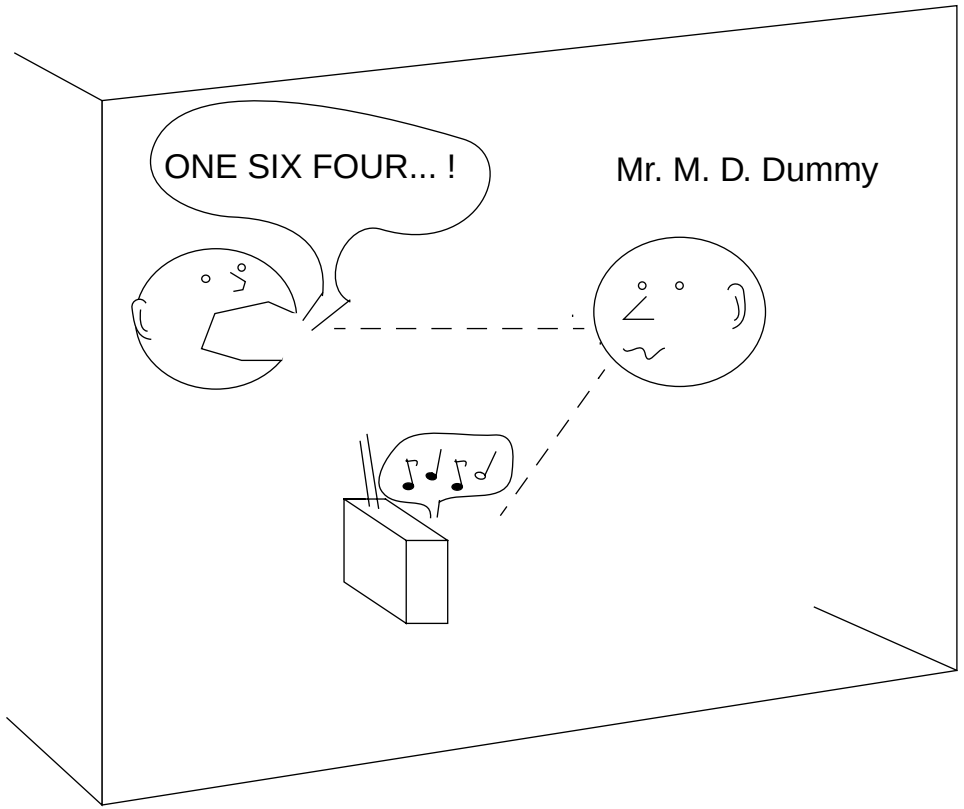
Missing data speech recognition in reverberant acoustic conditions

Kalle, Guy and Jon



Content

1. Introduction to reverberation
2. Speech modulation frequencies
3. Reverberation masking model
4. Test conditions
5. Results
6. Discussion

1. Introduction to reverberation

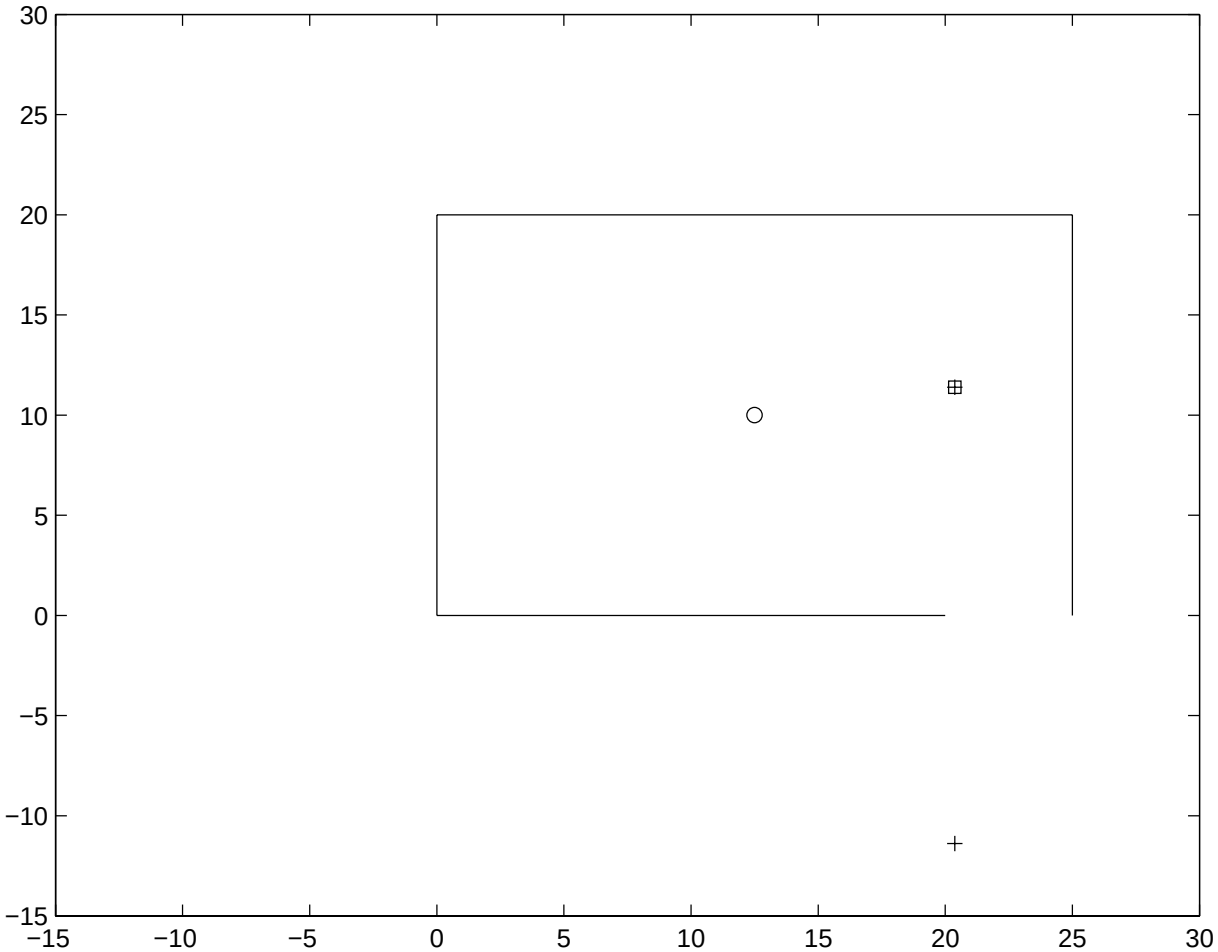


Fig1. Image expansion. Direct sound.

1. Introduction to reverberation

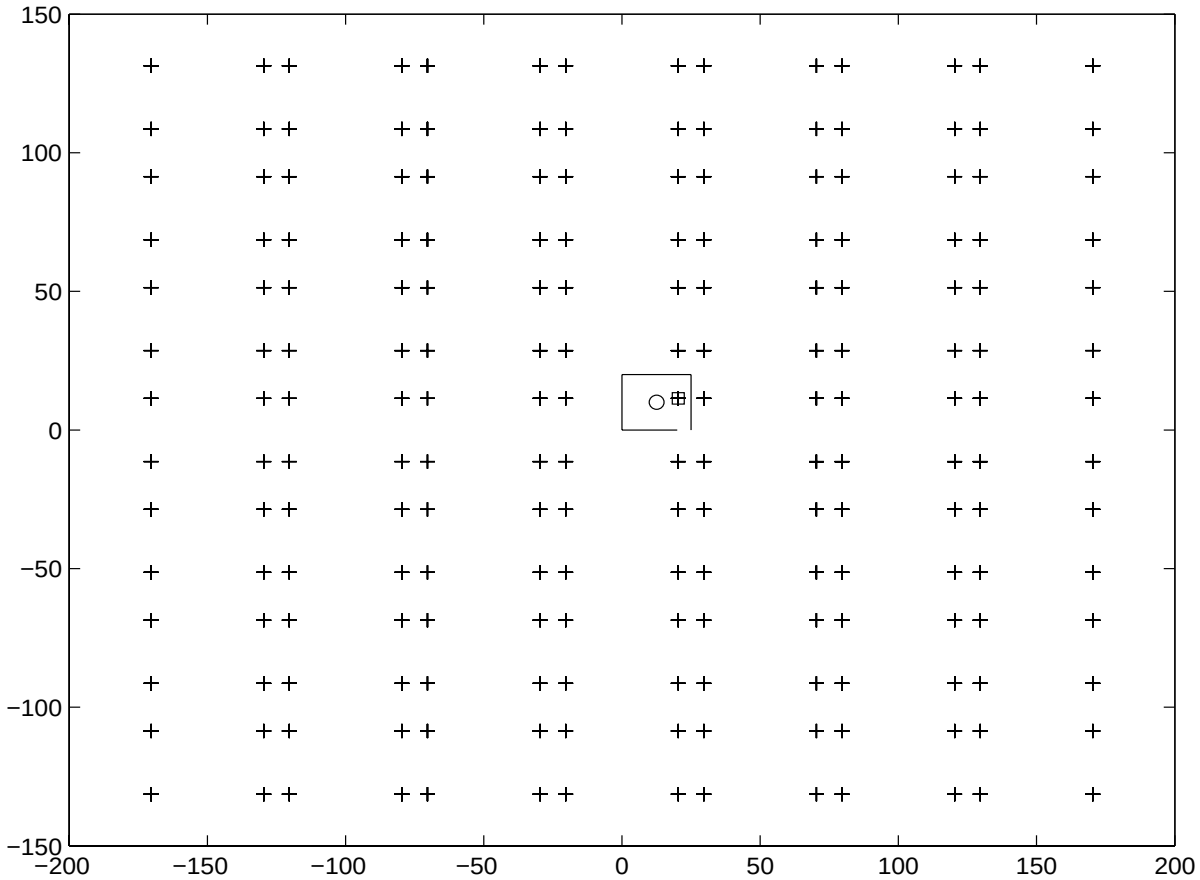


Fig2. Image expansion. 3rd order reflections.

1. Introduction to reverberation

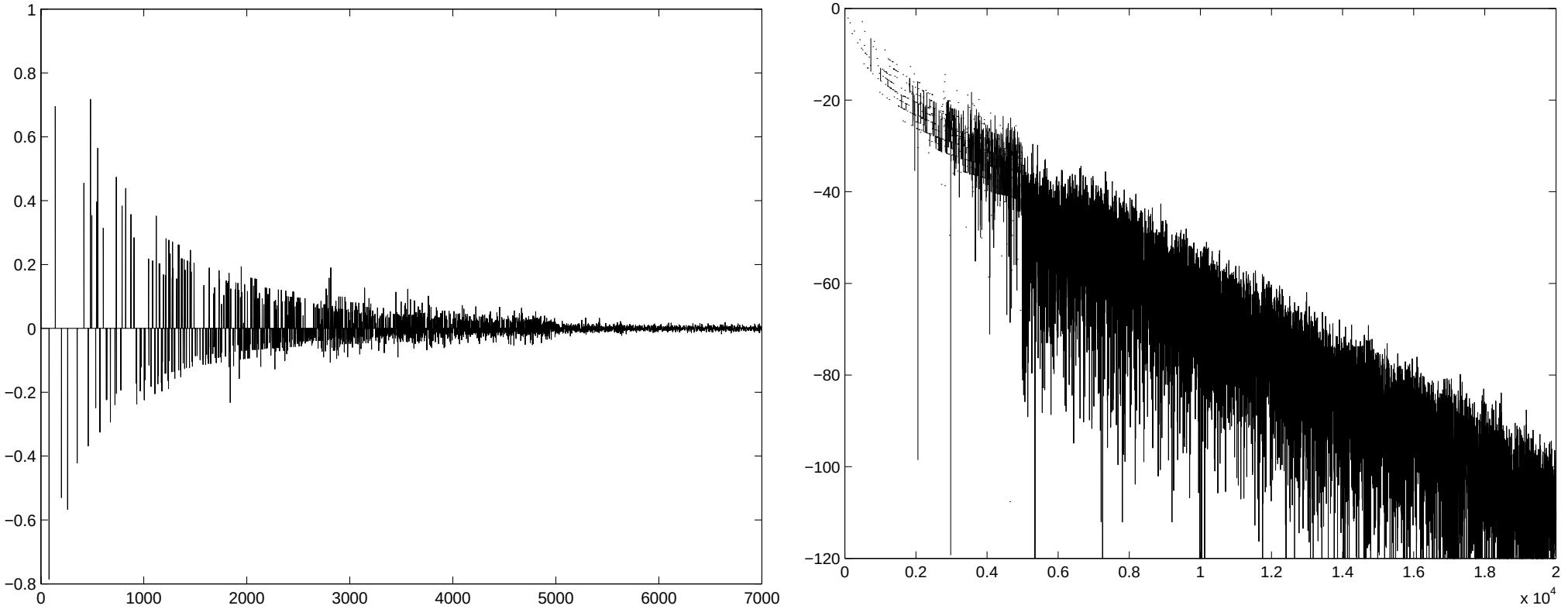


Fig3. Room impulse responses. Left: linear amplitude. Right: log-amplitude.

2. Speech modulation frequencies

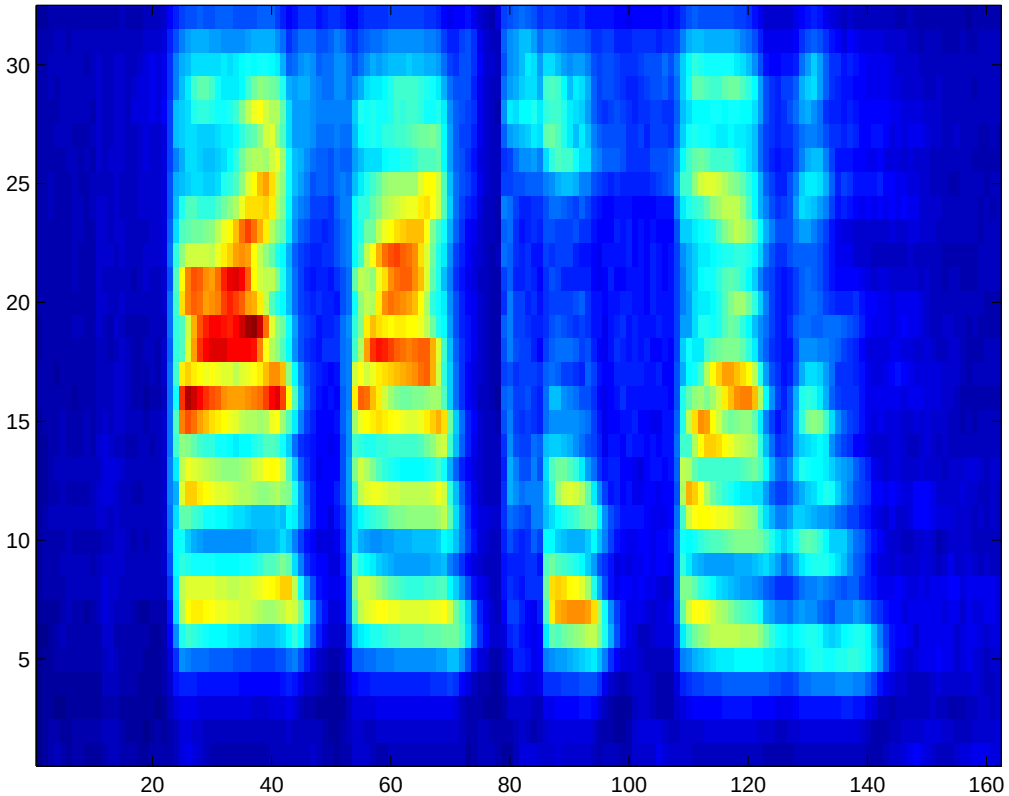


Fig4. Ratemap for utterance 5527.

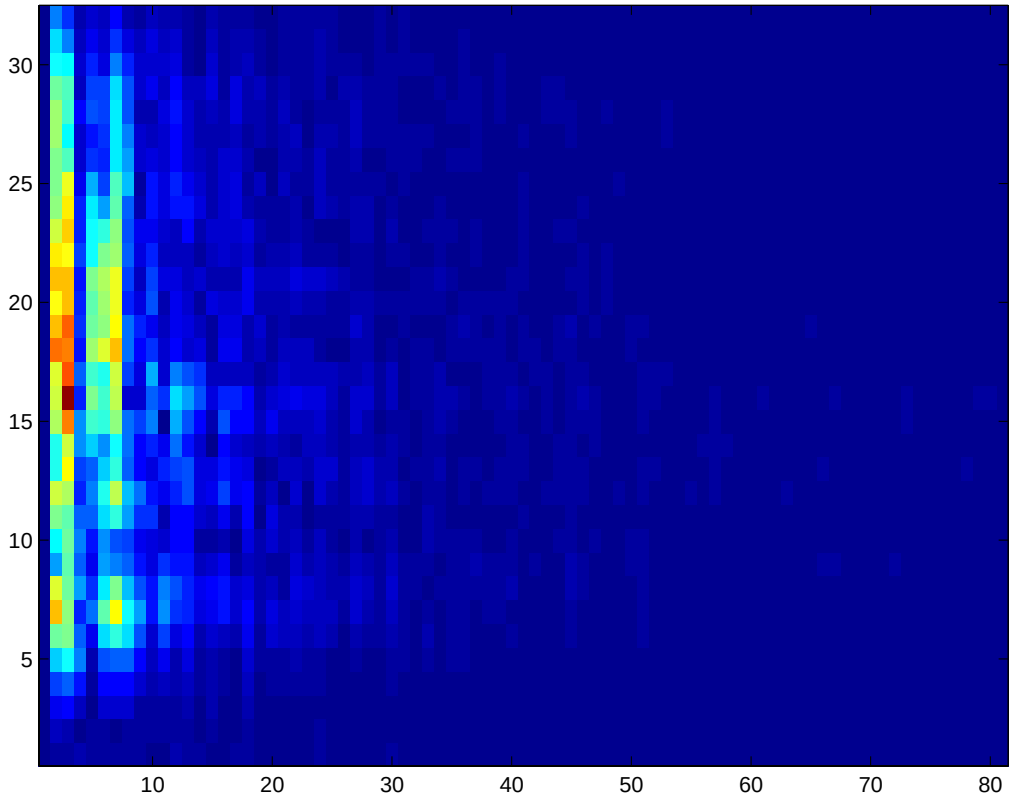


Fig5. FFT of the ratemap of Fig3.

3. Reverberation masking model

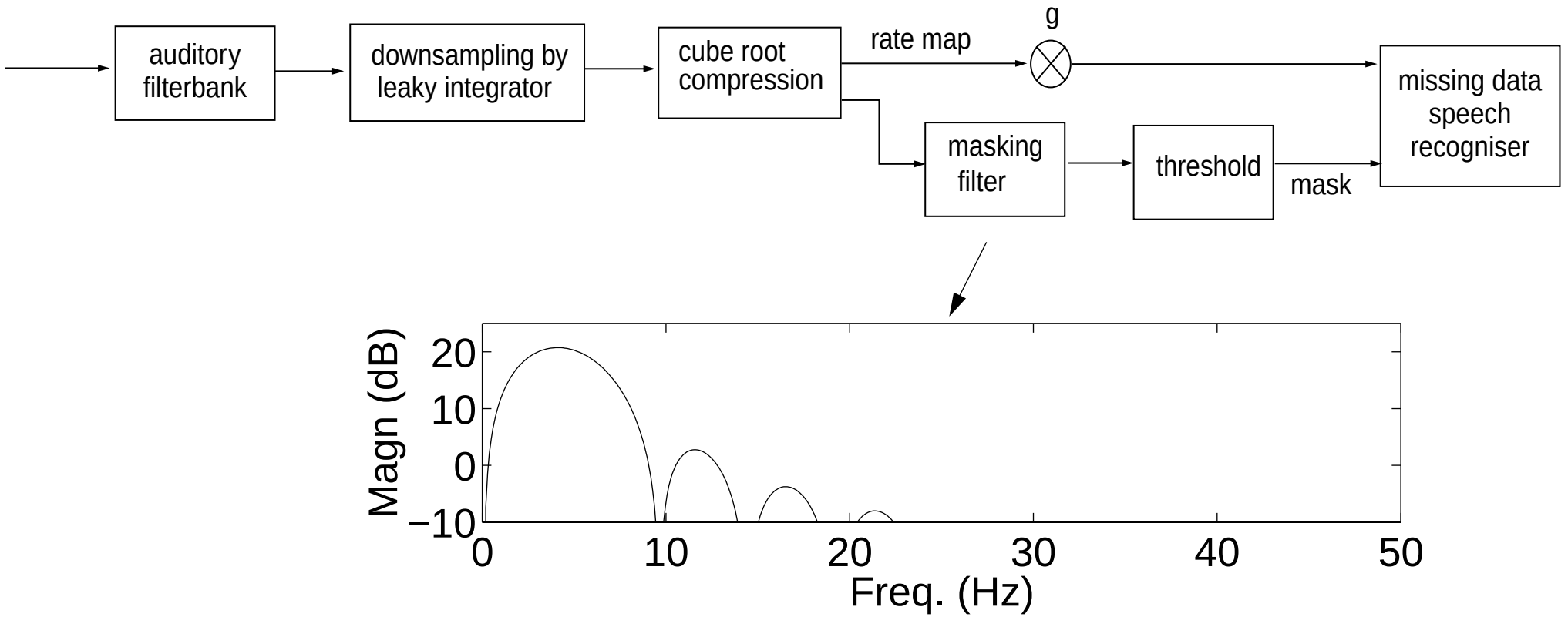


Fig6. Diagram of the model.

3. Reverberation masking model

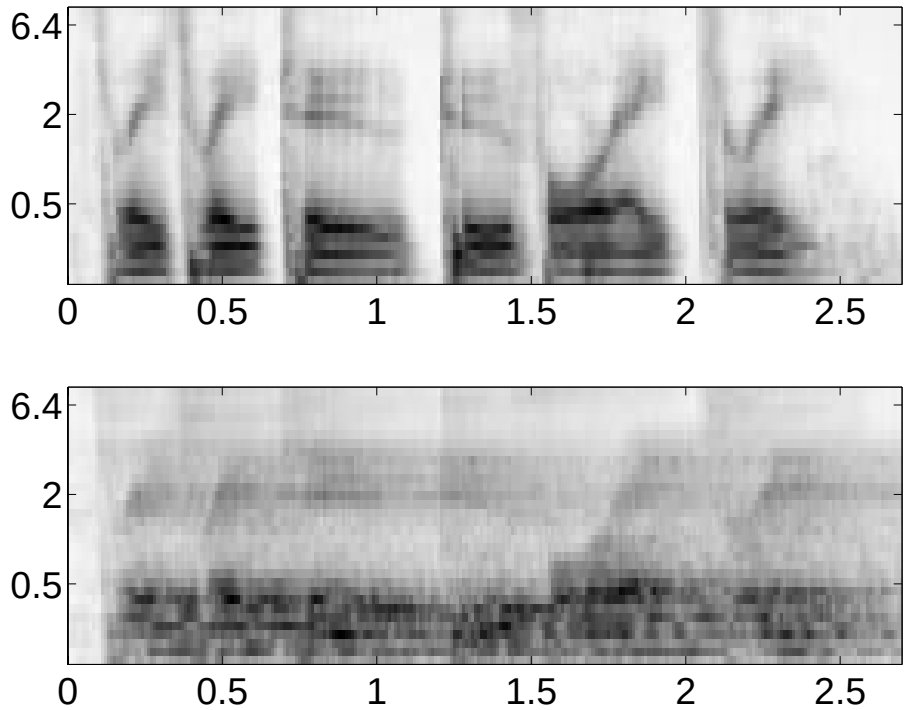


Fig7. Rate maps; Top: clean. Bottom: reverberated

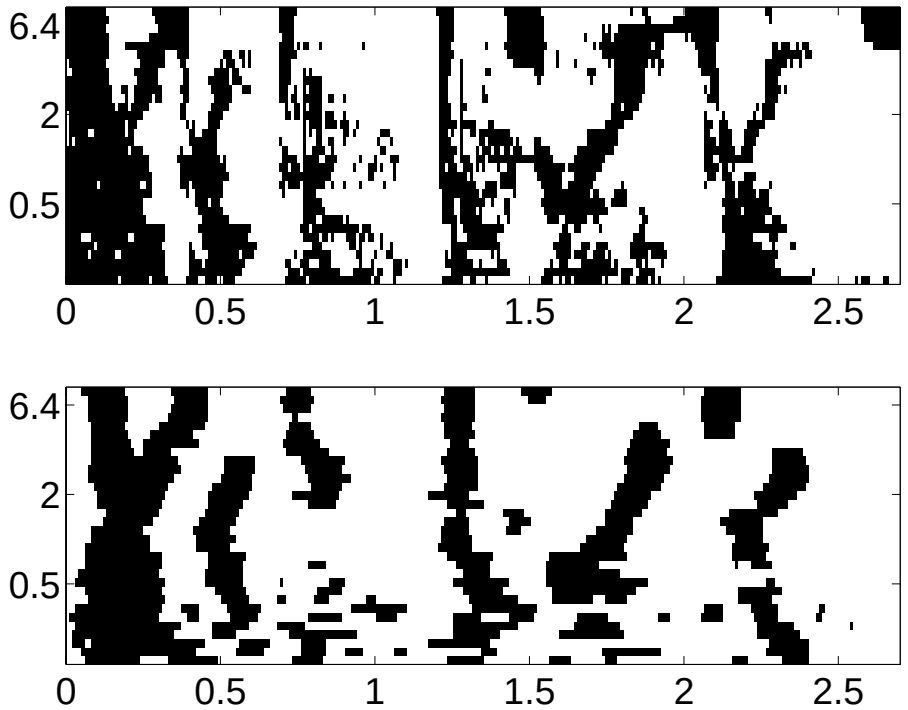


Fig8. Masks; Top: a-priori. Bottom: reverb. masking

4. Test conditions

Room 1: 15 x 13 x 6.5 m^3,
T60=0.7 sec., D/R=0 & -10 dB

Room 2: 25 x 20 x 6 m^3,
T60=1.7 sec., D/R=0 & -10 dB

Room 3: 55 x 35 x 14 m^3,
T60=2.7 sec., D/R=0 & -10 dB

5. Results

Recognition Technique	Clean	T_{60}=0.7s D/R= 0dB	T_{60}=0.7s D/R= -10db	T_{60}=1.7s D/R=0dB	T_{60}=1.7s D/R= -10dB	T_{60}=2.7s D/R= 0dB	T_{60}=2.7s D/R=- 10dB
Unity mask	98.26	62.48	46.39	42.82	29.24	34.90	24.28
MFCC	99.65	60.40	47.08	47.35	34.46	40.73	28.55
Reverb. masking		90.33	85.03	82.25	71.11	66.05	45.52
a priori mask		95.82	93.73	92.42	89.12	90.60	87.99

T

6. Discussion

- Advantage over robust feature techniques:

+ Mask estimation can be changed on the fly when conditions change -> no-retraining required when the rule is changed

+ Better performance ???

- Currently all thresholds are hand tuned, an adaptive system is under work