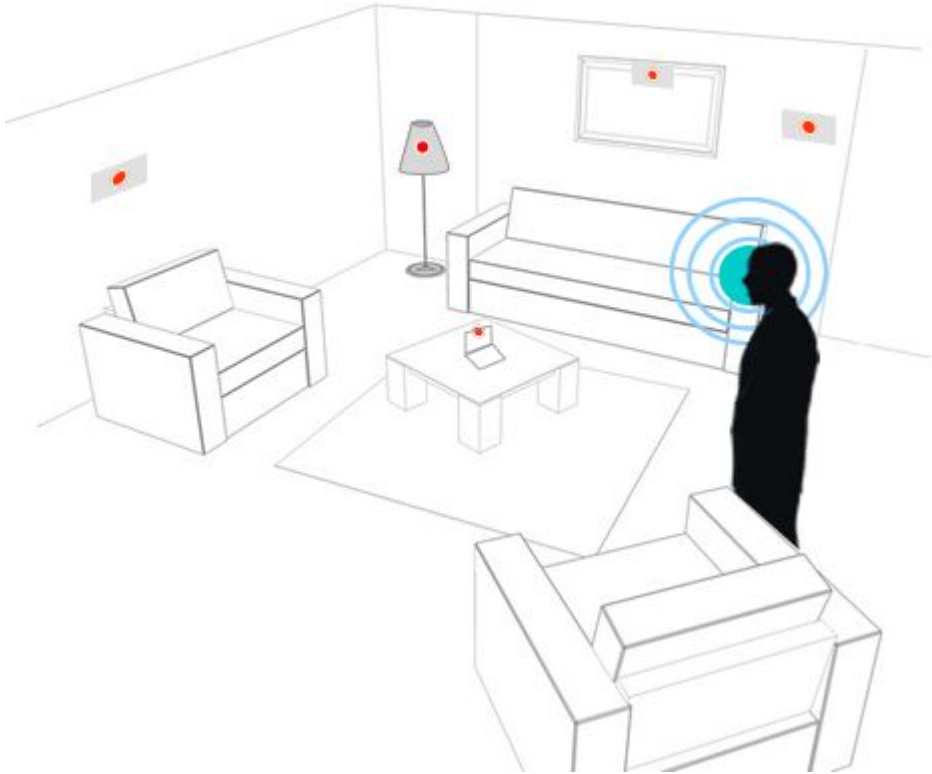
Exploiting Inter-Microphone Agreement for Hypothesis Combination in Distant Speech Recognition

Cristina Guerrero[1,2] and Maurizio Omologo[1]

1 - Fondazione Bruno Kessler (FBK)-Irst, Trento, Italy

2 - PhD student at University of Trento, Trento, Italy

EUSIPCO 2014 - Lisbon, Portugal



reverberation
attenuation
noise
simultaneous speech
what did he say?
sensing
coding
communication
of devices

Multi-microphone
Distant Speech Recognition (DSR)

Scenario - DIRHA

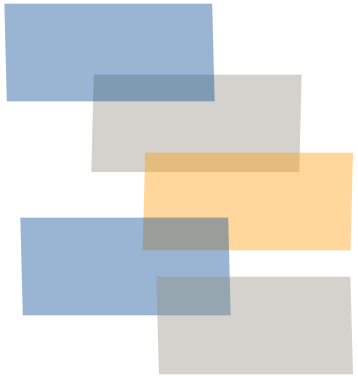
Goal: Voice-enabled Home Automation

- Distant Speech Recognition (DSR)
- Robustness to domestic-context conditions
- Always listening system
- Multi-room multi-sound sources
- Distributed microphone network

Some of the results presented in this talk represent achievements obtained under DIRHA (Distant-speech Interaction for Robust Home Applications). This project is funded by the European Union, Seventh Framework Programme for research, technological development and demonstration, grant agreement no. FP7-288121. For more details, see: http://**dirha.fbk.eu**

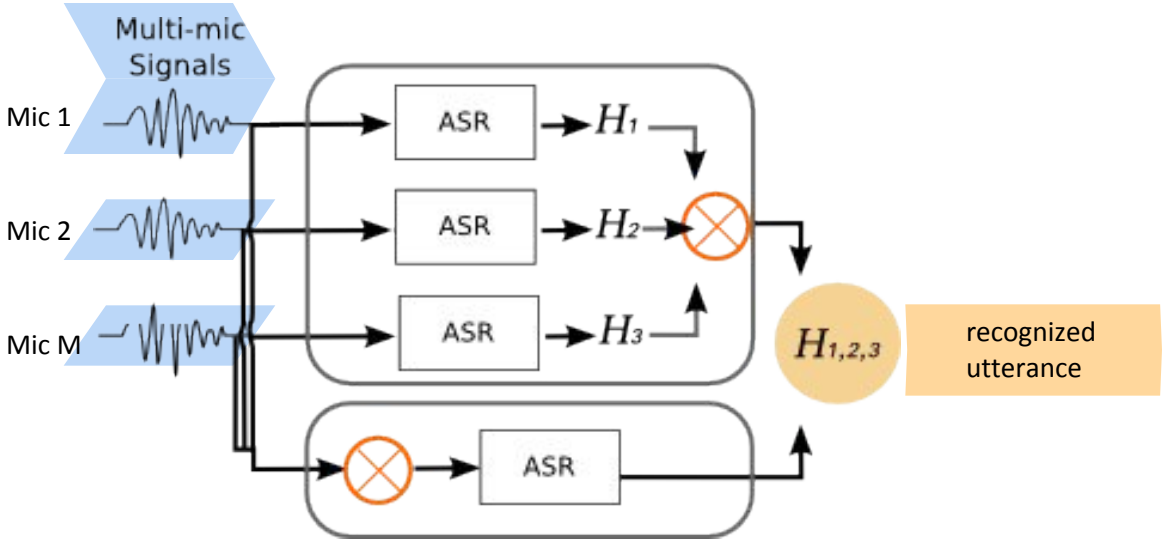
Outline

- Problem: Multi-Mic DSR
- Hypothesis Combination
- Multi-Mic Confusion Network
- Experiments and Results
- Conclusions



Multi-microphone DSR

- Multiple signals => one recognition hypothesis?



Multi-microphone DSR

- Multiple signals => one recognition hypothesis?
- Combination: all (or a subset of) info pieces

or Selection [Wolf-Nadeu 2014]

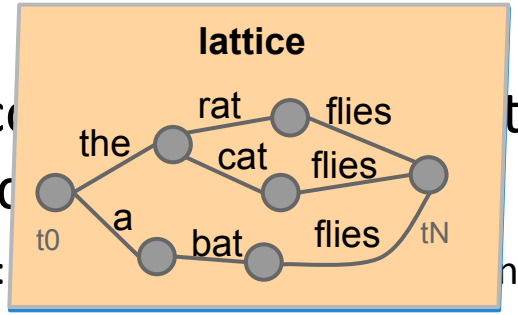
- ❏ Signal processing: beamforming
- ❏ Feature level: models
- ❏ Hypothesis combination: ROVER [Fiscus 1997], Confusion Network Combination (CNC) [Evermann-Woodland 2000, Stolcke et al. 2000]

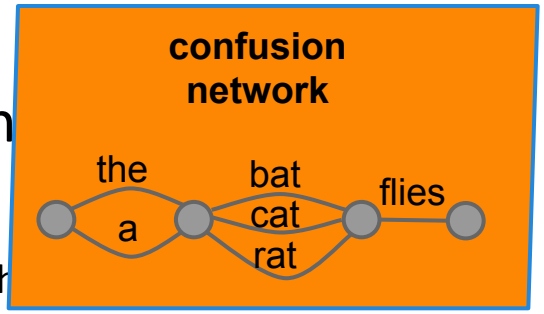
Hypothesis Combination

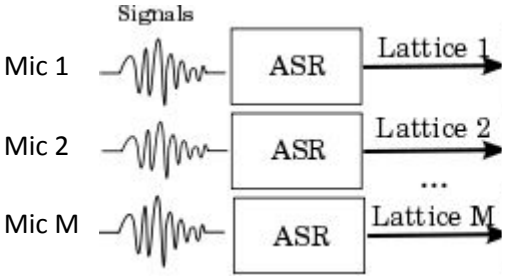
ROVER: Voting

CNC: Combine c tion of the decoded

[Multimic: hough

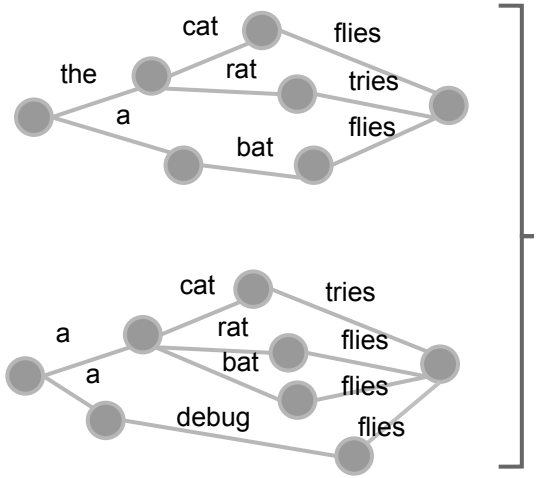






Multi-Mic Confusion Network (MMCN)

lattices decoded from multiple mics



Multi-Mic Confusion Network (MMCN)

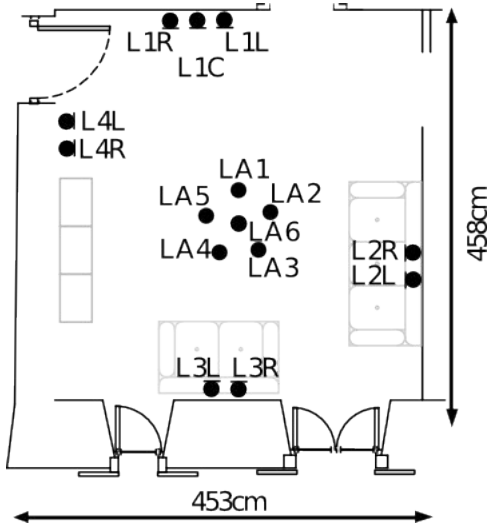
[Guerrero C., Omologo M., "Word Boundary Agreement to Combine Multi-Microphone Hypothesis in Distant Speech Recognition". HSCMA 14]

- Temporal agreement (word-boundaries)
- Information within these boundaries
  - Posterior Probabilities
  - Candidates in a Confusion Set
- No particular order of lattices

Q: Would additional mics benefit MMCN?

Experiments

- Full set of mics(15) in a room
- Recognize Continuous Spoken Commands
- Tested on: Simulated [2245 read commands], and Real data [278 spontaneous commands]
- Acoustic models:
  trained on contaminated dataset
  APASCI (phonetically rich sentences)
- **Comparison to other techniques**
  - ❏ BeamformIt / ROVER/ CNC
- **Different mic-group combinations**



Results

MMCN vs other approaches:

Beamforming (BF), ROVER, CNC

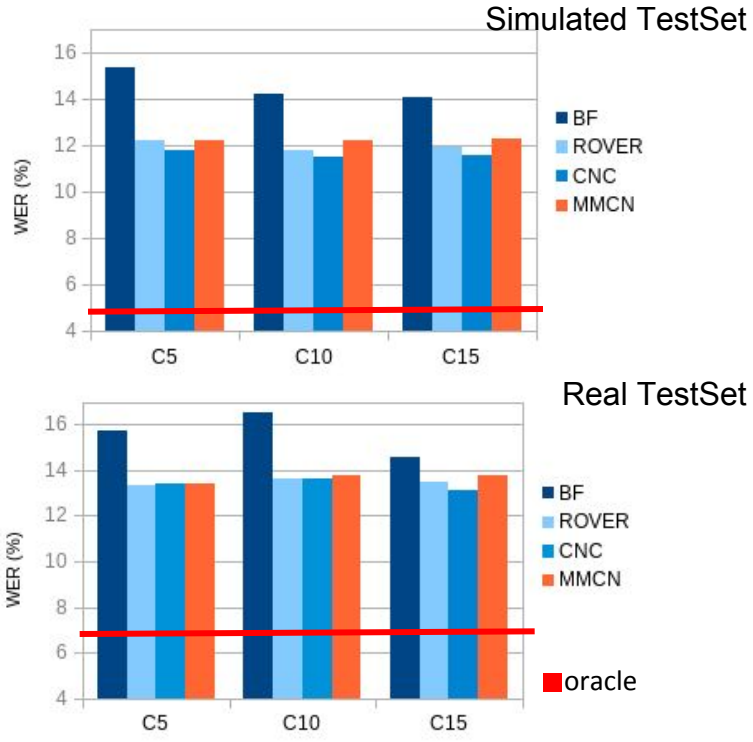
- Cx: configuration # mics

- No order for MMCN

- Alignment based approaches averaged over permutations

- Tested on simulated/real sets

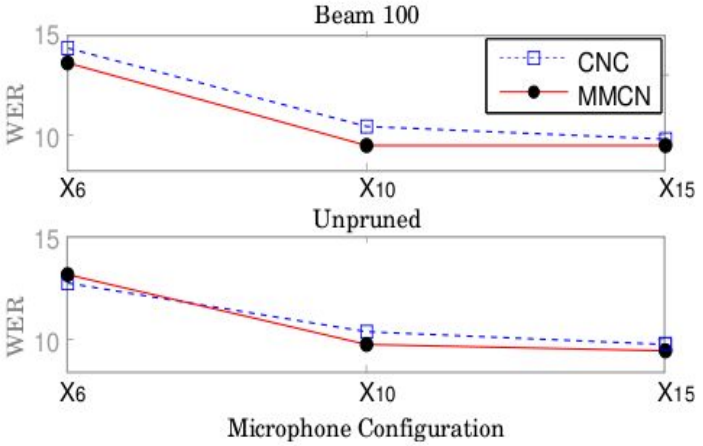
- Oracle: best mic per utterance



Results

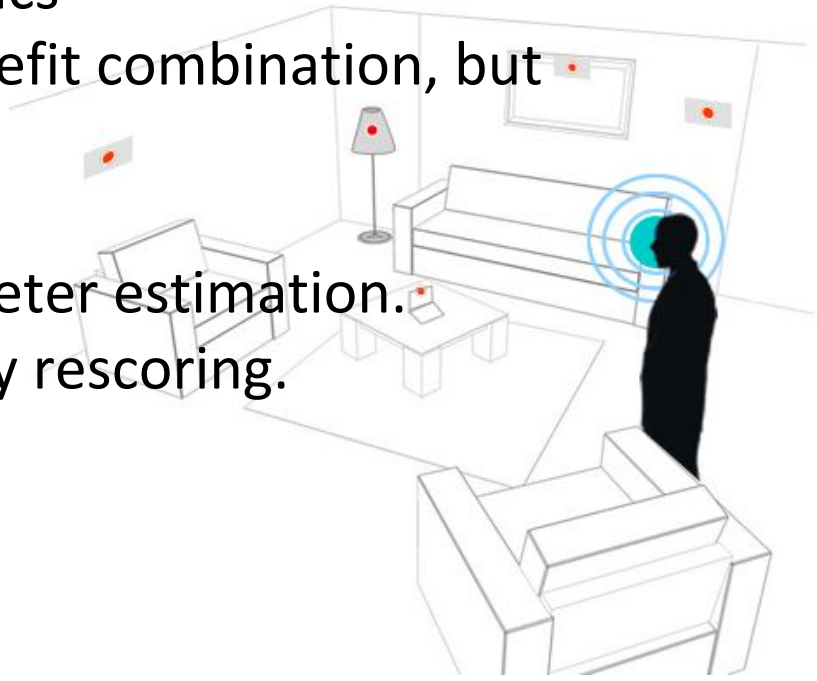
Analysis of a Specific Mic Ordering:
alignment based approaches subject to a specific combination of elements
(hypotheses, confusion networks)

- MMCN vs Rover/CNC
- Effect of Number of Mics



Conclusions

- Comparable performance to state of the art techniques
- MMCN not affected by order of mics
- Balanced increase of mics can benefit combination, but quality of lattices is important
- Currently:
  Improvement of automatic parameter estimation.
  Incorporate context into MMCN by rescoring.
  Lattices evaluation.



FONDAZIONE BRUNO KESSLER

Thank you for your attention.

Cristina Guerrero
guerrero@fbk.eu

FONDAZIONE
BRUNO KESSLER

DIRHA

Experimental details:

- Acoustic Models
  - Trained on contaminated APASCI (It) 16kHz
  - 27 context independent phone units (of the Italian language)
  - Features: MFCC_0_D_A_Z (w=25 ms, o=10ms)
- Language Model
  - Bigram - Read & spoken commands
- Systems
  - HTK, SRILM (posterior prob.), NIST Scoring Toolkit 2.4.0, BeamformIt 3.4.1.

Oracle

- Different on each dataset (more complextask for RealSet)
- Changes on setof microphones and beam threshold

  Beam 0

  WER
  - 5 mics (Dev 2.7, Test 2.35)
  - 15mics (Dev 6.06, Test 4.51)

MMCN vs BF/ROVER/CNC

TestSet b0
Oracle: *4.51*

	BF	ROVER	CNC	MMCN
C5	15.38	12.20	**11.82**	12.22
C10	14.23	11.76	**11.50**	12.20
C15	14.07	11.93	**11.57**	12.26

RealSet b0
Oracle: *7.28*

	BF	ROVER	CNC	MMCN
C5	15.71	**13.34**	13.42	13.42
C10	16.55	13.63	**13.59**	13.74
C15	14.57	13.46	**13.12**	13.74

FONDAZIONE BRUNO KESSLER

ROVER devset / testset (sim)

Mics		B80	B100	B0
	avg	**8.32**	**7.38**	**7.38**
5	min	7.72	6.71	6.71
	max	9.4	8.05	8.05
	avg	**9.57**	**8.66**	**8.69**
10	min	9.06	8.05	8.05
	max	10.4	9.4	9.4
	avg	**10**	**9.08**	**8.99**
15	min	9.06	8.39	8.05
	max	10.4	9.73	9.4

Mics		B80	B100	B0
	avg	**14.32**	**12.69**	**12.2**
5	min	14.24	12.56	12.08
	max	14.41	12.87	12.37
	avg	**13.47**	**12.21**	**11.76**
10	min	13.38	12.06	11.66
	max	13.65	12.4	11.91
	avg	**13.53**	**12.38**	**11.93**
15	min	13.39	12.27	11.85
	max	13.64	12.46	11.99

CNC

	Dev		Test	
Config	b100	b0	b100	b0
C5	8.05	8.05	12.18	11.82
C10	8.92	9.26	11.87	11.50
C15	9.37	9.44 **min 8.72** **max 9.73**	11.99	11.57 **min 11.48** **max 11.63**