Paper ID: 9317

The effect of Partial Time-Frequency Masking of the Direct Sound on the Perception of Reverberant Speech

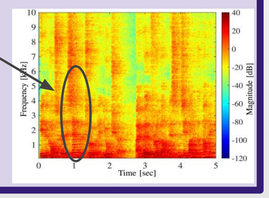
Lior Madmoni[1] , Shir Tibour [2], Israel Nelken [2], and Boaz Rafaely [1]
(1) Ben-Gurion University of the Negev
(2) The Hebrew University of Jerusalem

INTRODUCTION

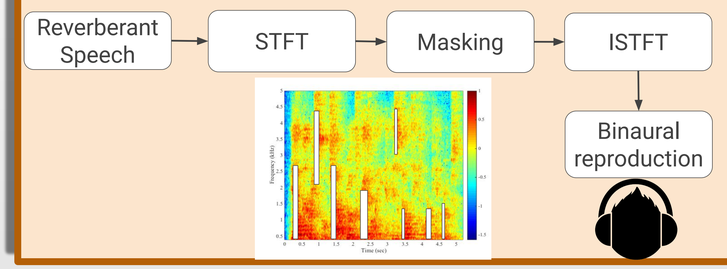
- **Spatial perception** of **speech**
  - affected by **reverberation**
  - **Direct sound** is important
- Spatial audio applications - **time-frequency (TF)** analysis
  - Detect TF bins of direct sound
  - Direct-to-reverberant ratio (**DRR**)
  - Enhance or reproduce

Hypothesis: spatial perception depends on DRR value

- What DRR values are "direct" ?
- Test hypothesis!



DIRECT SOUND MASKING



Processing stages:

$$a(k,\theta,\phi) = \sum_{n=0}^{N_a} \sum_{m=-n}^{n} a_{nm}(k) Y_n^m(\theta,\phi)$$

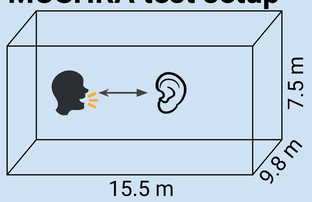
$$M(\tau,\omega) = \begin{cases} 0, & \mathrm{DRR}(\tau,\omega) > \mathrm{Thr} \\ 1, & \text{else} \end{cases}$$

$$a_{nm}^{\mathrm{masked}}(\tau,\omega) = M(\tau,\omega) \times a_{nm}(\tau,\omega)$$

$$p^{l,r}(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \left[\bar{a}_{nm}^{\mathrm{masked}}(k)\right]^* H_{n.m}^{l,r}(k)$$

LISTENING EXPERIMENTS

MUSHRA test setup

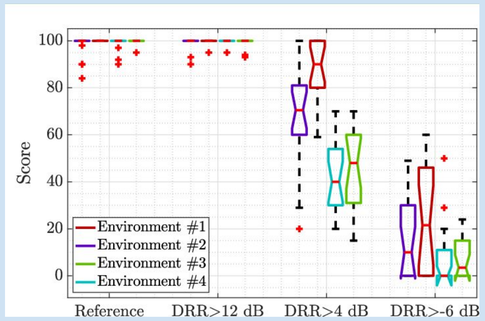


Acoustic environments

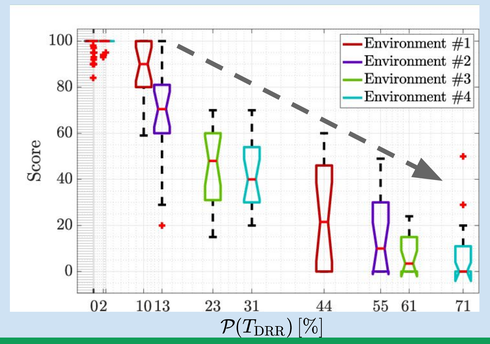
#	T60 [s]	r_s/r_d
1	1	2.7
2	0.6	2.7
3	1	1.5
4	0.6	1.5

Test signals
Exp. 1 - DRR threshold masking
Exp. 2 - FOA upscaling

Experiment 1 Results

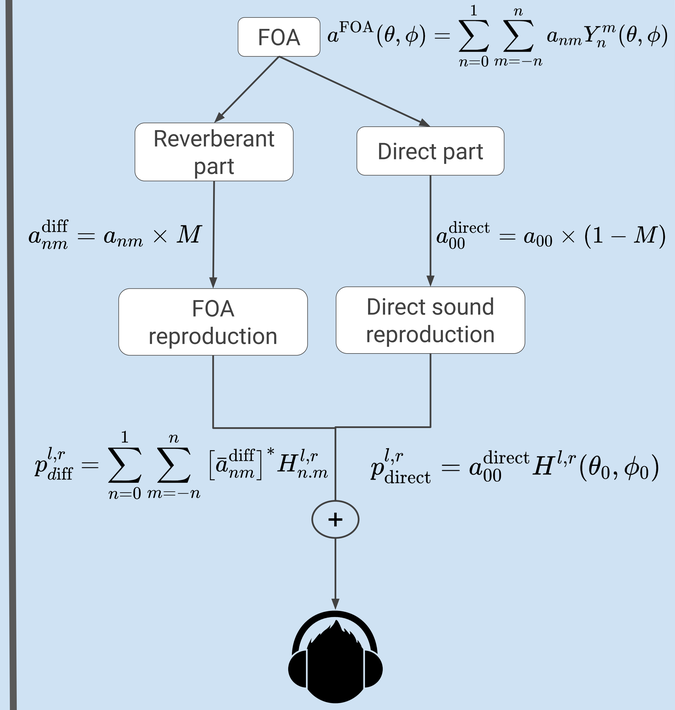


Score varies with acoustic environment

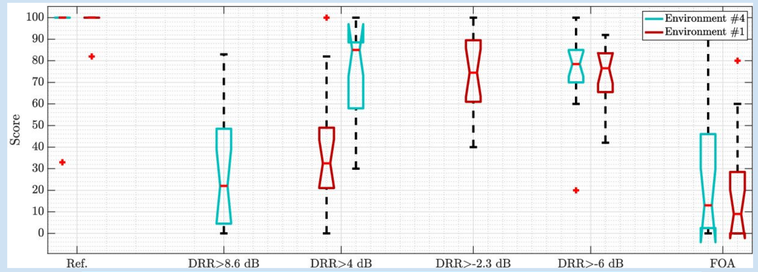


Percentage masked bins better correlates with scores!

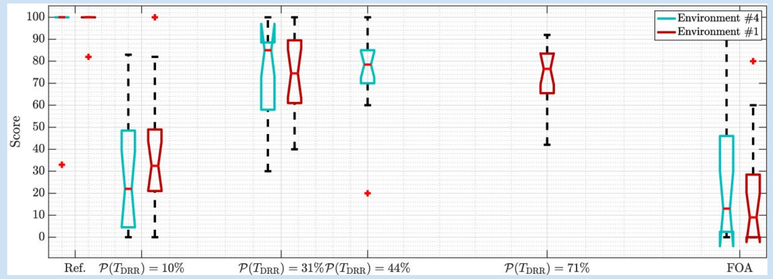
Experiment 2 - upscaling FOA



Experiment 2 Results



Similar DRR threshold - different perceived quality



Common percentage of masked bins - relatively similar perceived quality

CONCLUSIONS

- **Perceived quality** of reverberant speech depends on **DRR**
- Quality better indicated by **percentage** rather than **value** of DRR masking
- Insights may **improve design** of **spatial audio** algorithms

FUTURE WORK

- Extend independent variables - source directions, acoustic environments
- Multiple speakers
- Incorporate insights into spatial audio system

Published in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2037-2047, 2021.