The Kaldi Speech Recognition Toolkit

Daniel Povey[1], Arnab Ghoshal[2,3], Gilles Boulianne[4], Lukáš Burget[5,6], Ondřej Glembek[5], Nagendra Goel[6], Mirko Hannemann[5], Petr Motlíček[8], Yanmin Qian[9], Petr Schwarz[5], Jan Silovský[10], Georg Stemmer[11], Karel Veselý[5]

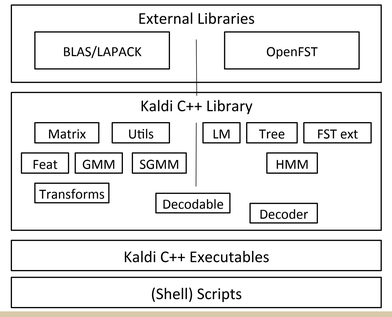
[1] Microsoft Research, USA; [2] University of Edinburgh, UK; [3] Saarland University, Germany; [4] Centre de Recherche Informatique de Montréal, Canada; [5] Brno University of Technology, Czech Republic; [6] SRI International, USA; [7] Go-Vivace Inc., USA; [8] IDIAP Research Institute, Switzerland; [9] Tsinghua University, China; [10] Technical University of Liberec, Czech Republic; [11] University of Erlangen-Nuremberg, Germany

KALDI

http://kaldi.sf.net

Features of Kaldi

- Integration with Finite State Transducers
- Extensive linear algebra support
- Extensible design
- Open license
- Complete recipes
- Thorough testing



Standard ASR techniques supported in Kaldi

- Acoustic front-end supports MFCC and PLP features, with cepstral mean and variance normalization, LDA, STC/MLLT, HLDA, VTLN, etc.
- HMM/GMM acoustic models
- No language modeling code, but support converting ARPA format LMs to FSTSs.

Features unique to Kaldi

- SGMM acoustic models
- Exponential transform

The Kaldi Decoder

Databases

- Resource Management (RM)
  - ▷
  - ▷
- Wall Street Journal (WSJ)
  - ▷

Comparison with previously published results

	Test set				
	Feb'89	Oct'89	Feb'91	Sep'92	Avg
HTK	2.77	4.02	3.30	6.29	4.10
Kaldi	3.20	4.21	3.50	5.86	4.06

	Test set	
	Nov'92	Nov'93
Bell	11.9	15.4
HTK (+GD)	11.1	14.5
KALDI	11.8	15.0

Other Results

- AR-Face: 110 classes, 110 train ("one-shot" training), 550 test

	RM (Avg)	WSJ Nov'92	WSJ Nov'93
Triphone	3.97	12.5	18.3
+ fMLLR	3.59	11.4	15.5
+ LVTLN	3.30	11.1	16.4
Splice-9 + LDA + MLLT	3.88	12.2	17.7
+ SAT (fMLLR)	2.70	9.6	13.7
+ SGMM + spk-vecs	2.45	10.0	13.4
+ fMLLR	2.31	9.8	12.9
+ ET	2.15	9.0	12.3

Conclusions

-