SINGLE-CHANNEL reverberant speech recognition using C50 estimation

Pablo Peso Parada[1], Dushyant Sharma[1],
Patrick A. Naylor[2] and Toon van Waterschoot[3]

Introduction

- We present several single-channel approaches to robust speech recognition in reverberant environments based on single-channel estimation of C50
- Our best method outperforms the best baseline of the challenge, reducing the word error rate by **5.7%** which corresponds to a **16.8%** relative word error rate reduction

Measures of reverberation

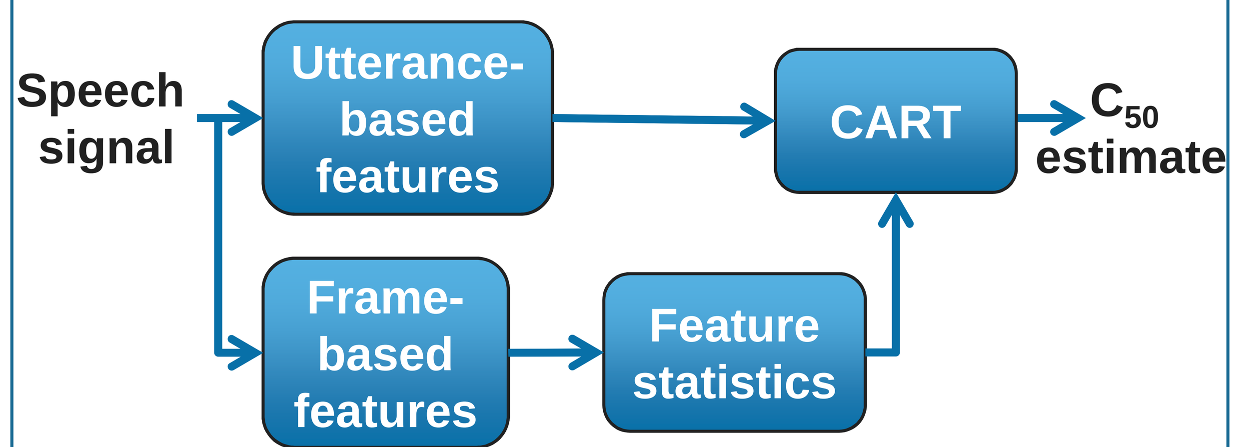
- T_{60}, DRR, T_s, D_{50} and C_{50} are parameters used to characterize the effect of reverberation from the room impulse response

	T_{60}	DRR	T_s	D_{50}	C_{50}
Acc.	0.64	0.69	0.79	0.64	**0.80**
PESQ	0.7	0.83	0.95	0.71	**0.96**

Correlation of various measures of reverberation with ASR accuracy (Acc.) and speech quality (PESQ)

C_{50} estimation [1]

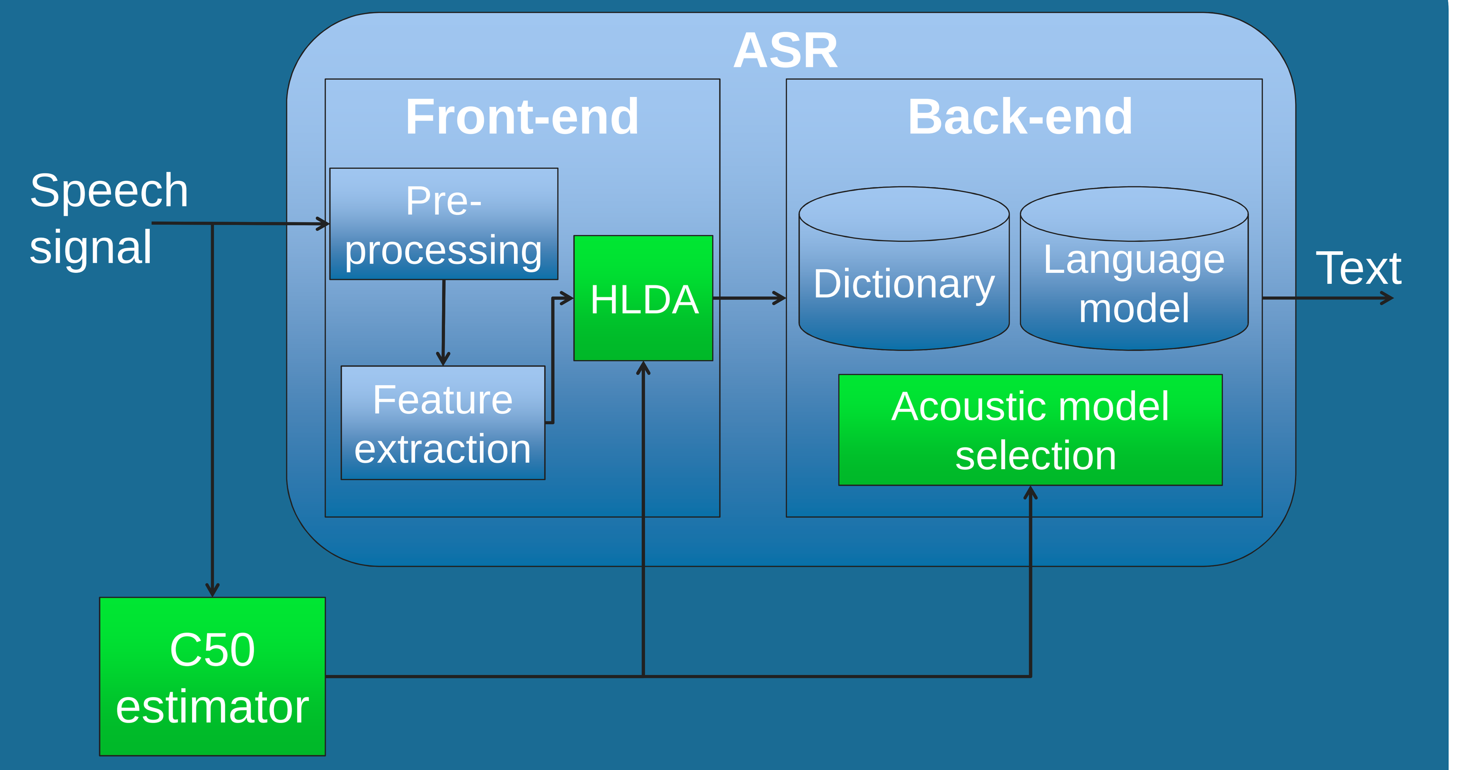
$$C_{50} = 10\log_{10}\left(\frac{\sum_{n=0}^{N_{50}} h^2(n)}{\sum_{n=N_{50}+1}^{\infty} h^2(n)}\right)$$



[1] P. Peso Parada, D. Sharma and P. A. Naylor, "Non-intrusive estimation of the level of reverberation in speech", ICASSP 2014.

Single-channel reverberant speech recognition approaches

- Motivation: use a C_{50} estimator to provide reverberation robustness to automatic speech recognition (ASR)
- Approaches:
  - Include C_{50} in the input feature vector (**frond-end**)
  - Use C_{50} to create different reverberant acoustic models and select the most adequate in the recognition stage (**back-end**)
  - Combination of both previous approaches (**hybrid**)



Front-end

- Techniques:
  1. Add C_{50} estimate to the 39 dimension MFCC_0_D_A feature vector ($C_{50}FV$)
  2. Apply heteroscedastic discriminant analysis transformation (HLDA) to reduce the final feature dimension by 1 (i.e. to 39) ($C_{50}HLDA$)
- Results:

Method	Clean	Sim.	Real
$C_{50}FV$	29.01	30.36	56.96
$C_{50}HLDA$	**26.41**	**28.02**	**56.12**

WER (%) averages w/o adaptation (CMLLR)

Back-end

- Techniques:
  - Select the optimal acoustic model according to the reverberation level (*MSx*, where *x* represents the number of trained acoustic models)
  - During training, the acoustic models can be built with overlapped data which provides a smoother transition
- Results:

Method	Clean	Sim.	Real
MS3 (no overlap)	28.00	27.93	59.59
MS5	23.22	26.81	57.88
MS8	23.14	**26.17**	**56.40**
MS11	**22.07**	26.40	56.80
MS14	22.85	26.31	57.48
MS18	23.95	26.51	58.06

WER (%) averages w/o adaptation (CMLLR)

Hybrid

- Techniques:
  - Merge the best method in the front-end ($C_{50}HLDA$) with the different back-end approaches to exploit the advantages of each of the methods (*MSx+C_{50}HLDA*, where *x* represents the number of trained acoustic models).
  - Therefore, *x* acoustic models are trained by using the modified feature vector $C_{50}HLDA$
- Results:

Method	Clean	Sim.	Real
MS3+C_{50}HLDA	24.41	25.70	57.00
MS5+C_{50}HLDA	20.93	25.22	55.97
MS11+C_{50}HLDA	**20.55**	**24.52**	**54.21**

WER (%) averages w/o adaptation (CMLLR)

Complete results for our best method (*MS11+C_{50}HLDA*)

Recordings			*MS11+ C_{50}HLDA*	*Clean-cond.*	*Multi-cond.*
Clean	R.1		20.69	**10.50**	30.29
	R.2		20.73	**11.51**	30.07
	R.3		20.22	**10.81**	30.11
	Avg.		*20.55*	***10.94***	*30.16*
Sim.	R.1	N.	15.54	**15.29**	20.60
		F.	**17.10**	25.29	21.15
	R.2	N.	**19.63**	43.90	23.70
		F.	**33.00**	85.80	38.72
	R.3	N.	**25.39**	51.95	28.08
		F.	**36.43**	88.90	44.86
	Avg.		***24.52***	*51.86*	*29.52*
Real	R.1	N.	**55.57**	88.71	58.44
		F.	**52.84**	88.31	55.44
	Avg.		***54.21***	*88.51*	*56.95*

WER (%) table w/o adaptation (CMLLR) where R.X is the room number and N. and F. stand for near and far recordings respectively

Conclusions

- C_{50} estimate successfully applied to reverberant ASR
- Overlapping training data for acoustic model creation gives WER improvement
- Best front end method gives 5.7% WERR
- Best Back-end method gives 11.3% WERR

[1]Nuance Communications Inc. Marlow, UK; [2]Department of Electrical and Electronic Engineering, Imperial College London, UK; [3]Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium