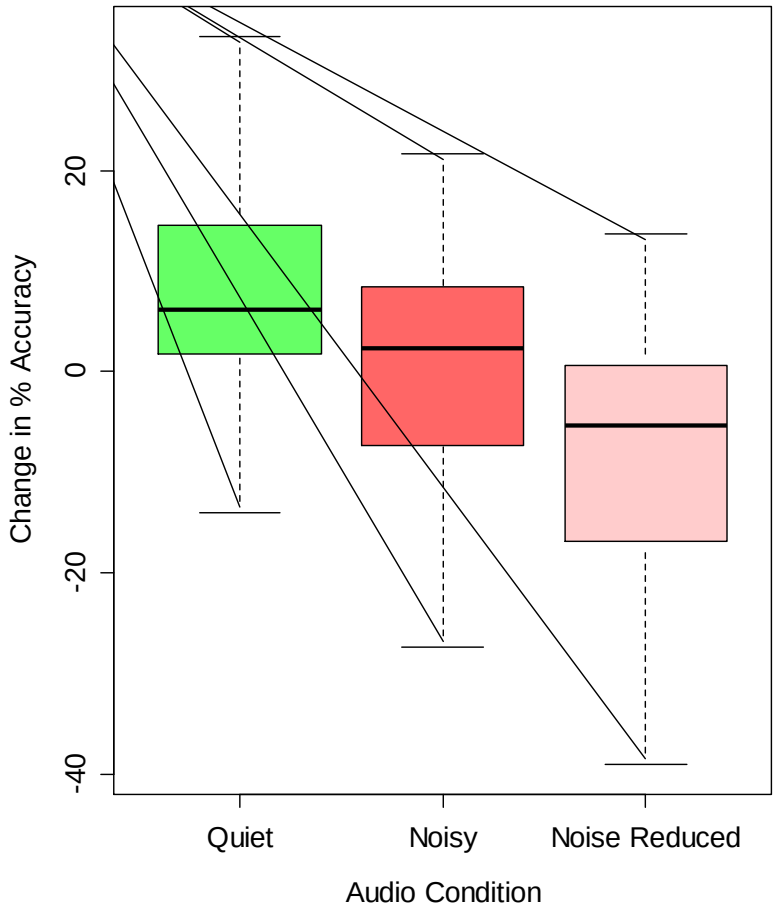
- Acoustic analysis shows masking of word onsets in the noise conditions (b)
- But although masking is significantly reduced in NR conditions (c) there is no significant reduction in RT
- We assume that RT is affected by more than energetic masking

Proofometer - Design



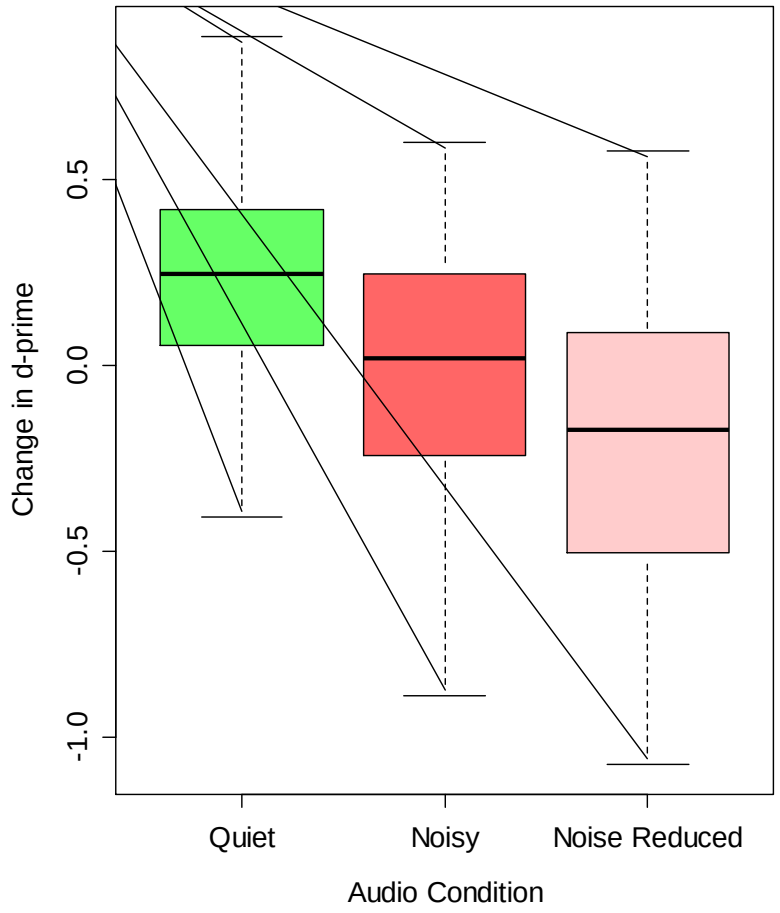
- 5 min audio recording of spontaneous dialogue (picture task)
- Listeners must identify 50 “typical” errors in a transcript of the audio as it is playing
- Measure % errors identified in Quiet, Noisy and Noise-reduced conditions

Proofometer – Task Accuracy



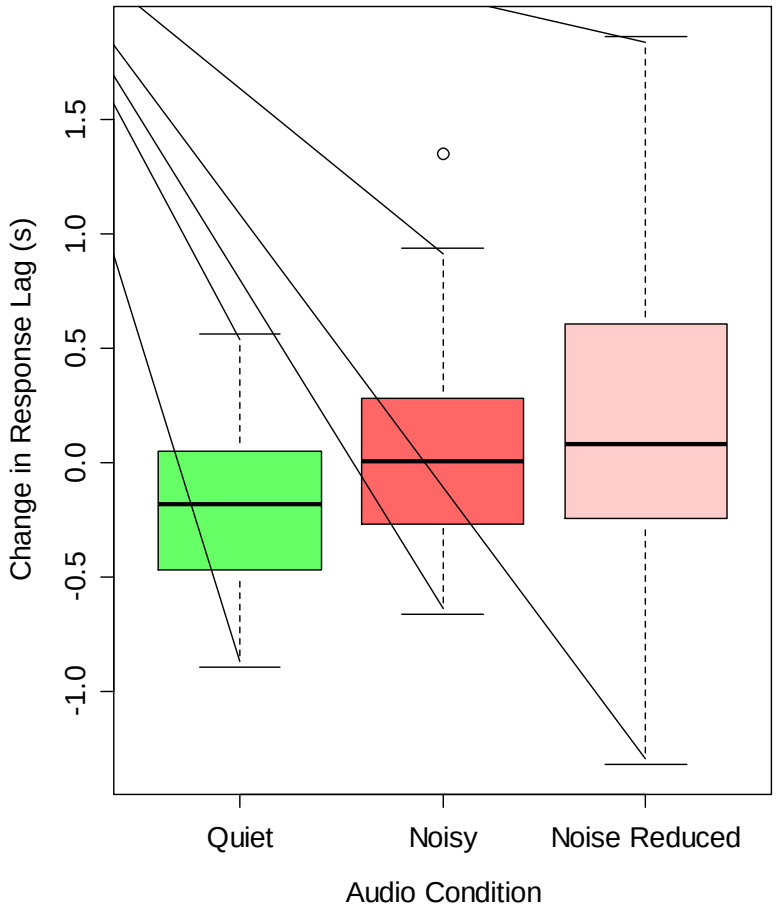
- Considerable inter-speaker variability on this task
- Number of errors identified decreased in the presence of noise
- Number of errors identified did not significantly improve after noise reduction

Proofometer – Task d-prime



- Considerable inter-speaker variability on this task
- Number of errors identified decreased in the presence of noise
- Number of errors identified did not significantly improve after noise reduction

Proofometer – Response Time



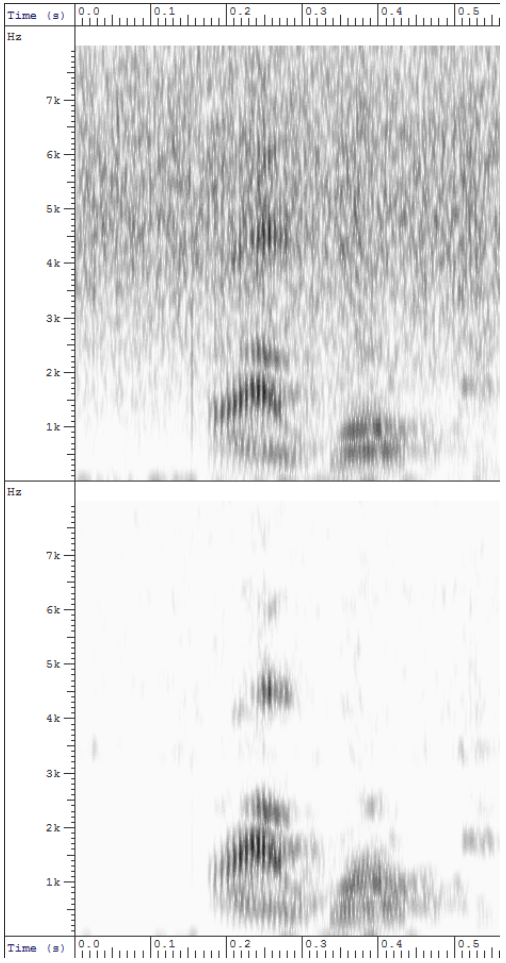
- We can also measure the mean delay listeners took to identify an error
- We see that responses took longer in both the noisy and noise-reduced conditions
- Implication is that cognitive effort is not improved by NR

Implications – Speech Quality

- Signal quality affects Proofometer task accuracy and response time *even for signals of similar intelligibility*
- Similar results to Durin (2008), but for spontaneous materials rather than digits
- But still too much variability, design needs:
  - Improved generation of transcript errors
  - Improved training of subjects
  - Improved motivation of subjects

Implications – Noise Reduction

- Found the degradation in task accuracy and response time caused by noise not subsequently improved by noise reduction
- Contradicts MOS Quality result
- Significantly more false alarms in NR conditions, could be an indicator of effect of processing on attention
- Objective quality tasks are required to justify use of speech signal enhancement processes



Conclusions

- Effects of changes in speech signal quality alone can be measured using a speech communication task
- Audio Proof-reading is a complex task that operates with realistic materials which assesses effort in terms of both accuracy and speed
- Noise reduction processing is an example of a technique that improves the opinion of quality but not quality itself

Discussion - Noise Reduction

- A noisy speech signal is more demanding to process at both auditory and cognitive levels
- For noise-reduction to be successful it has to improve processing at both levels
- Not good enough to improve SNR if as a consequence the speech left behind is distorted, or if the noise left behind becomes more speech-like.