IMPROVING SPEECH PRIVACY IN PERSONAL SOUND ZONES

Jacob Donley*, Christian Ritz* and W. Bastiaan Kleijn†
* School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia
† School of Engineering and Computer Science, Victoria University of Wellington, New Zealand
jrd089@uowmail.edu.au, critz@uow.edu.au, bastiaan.kleijn@vuw.ac.nz

Shanghai, China
ICASSP·2016
The 41st IEEE International Conference on Acoustics, Speech and Signal Processing, 20-25 March 2016

1 ABSTRACT

This poster proposes two methods for providing speech privacy between spatial zones in anechoic and reverberant environments. The methods are based on masking the content leaked between regions. The masking is optimised to maximise the speech intelligibility contrast (SIC) between the zones. The first method uses a uniform masker signal that is combined with desired multizone loudspeaker signals and requires acoustic contrast between zones. The second method computes a space-time domain masker signal in parallel with the loudspeaker signals so that the combination of the two emphasises the spectral masking in the targeted quiet zone. Simulations show that it is possible to achieve a significant SIC in anechoic environments whilst maintaining speech quality in the bright zone.

2 WEIGHTED MULTIZONE SPEECH SOUNDFIELDS

- Orthogonal basis expansion multizone soundfield function

$$S\left(\mathbf{x},k\right)=\sum_{j}P_j\left(k\right)F_j\left(\mathbf{x},k\right) \tag{1}$$

- Complex loudspeaker weights in the time-frequency domain

$$\tilde{Q}_l(k)=\sum_{m=-M}^{M}\frac{2e^{im\phi_l}\Delta\phi_s\sum_j\left(P_j\left(k\right)i^m e^{-im\phi_p}\right)}{i\pi H_m^{(1)}\left(kR_l\right)} \tag{2}$$

- Loudspeaker signal of the a^{th} frame in the time-domain

$$\tilde{q}_{al}\left(n\right)=\frac{1}{2K}\sum_{m=0}^{K-1}\tilde{Q}_l(m\Delta k)\tilde{Y}_a(m\Delta k)e^{i\pi mn/K} \tag{3}$$

- Loudspeaker signals, $q_l(n)$, constructed from (3) using overlap-add.
- Signals can be observed at any arbitrary point in the soundfield

$$p(\mathbf{x},n)=\frac{1}{2K}\sum_{l}\sum_{m=0}^{K-1}Q_l(m\Delta k)H(\mathbf{x},\mathbf{x}_l,m\Delta k)e^{i\pi mn/K}. \tag{4}$$

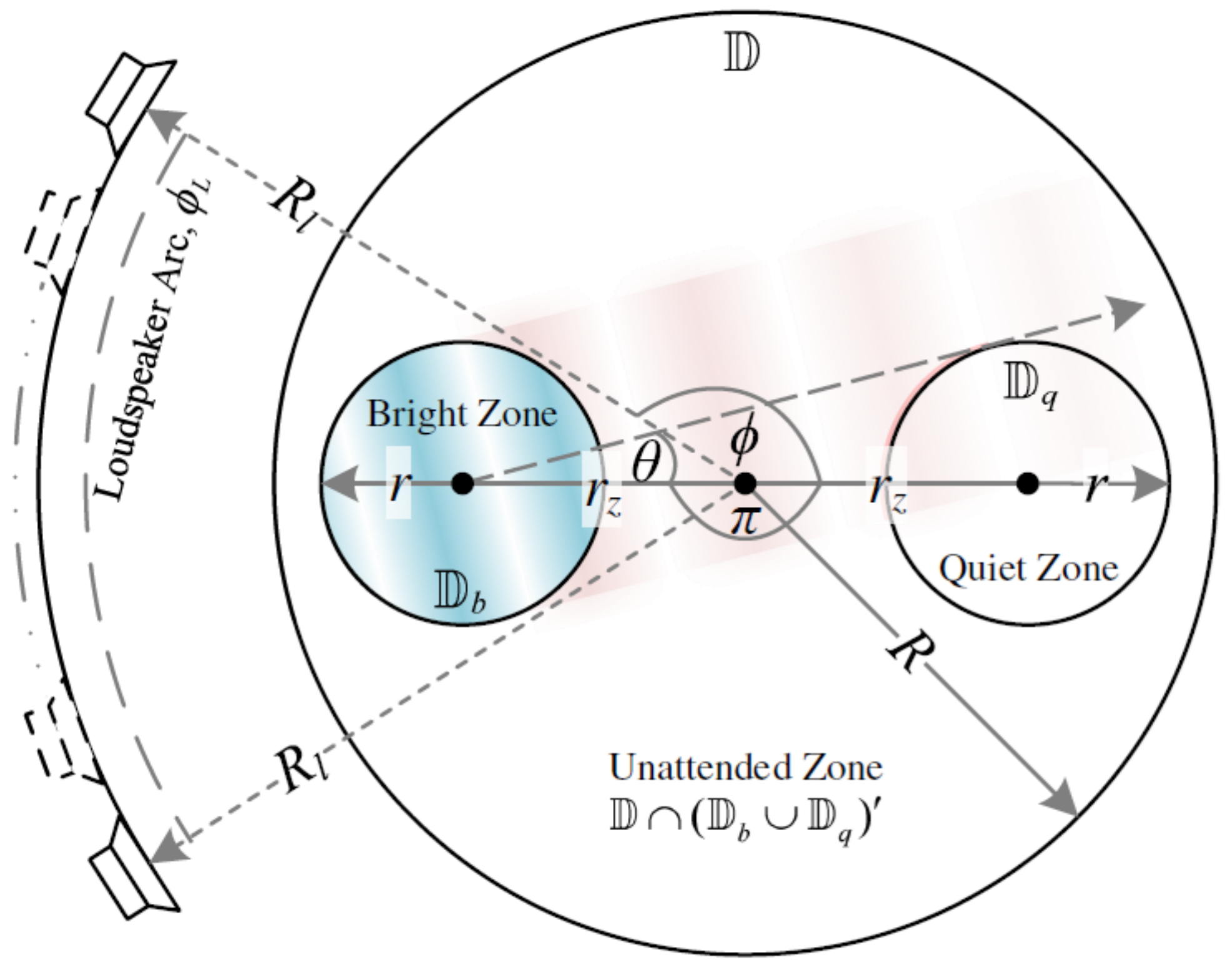

Fig. 1. – A weighted multizone soundfield reproduction layout is shown. The shading depicts the bright zone soundfield partially directed towards the quiet zone causing the occlusion problem.

3 PRIVATE SOUND ZONES

We propose a new measure: Speech Intelligibility Contrast (SIC)
It can be used to improve speech privacy in personal sound zones.

3.1 – Speech Privacy and Intelligibility Contrast

- Speech intelligibility and privacy are highly correlated.
- Objective (instrumental) and subjective intelligibility measures are highly correlated.
- A proxy of mutual information, such as STOI or STI, of two signals $x_1(n)$ and $x_2(n)$ from two zones is $\mathcal{I}_{\mathcal{M}}(x_1;x_2)$.
- SIC defined as

$$\mathrm{SIC}_{\mathcal{M}}=\frac{1}{||\mathbb{D}_b||}\int_{\mathbb{D}_b}\mathcal{I}_{\mathcal{M}}\,d\mathbf{x}-\frac{1}{||\mathbb{D}_q||}\int_{\mathbb{D}_q}\mathcal{I}_{\mathcal{M}}\,d\mathbf{x}. \tag{5}$$

3.2 – Improving Multizone Privacy

- Maximise the SIC by optimising the gain, G, of noise added to the loudspeaker signals.
- The optimisation becomes

$$\max_{G\in\mathbb{R}}\mathrm{SIC}_{\mathcal{M}}. \tag{6}$$

- New <u>'Flat Mask' (spatially uniform)</u> loudspeaker signals are

$$q_l'(n)=q_l(n)+u(n)A10^{\frac{G_{\mathrm{dB}}}{20\,\mathrm{dB}}}. \tag{7}$$

  - ➢ $A=\max(\{q_l(n):l=1,...,L\})$ is the maximum amplitude among L loudspeakers.
  - ➢ $u(n)$ is a time-domain noise masker.
  - ➢ G_{dB} is the noise level G in decibels.

3.3 – Improving Multizone Privacy and Quality

- Adding $u(n)$ to $q_l(n)$ reduces bright zone speech quality.
- Speech quality in the bright zone is $\mathcal{B}_{\acute{\mathcal{M}}}(p(\mathbf{x},\cdot);y)\in\{0,...,1\}$
- The optimisation becomes

$$\max_{G_{\mathrm{dB}}\in\mathbb{R}}\mathrm{SIC}_{\mathcal{M}}+\frac{\lambda}{||\mathbb{D}_b||}\int_{\mathbb{D}_b}\mathcal{B}_{\acute{\mathcal{M}}}\,d\mathbf{x}. \tag{8}$$

  - ➢ where λ is a weighting parameter for importance of quality.

- New <u>'Zone Weighted Mask' (spatially non-uniform)</u> loudspeaker signals to be optimised for SIC and quality are

$$q_l''(n)=q_l(n)+\hat{q}_l(n)A10^{\frac{G_{\mathrm{dB}}}{20\,\mathrm{dB}}}. \tag{9}$$

  - ➢ $\hat{q}_l(n)$ are a new set of loudspeaker signals for a multizone reproduction of $u(n)$ using (2), (3) and overlap-add.

4 RESULTS

Objective intelligibility and quality results including corresponding SIC and quality trade-off are presented.

4.1 – Multizone Reproduction Evaluation

$r=0.3\text{m}$
$r_z=0.6\text{m}$
$R=1.0\text{m}$
$R_l=1.5\text{m}$
$\theta=\{0°,15°,90°\}$

$L=295$
$\phi_L=2\pi$
$G_{\text{dB}}=$
$-40\text{dB to }20\text{dB}$

4 rooms =
Anechoic,
4m × 9m × 3m,
8m × 10m × 3m,
9m × 14m × 3m

Absorb Coeff = 0.3
20 Speech files
M:F = 50:50
≈ **332,800 data points**

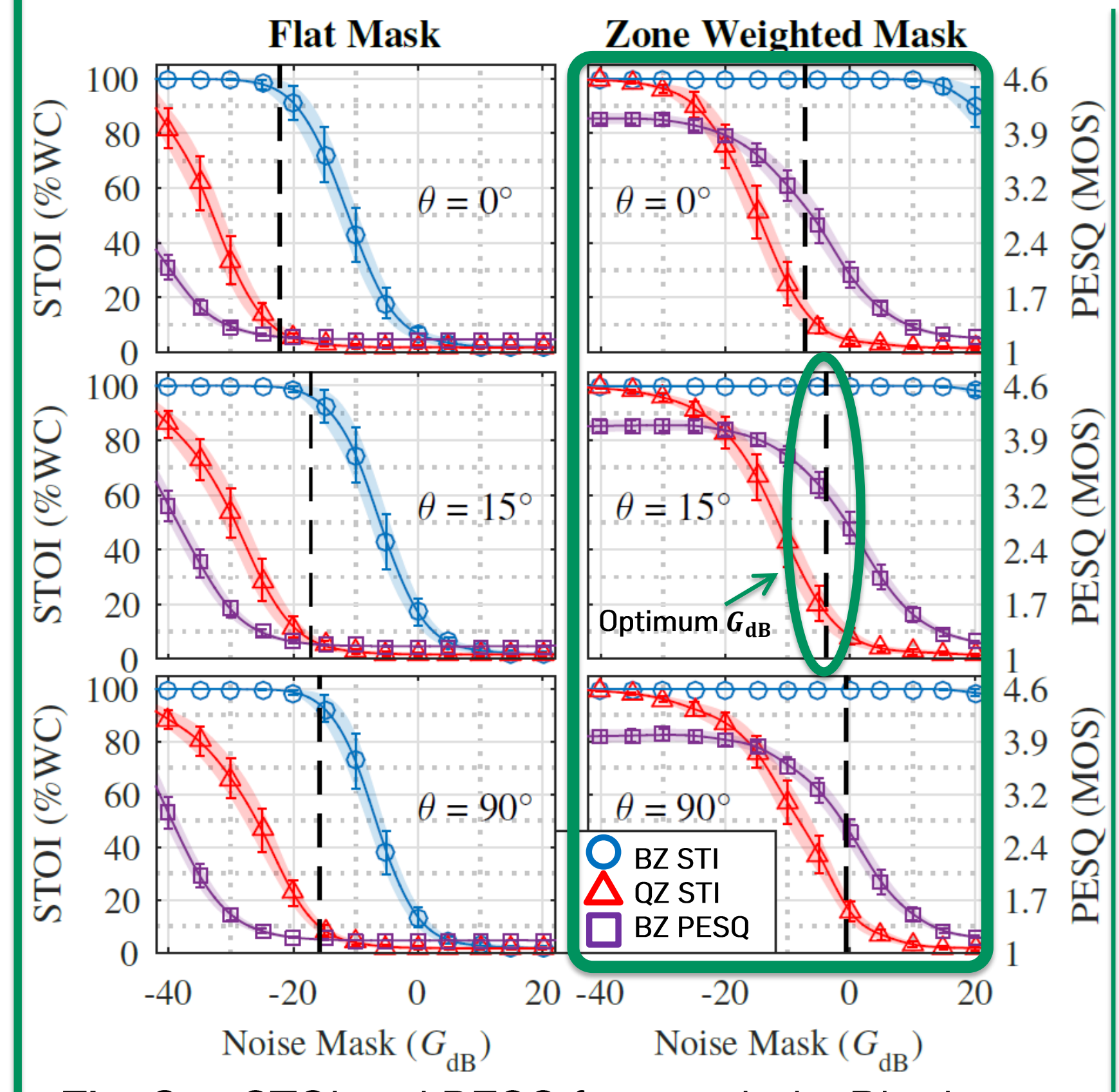

Fig. 2. – STOI and PESQ for anechoic. Black dashed lines are optimum G_{dB} and $\lambda=1$.

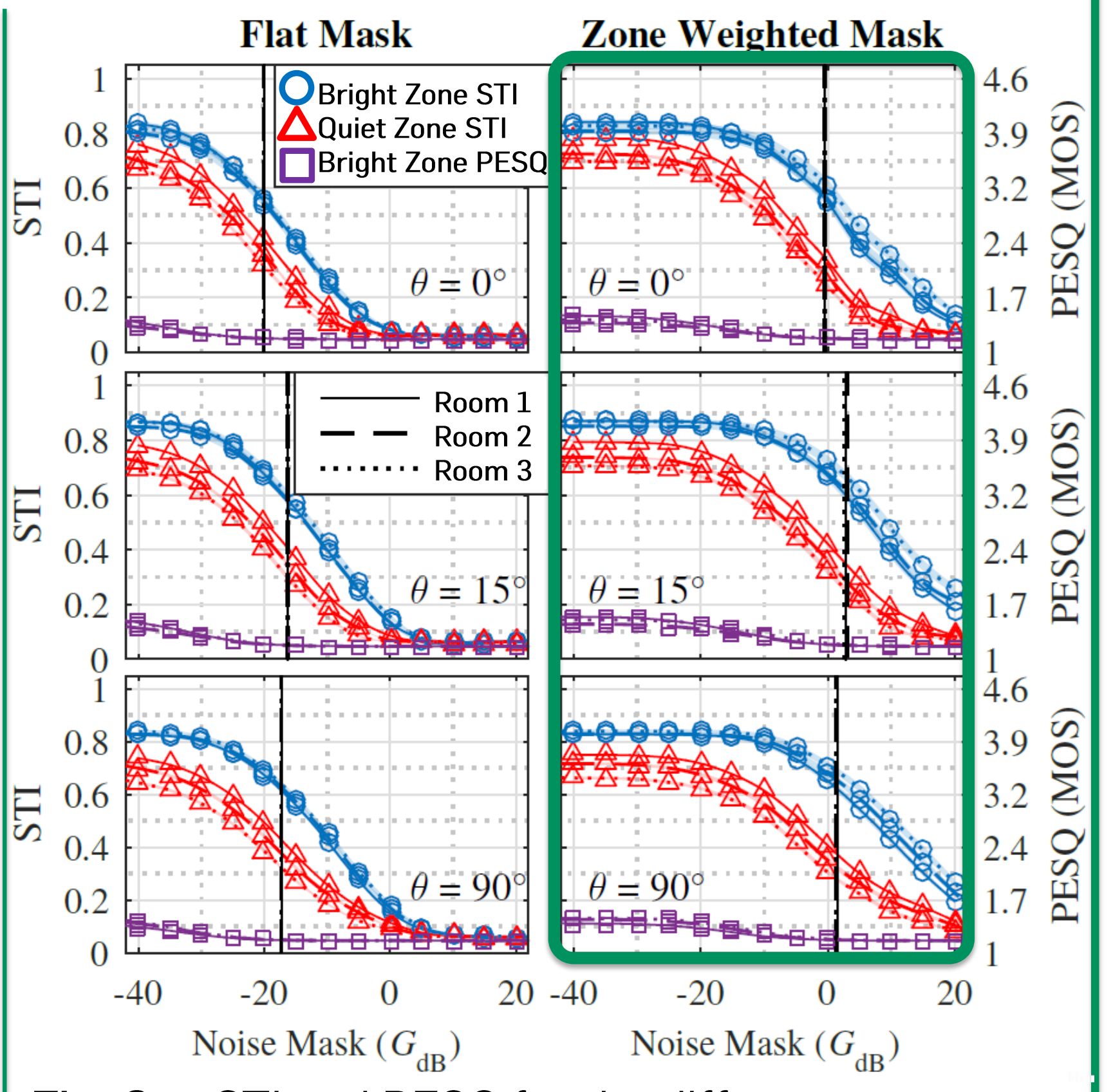

Fig. 3. – STI and PESQ for the different rooms. Vertical black lines are optimum G_{dB} and $\lambda=1$.

4.2 – Intelligibility Contrast from Noise-Based Sound Masking

Anechoic
Flat Mask:
$\mathrm{SIC}_{\mathrm{STOI}}=85\%$ 👌
$\mathrm{PESQ}<1.2$ 👎

Anechoic
Zone Weighted Mask:
$\mathrm{SIC}_{\mathrm{STOI}}=70\text{ to }90\%$ 👌
$\mathrm{PESQ}=3.4\text{ to }2.8$ 👍

Reverberant
Zone Weighted Mask:
$\mathrm{SIC}_{\mathrm{STI}}=40\%$ 👍
$\mathrm{PESQ}<1.2$ 👎
(Not optimised for room)

5 CONCLUSIONS

- Proposed and evaluated methods for increasing speech intelligibility contrast in anechoic and reverberant environments.
- Shown possible to obtain a SIC value higher than 95%.
- Shown possible to maintain quality in the bright zone with a PESQ MOS of 3.2 and SIC above 80%.
- Shown possible to obtain a SIC value up to 40% in reverberant environments with no dereverberation/equalisation.
- Future work:
  - ➢ Further improve quality and privacy with less loudspeakers
  - ➢ Optimise for reverberant environments

RESEARCH INSTITUTE
GLOBAL CHALLENGES
UNIVERSITY OF WOLLONGONG AUSTRALIA
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA UNIVERSITY OF WELLINGTON