2011 NIST Language Recognition Evaluation (LRE11)

Alvin Martin and Craig Greenberg

NIST/ITL/Information Access Division
Multimodal Information Group

http://www.nist.gov/itl/iad/mig

LRE 11 Workshop

6-7 December 2011

Atlanta, Georgia, USA

Outline

- Task
- Languages
- Data
- Trials
- Metric
- Participants
- Performance Results
- Summary

Task

- Language detection in the context of a given language pair:

> ***Given a segment of speech and a pair of languages, decide which of these two languages is spoken in the speech segment***

- This was an optional task in LRE09
- Included in earlier LRE's in the context of dialect recognition tests

The 24 LRE11 Target Languages

Polish	Dari	***Arabic Iraqi***
Czech	Farsi/Persian	***Arabic Levantine***
Russian	Hindi	***Arabic Maghrebi***
Slovak	Bengali	***Arabic MSA***
Ukrainian	Urdu	English American
Thai	***Panjabi***	English Indian
Lao	Tamil	Spanish
Turkish	Pashto	Mandarin

- Color coded for confusable clusters
- New Languages to LRE in ***bold italics***

LRE11 Data – Source

- Two source types
  - *Telephone* conversations of native speakers
  - Narrowband *broadcasts*,
    often from several different sources

- Both source types were generally used for each language
- Data collected and audited by the LDC

LRE11 Data – Language Statistics

Language	Number of Telephone Segments	Number of Broadcast Segments	Number of Broadcast Sources
arabic_iraqi	1224	0	0
arabic_levantine	1224	0	0
arabic_maghrebi	1215	0	0
arabic_msa	0	1218	49
bengali	660	681	19
czech	537	837	4
dari	858	1647	27
english_american	363	993	8
english_indian	150	1098	45
farsi	591	624	21
hindi	210	1047	32
lao	379	379	5
mandarin	777	519	9
panjabi	1191	33	5
pashto	465	771	12
polish	726	717	1
russian	417	906	4
slovak	516	726	3
spanish	693	564	16
tamil	600	642	9
thai	195	1014	5
turkish	501	915	9
ukrainian	357	201	4
urdu	672	771	7

LRE11 Data – Encoding and Duration

- Down/Re-sampled to a common encoding:
  8kHz-16bit PCM

- Three evaluation conditions based on duration:
  - ***3*** sec. of speech, nominal (*2-4* sec. actual)
  - ***10*** sec. of speech, nominal (*7-13* sec. actual)
  - ***30*** sec. of speech, nominal (*25-35* sec. actual)

- Segments of *13-25* sec. included but not scored

Trials

Given N languages, each speech segment is used in
N * (N-1) / 2 trials (one trial per language pair)
N = 24 => 276 trials per segment
Only N-1 (8.3 %) of these trials are scored

Input for each trial

- Audio segment
- Language pair *L1* and *L2*

Output for each trial

- Language ***decision*** (either *L1* or *L2*)
- Confidence ***score***

Language Pair Cost Function

$$C(L_1, L_2) = C_{L_1} * P_{L_1} * P_{Miss}(L_1) + C_{L_2} * (1 - P_{L_1}) * P_{Miss}(L_2)$$

$$C(L_1), = C(L_2) = 1; P_{L_1} = 0.5$$

- Compute separate cost for each duration condition
- Compute both cost of actual language decisions, as well as the minimum cost obtained by varying decision threshold
- Difference of the actual and minimum decision costs may be viewed as calibration error

Overall Performance Measure

- Find N (= 24) language pairs with the greatest minimum[1] costs for the 30-second segments

Minimum score was chosen to avoid excessive calibration penalty.
To avoid gaming, minimum score = min(minimum score, actual score)

- For each duration:

  Cost = the mean of the actual decision costs for these N language pairs

Participating Sites/Teams (1)

System	Site(s)/Team	Location
ATVS	Universidad Autonoma de Madrid,	Madrid, Spain
BLZ	University of the Basque Country Spoken Language Systems Lab, INESC-ID University of Zaragoza	Brno, Czech Republic Lisbon, Portugal Zaragoza, Spain
BRNO276	Brno University of Technology Agnitio Politecnico di Torino	Brno, Czech Republic South Africa Torino, Italy
CHULA	Chulalongkorn University	Bangkok, Thailand
CRSS	University of Texas at Dallas	Richardson, Texas, USA
EHU	University of the Basque Country	Bizkaia, Spain
I3A	University of Zaragoza	Zaragoza, Spain
IACAS	Chinese Academy of Sciences	Beijing, China
IFLYTEK	iFlyTek Speech Lab, EEIS University of Science and Technology of China	HeFei, AnHui, China
IIR	Institute for Infocomm Research	Fusionopolis, Singapore

Participating Sites/Teams (2)

System	Site(s)/Team	Location
IITKGP	Indian Institute of Technology, Kharagpur	Kharagpur, India
L2F	Spoken Language Systems Lab, INESC-ID	Lisbon, Portugal
LABRI	LABRI-Universitée Bordeaux	Talence, France
LIMSI	CNRS-LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur)	Orsay, France
MITLL	MIT Lincoln Laboratory MIT Computer Science and Artificial Intelligence Laboratory	Lexington, MA, USA Cambridge, MA, USA
NTUT	National Taipei University of Technology	Taipei, Taiwan
THUEE	Tsinghua University Department of Electronic Engineering	Beijing, China
UEKAE	TŰBİTAK BİLGEM, UEKAE	Gebze, Turkey
ULTRA-SWAN-LIA	Ultra-Electronics Audiosoft Swansea University Laboratoire Informatique d'Avignon	Cirencester, United Kingdom Swansea, United Kingdom Avignon, France