The RWTH/UPB/FORTH System Combination for the 4th CHiME Challenge Evaluation

HLT | RWTH AACHEN UNIVERSITY
Computer Science, Electrical Engineering and Mathematics
NT Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach
FORTH Institute of Computer Science

Overview

Content

- Motivation
- Front ends
- Back ends
- System combination
- Outlook

Overview

Objectives

- Combination of multiple front ends
- Use of BLSTM acoustic model
- Optimization for results on real data

HLT | RWTH AACHEN UNIVERSITY
Computer Science, Electrical Engineering and Mathematics
NT Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach
FORTH Institute of Computer Science

MVDR beamformer

$$\boldsymbol{w}_{\mathrm{MVDR}}(k,l) = \frac{\left[\boldsymbol{R}_{\mathrm{nn}}(k) + \epsilon\ \mathrm{diag}(|\boldsymbol{X}(k,l)|^2)\right]^{-1}\boldsymbol{d}(k)}{\boldsymbol{d}(k)^H\left[\boldsymbol{R}_{\mathrm{nn}}(k) + \epsilon\ \mathrm{diag}(|\boldsymbol{X}(k,l)|^2)\right]^{-1}\boldsymbol{d}(k)}$$

- Estimation based on Time-Frequency masks for *speech+noise* & *noise* components
- Noise and speech spatial correlation matrices are computed from the signals using the estimated masks

Estimation of $\boldsymbol{d}(k)$

As the principal component of the speech spatial correlation matrix

Estimation of $\boldsymbol{R}_{\mathrm{nn}}(k)$

- **6-channel track:** From the estimated TF-masks of the noise component
- **2-channel track:** From 400 ms to 800 ms of context immediately before the utterance

HLT | RWTH AACHEN UNIVERSITY
Computer Science, Electrical Engineering and Mathematics
NT Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach
FORTH Institute of Computer Science

Superdirective beamformer

- Conventional superdirective beamformer based on time-delays

$$\boldsymbol{w}_{\mathrm{SD}}(k,l) = \frac{\left[\boldsymbol{\Gamma}(k) + \epsilon \boldsymbol{I}\right]^{-1} \boldsymbol{d}(k,l)}{\boldsymbol{d}(k,l)^{H} \left[\boldsymbol{\Gamma}(k) + \epsilon \boldsymbol{I}\right]^{-1} \boldsymbol{d}(k,l)}$$

Estimation of $\boldsymbol{d}(k)$

$$\boldsymbol{d}(k,l) = \left[e^{-j2\pi f \tau_1(l)} \quad \dots \quad e^{-j2\pi f \tau_M(l)}\right]$$

Time-delays estimated using the SRP-PHAT localization method of the baseline system

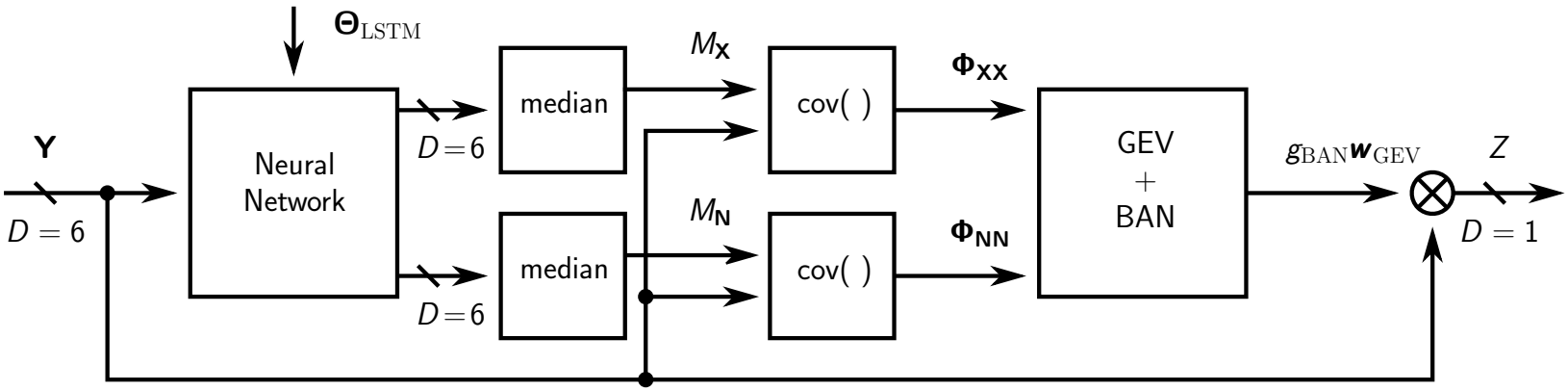
Estimation of $\boldsymbol{\Gamma}(k)$

Assumed to be spherically isotropic diffuse:

$$\Gamma_{ij}(k) = \mathrm{sinc}\left(\frac{2\pi f d_{ij}}{c}\right)$$

HLT | RWTH AACHEN UNIVERSITY
Computer Science, Electrical Engineering and Mathematics
NT Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach
FORTH Institute of Computer Science

BLSTM supported GEV beamformer

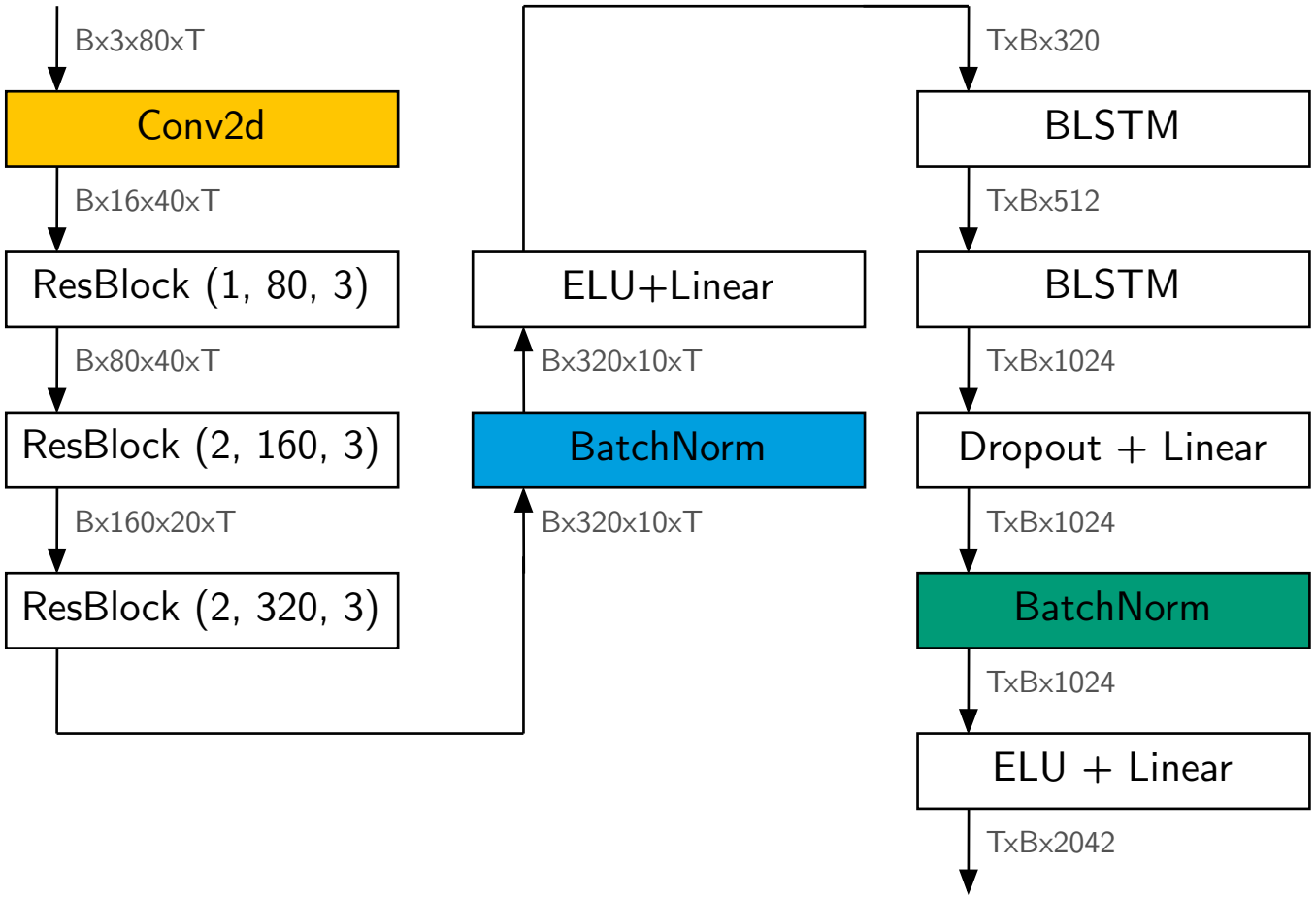


- GEV beamformer optimizes SNR for each frequency:

$$\boldsymbol{w}_{\mathrm{GEV}}(f) = \underset{\boldsymbol{d}}{\operatorname{argmax}} \frac{\boldsymbol{d}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{XX}}(f)\boldsymbol{d}}{\boldsymbol{d}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{NN}}(f)\boldsymbol{d}}$$

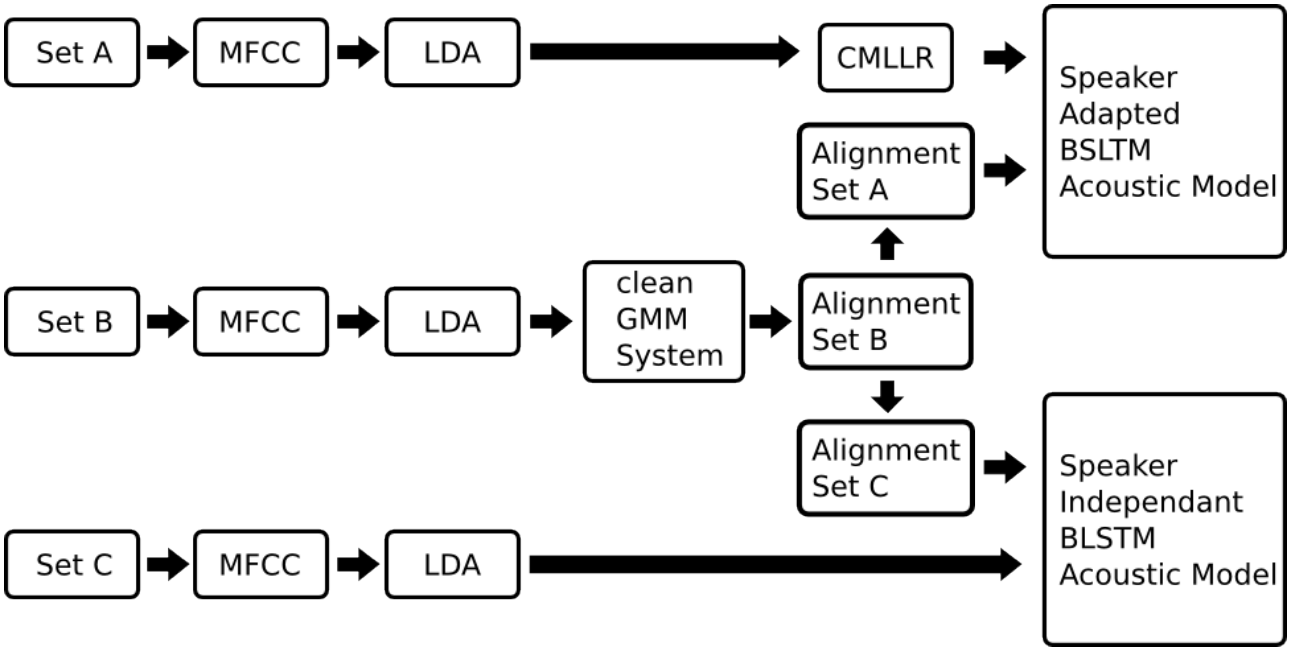
- Independent of sensor array geometry & speaker location

Wide residual acoustic model (WRN)



Back ends

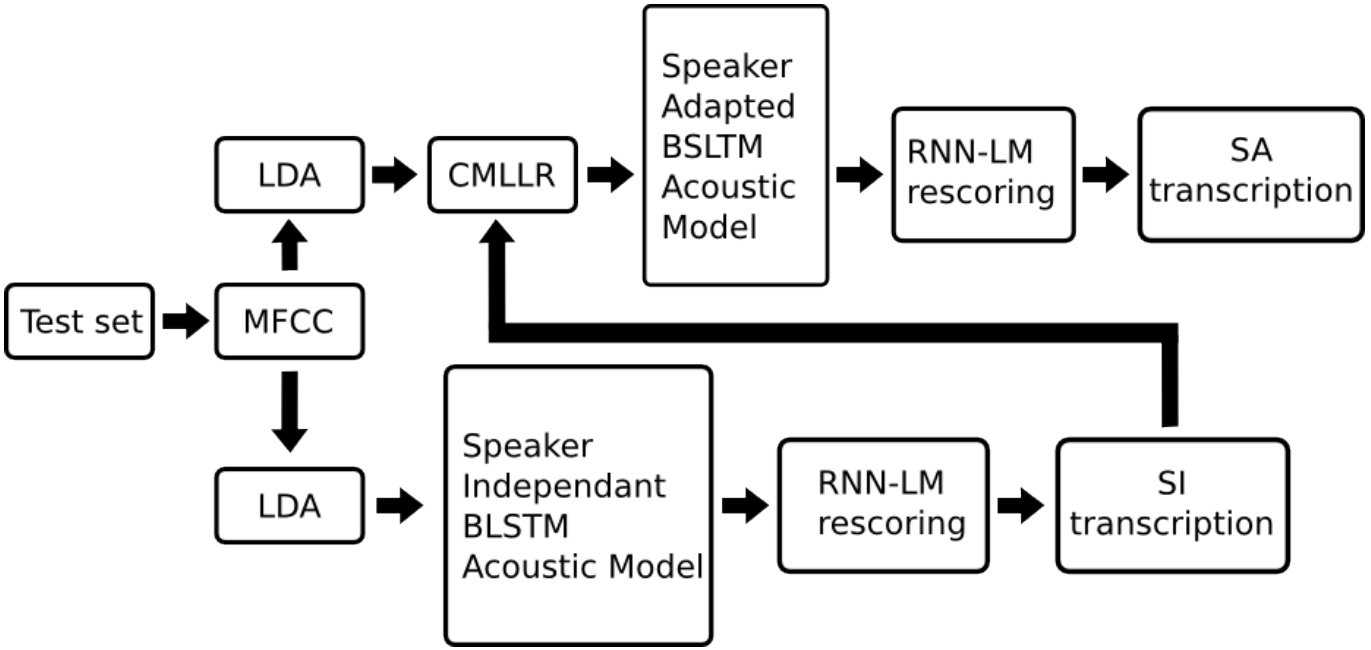
BLSTM acoustic model (RASR)



- 5 BLSTM layer
- 600 units per layer
- Dropout, L2-Regularization
- Gradient noise, NADAM

- Set A: Channel 1,3-6 + optinal extension
- Set B: Quasi clean data
- Set C: Channel 1,3-6

BLSTM acoustic model (RASR)



- 5 BLSTM layer
- 600 units per layer
- Dropout, L2-Regularization
- Gradient noise, NADAM

Back ends

Performance in various stages of BLSTM backend

System				Dev		Test	
Front end	Back end	Speaker adaptation	RNN-LM rescoring	real	simu	real	simu
Baseline	Kaldi	X	X	5.75	6.76	11.49	10.89

Back ends

Performance in various stages of BLSTM backend

System				Dev		Test	
Front end	Back end	Speaker adaptation	RNN-LM rescoring	real	simu	real	simu
Baseline	Kaldi	X	X	5.75	6.76	11.49	10.89
	RASR			7.80	8.71	11.81	13.89

Back ends

Performance in various stages of BLSTM backend

System				Dev		Test	
Front end	Back end	Speaker adaptation	RNN-LM rescoring	real	simu	real	simu
Baseline	Kaldi	X	X	5.75	6.76	11.49	10.89
	RASR		X	7.80	8.71	11.81	13.89
				5.87	6.43	9.35	10.55

Back ends

Performance in various stages of BLSTM backend

System				Dev		Test	
Front end	Back end	Speaker adaptation	RNN-LM rescoring	real	simu	real	simu
Baseline	Kaldi	X	X	5.75	6.76	11.49	10.89
	RASR	X		7.80	8.71	11.81	13.89
			X	5.87	6.43	9.35	10.55
				6.22	8.10	9.69	11.70

Back ends

Performance in various stages of BLSTM backend

System				Dev		Test	
Front end	Back end	Speaker adaptation	RNN-LM rescoring	real	simu	real	simu
Baseline	Kaldi	X	X	5.75	6.76	11.49	10.89
	RASR			7.80	8.71	11.81	13.89
			X	5.87	6.43	9.35	10.55
		X		6.22	8.10	9.69	11.70
		X	X	4.34	5.65	6.83	8.16

Performance of different front ends

System		Dev		Test	
Front end	Back end	real	simu	real	simu
Baseline	Kaldi	5.75	6.76	11.49	10.89
Baseline	RASR	4.34	5.65	6.83	8.16

Performance of different front ends

System		Dev		Test	
Front end	Back end	real	simu	real	simu
Baseline	Kaldi	5.75	6.76	11.49	10.89
Superdirective		5.47	6.34	11.47	10.42
Baseline	RASR	4.34	5.65	6.83	8.16
Superdirective		3.89	5.14	6.59	7.99

Back ends

Performance of different front ends

System		Dev		Test	
Front end	Back end	real	simu	real	simu
Baseline	Kaldi	5.75	6.76	11.49	10.89
Superdirective		5.47	6.34	11.47	10.42
MVDR		4.63	5.44	8.73	8.62
Baseline	RASR	4.34	5.65	6.83	8.16
Superdirective		3.89	5.14	6.59	7.99
MVDR		3.90	5.23	5.65	8.36

Back ends

Performance of different front ends

System		Dev		Test	
Front end	Back end	real	simu	real	simu
Baseline	Kaldi	5.75	6.76	11.49	10.89
Superdirective		5.47	6.34	11.47	10.42
MVDR		4.63	5.44	8.73	8.62
GEV		3.70	3.72	5.76	4.24
Baseline	RASR	4.34	5.65	6.83	8.16
Superdirective		3.89	5.14	6.59	7.99
MVDR		3.90	5.23	5.65	8.36
GEV		3.27	3.41	4.02	3.93

Variation of training set for speaker adapted RASR model

System		Dev		Test	
Front end	Trainingset extension	real	simu	real	simu
Baseline	-	4.34	5.65	6.83	8.16
	Baseline	4.11	5.77	6.82	8.53

Back ends

Variation of training set for speaker adapted RASR model

System		Dev		Test	
Front end	Trainingset extension	real	simu	real	simu
Baseline	-	4.34	5.65	6.83	8.16
	Baseline	4.11	5.77	6.82	8.53
Superdirective	-	3.89	5.14	6.59	7.99
	Superdirective	3.74	5.03	6.52	7.84

Variation of training set for speaker adapted RASR model

System		Dev		Test	
Front end	Trainingset extension	real	simu	real	simu
Baseline	-	4.34	5.65	6.83	8.16
	Baseline	4.11	5.77	6.82	8.53
Superdirective	-	3.89	5.14	6.59	7.99
	Superdirective	3.74	5.03	6.52	7.84
MVDR	-	3.90	5.23	5.65	8.36
	MVDR	3.57	4.73	5.58	6.28

Variation of training set for speaker adapted RASR model

System		Dev		Test	
Front end	Trainingset extension	real	simu	real	simu
Baseline	-	4.34	5.65	6.83	8.16
	Baseline	4.11	5.77	6.82	8.53
Superdirective	-	3.89	5.14	6.59	7.99
	Superdirective	3.74	5.03	6.52	7.84
MVDR	-	3.90	5.23	5.65	8.36
	MVDR	3.57	4.73	5.58	6.28
GEV	-	3.27	3.41	4.02	3.93
	GEV	3.05	2.79	3.77	2.67

System combination

Combination of multiple beamformers and fixed back end (RASR)

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93

System weights				Dev		Test	
4	3	2	1	real	simu	real	simu

System combination

Combination of multiple beamformers and fixed back end (RASR)

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93

System weights				Dev		Test	
4	3	2	1	real	simu	real	simu
0.50	0.50			2.75	3.13	3.57	3.56

System combination

Combination of multiple beamformers and fixed back end (RASR)

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93

System weights				Dev		Test	
4	3	2	1	real	simu	real	simu
0.50	0.50			2.75	3.13	3.57	3.56
0.33	0.33	0.33		2.70	3.18	3.55	3.77

System combination

Combination of multiple beamformers and fixed back end (RASR)

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93

System weights				Dev		Test	
4	3	2	1	real	simu	real	simu
0.50	0.50			2.75	3.13	3.57	3.56
0.33	0.33	0.33		2.70	3.18	3.55	3.77
0.40	0.25	0.25	0.10	2.61	3.07	3.40	3.46

System combination

System combination of submitted system

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93
5	GEV	WRN	CH1-6	2.73	2.34	3.48	2.76

System weights					Dev		Test	
5	4	3	2	1	real	simu	real	simu

System combination

System combination of submitted system

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93
5	GEV	WRN	CH1-6	2.73	2.34	3.48	2.76

System weights					Dev		Test	
5	4	3	2	1	real	simu	real	simu
0.45	0.55				2.48	2.47	3.12	2.90

System combination

System combination of submitted system

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93
5	GEV	WRN	CH1-6	2.73	2.34	3.48	2.76

System weights					Dev		Test	
5	4	3	2	1	real	simu	real	simu
0.45	0.55				2.48	2.47	3.12	2.90
0.43	0.33	0.34			2.25	2.30	2.98	2.61

System combination

System combination of submitted system

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93
5	GEV	WRN	CH1-6	2.73	2.34	3.48	2.76

System weights					Dev		Test	
5	4	3	2	1	real	simu	real	simu
0.45	0.55				2.48	2.47	3.12	2.90
0.43	0.33	0.34			2.25	2.30	2.98	2.61
0.35	0.20	0.20	0.25		2.19	2.34	2.91	2.68

System combination

System combination of submitted system

System				Dev		Test	
ID	Front end	Back end	Training set	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective	3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR	3.57	4.73	5.58	6.28
4	GEV		CH1,3-6	3.27	3.41	4.02	3.93
5	GEV	WRN	CH1-6	2.73	2.34	3.48	2.76

System weights					Dev		Test	
5	4	3	2	1	real	simu	real	simu
0.45	0.55				2.48	2.47	3.12	2.90
0.43	0.33	0.34			2.25	2.30	2.98	2.61
0.35	0.20	0.20	0.25		2.19	2.34	2.91	2.68
0.35	0.20	0.20	0.25	0.00	2.19	2.34	2.91	2.68

System combination

Overview over submitted results

Tr.	System			Dev		Test	
	Front end	Back end	Training set	real	simu	real	simu
1ch	-	RASR	CH1,3-6	7.42	9.86	12.02	15.22
	-	WRN	CH1-6	5.19	6.69	9.34	11.11
	COM_1ch			5.14	7.40	**9.29**	12.36
2ch	COM_2ch			3.02	4.04	**5.32**	5.27
6ch	COM_6ch			2.19	2.34	**2.91**	2.68

Post evaluation - Sequence training

System					Dev		Test	
ID	Front end	Back end	Training set	Sequence training	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline		4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective		3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR		3.57	4.73	5.58	6.28
4	GEV		CH1,3-6		3.27	3.41	4.02	3.93
4.1	GEV		CH1,3-6+GEV		3.05	2.79	3.77	2.67
5	GEV	WRN	CH1-6		2.73	2.34	3.48	2.76

System weights						Dev		Test	
5	4	4.2	3	2	1	real	simu	real	simu
0.35	0.20		0.20	0.25	0.00	2.19	2.34	2.91	2.68

Post evaluation - Sequence training

System					Dev		Test	
ID	Front end	Back end	Training set	Sequence training	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline	X	4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective		3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR		3.57	4.73	5.58	6.28
4	GEV		CH1,3-6		3.27	3.41	4.02	3.93
4.1	GEV		CH1,3-6+GEV		3.05	2.79	3.77	2.67
4.2	GEV		CH1,3-6		2.77	3.11	3.43	3.30
5	GEV	WRN	CH1-6		2.73	2.34	3.48	2.76

System weights						Dev		Test	
5	4	4.2	3	2	1	real	simu	real	simu
0.35	0.20		0.20	0.25	0.00	2.19	2.34	2.91	2.68

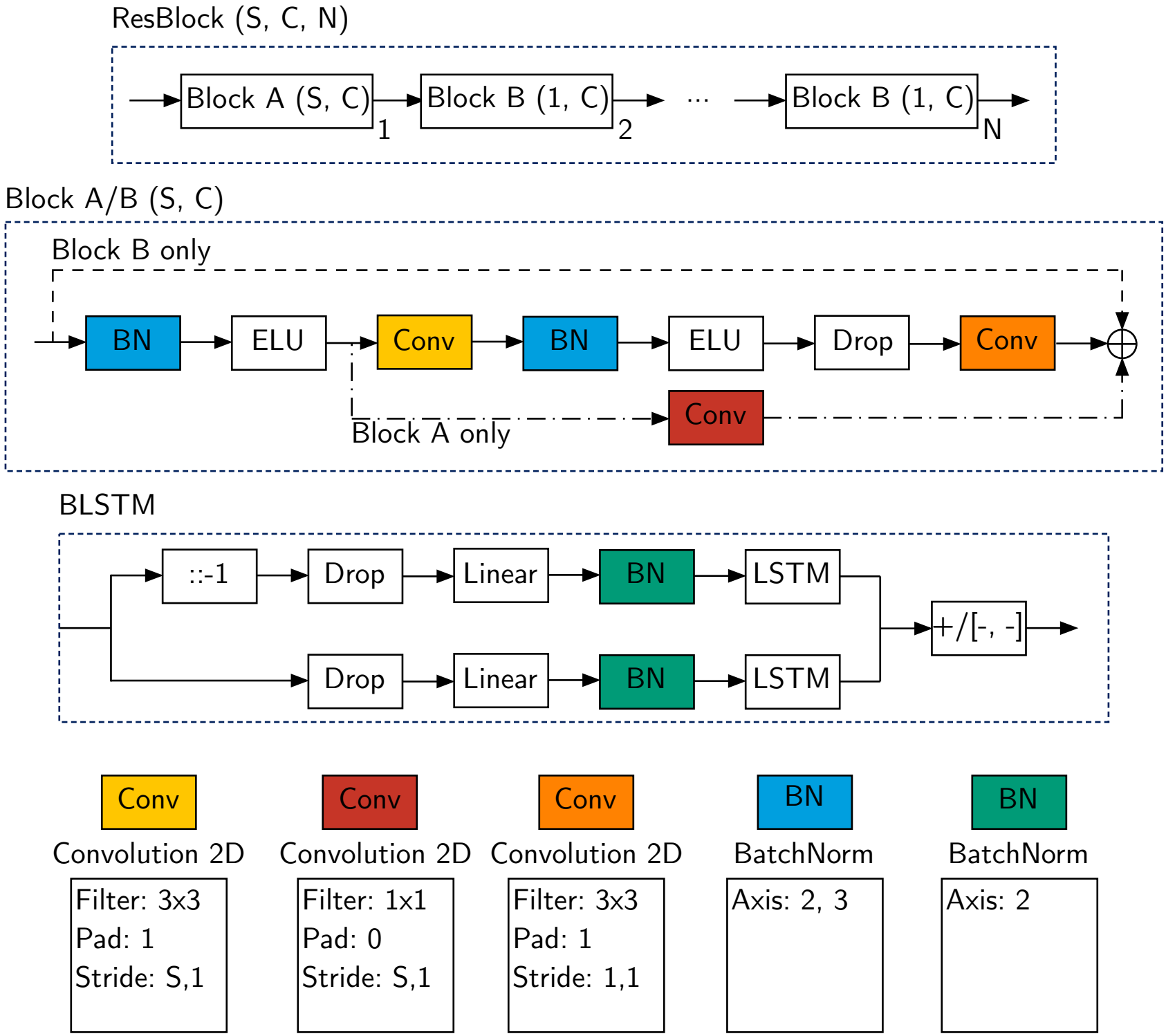
Post evaluation - Sequence training

System					Dev		Test	
ID	Front end	Back end	Training set	Sequence training	real	simu	real	simu
1	Baseline	RASR	CH1,3-6+Baseline		4.11	5.77	6.82	8.53
2	Superdirective		CH1,3-6+Superdirective		3.74	5.03	6.52	7.84
3	MVDR		CH1,3-6+MVDR		3.57	4.73	5.58	6.28
4	GEV		CH1,3-6		3.27	3.41	4.02	3.93
4.1	GEV		CH1,3-6+GEV		3.05	2.79	3.77	2.67
4.2	GEV		CH1,3-6	X	2.77	3.11	3.43	3.30
5	GEV	WRN	CH1-6		2.73	2.34	3.48	2.76

System weights						Dev		Test	
5	4	4.2	3	2	1	real	simu	real	simu
0.35	0.20		0.20	0.25	0.00	2.19	2.34	2.91	2.68
0.30		0.25	0.20	0.10	0.15	2.09	2.32	2.71	2.47

Thank you for your attention

Any questions?


ResBlock (S, C, N)
Block A (S, C)
1
Block B (1, C)
2
···
Block B (1, C)
N
Block A/B (S, C)
Block B only
BN
ELU
Conv
BN
ELU
Drop
Conv
Conv
Block A only
BLSTM
::-1
Drop
Linear
BN
LSTM
+/[-, -]
Drop
Linear
BN
LSTM
Conv
Convolution 2D
Filter: 3x3
Pad: 1
Stride: S,1
Conv
Convolution 2D
Filter: 1x1
Pad: 0
Stride: S,1
Conv
Convolution 2D
Filter: 3x3
Pad: 1
Stride: 1,1
BN
BatchNorm
Axis: 2, 3
BN
BatchNorm
Axis: 2