The SRI NIST SRE08 Speaker Verification System

M. Graciarena, S. Kajarekar, N. Scheffer
E. Shriberg, A. Stolcke

SRI International

L. Ferrer, *Stanford U. & SRI*

T. Bocklet, *U. Erlangen & SRI*

Talk Outline

- Introduction
  - SRI approach to SRE08
  - Overview of systems
  - Development data and submissions
- System descriptions
  - ASR updates
  - Cepstral systems
  - Prosodic systems
  - Combiner
- Results and analyses
- Conclusions

Introduction: SRI Approach

- Historical focus
  - Higher-level speaker modeling using ASR
  - Modeling many aspects of speaker acoustics & style
- For SRE08:
  - 14 systems (though some are expected to be redundant)
  - Some systems have ASR-dependent and –independent versions
  - System selection would have required more development data
  - Relied on LLR combiner to be robust to large number of inputs
  - Also: joint submission with ICSI and TNO (see David v. L. talk)
- Effort to do well on non-English and on altmic conditions
  - However, oversight for non-English: system lacked proper across-language calibration. Big improvement in Condition 6 once fixed.
  - Excellent telephone altmic results

Overview of Systems

Feature	ASR-independent	ASR-dependent
Cepstral	**MFCC GMM-LLR**	**Constrained GMM-LLR***
	MFCC GMM-SV	
	PLP GMM-SV	
	MFCC Poly-SVM	
	PLP Poly-SVM	
MLLR	**Phoneloop MLLR**	MLLR
Prosodic	**Poly coeff SV**	**SNERF+GNERF SVM**
	Poly coeff GMM-wts	
Duration		Word, state duration GMM-LLR
Lexical		Word N-gram SVM

- Systems in **red/bold** are new* or have improved features

Interview Data Processing

- Development data
  - Small number of speakers
  - Samples not segmented according to eval conditions; contain read speech
- VAD choices
  - **NIST VAD – uses interviewer channel and lapel mic (too optimistic?)**
  - NIST ASR – should be even better than NIST VAD, but dev results were similar
  - SRI VAD – uses subject target mic data only, results would not be comparable with other sites
  - Hybrid – successful for other sites; not investigated due to lack of time
- ASR choices
  - NIST ASR obtained from lapel mic
  - **SRI ASR obtained from interviewee side** – needed for intermediate output and feature consistency with telephone data
- Despite not training or tuning on interview data, performance was quite good
  - Compared to other sites that did no special interview processing
- Separate SRI study varying style, vocal effort, and microphone, shows cepstral systems don't suffer from style mismatch between interviews and conversations if channel constant (Interspeech 2008)

Development Data and Submissions

- SRE08 conditions 5-8 had dev data from SRE06
- For conditions 1-4, used altmic as a surrogate for interview data
  - MIT kindly provided dev data key for all altmic/phone combinations

Conversation		Phonecall (test)		Interview (test)
Type	**Mic type**	**phn**	**mic**	**mic**
Phonecall (train)	**phn**	**1conv4w-1conv4w *(condition 6,7,8)***	**1conv4w-1convmic *(condition 5)***	
	mic	***(not evaluated in SRE08)***		
Interview (train)	**mic**	**1convmic-1conv4w *(condition 4)***		**1convmic-1convmic *(condition 1,2,3)***

- Submissions
  - **short2-short3** (main focus of development)
  - **8conv-short3**
  - **long-short3** and **long-long** (submitted "blindly", not discussed here)

System Descriptions: ASR Update

- Same system architecture as in SRE06
  1. Lattice generation (MFCC+MLP features)
  2. N-best generation (PLP features)
  3. LM and prosodic model rescoring; confusion network decoding
- Improved acoustic and language modeling
  - Added Fisher Phase 1 as training data; web data for LM training
  - Extra weight given to nonnative speakers in training
  - State-of-the-art discriminative techniques: MLP features, fMPE, MPE
- Experimented with special processing for altmic data
  - Apply Wiener filtering (ICSI Aurora implementation) before segmentation
  - Distant-microphone acoustic models gave no tangible gains over telephone models
- Runs in 1xRT on 4-core machine

Results with New ASR

- Word error rates (transcripts from LDC and ICSI)

ASR System	Fisher 1 native	Mixer 1 native	Mixer 1 nonnative	SRE06 altmic
SRE06	23.3	29.4	49.5	35.3
SRE08	17.0	23.0	36.1	28.8
Rel. WER reduction	**27%**	**22%**	**27%**	**18%**

- Effect on ASR-based speaker verification
  - Identical SID systems on SRE06 English data (minDCF/EER)
  - No NAP or score normalization

ASR System	MLLR tel	MLLR altmic	SNERF altmic	Word N-gram tel
SRE06	.156/3.47	.250/6.46	.645/16.46	.831/24.1
SRE08	.147/2.82	.228/6.25	.613/15.79	.818/23.5
Rel. DCF reduction	**5.8%**	**8.8%**	**5.0%**	**1.6%**

- Nativeness ID (using MLLR-SVM): 12.5% ⇒10.9% EER

Cepstral Systems: GMMs

- Front-end for GMM-based cepstral systems
  - 12 cepstrum + c0, delta, double and triple (52)
  - 3 GMM based systems submitted, 1 LLR, 2 SVs
- GMM-LLR system
  - MFCCs, 2048 Gaussian, Eigenchannel MAP
  - Gender-independent system, but **gender-DEPENDENT ZTnorm**
  - ISV and Score normalization data: SRE04 and SRE05 altmic.
  - Background data: Fisher-1, Switchboard-2 phase 2,3 and 5
- GMM-SVs system
  - 1024 Gaussian gender-dependant systems
  - MFCC : use HLDA to get from 52 to 39
  - PLP : use MLLT + LDA to get from 52 to 39
  - Score-level combination (feature level gives similar performances)
  - PLP is optimized for phonecall conditions

Cepstral systems: GMMs (2)

- ISVs for GMM-SVs:
  - Factor Analysis estimators: 4 ML iterations, 1 MDE final iteration
  - MFCC
    - Concatenation of 50 EC from SRE04 + 50 EC from SWB2 phase 2,3,5 + 50 EC from SRE05 altmic
    - Surprising results on altmic conditions (8conv)
  - PLP
    - Concatenation of 80 EC from SRE04 + 80 EC from SRE05 altmic
- Combination
  - GMM-LLR and GMM-SVs have equivalent performances
  - Combination of gender-independent and -dependent was good strategy
- Particularities
  - PLP-based systems use VTLN and SAT transforms (borrowed from ASR front-end)
  - Should remove speaker information but gives better results in practice
  - Did not find any improvement on “short” conditions when using JFA instead of Eigenchannel MAP

Cepstral Systems: MLLR SVM

- ASR-dependent system (for English)
  - PLP features, 8 male + 8 female transforms, rank-normalized
  - Same features as in 2006, but better ASR
  - NAP [32 d] trained using combined SRE04 + SRE05-altmic data
- ASR-independent system (for all languages)
  - Based on (English) phone loop model
  - NAP [64 d] on SRE04 + SRE05-altmic + non-English data
  - Improved since '06 by making features same as ASR-dep. MLLR:
    MFCC $\Rightarrow$ PLP and 2 + 2 transforms $\Rightarrow$ 8 + 8 transforms

Feature	Transforms	ASR?	SRE06 English	SRE06 All *
MFCC	2+2	no	.189 / 3.90	.270 / 5.92
PLP	2+2	no	.154 / 3.36	.266 / 5.42
PLP	8+8	no	.138 / 2.87	.260 / 5.23
PLP	8+8	yes	.111 / 2.22	n/a

* No language calibration used

Constrained Cepstral GMM (1)

- New system for English. Submitted for 1conv ("short") training only
- Best among all SRI systems for short2-short3 condition
- Combines 8 subsystems that use frames matching 8 constraints:
  - Syllable onsets (1), nuclei (2), codas (3)
  - Syllables following pauses (4), one-syllable words (5)
  - Syllables containing [N] (6), or [T] (7), or [B,P,V,F] (8)
- Unlike previous word- or phone-conditioned cepstral systems:
  - Uses automatic syllabification of phone output from ASR
  - Model does not cover all frames, and subsets can reuse frames
- Modeling:
  - GMMs, background models trained on SRE04, no altmic data
  - ISV: 50 eigenchannels matrix trained on SRE04+05 altmic data
  - Score combination via logistic regression, no side information
  - ZT-Norm used for score normalization (trained on e04)

Constrained Cepstral GMM (2)

- Post-eval analyses show that across SRE08 conditions:
  - 4 or 5 constraints give similar performance to 8
  - Best systems include nuclei, onset, and [N]-in-syllable constraints
- After evaluation, finished 8conv training and testing. This is the best system among all SRI systems on this condition.
- Future Work:
  - Better explore candidate constraint combinations. (Used crude forward search on pre-ISV constraints for evaluation.)
  - Port to language-independent system that uses phone recognition
  - Combine constraints into a single supervector system
  - Include altmic data in background model, improve altmic robustness
  - Publication in preparation

Prosodic Systems (1)

- Pitch and energy signals obtained with get_f0
  - Waveforms preprocessed with a bandpass filter (250-3500)
- ASR-independent systems
  - Features:
    - Polynomial approximation of pitch and energy profiles over pseudo-syllables + region length (Dehak '07)
  - GMM supervector modeling (Dehak '07):
    - Order 5 polynomial coefficients with mean-variance norm. applied
    - Joint Factor Analysis on gender-dependent 256-mixture GMM models
    - Eigenvoice (70 EV on Fisher2 + NIST SRE 04 + NIST SRE 05 altmic)
    - Eigenchannel + Diagonal model (50 EC on e04+e05), same for diagonal d)
  - Weight modeling + SVM:
    - All polynomial orders from 0 to 5 used
    - One GMM trained for each individual feature, certain subsets and their sequences. Features are adapted weights
    - Transformed vectors are rank-normed, 16 NAP directions subtracted
    - Model these features with SVM regression and perform TZ-norm.

Prosodic Systems (2)

- ASR-dependent system
  - Features: Prosodic polynomial features plus two more sets
    - SNERFs (syllable NERFs): extracted from all (real) syllables
    - GNERFS (grammar-constrained NERFs): extraction location constrained to specific “wordlists”
    - Extract features over those regions
    - Features reflect characteristics about the pitch, energy and duration patterns
  - Weight modeling + SVM:
    - Transform features and model them using same method as language independent system (except use 32 NAP directions)
  - Performance is 50% better than language independent prosodic systems
  - 25% improvement in this system since 2006 evaluation from
    - Improvements in the feature transform
    - Use of eval04 data
    - Addition of polynomial features
  - Combination of ASR-dependant and ASR-independent features gives a high performance prosodic system

SRE06 Results (1conv4w English)

Systems (by approach) filled=ASR-dep.	Tel-Tel		Tel-Altmic		Altmic-Tel		Altmic-Altmic	
	%EER	DCF	%EER	DCF	%EER	DCF	%EER	DCF
Constrained **CEP**	**1.30**	0.075	2.48	0.111	3.31	0.150	5.76	0.392
CEP	1.90	0.095	2.19	0.100	4.05	0.149	3.87	0.259
SV-PLP	1.79	**0.074**	2.36	**0.080**	**2.67**	**0.111**	**3.05**	**0.170**
SV-MFCC	1.84	0.089	**1.90**	0.083	3.13	0.136	3.20	0.193
MLLR	2.38	0.108	4.01	0.140	4.55	0.167	4.84	0.204
MLLR-PL	2.76	0.136	5.84	0.199	6.11	0.240	6.95	0.279
POLY-MFCC	3.95	0.188	6.95	0.299	8.87	0.327	10.43	0.560
POLY-PLP	4.06	0.183	7.57	0.307	9.56	0.375	12.02	0.652
PROSODIC	7.64	0.350	10.72	0.444	12.41	0.547	13.31	0.604
POLY-**PROSODIC**	16.47	0.650	21.31	0.779	23.90	0.834	19.33	0.744
SV-**PROSODIC**	16.36	0.715	23.30	0.860	22.62	0.880	20.06	0.812
STATE-**DUR**	13.98	0.633	18.50	0.761	22.94	0.849	20.95	0.932
WORD-**DUR**	17.93	0.734	22.64	0.828	25.47	0.894	26.62	0.887
WORD-NG	23.35	0.803	25.29	0.845	26.62	0.901	24.68	0.845

Combination Procedure

- Linear logistic regression with auxiliary information (ICASSP'08)
  - Auxiliary information conditions weights applied to each system
  - Weights obtained using a modified logistic regression procedure
  - Uses scores from a nativeness classifier for English speakers
- Combination strategy
  - Split each condition into two splits – English-English and others (*)
  - Train combiner separately for each split
  - Subtract threshold from each split
  - Pool scores for the two splits

<table>
<tr><td colspan="2">Conversation</td><td colspan="2">Phonecall (DEV data)</td><td>Interview (no DEV data)</td></tr>
<tr><td>Type</td><td>Mic type</td><td>phn</td><td>mic</td><td>mic</td></tr>
<tr><td rowspan="2">phonecall</td><td>phn</td><td>1conv4w-1conv4w
(condition 6,7,8)</td><td>1conv4w-1convmic
(condition 5)</td><td rowspan="2"></td></tr>
<tr><td>mic</td><td colspan="2">(not evaluated in SRE08)</td></tr>
<tr><td>Interview</td><td>mic</td><td>1convmic-1conv4w
(condition 4)</td><td></td><td>1convmic-1convmic
(condition 1,2,3)</td></tr>
</table>

Combination Analysis

- Submission results
  - **SRI_1:** 13 ASR-dependent systems for English and 8 ASR-independent systems for non-English (SNERF SVM system subsumes poly-coeff SVM system)
  - **SRI_2:** 8 ASR-independent systems for both English and non-English
- Combination results (based on SRE06) are presented as
  - **1BEST:** Best single system based on SRE06
  - **4BEST:** 4-best results obtained separately for English and non-English
  - **4CEP:** GMM-LLR + MLLR_PL + SV_PLP + SV_MFCC
    - ASR-independent cepstral systems, comparable to other sites

Results – Condition 7

English telephone in training and test

- *Constrained GMM not ready for SRI_1 ***8conv*** submission; was run later
- Up to 4 times reduction in EER and DCF from short2 → 8conv
  - Ordering is fairly consistent
- 8conv-short3 has very few errors. Best system has
  - EER - 3 FA, 49 FR
  - DCF - 7 FA, 17 FR
- Detailed analysis is presented for only short2-short3

Systems (filled rows = ASR-dep)	Short2-short3 (17761)		8conv-short3 (7408)	
	%EER	mDCF	%EER	mDCF
Constrained GMM	2.769	0.1342	*0.658**	*0.0396**
CEP GMM	2.914	0.1395	1.277	0.0565
SV-PLP	3.419	0.1424	1.095	0.0500
SV-MFCC	3.683	0.1427	1.312	0.0633
MLLR	4.154	0.1887	1.312	0.0639
MLLR_PL	4.154	0.1808	1.972	0.0839
POLY-MFCC	6.194	0.2452	2.190	0.1024
POLY-PLP	6.351	0.2496	2.632	0.1060
PROSODIC	10.016	0.4321	3.502	0.1614
STATE-DUR	14.820	0.6984	9.208	0.5091
POLY-PROSODIC	17.180	0.6939	10.253	0.4070
SV-PROSODIC	17.765	0.7532	12.282	0.5120
WORD-DUR	19.626	0.7793	8.113	0.3725
WORD-NG	20.685	0.7622	7.714	0.3992

Combination – Condition 7

- Short2-short3 English telephone

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4BEST	0.045	[illegible]	0.104	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

With nativeness calibration

- 4BEST = Constrained GMM + SV-PLP + PROS + MLLR (in order of importance)
  - ASR-based and prosodic systems are important
- Combinations give different relative performance on SRE06 than on SRE08
- Nativeness calibration gives small but consistent improvements
  - Individual systems are robust to nativeness variation

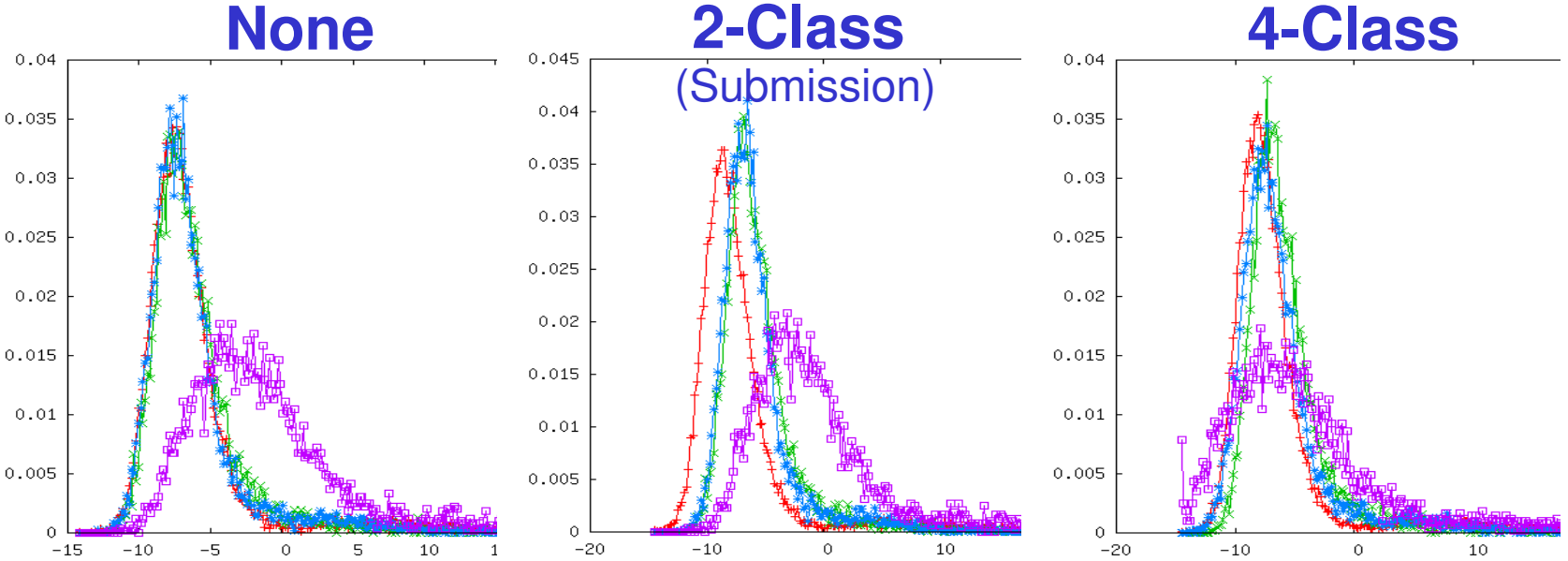
Results – Condition 6

ASR Independent systems - Telephone data in training and test

Systems	Short2-short3 (35896)		8conv-short3 (11849)	
	%EER	DCF	%EER	DCF
CEP GMM	7.178	0.3952	3.747	0.2490
POLY-MFCC	9.559	0.4508	4.439	0.2461
SV-MFCC	8.029	0.4541	4.866	0.2997
SV-PLP	8.209	0.4644	5.176	0.2924
POLY-PLP	9.934	0.4694	4.898	0.2475
MLLR_PL	9.410	0.5294	6.021	0.3767
SV-PROSODIC	20.545	0.8448	13.399	0.6252
POLY-PROSODIC	20.799	0.8947	12.248	0.6553

- Without nativeness calibration
- All systems are without language calibration
- Reduction by factor of 2 in EER and DCF with more data

Language calibration



- No calibration: surprisingly, trials with English in either train or test are more similar to trials with English in both train and test
  - Trials with non-English in both train and test have a bias
- In submission, we compensated language by splitting trials into English-English and rest. This left overall distribution with 3 peaks
- Post submission – We compensate trials with 4 classes – Train-Test, English-nonEnglish
- Does not affect English-English trials

Combination – Condition 6

- Short2-short3 – Telephone speech

Before language calibration (as submitted)

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
4CEP	0.140	2.821	0.547	0.408	7.095
SRI_1 (Nativeness)	0.124	2.574	0.503	0.372	6.834
SRI_2	0.137	2.738	0.538	0.397	6.871

After language calibration

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
4CEP	0.116	2.378	0.310	0.276	5.303
SRI_1 (Nativeness)	0.110	2.015	0.317	0.274	5.302
SRI_2	0.113	2.185	0.309	0.279	5.228

- Similar improvements as for non English results – better generalization of DCF values

Results - Condition 5

Telephone in training and Altmic in test

- 12 non-English trials are ignored
- Ordering of systems is fairly consistent
- More data reduces EER and DCF by a factor of 3
- Very few errors in 8conv-short3. Best system has
  - EER - 16 FA, 75 FR
  - DCF - 43 FA, 6 FR
- Detailed analysis is presented only for short2-short3

Systems (filled rows = ASR-dep system)	Short2-short3(8442)		8conv-short3(4308)	
	%EER	DCF	%EER	DCF
SV-MFCC	5.756	0.1914	2.110	0.0733
CEP GMM	7.394	0.2422	2.612	0.1009
SV-PLP	7.345	0.2465	4.341	0.1345
Constrained GMM	7.331	0.2549	4.083	0.0926
MLLR	9.929	0.3204	5.267	0.1350
MLLR_PL	9.655	0.3494	6.315	0.2064
POLY-MFCC	12.330	0.4207	5.920	0.2141
POLY-PLP	12.316	0.4525	7.362	0.2624
PROSODIC	13.891	0.5305	11.036	0.3733
WORD-NG	19.311	0.6359	12.629	0.4310
WORD-DUR	25.697	0.8011	18.032	0.6750
POLY-PROSODIC	25.550	0.8581	18.822	0.7278
SV-PROSODIC	28.287	0.8971	23.163	0.8577
STATE-DUR	25.675	0.9267	19.625	0.8002

Combination – Condition 5

- Short2-short3 common condition 5: Telephone training, Altmic test
  - 12 non-English trials are ignored in these results

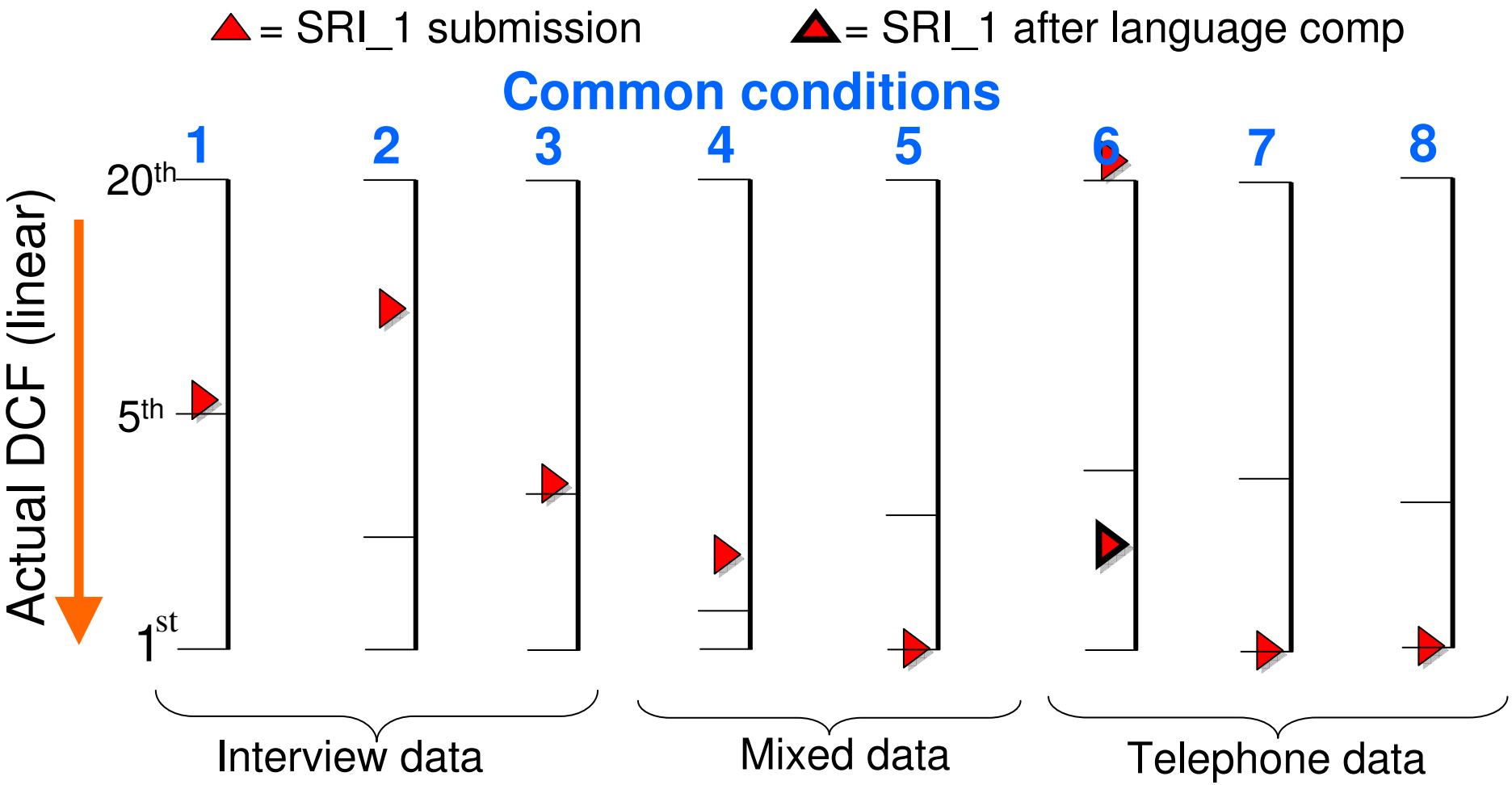
System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST (SV-MFCC)	0.077	1.780	0.209	0.193	5.685
4BEST	0.043	1.407	0.186	0.157	4.863
SRI_1 (14)	0.039	0.993	0.175	0.150	4.726
4CEP	0.047	1.407	0.197	0.153	4.795
SRI_2 (8)	0.045	1.117	0.200	0.161	4.863
SRI_1 (14)	0.044	1.117	0.177	0.151	4.863

With nativeness calibration

- 4BEST = SV-MFCC + SV-PLP + MLLR + PROSODIC (in order of importance)
  - Prosodic systems are important for this task
- Combinations give different relative performance on SRE06 than on SRE08
- Nativeness calibration gives small but consistent improvement

SRI Performance in Context

Actual DCF of the SRE08 primary submissions ranked 1st, 5th and 20th for short2-short3 common conditions



Summary and Conclusions (1)

- Achieved highly competitive performance with a combination of frame-level and higher-level systems
- ASR significantly improved, especially for nonnatives, altmic data
- Single best-performing subsystem: novel cepstral GMM variant using syllable-level constraints
- Newly developed and/or improved ASR independent systems:
  - Various ASR-independent cepstral GMM-LLR and GMM-SV systems
  - ASR-independent MLLR
  - Prosodic (added ASR-independent version)
- Performance on interview data relatively good
  - Despite the fact that we chose not to use the sample interview data, and that we used suboptimal VAD
  - Other teams showed that clear improvements are possible by investing in question of how to best use the sample data

Summary and Conclusions (2)

- Four system combination gives comparable performance to our primary submission (14 systems)
  - Found 4-best combinations typically use higher-level information (constrained GMM, MLLR, prosody)
  - But 4-way low-level cepstral system combination not far behind
- Order of importance of systems is fairly consistent with more training data
  - Errors reduced by a factor of up to 3 with 8conv training data
  - Low error count on 8conv condition prevents detailed analysis
- Found nativeness calibration for English speakers more important in SRE06 data than in SRE08 data
  - More analysis necessary with native labels from SRE08 data
  - May reflect distribution of L1s (*cf.* Odyssey 2008 paper)
- Language calibration is critical for good performance
  - Eng-nonEng trials more similar to all-Eng than to all-nonEng

Thank You

http://www.speech.sri.com/projects/verification/SRI-SRE08-presentation.pdf

Results for Other Conditions

Results – Condition 8

Native English Telephone Training and Testing

- *Constrained cepstral system not in 8conv submission (lack of time), finished later
- Up to 3 times reduction in EER and DCF from short2 → 8conv
- Very few errors in 8conv-short3. Best system has
  - EER – 3 FA, 43 FR
  - DCF – 6 FA, 12 FR
- Detailed analysis is presented only for short2-short3

Systems (filled rows = ASR-dep)	Short2-short3 (8489)		8conv-short3 (3993)	
	%EER	DCF	%EER	DCF
Constrained GMM	2.629	0.1156	*1.129**	*0.0545**
CEP GMM	2.629	0.1291	1.452	0.0616
SV-MFCC	3.453	0.1319	1.506	0.0583
SV-PLP	3.782	0.1453	1.559	0.0612
MLLR	4.441	0.1762	1.882	0.0597
MLLR_PL	4.606	0.1989	2.635	0.0696
POLY-MFCC	6.113	0.2423	1.882	0.1006
POLY-PLP	5.923	0.2695	3.025	0.1111
PROSODIC	10.694	0.4532	3.401	0.1482
STATE-DUR	16.281	0.7074	10.191	0.5242
POLY-PROSODIC	19.081	0.7256	10.957	0.4739
SV-PROSODIC	18.752	0.8104	15.004	0.5923
WORD-DUR	20.241	0.8027	8.685	0.3797
WORD-NG	22.205	0.7910	8.685	0.3709

Combination – Condition 8

- Short2-short3 common condition 8 – Native English in training and test
  - Nativeness calibration not applicable

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST (SV-PLP)*	0.074	1.788	0.166	0.145	3.783
4BEST	0.050	0.975	0.104	0.095	1.809
4CEP	0.064	1.246	0.116	0.106	2.126
SRI_2 (8)	0.063	1.192	0.123	0.111	2.138
SRI_1 (14)	0.052	0.867	0.105	0.099	1.809

- Although Constrained GMM is the best system on SRE08, the systems here are chosen based on SRE06 performance so 1BEST system is SV-PLP
- 4BEST = SV-PLP + Constrained GMM + Prosodic + Poly-PLP

Results – Condition 1

Interview Training and Testing

- SRE06 alt-alt performance significantly differs from SRE08 short2-short3, common condition=1
  - Mic v/s Mode
- ASR dependent systems are more affected by altmic and interview data
  - Segmentation issues

Systems (filled rows = ASR dep)	SRE06 alt-alt (132341)		SRE08 short2-short3 (34181)	
	%EER	DCF	%EER	DCF
SV-PLP	3.054	0.170	8.622	0.358
SV-MFCC	3.204	0.196	6.387	0.271
MLLR	4.839	0.204	12.929	0.446
CEP GMM	3.871	0.259	8.561	0.366
MLLR_PL	6.946	0.271	12.730	0.453
Constrained GMM	5.763	0.392	12.868	0.529
POLY-MFCC	10.430	0.560	15.139	0.668
PROSODIC	13.312	0.604	21.543	0.772
POLY-PLP	12.021	0.652	18.128	0.752
SV-PROSODIC	20.064	0.812	25.329	0.926
WORD-NG	24.688	0.866	33.267	0.999
WORD-DUR	24.172	0.887	35.797	1.000
STATE-DUR	20.946	0.932	37.461	0.999

Combination – Condition 1

- Short2-short3 – Interview Train and Test

System/ Combination (w/o nativeness comp)	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST (SV-PLP)	0.170	3.054	0.369	0.358	8.622
4BEST	0.121	2.193	0.285	0.278	7.036
4CEP	0.153	2.495	0.279	0.278	6.542
SRI_2 (8)	0.113	2.129	0.275	0.264	6.516
SRI_1 (13)	0.099	1.871	0.271	0.254	6.482

- SV-PLP is the best min DCF system based on SRE06
  - SV-MFCC is the best min DCF system based on SRE08
- DCF values are calibrated well given difference in performance
- 4BEST systems – SV-PLP, SV-MFCC, POLY-MFCC, MLLR

Results – Condition 4 (English)

Interview Training and Telephone Testing

- Results reported on English trials
  - About 1000 (10%) trials are non-English
- SRE08 performance is significantly worse than SRE06
  - DCF ranking is consistent

Systems (filled rows = ASR dep)	SRE06 alt-tel (19223)		SRE08 short2-short3 (10719)	
	%EER	DCF	%EER	DCF
SV-PLP	2.667	0.111	8.359	0.294
SV-MFCC	3.126	0.136	8.461	0.286
CEP GMM	4.046	0.149	7.747	0.363
Constrained GMM	3.310	0.150	9.582	0.399
MLLR	4.552	0.167	11.417	0.445
MLLR_PL	6.115	0.240	13.761	0.540
POLY-MFCC	8.874	0.327	14.067	0.611
POLY-PLP	9.563	0.375	16.106	0.806
PROSODIC	12.414	0.547	21.407	1.001
STATE-DUR	22.942	0.849	30.479	0.972
SV-PROSODIC	22.621	0.880	29.154	0.967
WORD-DUR	25.471	0.894	31.702	0.951
WORD-NG	26.621	0.901	33.945	0.294

Combination – Condition 4 (English)

- Short2-short3 – Interview Train and Telephone Test (English trials)

System/ Combination (w/o nativeness comp)	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST (SV-MFCC)	0.111	2.667	0.321	0.286	8.359
4BEST	0.066	1.563	0.297	0.215	5.505
4CEP	0.079	1.839	0.263	0.216	5.301
SRI_2 (8)	0.075	1.885	0.271	0.221	5.097
SRI_1 (13)	0.057	1.241	0.269	0.194	4.791

- 4BEST – SV-MFCC, SV-PLP, MLLR, PROSODIC
- Significantly better performance with 13 systems than 4 systems
- Calibration issue with SRE08 DCF values

Combination – Condition 6 (Non-English subset)

- Short2-short3 – "Non English telephone" subset

Suboptimal 2-class language calibration (as submitted)

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST(SV-PLP)	0.247	5.254	1.121	0.639	13.034
4CEP	0.209	4.294	0.998	0.596	11.655
SRI_1, SRI_2	0.199	4.124	0.888	0.564	11.103

"Corrected" (4-class language calibration)

System/ Combination	SRE06		SRE08		
	DCF(M)	%EER	DCF(A)	DCF(M)	%EER
1BEST(SV-PLP)	0.201	4.294	0.618	0.495	10.069
4CEP	0.166	3.277	0.503	0.417	8.207
SRI_1, SRI_2	0.160	3.051	0.471	0.420	8.000

- Overall about 30% improvement with correct language calibration
  - Actual DCF is closer to minimum DCF