2008 NIST Speaker Recognition Evaluation

Slides From Evaluation Workshop Presentation

Alvin Martin, Craig Greenberg
www.nist.gov/speech/tests/sre/2008/

June 17-18th, 2008

McGill University

Montreal, Quebec, Canada

Outline

- Introduction
- Participants
- Tests and Data Used
- Core Test
- Summary

NIST
National Institute of
Standards and Technology

Introduction

- Task
- Metrics
- Rules

NIST
National Institute of
Standards and Technology

Speaker Detection Task

- Given a target speaker and a test speech segment, determine if the target is speaking in the test segment
  - Each trial is defined by a *model* (target training data) and a *test segment*
  - Outputs required for each trial are a *decision* (*'T'* or *'F'*) and a *likelihood score* (preferably a *log-likelihood ratio*)

NIST
National Institute of Standards and Technology

Evaluation Metric

$$C_{Norm} = ((C_{Miss} * P_{Miss/Target} * P_{Target}) + (C_{FA} * P_{FA/NonTarget} * P_{NonTarget})) / C_{Default}$$

Cost of a miss	$C_{Miss} = 10$
Cost of a false alarm	$C_{FA} = 1$
Probability of a target	$P_{Target} = 0.01$
Probability of a non-target	$P_{Nontarget} = 1 - P_{Target} = 0.99$

A normalization factor ($C_{Default}$) is defined to make 1.0 the score of a knowledge-free system that always decides "False".

$$C_{Default} = \min(C_{Miss} * P_{Target}, C_{FA} * P_{Nontarget}) = .1$$

NIST
National Institute of Standards and Technology

Performance Representation

- Detection Error Tradeoff (DET) Plots
  - Shows the tradeoff of False Alarm and Miss error rates on a normal deviate scale
  - Actual decision points marked with a cross, minimum detection point marked with a circle
- Bar Graphs
  - Shows the contribution of two error types to C_{Norm} values

NIST
National Institute of Standards and Technology

Alternative Metric

- Log-likelihood-ratio based cost function

$$C_{llr} = 1 / (2 * \log 2) * (\sum \log(1+1/s)/N_{TT}) + (\sum \log(1+s))/ N_{NT})$$

  - The first summation is over all target trials, the second is over all non-target trials, N_{TT} *and* N_{NT} *are the total numbers of target* and non-target trials, respectively, and *s represents a trial's* likelihood ratio, = prob (data | target hyp.) / prob (data | non-target hyp.).
- Measures the effective amount of information that the system delivers to the user
- Is an application independent metric
- Requires likelihood scores to be estimated llr's
- Reference
  - *"Application-Independent Evaluation of Speaker Detection" in Computer Speech & Language, volume 20, issues 2-3, April-July 2006, pp. 230-275, by Niko Brummer and Johan du Preez*

NIST
National Institute of Standards and Technology

Evaluation Rules

- Each trial decision to be made independently
  - Based only on the specified segment and the speaker model
- Normalization over multiple test segments NOT allowed
  - Except for systems using unsupervised adaptation
- Normalization over multiple target speakers NOT allowed
- Use of evaluation data for impostor modeling NOT allowed
- Use of manually produced transcripts or any other human interaction with the data NOT allowed
- Knowledge of the model speaker gender ALLOWED
  - No cross sex trials

Outline

- Introduction
- **Participants**
- Tests and Data Used
- Core Test
- Summary

NIST
National Institute of
Standards and Technology

Participants

- 46 sites

Australia (2)	Canada	China (9)
Czech Republic (2)	Finland	France (5)
Germany	Israel	Italy (2)
Lebanon	Lithuania	Mexico
Netherlands	Singapore (2)	Slovenia
South Africa (2)	Spain (5)	Switzerland
United Kingdom	United States (6)	

- 46 sites or collaborations with submissions
- 107 total systems
  - 6 unsupervised adaptation systems
- 1 "mothballed" system
- 246 test condition/system combinations

NIST
National Institute of Standards and Technology

Participants – Africa, Middle East

NIST ID	Site	Location
PRS	Persay Ltd	Israel
	Spescom DataVoice	South Africa
	Stellenbosch University, DSP Group	South Africa
UOB	University of Balamand	Lebanon

NIST
National Institute of Standards and Technology

Participants – Asia

NIST ID	Site	Location
CASIA	Inst. of Automation, Chinese Acad. of Sciences*	China
CSLT	Centre for Speech and Language Technologies, Tsinghua Univ.	China
FTRD	France Telecom Orange Labs	China
iFly	IFlyTek Speech Lab, USTC*	China
IIR	Institute for Infocomm Research	Singapore
IOA	Inst, of Acoustics, Chinese Acad. of Sciences	China
MCRC	Motorola China Research Center, Shanghai*	China
THU	Dept. of Electronic Engineering, Tsinghua Univ.	China
USTC	USTC SSIP Laboratory	China
	Beijing Univ. of Posts and Telecommunications*	China
	Nanyang Technological University*	China

* denotes first time participant

NIST National Institute of Standards and Technology

Participants – Australia

NIST ID	Site	Location
QUT	Queensland Univ. of Technology	Australia
	Univ. of New South Wales*	Australia

* denotes first time participant

NIST
National Institute of Standards and Technology

Participants – Europe

NIST ID	Site	Location
AGN	AGNITIO, S. L.*	Spain
ATVS	Universidad Autonoma de Madrid	Spain
BUT	Brno Univ. of Technology	Czech Republic
ENST	Ecole Nationale Superieure des Telecommunications, IRCGN	France
I3A	Aragon Institute for Engineering Research, University of Zaragoza	Spain
IDI	IDIAP Research Institute	Switzerland
IESK	IESK Cognitives Systems, University of Magdeburg	Germany
JoY	Univ. of Joensuu	Finland
LIA	Laboratorie d'Informatique d'Avignon, University of Avignon	France

* denotes first time participant

NIST National Institute of Standards and Technology

Participants – Europe (cont'd)

NIST ID	Site	Location
LIM	LIMSI, CNRS	France
TNO	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek	The Netherlands
TUL	Technical Univ. of Liberec*	Czech Republic
ULJ	Univ. of Ljubljana	Slovenia
VIL	Vilnius Univ.*	Lithuania
	IKERLAN Technological Research Center*	Spain
	Laboratoire de Recherche et Developpement de l'EPITA	France
	Loquendo	Italy
	Politecnico di Torino	Italy
	Swansea Univ.	United Kingdom
	Thales Communications	France
	Univ. of the Basque Country*	Spain

* denotes first time participant

NIST National Institute of Standards and Technology

Participants – N. America

NIST ID	Site	Location
CMU	Carnegie Mellon Univ.*	USA
CRIM	Centre de Recherche Informatique de Montreal	Canada
CRSS	Center for Robust Speech Systems, Univ. of Texas at Dallas	USA
IBM	IBM T.J. Watson Research Center	USA
ICSI	International Computer Science Inst.	USA
MITLL	MIT Lincoln Laboratory	USA
SRI	SRI International	USA
TEC	Tecnologico de Monterrey*	Mexico

* denotes first time participant

NIST
National Institute of Standards and Technology

System Collaborations

- ASA: ATVS + SUNSDV + AGN
- BUFT: Beijing University of Posts and Telecommunications + FTRD
- EHU: University of the Basque Country + IKERLAN Technological Research Center
- I3ACMU: I3A + CMU
- I4U: IIR + iFly + Univ. of New South Wales + Nanyang Technological Univ. + CMU
- ICSRI: ICSI + SRI
- LIMA: LIM + LIA
- LPT: Loquendo + Politecnico di Torino
- LRDECR: Laboratoire de Recherche et Developpement de l'EPITA + CRIM
- PRS_1 short-short: PRS + TNO
- SUNSDV: Stellenbosch University DSP Group + Spescom DataVoice
- SUNSDV-1: SUNSDV + TNO
- THL: Thales Communications + LIA
- TNO-1: ICSI + PRS + SUNSDV + TNO
- UWS: Swansea University + IDIAP + LIA

NIST
National Institute of Standards and Technology

Shared Resources

- ALIZE Users:
  - ENST, IESK, LIA, LIMSI, Thales, USTC, UWS
- FoCal Users:
  - ASA, ATVS, BUT, CASIA, FTRD, I3A, I3ACMU, I4U, ICSI, IOA, JoY, LPT, LRDE, MCRC, PRS, QUT, SUNSDV, THU, TUL, UWS
- Google Group:
  - BUT, CRIM, IBM, LPT, MITLL, PRS, QUT, SRI, SUNSDV, TNO

NIST
National Institute of Standards and Technology

Outline

- Introduction
- Participants
- **Tests and Data Used**
- Core Test
- Summary

NIST
National Institute of Standards and Technology

Data Sources: LDC Corpora

- Mixer 3
  - Recorded telephone conversations of ~10 min.
- Mixer 4
  - Subset of Mixer 3 conversations recorded (simultaneously) over multiple microphone channels placed in room
- Mixer 5
  - Half hour interview sessions including portions of conversational and of read speech

NIST
National Institute of Standards and Technology

Additional Data Provided

- ASR Transcripts: BBN ran fast state of the art ASR system on both the telephone and interview data
  - Produced English output regardless of segment language
  - Run only on telephone channel for mixer 3 and mixer 4 and interviewer and subject lavalier channels for mixer 5, not the other channels.
    - ASR output provided may have been unrealistically good for mx4 and non-lavalier mixer 5
- VAD (Voice Activity Detection) Files
  - Provided energy-based software did speaker diarization using the two lavalier channels.

NIST
National Institute of Standards and Technology

Training Conditions

- **10-sec**: Two-channel excerpt of a Mixer 3 call with ~10s of actual speech by speaker of interest
- **short2**: Two-channel Mixer 3 excerpt of ~5 min., *or* Mixer 5 conversational segment of ~3 min.
- **3conv**: 3 two-channel Mixer 3 excerpts, ~5 min. each
- **8conv**: 8 two-channel Mixer 3 excerpts, ~5 min. each
- **long**: Mixer 5 conversational segment of ~12 min. or more
- **3summed**: Summed channel version of 3conv

NIST
National Institute of Standards and Technology

Test Conditions

- **10-sec**: Two-channel excerpt of a Mixer 3 call with ~10s of actual speech by speaker of interest
- **short3**: Two-channel Mixer 3 excerpt of ~5 min., *or* Mixer 4 excerpt of ~5 min., *or* Mixer 5 conversation segment of ~3 min.
- **long**: Mixer 5 conversational segment of ~8 min.
- **summed**: Summed channel Mixer 3 excerpt of ~5 min.

NIST
National Institute of Standards and Technology

13 Evaluation Tests

Test→ Train↓	10-sec	short3	long	summed
10-sec	optional			
short2	optional	**required (core)**		optional
3conv		optional		optional
8conv	optional	optional		optional
long		optional	optional	
3summed		optional		optional

NIST
National Institute of Standards and Technology

Outline

- Introduction
- Participants
- Tests and Data Used
- **Core Test**
- Summary

NIST
National Institute of Standards and Technology

Core Test (Required)

- Training on Mixer3 (5 min.) and Mixer 5 (3 min.)
- Test on Mixer 3 (5 min.), Mixer4 (5 min.) and Mixer5 (3 min.)
- Numbers of trials (target, nontarget) in five subtests:

Test: / Train	Mixer3 2573 segments	Mixer4 1460 segments	Mixer5 2344 segments
Mixer3 (1788 models)	(3832*, 33218)	(1472, 6982)	(2500, 4850)
Mixer5 (1475 models)	(1105, 10636)	(Not tested)	(11540, 22641)

* 70% of the target trials use a different phone number than that used in training

NIST National Institute of Standards and Technology

Core Test – Common Conditions

- Plan specified 8 common conditions of interest
  - 1) Interview, training and test
  - 2) Interview, same microphone
  - 3) Interview, different microphone
  - 4) Interview training, telephone test
  - 5) Telephone training, mic. recorded telephone test
  - 6) Telephone, training and test
  - 7) Telephone, English only
  - 8) Telephone, native U.S. English only
- See comparative plots in poster session

NIST
National Institute of Standards and Technology

Mixer 3 Core Test Language Mix

- 1336 speakers, 53% multi-lingual
- 1788 models, 29% non-English
- 2573 test segments, 29% non-English
- Examined effect of limiting trials to those that involve only English, or only native U.S. English speech

	All	English	USE
Target	3832	1827	992
Non-target	33218	16533	7881

NIST
National Institute of Standards and Technology

Summary

- 2008 saw record evaluation participation
- 2008 evaluation was larger in data size and more complicated than prior evaluations
  - Core test, in particular, was more extensive, requiring processing by all participants of interview data and microphone recorded telephone data, as well as conversational telephone data as previously
- Significant performance improvement seen on Mixer 3 and Mixer 4 data compared with prior evaluations
- Performance on Mixer 5 interview data was encouraging

NIST
National Institute of Standards and Technology