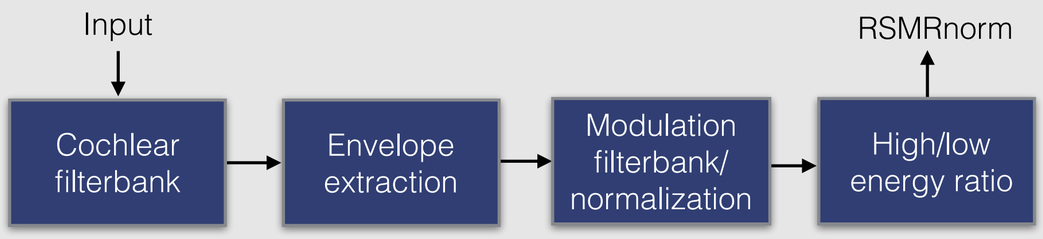
Improving blind reverberation time estimation on a two-microphone portable device by using speech source distance information

João F. Santos[1], Anderson Ávila[1], Rachel Bou-Serhal[2], Tiago H. Falk[1]
[1] INRS-EMT, University of Quebec, Montreal, QC, Canada
[2] École de Technologie Superieure, University of Quebec, Montreal, QC, Canada

Introduction

- Room reverberation has a significant impact in speech intelligibility and ASR performance, especially in the far field case.
- Blind RT estimation (based on reverberant speech only) can be a useful tool to improve performance.
- Blind estimates suffer from variability issues due to many factors, including source distance.
- We propose using a magnitude-squared coherence (MSC) based metric to integrate source distance into our model.

Materials and methods



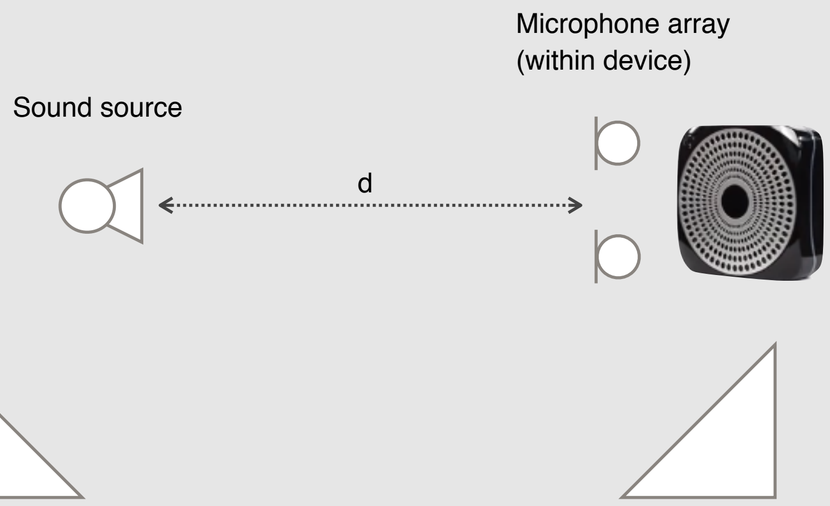
Block diagram of the RSMRnorm metric

$$\mathrm{MSC}_{xy}(f) = \frac{|X_{xy}(f)|^2}{X_{xx}(f)X_{yy}(f)}$$

$$\mathrm{MSCD}_{xy} = \sqrt{\sum_{n}^{N}(\mathrm{MSC}_{xy}(f_n) - 1)^2}$$

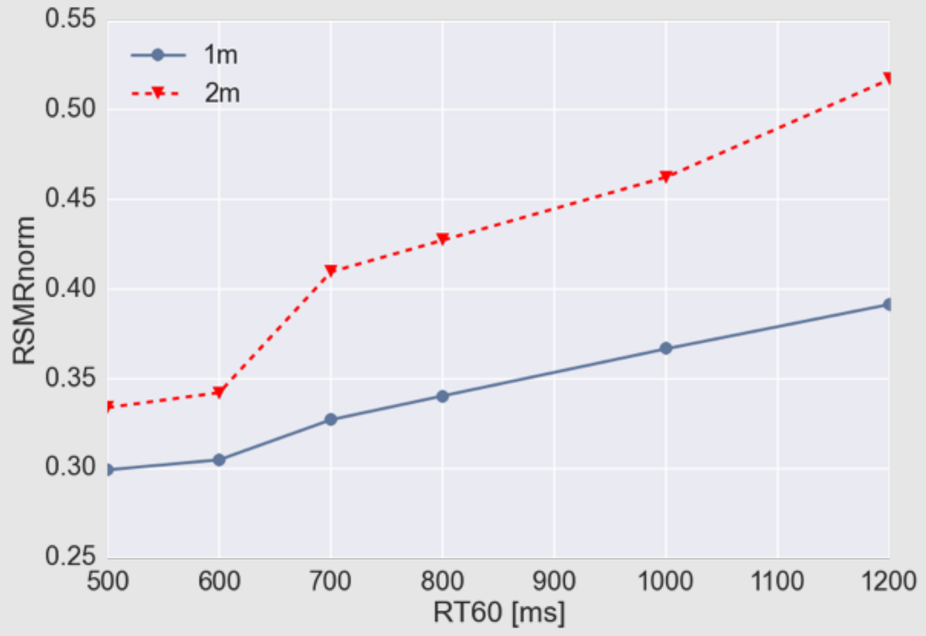
MSC and MSC distance measures

Experimental setup

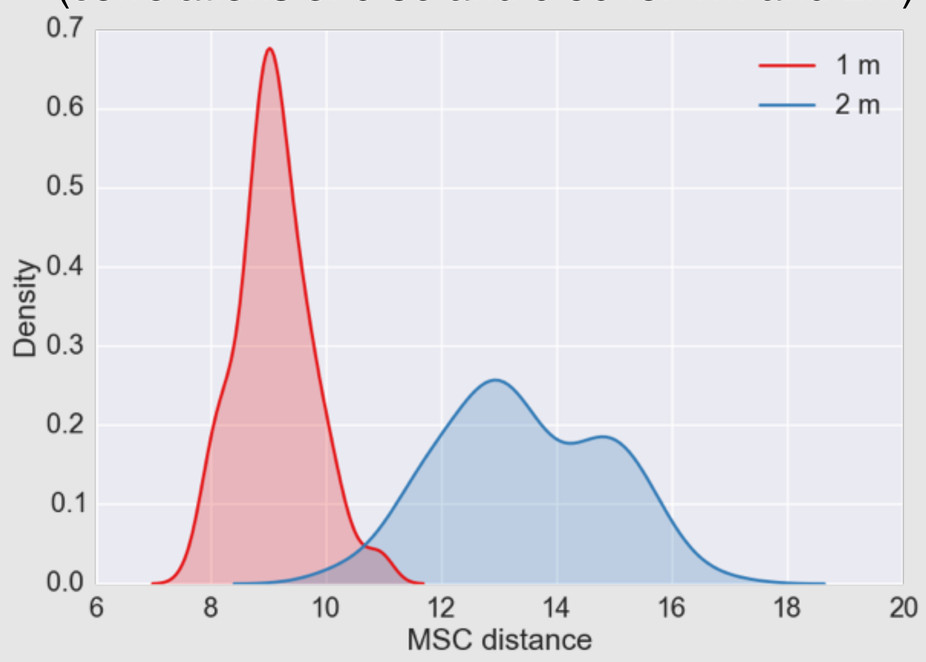


- 120 sentence lists (approx. 90 seconds each)
- 6 different setups in the reverberant chamber - 6 RTs and 2 distances
- Reverberant speech recorded with 2 microphones, using a prototype of a voice-activated device (the Ubi).

Experimental results



Averages of the RSMRnorm metric per condition (correlations of 0.86 and 0.90 for 1m and 2m)



Distribution of the MSCD for different source distances

Results of the estimated RT60 using the RSMRnorm only (left) and incorporating interactions with the MSCD. Both models were fit with 2nd order polynomials on RSMRnorm.

Model	Correlation (per sample)
RSMRnorm only	0.707
RSMRnorm + MSCD	0.847

Acknowledgments: The authors thank the Unified Computer Intelligence Corporation (UCIC) for providing the Ubi device for research purposes.