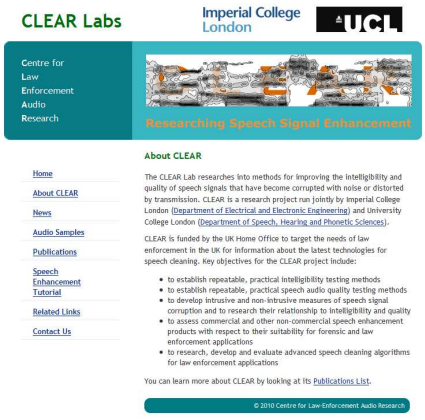
Measurement, modelling and enhancement of the intelligibility of speech in noise at the Centre for Law-Enforcement Audio Research (CLEAR)

Mark Huckvale
Speech, Hearing & Phonetic Sciences
University College London

The CLEAR Project

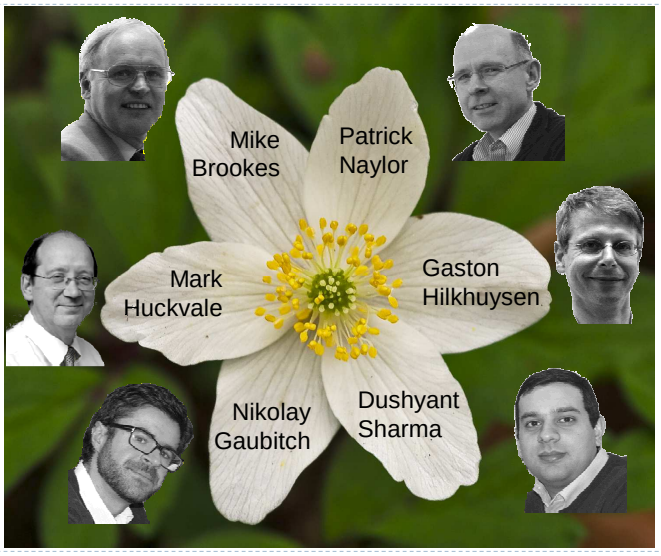
- Centre for Law-Enforcement Audio Research
  - Funded by UK Home Office
  - Joint Imperial College / University College London
  - Establish reliable techniques for testing the quality and intelligibility of speech signals after enhancement
  - Evaluate commercial products for speech enhancement
  -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



www.clear-labs.com

June 2012

The CLEAR Team



June 2012

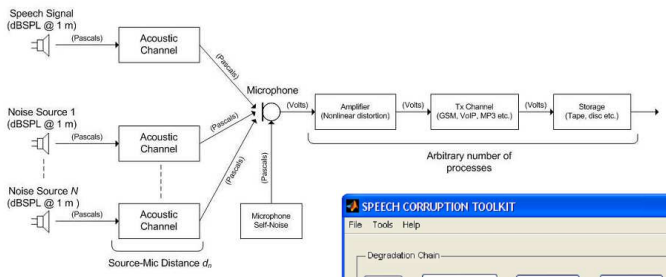
The CLEAR Approach

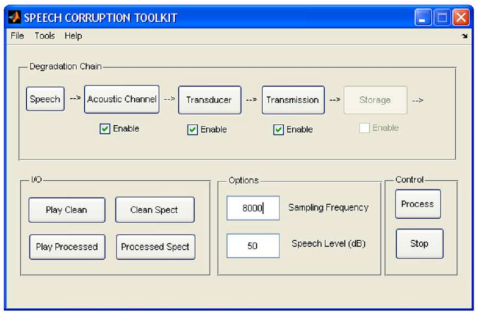
- Measurement
  - get a good understanding of how well existing systems work on a variety of signal types
- Modelling
  - build models which predict how well systems would work on specific signals
- Enhancement
  - choose the methods which we predict would have the best performance on a given signal
  - develop new methods based on improved understanding of speech perception in noise

June 2012

Measurement

Speech Corruption Toolkit

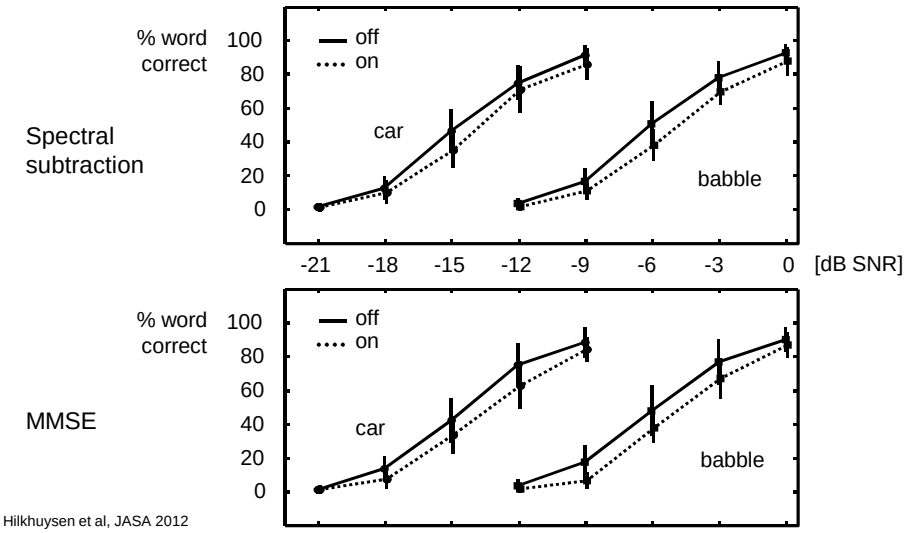




AUDIO!

June 2012

Noise Reduction Technology Evaluation



Hilkhuysen et al, JASA 2012

June 2012

Commercial Systems Evaluation

- 12 modules from 5 commercial systems, up to 5 parameter settings per module
- 5 recording scenarios, each with/without CODEC
- Adaptive procedure for estimating SRT
- Closed response set task
- 2400 SRTs estimated, >900hrs of subject tests

June 2012

Commercial Systems Evaluation

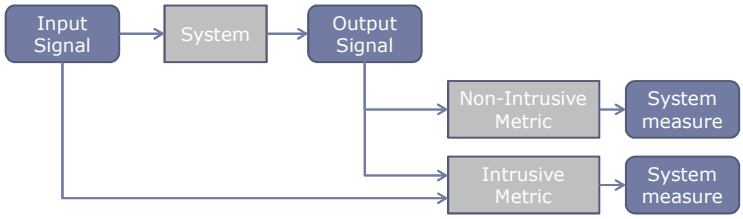
SRT shift (dB)		Scenario				
		S1	S2	S3	S4	S5
Product	P1	-3.7	-3.4	-3.4	-3.4	-3.4
	P2	-1.3	0	0	0	0
	P3	-1.2	-3	-1.2	-1.2	-1.2
	P4	-1.6	-1.6	-1.6	-1.6	-1.6
	P5	-2.5	-1.6	-1.6	-1.6	-3
	P6	-2.3	-0.9	-0.9	-2.1	-0.9
	P7	-1.7	-1.7	-1.7	-1.7	-1.7
	P8	-3.3	-1.9	-1.9	-1.9	-1.9
	P9	-2.7	-2.7	-2.7	-2.7	-2.7
	P10	-3.8	-1.3	-1.3	-1.3	-1.3
	P11	-0.2	0	0	0	0

- All scenarios could be made better
  - SRT improved by up to 2.0 dB
- All scenarios could be made worse
  - SRT degraded by up to 3.8 dB
- No module consistently best
- No parameter settings consistently best

June 2012

Modelling

Estimation of Signal Properties

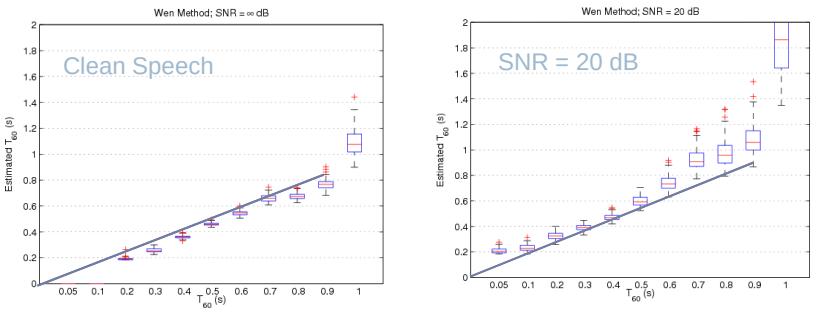


- Analyse signals to estimate nature of corruption:
  - Noise, SNR, Reverberation, Channel, CODEC
- Estimate intelligibility/quality of signal regions
  - Non-intrusive prediction of STOI
  - Non-intrusive prediction of PESQ

June 2012

Example: T60 Estimation

- T60 = Reverberation time = time for energy to decay by 60 dB
- Estimate by looking at peak decay rates in speech
- Excellent results in clean speech (SNR > 20 dB)
- T60 is overestimated in very reverberant or noisy conditions

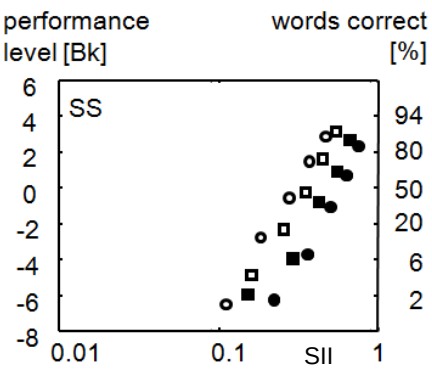


Gaubitch et al, 2012

June 2012

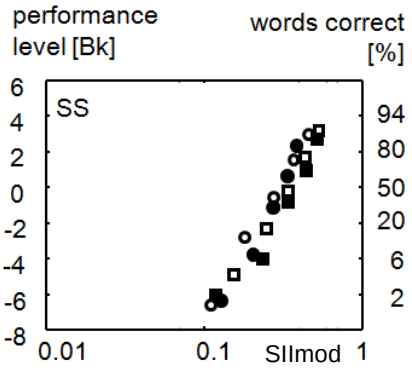
Intelligibility Modelling and Prediction

Speech Intelligibility Index



SII incorrectly predicts that NR improves intelligibility

SII from Envelope Modulations



SIImod provides a better, though still imperfect prediction

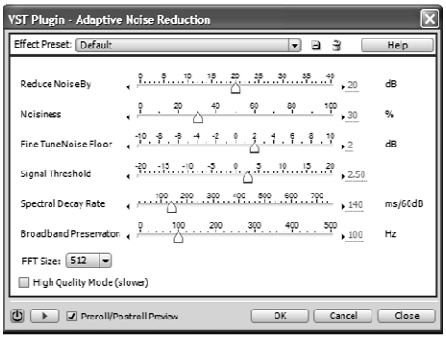
Hilkhuysen et al, in preparation

June 2012

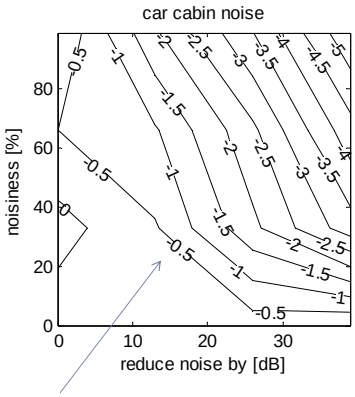
Enhancement

Optimising use of commercial system

Adobe Audition Noise Reduction



Change in intelligibility



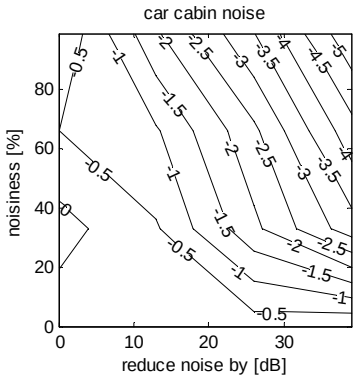
Change measured in log odds ratio
(-1 = double errors for same # correct)

Hilkhuysen & Huckvale
AES Audio Forensics 2010

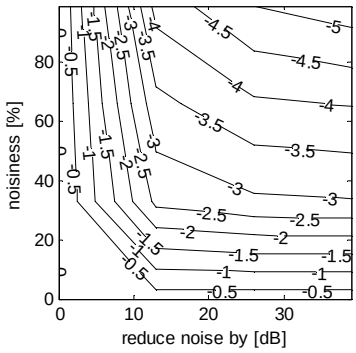
June 2012

Optimising use of commercial system

Observed



Predicted



Hilkhuysen & Huckvale
AES Audio Forensics 2010

June 2012

Blind channel estimation

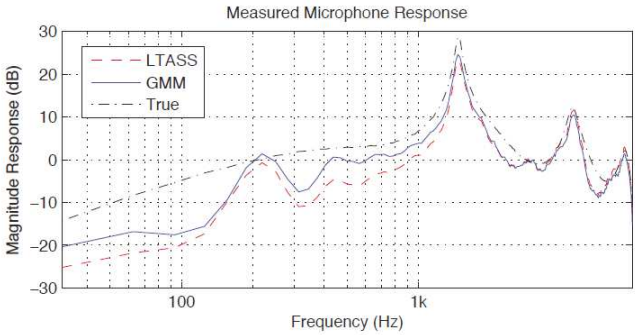


Figure 5: Estimates of a measured microphone response in the noise-free case. The WSD is 5.5 dB with LTASS based estimation and 2.9 dB with GMM based estimation.

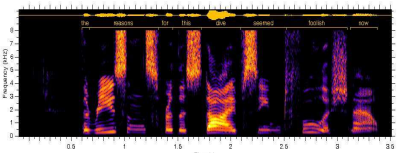
Gaubitch et al, 2011

June 2012

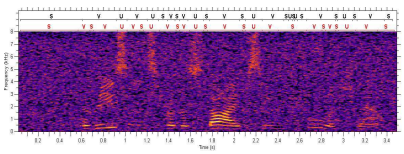
Binary Masks

- Noise-robust pitch estimation (PEFAC)
- HMM segmentation into Voiced/Unvoiced/Silent
  - inputs are voiced and sibilant detector outputs
  - imposes temporal continuity
- Mask generation:
  - Voiced: 10 harmonics
  - Unvoiced: Estimated sibilant freqs
  - Silent: None
  - Add original at -20 dB
- Segmentation good
  - 80% @ 0 dB SNR
- Working on improved mask generation

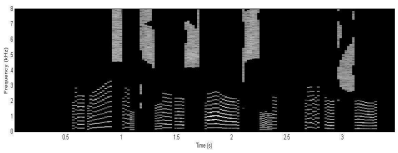
Clean Speech



SNR = 0 dB white noise



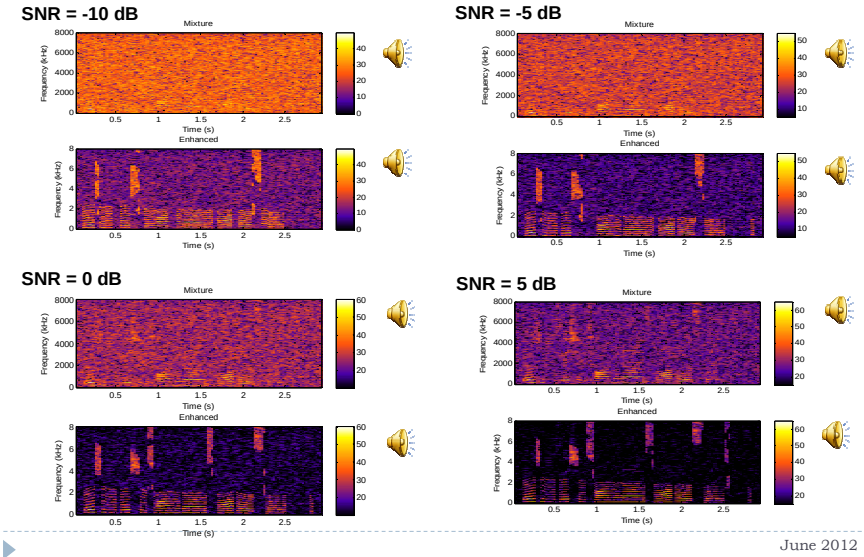
Estimated Mask



Gonzales and Brookes, 2011

June 2012

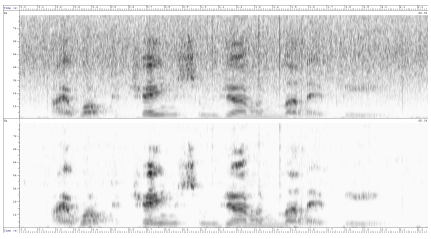
Binary Mask Examples



June 2012

Conclusions

- The CLEAR Project seeks to advance the science of speech signal enhancement through application of models based on quantitative measurement.



- Contact us for more information
- Visit www.clear-labs.com for publications

June 2012