A Trajectory-based Parallel Model Combination with a Unified Static and Dynamic Parameter Compensation For Noisy Speech Recognition

Khe Chai SIM, Minh-Thang LUONG

simkc@comp.nus.edu.sg, luongmin@comp.nus.edu.sg

NUS National University of Singapore — School of Computing

Abstract:

Parallel Model Combination (PMC) is widely used as a technique to compensate Gaussian parameters of a clean speech model for noisy speech recognition. The basic principle of PMC uses a log normal approximation to transform statistics of the data distribution between the cepstral domain and the linear spectral domain. Typically, further approximations are needed to compensate the dynamic parameters separately. In this paper, Trajectory PMC (TPMC) is proposed to compensate both the static and dynamic parameters. TPMC uses the explicit relationships between the static and dynamic features to transform the static and dynamic parameters into a sequence (trajectory) of static parameters, so that the log normal approximation can be applied. Experimental results on WSJCAM0 database corrupted with additive babble noise reveals that the proposed TPMC method gives promising improvements over PMC and VTS.

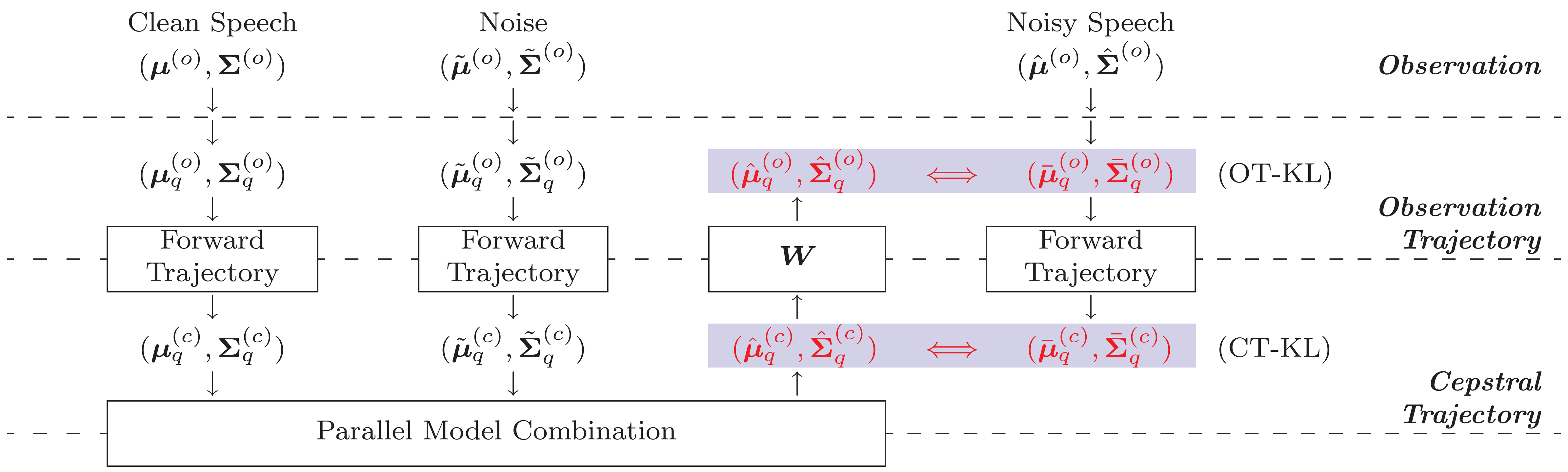
Parallel Model Combination:

- Clean Model + Noise Model = Noisy Model
- Use *log normal approximation*
- Not directly applicable to dynamic parameters

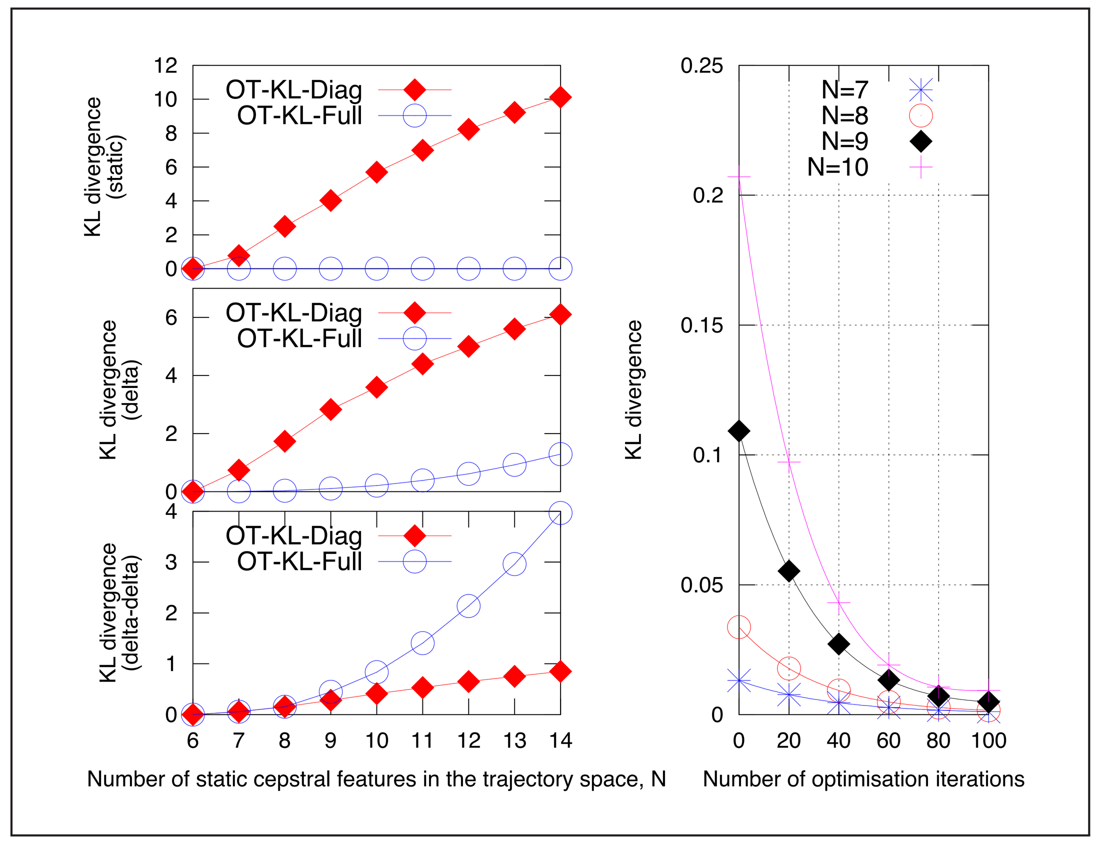
Trajectory HMM [3]:

- Models likelihood of *cepstral trajectory*
- *Cepstral trajectory* contains dynamic information
- Explicit static/dynamic relationship

Trajectory-based PMC (TPMC):



1. ***Observation → Cepstral Trajectory***
   - Constant statistics within state (N times)
   - Linear transformation: $\boldsymbol{o} = \boldsymbol{W}\boldsymbol{c}$
2. ***Parallel Model Combination***
   - Apply PMC in *cepstral trajectory* domain
3. ***Cepstral Trajectory → Observation***
   - **OT-KL:** Min KL-div in observation trajectory
   - **CT-KL:** Min KL-div in cepstral trajectory



Reversibility of TPMC in noise-free condition

Experimental Results:

- Corpus: WSJCAM0 corrupted with babble noise from NOISEX
- 5k task (Bigram full decoding + Trigram lattice rescoring): average of `si_dt5a` & `si_dt5b`
- Cross-word triphone HMM models (4000 physical states) with `MFCC_0_D_A`
- WER performance of baseline models:

No. of Components	Model	Clean	20dB	10dB	0dB
1	Clean	17.21	27.30	57.40	95.28
	Noisy (SPR)	–	19.32	29.64	55.83
16	Clean	8.49	17.84	50.59	93.65
	Noisy (SPR)	–	9.48	17.59	44.11

- WER performance of 1-component TPMC using different trajectory length, N:

SNR (dB)	OT-KL					CT-KL	
	6	7	8	9	10	8	10
20	**21.61**	22.14	23.83	26.38	29.60	**22.27**	23.55
10	39.67	37.71	**37.51**	40.45	45.01	36.86	**36.58**
0	71.57	68.67	**66.70**	67.75	71.17	66.44	**64.93**

- KL-div of various 1-component models *w.r.t.* the baseline Noisy (SPR) model:

SNR (dB)	PMC [1]			VTS [4]			TPMC		
	$\Delta^{(0)}$	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(0)}$	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(0)}$	$\Delta^{(1)}$	$\Delta^{(2)}$
20	**0.033**	**0.061**	**0.059**	0.051	0.083	0.104	0.037	0.088	0.113
10	0.071	0.218	0.202	0.091	0.147	0.186	**0.069**	**0.101**	**0.123**
0	0.120	0.664	0.596	0.134	0.180	0.230	**0.111**	**0.110**	**0.151**

- WER performance comparison of various 16-component models:

SNR (dB)	PMC [1]	VTS [4]	DPMC [2] (+`wgt`)	TPMC
20	11.76	11.26	17.32 (10.68)	10.84
10	29.53	20.25	28.42 (19.10)	19.68
0	73.25	51.57	58.90 (49.86)	50.20

Note: `+wgt` means components weights are also updated

Conclusions:

This paper has presented an extension to the standard Parallel Model Combination (PMC) technique that offers a solution to compensate both the static and dynamic parameters in a unified manner. The proposed method is called Trajectory PMC (TPMC) as it is motivated by the trajectory HMM formulation. The explicit relationships between the static and dynamic features are used to derive the statistics in the *cepstral trajectory* domain such that log normal approximation can be applied. The proposed TPMC method was found to yield consistently better performance compared to the standard PMC and VTS methods, both in terms of the Kullback-Leibler divergence and word error rate evaluations.

References

[1] M. J. F. Gales, "Model-based techniques for noise robust speech recognition," Ph.D. dissertation, Gonville and Caius College, University of Cambridge, 1996.

[2] M. J. F. Gales and S. J. Young, "A fast and flexible implementation of parallel model combination," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 1, pp. 133–136, 1995.

[3] H. Zen, K. Tokuda, and T. Kitamura, "Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences," *Computer Speech & Language*, vol. 21, no. 1, pp. 153–173, 2007.

[4] A. Acero, L. Deng, T. Kristjansson, and J. Zhang, "HMM adaptation using vector Taylor series for noisy speech recognition," in *Proc. of ICSLP*, vol. 3, 2000, pp. 869–872.