Institute of Sound and
Vibration Research

UNIVERSITY OF Southampton

Combining Artificial and Natural Background Noise in Personal Audio Systems

Daniel Wallace, Dr. Jordan Cheer

9th July 2018

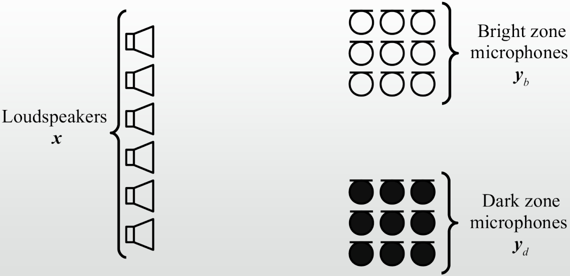
EPSRC
Engineering and Physical Sciences
Research Council

NGCM

Introduction

UNIVERSITY OF Southampton

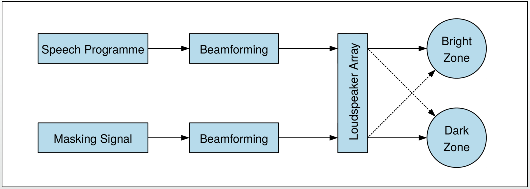
- Personal Audio refers to the production of sound zones, where individuals may listen to their own audio in a shared space.
- A number of methods exist for producing these zones using loudspeaker arrays.
- Traditionally, personal audio system performance has been measured in terms of energy difference between zones.
- This does not take into account the perceptual relevance of the sound leaked from one zone to the other.



Introduction

University of Southampton

- A consequence of not taking into account the *meaning* of the leaked sound is that private speech intended for one listener may remain intelligible outside of the bright zone [1].
- By radiating a masking signal into the dark zone, speech intelligibility can be reduced in sensitive locations.
- Perceptual optimisation must take into account both the intended listener of the system and others in the vicinity.



1. D. Wallace and J. Cheer, 'Optimisation of Personal Audio Systems for Intelligibility Contrast' in 144th AES Convention, Milan, 2018. Online: eprints.soton.ac.uk/420254/

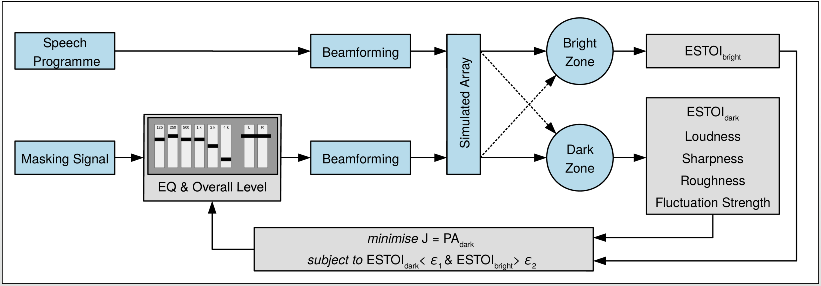
Optimisation

UNIVERSITY OF Southampton

To be perceptually acceptable, the system must satisfy a number of objectives:

- High intelligibility in the bright zone.
- Low intelligibility in the dark zone.
- Low annoyance in the dark zone (for passive listeners nearby).

As intelligibility can be objectively measured, meaningful constraints can be set for the intelligibility in the bright and dark zones.



Optimisation

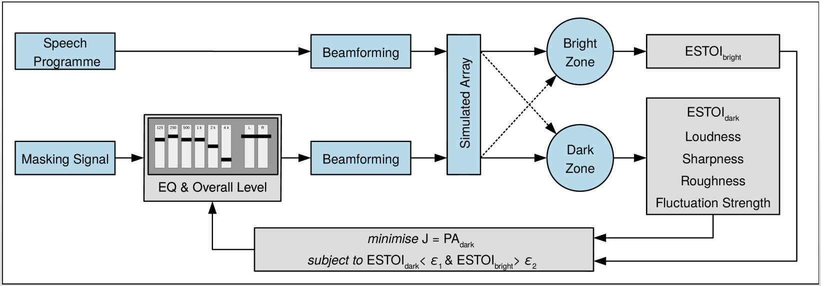
UNIVERSITY OF Southampton

Bright and dark zone sound fields are evaluated using two metrics:

- Extended Short Term Objective Intelligibility (Jensen and Taal, 2016)
- Psychoacoustic Annoyance (Zwicker and Fastl, 2007)

$$PA = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2}\right)$$

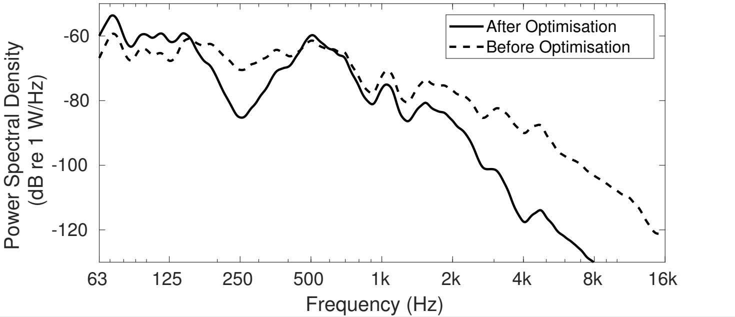
N_5 = Loudness exceeded 5% of signal duration;
w_S = Sharpness contribution;
w_{FR} = Fluctuation Strength and Roughness Contribution.



Optimal Masking Signal Spectrum

UNIVERSITY OF Southampton

- After optimisation, masking signal spectra take on a low-pass characteristic above 1 kHz.
- This equalisation decreases sharpness, while maintaining the ability to effectively mask speech.

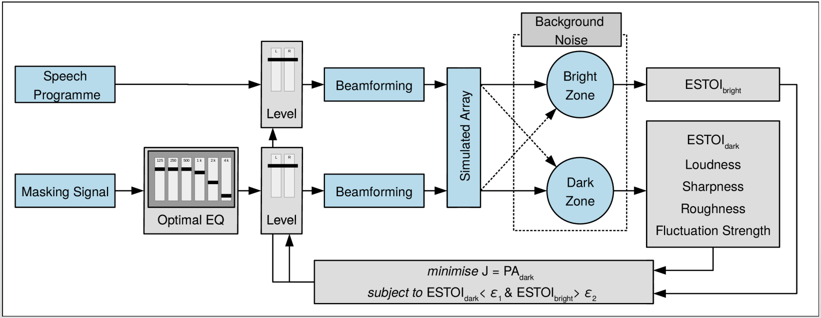


Power Spectral Density of speech shaped noise masker before and after optimisation

Incorporating Background Noise

UNIVERSITY OF Southampton

- Simulations used to find the optimal masking signal assumed the system was operating in quiet. In situations where privacy is a concern, this is not often the case.
- Background noise will reduce intelligibility in both zones.
- By including background noise in simulations and optimising as before, appropriate programme and masker levels relative to background can be found.



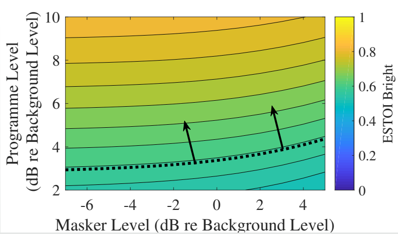
Simplifications to Simulations

- The principal difference between bright and dark zone sound fields are the relative levels of programme, masker and background noise.
- In order to facilitate general conclusions, the sound fields in each zone are simulated by summing monophonic programme, masker and background signals with appropriate weightings, given by a broadband acoustic contrast level.
- The effects of frequency dependent acoustic contrast, choice of sound zoning method, spatial masking effects and target quality are discussed later.

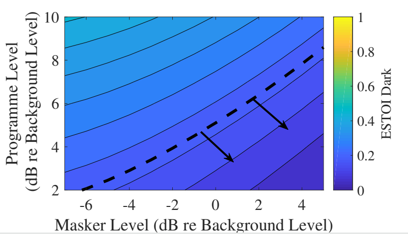
Results: Cost and Constraint Surfaces

UNIVERSITY OF Southampton

- The shape of the constraint functions as programme and masker levels are varied can be visualised as an ESTOI surface for each zone.
- Dashed contours show the position of the intelligibility constraints in each zone.
- In these plots, 10 dB of broadband acoustic contrast is assumed.



ESTOI in Bright Zone
Above contour, ESTOI > 0.6

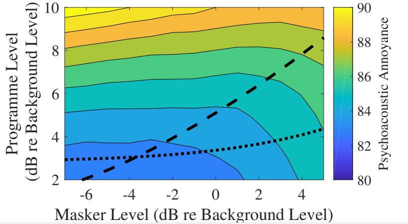


ESTOI in Dark Zone
Below contour, ESTOI < 0.2

Results: Cost and Constraint Surfaces

UNIVERSITY OF Southampton

- The two feasibility contours from the previous slide can be applied to the Psychoacoustic Annoyance cost surface.
- The optimal programme and masking signal levels for minimising annoyance occurs when both constraints on intelligibility are met.



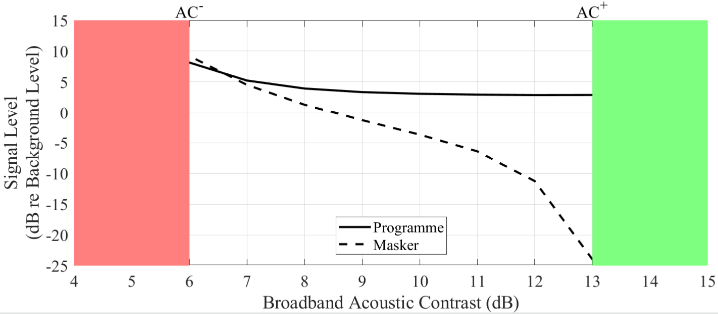
Psychoacoustic Annoyance in Dark Zone.
Region between contours
satisfies intelligibility constraints.

Parameter	Value
ESTOI Bright	0.6
ESTOI Dark	0.2
Acoustic Contrast	10 dB
Optimal Masker	-4 dB
Optimal Programme	3 dB

Optimal Programme and Masker Levels

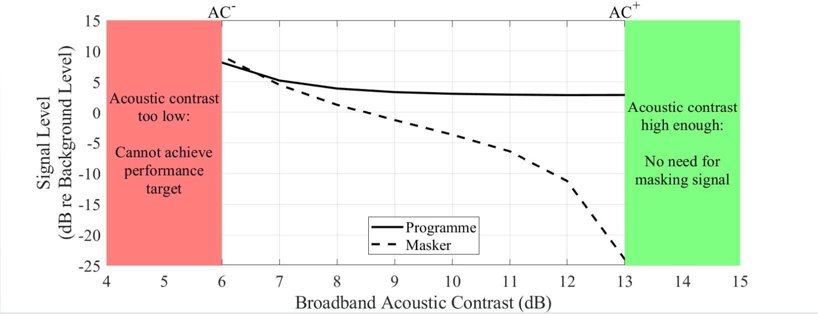
UNIVERSITY OF Southampton

- Optimal program and masker levels are dependent on the level of acoustic contrast delivered by the personal audio system.
- This allows us to see the advantage which may be gained by incorporating a masking signal into a design.
- For ESTOI Dark = 0.2, ESTOI Bright = 0.6:



Optimal Programme and Masker Levels

- Optimal program and masker levels are dependent on the level of acoustic contrast delivered by the personal audio system.
- This allows us to see the advantage which may be gained by incorporating a masking signal into a design.
- For ESTOI Dark = 0.2, ESTOI Bright = 0.6:



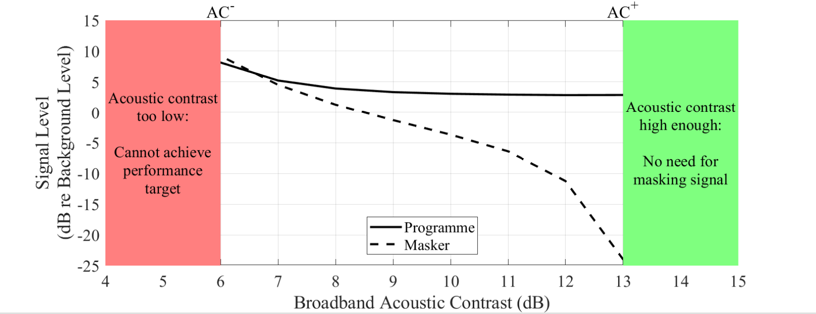
Adjusting Performance Requirements

UNIVERSITY OF Southampton

The position of AC^- and AC^+ is affected by the ESTOI constraints.

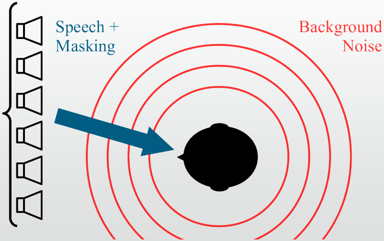
AC^- (dB)		ε_d		
		0.1	0.2	0.3
	0.6	**8**	**6**	**4**
ε_b	0.7	**9**	**7**	**6**
	0.8	**11**	**9**	**7**

AC^+ (dB)		ε_d		
		0.1	0.2	0.3
	0.6	**18**	**13**	**9**
ε_b	0.7	**21**	**16**	**12**
	0.8	**24**	**19**	**15**



Discussion: Spatial Effects

- There is a spatial difference between the diffuse background noise and the artificial masking noise emitted by the array - in poor SNR conditions humans can often still understand speech due to its constant direction of arrival.
- Intelligibility estimates for speech masked by background noise alone are likely to be under-predictions. This suggests there is an additional advantage to providing a masking signal via the array.
- Binaural intelligibility metrics which correct for the human ability to discern speech in diffuse noise may be included in future simulations.



Discussion: Frequency Dependencies

UNIVERSITY OF Southampton

- We have modelled acoustic contrast between zones using a single broadband index. In practice, different zoning methods produce frequency dependent levels of acoustic contrast.
- Acoustic contrast cannot be reliably produced over the spatial aliasing frequency of the array.
- For a given array and sound zone geometry, the position of high frequency grating lobes can be predicted[2] and the masking signal low-pass filtered to reduce the contribution of incorrectly reproduced content.
- In practice, the optimised masking signal is low-pass filtered below 1 kHz for the purposes of reducing annoyance.

2. J. Donley et. al., 'Multizone Soundfield Reproduction With Privacy and Quality Based Speech Masking Filters'. IEEE/ACM Trans. Audio, Speech, Language Process.

Conclusions

- Limitations in array directivity can lead to a loss of privacy, or increased potential for distraction, due to leaked speech remaining intelligible outside of its intended zone.
- Simulation results using objective intelligibility measures and subjective metrics suggest that a secondary masking signal can be used to improve privacy.
- In order to achieve a certain privacy performance, given by ε_b and ε_d, less acoustic contrast is required when a masking signal is applied.
- This indicates that a smaller, cheaper array, or one which is more robust to environmental changes may be used, as less overall energy difference between zones is required.
- Alternatively, if an array is already in use, it may be repurposed at no additional hardware cost to provide better privacy.

Any Questions?

We acknowledge financial support from the
EPSRC Centre for Doctoral Training in
Next Generation Computational Modelling grant EP/L015382/1.