Meeting Corpora Hardware Overview & ASR Accuracies

George Jose (153070011)
Guide : Dr. Preeti Rao

Indian Institute of Technology, Bombay

22 July, 2016

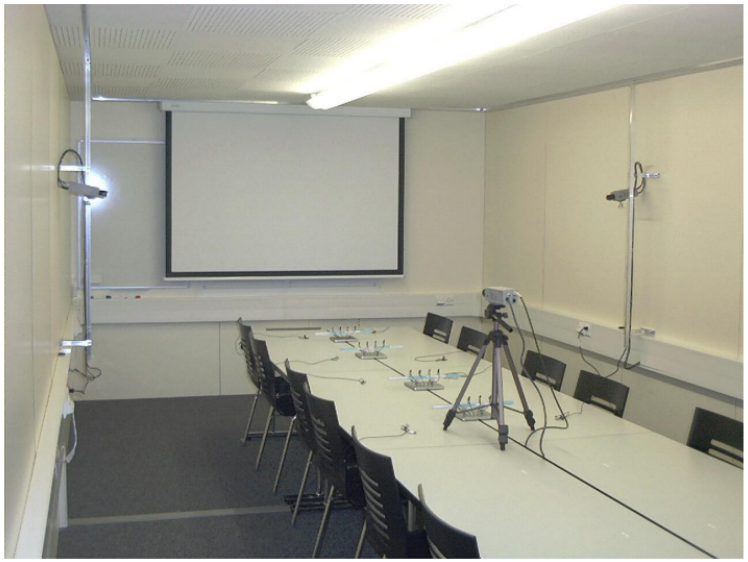
Outline

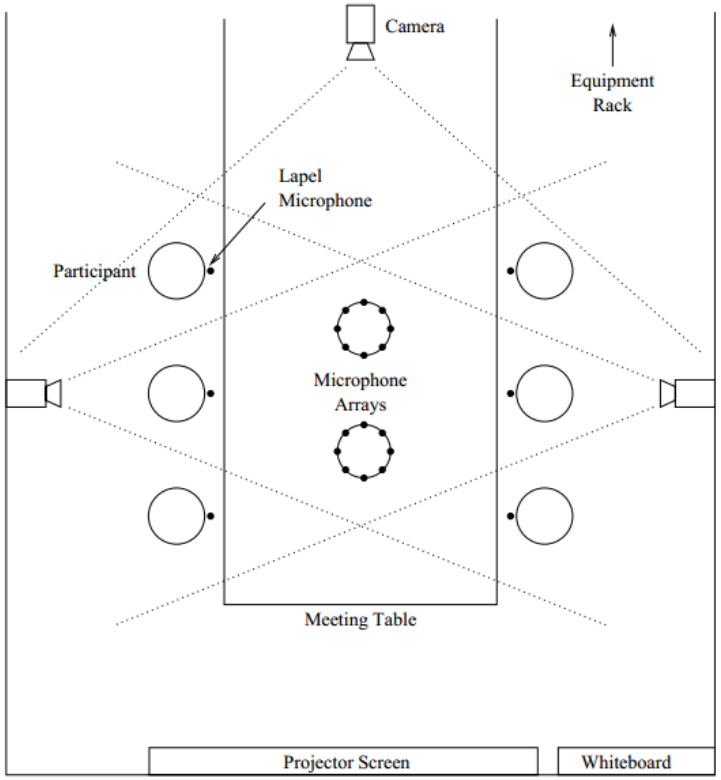
1 AMI Meeting Corpora

2 Kaldi Training

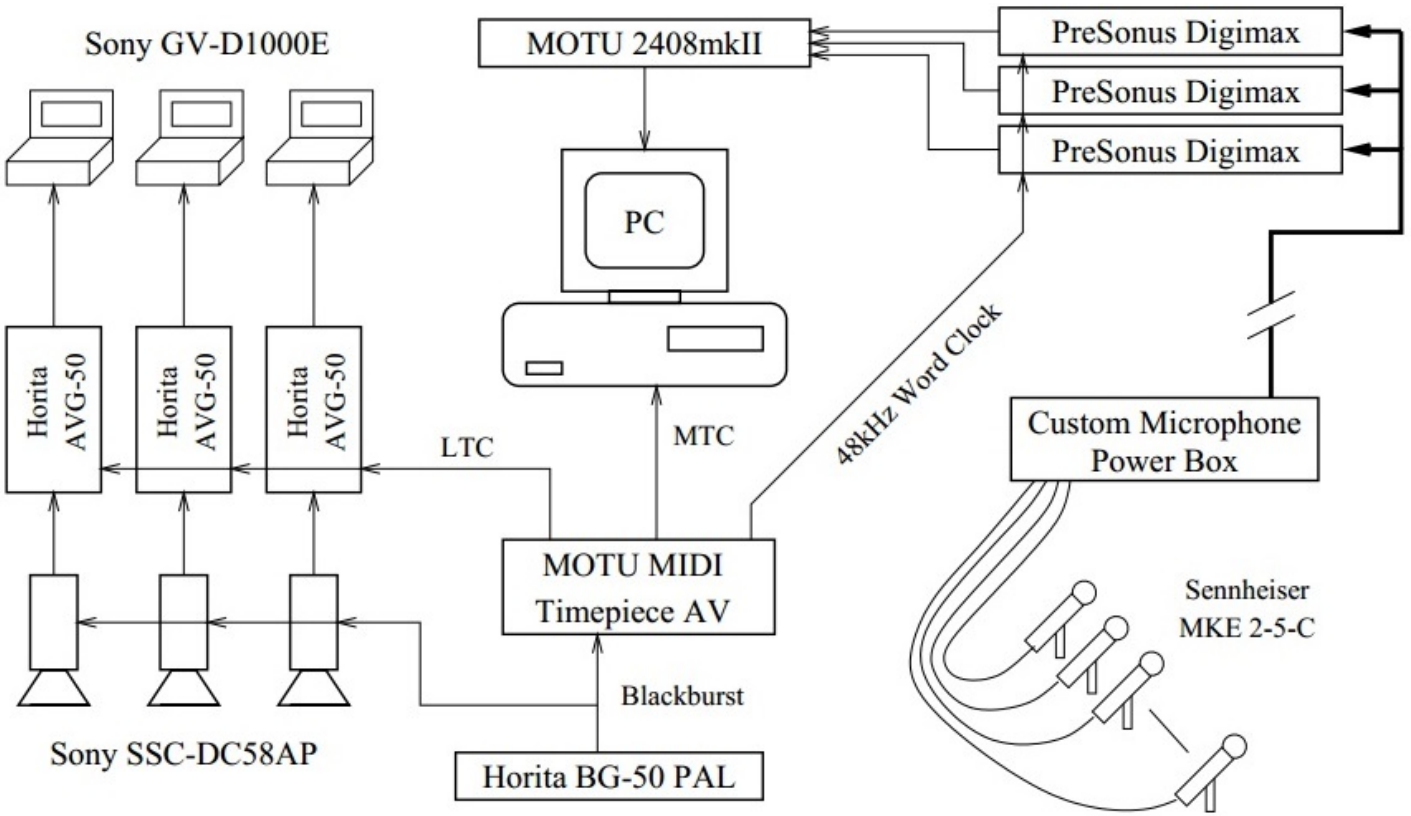
3 BeamformIt Results

AMI Meeting Layout





Hardware Block Diagram



Audio Acquisition

- 24 Sennheiser MKE 2-5-C miniature electret microphone
- Custom-built microphone power box
- 3 PreSonux Digimax preamplifier/digitizer
- 1 Mark of the Unicorn 2408mkII PC interface
- Cakewalk SONAR recording software

Audio Acquisition

- Sennheiser MKE 2-5-C miniature electret microphone

  - Linear frequency response between 20Hz and 20kHz
  - Omnidirectional characteristics
  - High sensitivity : 31mV/Pa
- Custom-built microphone power box
  - MKE 2-5-C requires separate DC bias voltage
  - Provides a biasing voltage for all microphones

Audio Acquisition

- PreSonux Digimax preamplifier/digitizer

  - 8 channel microphone preamplifier
  - 24bit digitization
  - Sample rates- 32kHz,44.1kHz,48kHz

- Mark of the Unicorn 2408mkII PC interface

  - Provides interface to PC for hard-disk based audio recording
  - Supports 72 simultaneous input and audio channels
  - Allows controlled acquisition through driver software on PC

Integrated Hardware

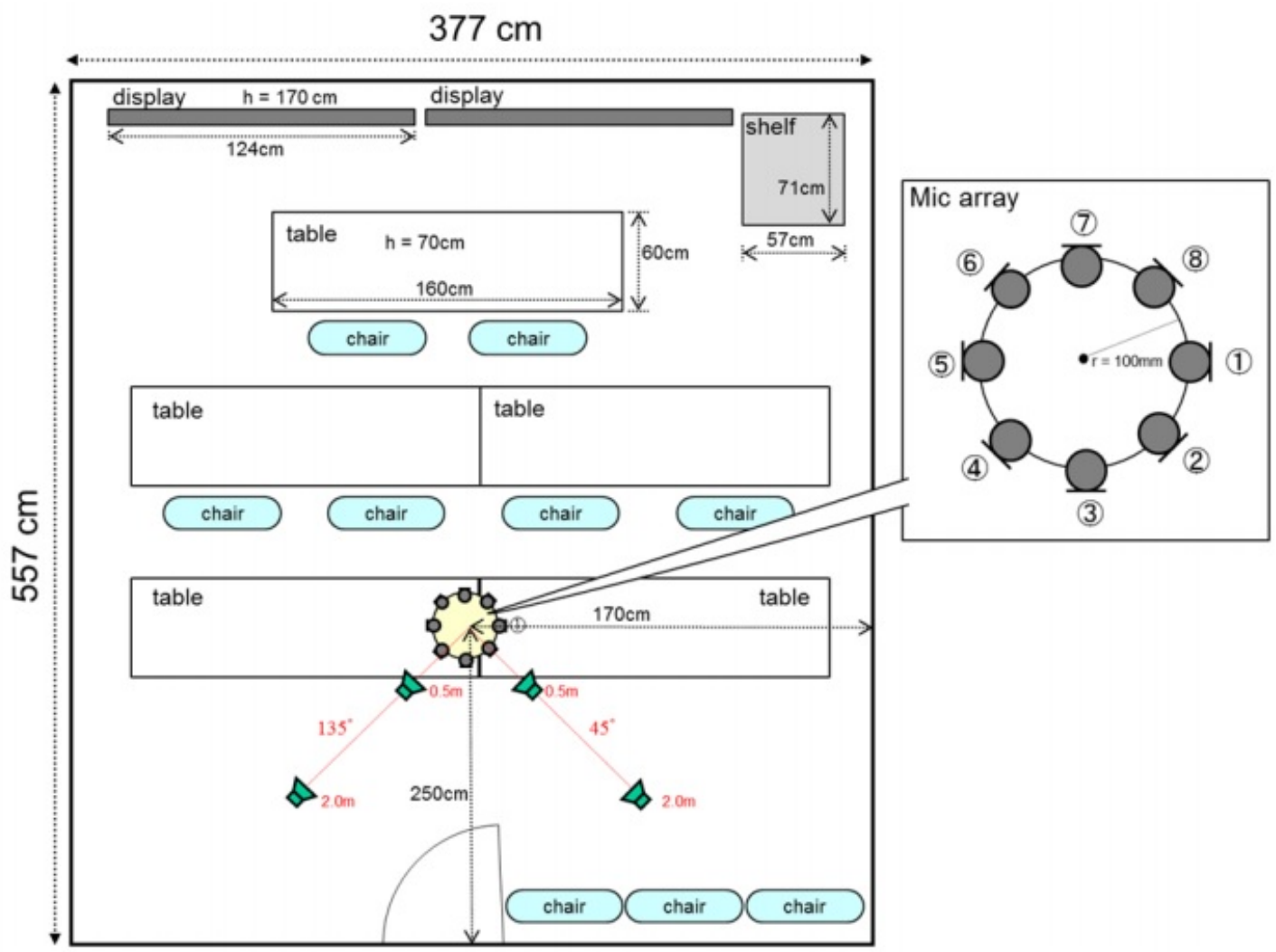
Kaldi Training & Testing

- Database : TIDigits (Adults)
- Training Data : 55M + 57W = 112 Speakers
- Vocabulary : Digits 0-9, "oh"
- Phones : 20
- Monophone & Triphone Model

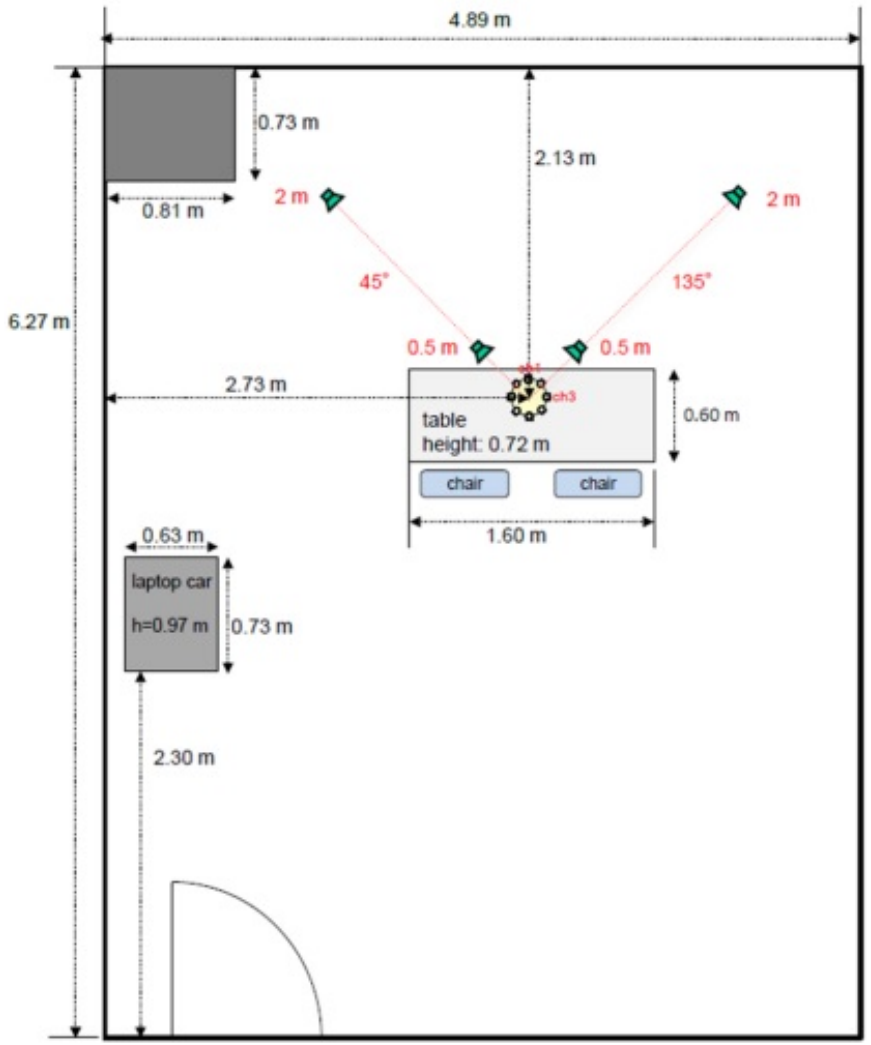
Test Data Simulation (REVERB Challenge Data)

- Testing Data : 3M + 3W = 6 Speakers
- RIR : 8ch circular array (Diameter = 20cm)
- Convolved with RIRs of 3 different rooms
  1. SimRoom1 : $T_{60} = 0.25s$
  2. SimRoom2 : $T_{60} = 0.68s$
  3. SimRoom3 : $T_{60} = 0.75s$

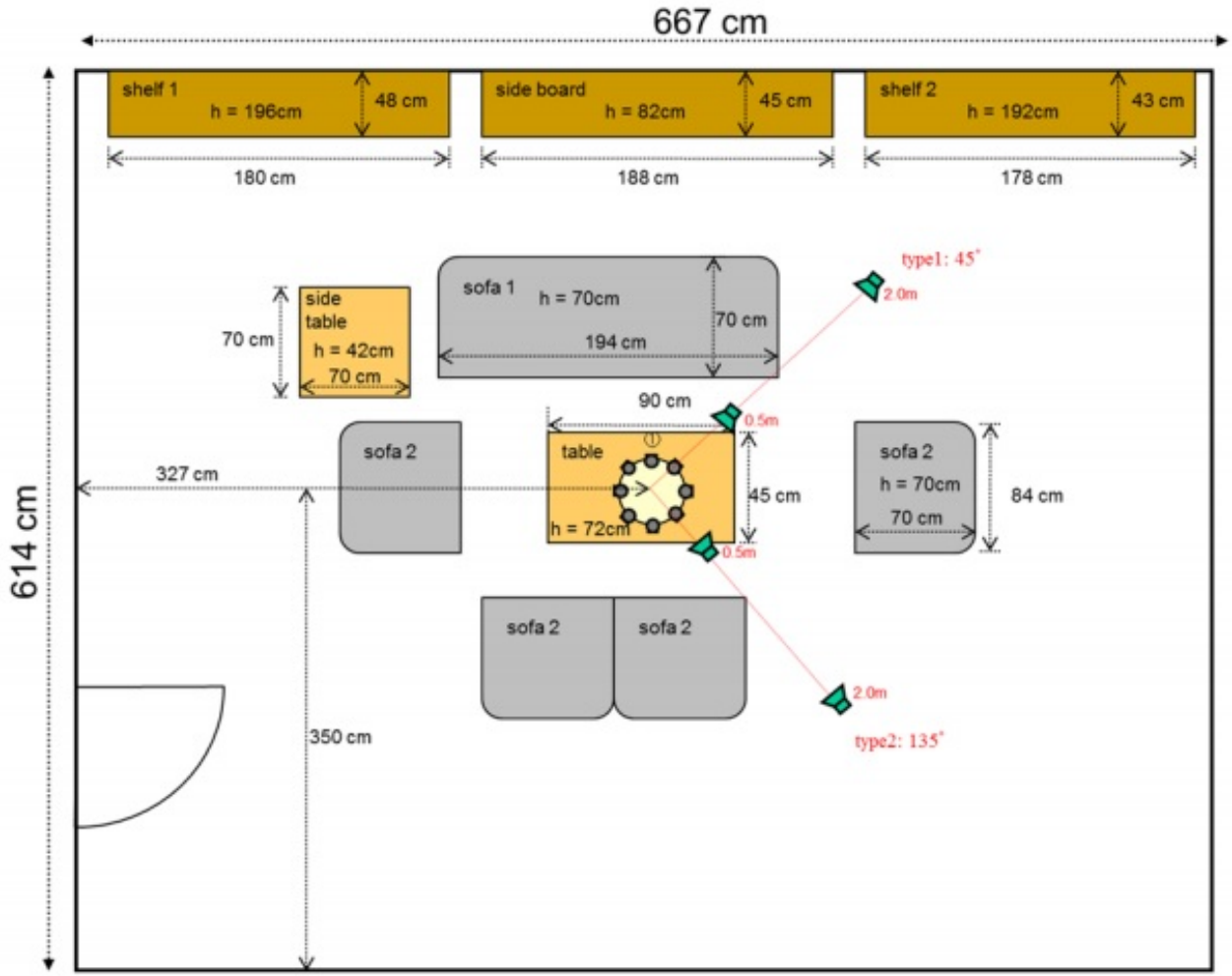
SimRoom1 ($T_{60} = 0.25s$)



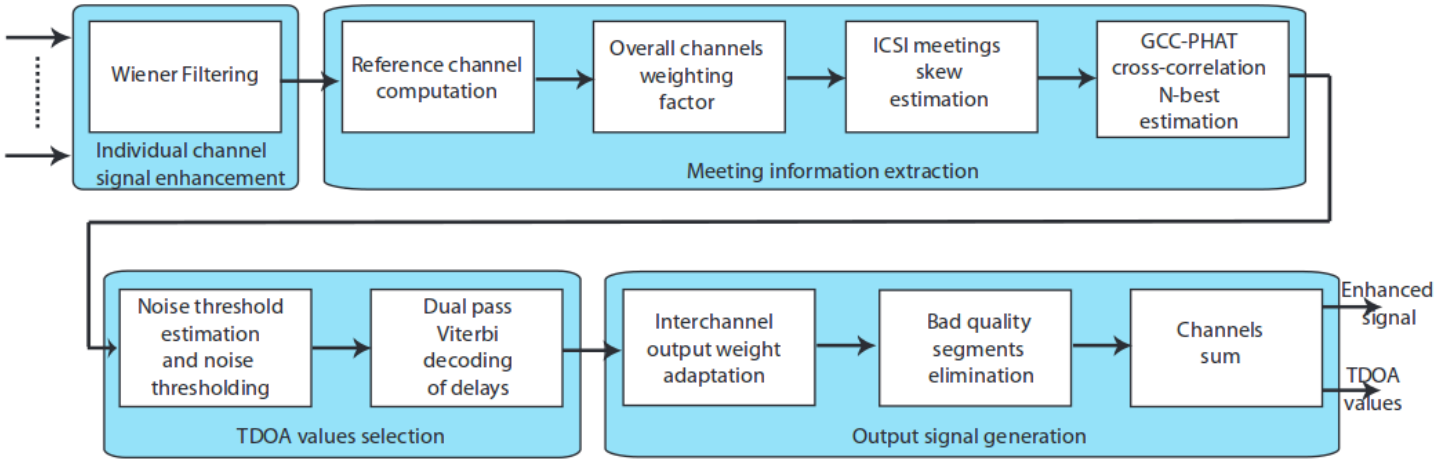
SimRoom2 ($T_{60} = 0.68s$)



SimRoom3 ($T_{60} = 0.75s$)



BeamformIt Block Diagram



- Wiener Filter to each channel for noise reduction
- Reference channel selection using cross correlation value
- GCC PHAT for TDOA estimation
- TDOA post processing to get better estimate

BeamformIt : Parameters

- Window Size : 64ms
- Hop Size : 32ms
- Reference Channel Selection : Based on cross correlation
- TDOA postprocessing : Noise Threshold & Viterbi Decoding
- Noise threshold : 10% of maximum cross correlation
- Performed Channel Elimination : Yes (To avoid bad frames)
- Performed Weight Adaptation : Yes (To reduce noise)

Word Error Rates : Before & After BeamformIt

- Scenario : Clean Speech

	Monophone	**Triphone**
Clean Speech	0.40	0.59

- Scenario : Noise Only

SNR	**Condition**	**Monophone**	**Triphone**
15dB	Before	2.83	2.64
	After	1.58	1.45
10dB	Before	8.04	6.26
	After	3.56	3.23

Word Error Rates : Before & After BeamformIt

- Scenario : Reverberation Only

T60	Condition	Monophone	Triphone
250ms	Before	2.31	2.31
	After	1.45	1.84
680ms	Before	9.68	15.35
	After	6.98	12.25
750ms	Before	14.89	18.18
	After	11.26	13.90

Word Error Rates : Before & After BeamformIt

- Scenario : SimRoom2(680ms)

SNR	Condition	Monophone	Triphone
15dB	Before	39.06	39.92
	After	16.67	15.88
10dB	Before	60.54	64.16
	After	33.20	33.99

- Scenario : SimRoom3(750ms)

SNR	Condition	Monophone	Triphone
15dB	Before	34.78	41.11
	After	15.15	20.82
10dB	Before	56.85	66.73
	After	31.09	40.12