Boosting speech intelligibility using spectral reweighting under a constant energy constraint

Yan Tang[1]
Martin Cooke[2,1]

[1]Language and Speech Lab, Universidad del País Vasco, Spain
[2]Ikerbasque (Basque Foundation for Science)

Introduction

Speech intelligibility in noise can be increased without changing overall RMS energy by reallocating energy to frequency regions which result in masking release. Different combinations of maskers and speech signals might be expected to require different spectral weights for optimal masking release. In a previous study [1], a genetic-algorithm based spectral weight optimisation procedure with a simple objective intelligibility measure led to gains of up to 15 percentage points. The current study examines the performance of weights discovered via a pattern search optimisation procedure using a more sophisticated intelligibility metric.

Spectral weighting and objective intelligibility

Spectral weighting

$$\log|S'(f)| = \log|S(f)| + \log|W(f)|$$

S, S' original and modified speech spectrum at frequency channel f
W weighting applied to frequency channel f

Energy and duration constraints

$$\sum_{t=1}^{T'} s'(t)^2 = \sum_{t=1}^{T} s(t)^2, \text{and } T' = T$$

s, s' original and modified speech waveform

Objective intelligibility prediction

The extended glimpse proportion (xGP) [2] augments the raw GP [3] with terms representing audibility, duration, detectability and glimpse redundancy:

$$xGP = v[\frac{1}{K_o F}\sum_{f=1}^{F}\sum_{k=1}^{K}\mathcal{H}(S'_{k,f} - (N_{k,f}+\alpha)) \wedge (Y_{k,f} > \max(HL, \bar{Y}_f))]$$

with compressive nonlinearity

$$v(x) = \frac{\log(1+x/\delta)}{\log(1+1/\delta)}$$

K_o, K number time frames in unmodified & modified speech
F number of frequency channels (34)
$\mathcal{H}(.)$ Heaviside unit step function
α local SNR threshold (3 dB)
$S'_{k,f}$ modified speech spectro-temporal excitation pattern at k and f
$N_{k,f}$ noise spectro-temporal excitation pattern at k and f
$Y_{k,f}$ speech+noise spectro-temporal excitation pattern at k and f
$\bar{Y}_f$ cross-time average of $Y_{k,f}$
HL hearing level (25 dB)
δ offset to avoid log zero (0.01)

Optimisation algorithm

Pattern search [4]

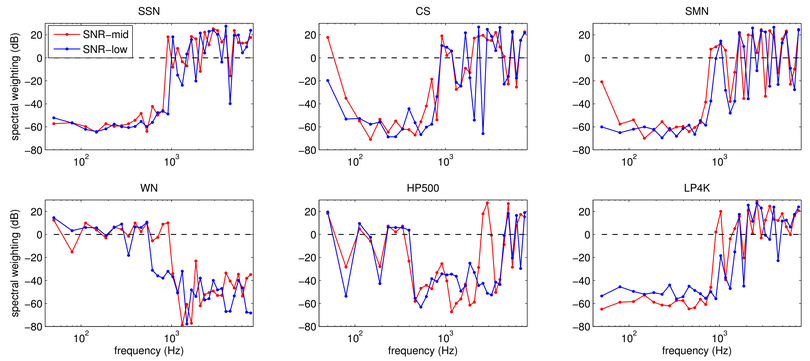
- objective function: xGP
- design variables: spectral wieghts, W
- boosting bounds: [-50, +50] dB
- stopping criterion: max iterations (200)
- number of trials: 2

Optimised and approximated spectral weightings

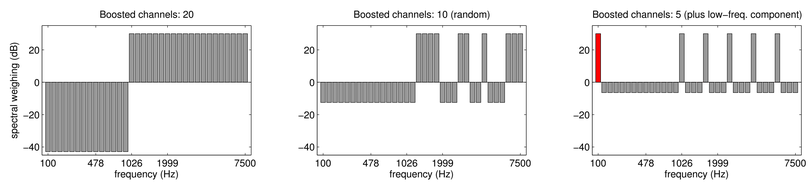
Maskers

- speech-shaped noise (SSN)
- competing speaker (CS)
- speech-modulated noise (SMN)
- white noise (WN)
- 500 Hz high-pass filtered SSN (HP500)
- 4 kHz low-pass filtered SSN (LP4K)

Noise- and level-dependent optimal spectral weighting (expt 1)



Noise- and level-independent approximated weightings (expt 2)



Evaluation

Keyword identification in phonemically-balanced Spanish sentences (Sharvard Corpus; [5])

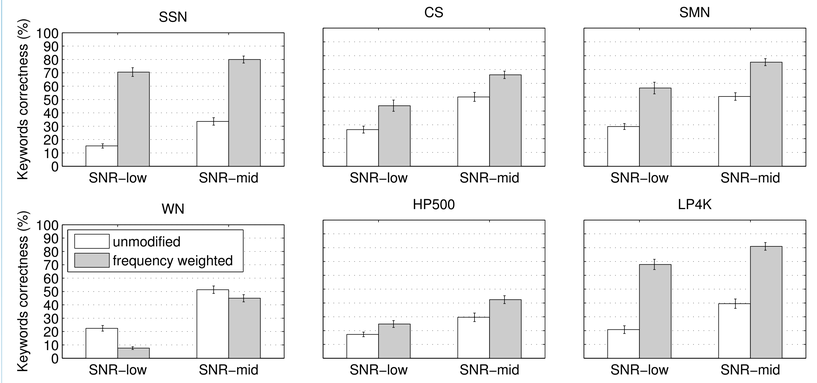
Expt 1 Effect of noise and level-dependent optimal weightings for 6 maskers

Expt 2 Effect of noise and level-**independent** approximated weighting; tested with CS, SSN and SMN only

- 22 native Spanish listeners for each experiment
- Two SNRs, different for each masker, predicted to produce 25 and 50% keywords correct using xGP metric

Results

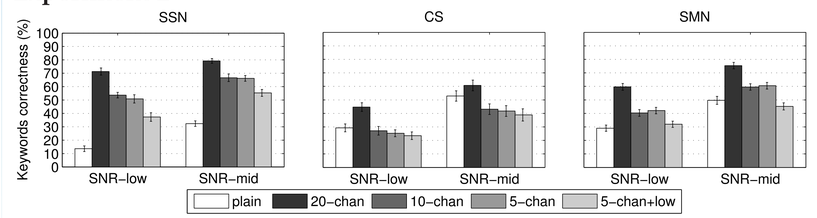
Experiment 1



Error bars for both expts indicate ± 1 standard error

- Keyword scores for reweighted speech increased by 8 to 55 percentage points for most maskers
- Decrease in intelligibility for the white noise masker is observed. Failure of objective intelligibility model?

Experiment 2



- Intelligibility gains for 20-channel boosting similar to those in expt. 1
- Decreasing the number of boosted channels leads to intelligibility reduction
- Very low frequency boost – observed in optimal patterns – proved unhelpful
- Statistically: 20-chan $>$ 10-chan $=$ 5-chan $>$ 5-chan+low (mainly)

Discussion

- Noise- and level-dependent optimal spectral weighting can lead to very substantial intelligibility gains.
- Noise- and level-**independent** spectral weighting is nearly as effective as those customised for specific maskers. These findings point to a practical mechanism for intelligibility enhancement in some common noise conditions.
- Further work will investigate boosting strategies for maskers with a uniform or high-pass characteristic, and evaluate alternative objective intelligibility measures more sensitive to frequency regions important for speech comprehension.

[1] Tang, Y. & Cooke, M. (2012). Optimised spectral weightings for noise-dependent speech intelligibility enhancement. Proc. Interspeech.
[2] Tang, Y. & Cooke, M. (in preparation). A glimpse-based intelligibility metric for plain, enhanced and synthetic speech in additive noise conditions.
[3] Cooke, M. (2006). A glimpsing model of speech perception in noise, J. Acoust. Soc. Am. 119, 1562-1573.
[4] Hooke, R. & Jeeves, T.A. (1961). "Direct search" solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 8(2): 212–229.
[5] Aubanel, V., Garcia Lecumberri, M. L., & Cooke, M. (in revision). The Sharvard corpus: A phonemically-balanced Spanish sentence resource for audiology. *Int. J. Audiology*.

This work was supported by the EU Future and Emerging Technology Project LISTA (The Listening Talker)