Performance-Based Measures of Speech Quality
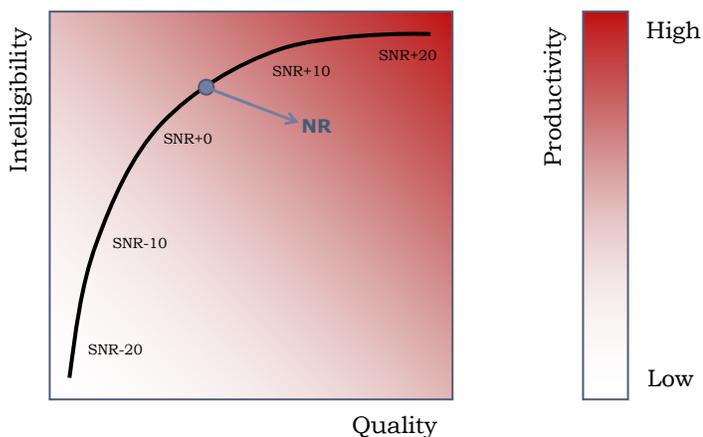
Mark Huckvale

Centre for Law-Enforcement Audio Research

Speech, Hearing and Phonetic Sciences

University College London

Intelligibility, Quality, Productivity & NR

- Speech Intelligibility
  - Typically measured as % words correct in articulation test
  - Robust performance measure
- Speech Quality
  - Typically measured on rating scale 1-5 (MOS)
  - Fragile opinion measure
- Speech Productivity
  - Performance in everyday speech communication tasks
- Noise Reduction / Speech Enhancement
  - Generally causes slight increase in speech quality ☺
  - Generally causes slight decrease in speech intelligibility ☹
  - But what is effect on productivity ?

Possible Impact of NR on Productivity

How to Measure Productivity

- Laboratory-based behavioural task
- Assessed through performance not opinion
- Clearly related to everyday speech communication
- Sensitive to changes in speech quality at high intelligibility
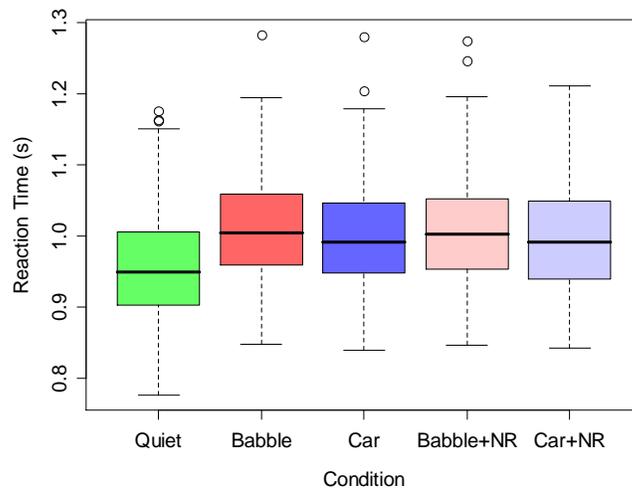- Easy for subjects to understand, perform, etc

Possible Productivity Tasks

- Choice reaction time
  - The Typometer (Huckvale & Leak, 2009)
- Serial recall
  - The Memometer (untested)
- Audio proof-reading
  - The Proofometer (Huckvale & Frasi, 2010)
- Other Ideas:
  - Lexical decision task
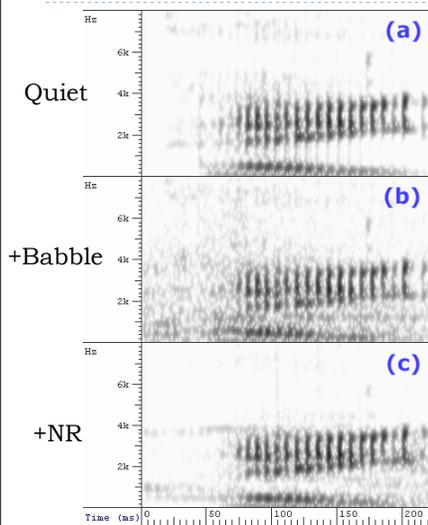 - Comprehension tests

---

Typometer - Design
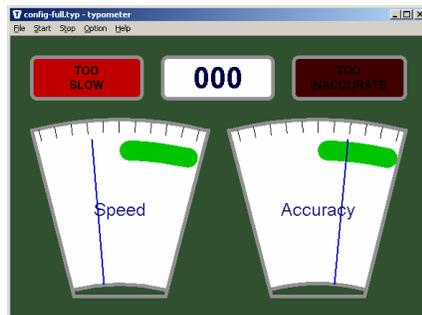
- Measures reaction time to spoken digits
- Subjects encouraged to be fast and accurate
- Quiet, Babble, Car noise, Babble+NR, Car noise+NR
- 20 subjects, 10 RTs per condition per subject per digit
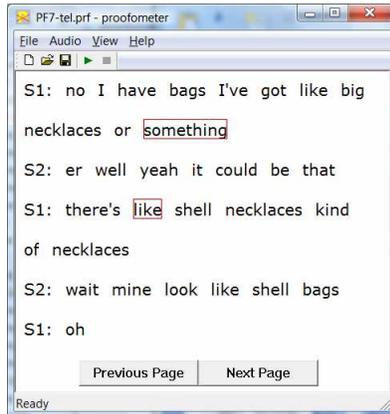
Typometer - Reaction Time Results

Typometer – Auditory & Cognitive Effects

▸ Acoustic analysis showed masking of word onsets in the noise conditions (b)

▸ But masking significantly reduced in NR conditions (c)

▸ Although no significant reduction in RT

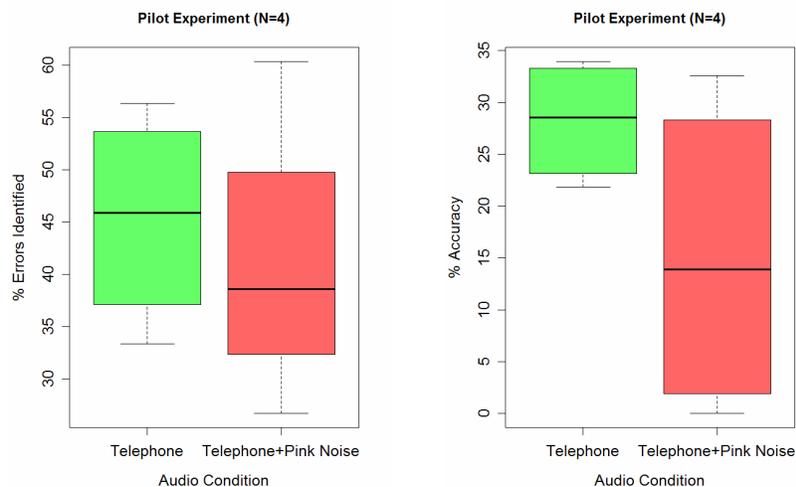▸ Implication is that there is also some cognitive level effect

Proofometer - Design

- 5 min audio recording of spontaneous dialogue (picture task)
- Transcript contains 50 "typical" errors
- Listeners must click on errors in real-time without pausing audio
- Rated for # errors and # false alarms

Proofometer – Pilot Results

Thoughts on Quality & Productivity

- Productivity would be reduced by mis-detection or mis-interpretation of speech events even if signal is supposedly intelligible.
- Auditory masking
  - Set of p(segment|signal) more similar, so greater weight on top-down prediction – slower to detect and failure to recognise low-probability events.
- Informational masking
  - Added speech-like noise increases # competing segment hypotheses - more time/effort spent evaluating utterance hypotheses & phonological memory disrupted
- Availability of cognitive resources
  - Reduction in available processing resources (tiredness, boredom, distractions) limits capability of speech system to draw on the extra processing required to deal with difficult signals.

Conclusions

- Ongoing interest in moving to measures of speech quality that can be related to performance in everyday tasks.
  - Is noise reduction or speech enhancement useful?
- Tools like the Typometer and the Proofometer are starting to explore how task performance is related to speech quality for signals of high intelligibility.
- In the long run, such experiments might help us better understand the impact of signal degradation on cognitive processing of speech at all levels.