

COCOSDA



International Coordinating Committee
for Speech Databases and Assessment

Dafydd Gibbon

COCOSDA Convenor

COCOSDA-2010, Valletta, Malta

COCOSDA

- **Objectives:**
 - **The International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques,**
 - Cocosda, has been established to encourage and promote international interaction and cooperation in the foundation areas of Spoken Language Processing, especially for Speech Input/Output.
 - **The importance of collaboration which transcends national boundaries is increasingly recognized.**
 - This is both because of the practical and scientific value attached to systematic work which encompasses a range of languages and analytic approaches and also because of the practical need to establish common methods of performance description and quantitative comparison.

COCOSDA background

- **Prehistory**
 - **March 1982, Taithersburg, USA**
 - **September 1989, Noordwijkerhout, NL**
 - **November 1990, Kobe, Japan**
 - **September 1991, Chiavari, Italy**

COCOSDA background

- **Workshops**
 - **October 1992, Banff, Canada, ICSLP 92**
 - **September 1993, Berlin, Germany, Eurospeech 93**
 - **September 1994, Yokohama, Japan, ICSLP 94**
 - **September 1995, Madrid, Spain, Eurospeech 97**
 - **October 1996, Philadelphia, USA, ICSLP 96**
 - **November 1998, Sydney, Australia, ICSLP 98**
 - **September 1999, Budapest, Hungary, Eurospeech 99**
- **Reorganisation meeting**
 - **May-June 2000, Athens LREC 2000**

COCOSDA background

- **Recent meetings**
 - **October 2000 Beijing China**
 - **September 2001 Aalborg Denmark**
 - **September 2002 Denver Colorado**
 - **September 2003 Eurospeech Geneva**
 - **May 2004 LREC Lisbon (COCOSDA/Write)**
 - **October 2004 Jeju Island Korea**
 - **May 2006 LREC Genoa**
 - **September 2007 Interspeech Antwerp (COCOSDA Prize Award)**
 - **May 2008 LREC Marrakesh**
 - **September 2009 Interspeech Brighton, informal discussions**

Central Coordinating Committee

- **Working Group**
 - organisation
- **Topic Domain Rapporteurs**
 - current topics - flexible
- **Regional Domain Rapporteurs**
 - Unevenly represented areas of the world
- **Advisory Committee**
 - Consultative input to Working Group & Rapporteurs

Working Group

- **Officers:**
 - **Convenor:** Dafydd Gibbon, Germany
 - **Deputy Convenor:** Khalid Choukri , France
 - **Secretary:** Nick Campbell, Ireland
- **Current state:**
 - **transitional: need for new representatives**
 - **change in organisational structure needed**
 - cf. Oriental COCOSDA

Topic Domain Rapporteurs

- **Evaluation of Speech Understanding / Dialogue Systems:**
 - Wolfgang Minker ,Germany
- **Multi-modal Corpora:**
 - Satoshi Nakamura , Japan
- **Corpus Annotation Tools:**
 - Steven Bird , Australia
- **Signal Processing Tools**
 - Inma Hermaez (Spain)
- **Local Languages:**
 - Dafydd Gibbon , Germany

Regional Domain Rapporteurs

- **Europe:**
 - **Khalid Choukri , France**
- **Asia:**
 - **Shuichi Itahashi , Japan**
- **Oceania:**
 - **Steve Cassidy , Australia**
- **Africa:**
 - **Justus Roux , South Africa**
- **Americas:**
 - **Chris Cieri, USA**

Data Centre Representatives

- **LDC:**
 - **Chris Cieri**
- **ELDA:**
 - **Khalid Choukri**

Advisory Committee

- **Adrian Fourcin, UK**
- **Hiroya Fujisaki, Japan**
- **Akira Kurematsu, Japan**
- **Mark Liberman, USA**
- **Joseph Mariani, France**
- **Louis Pols, Rob van Son, Netherlands**

Current events

- **Last COCOSDA prize award:**
 - Interspeech 2007, Antwerp
- **Last workshop:**
 - LREC 2008, Marrakesh, Morocco
- **Intervening years:**
 - informal meetings
- **Currently:**
 - transitional: unbalanced - currently entirely Europe
 - change in organisational structure needed
- **Next Oriental COCOSDA:**
 - 24-25 November 2010, Kathman
- **Next COCOSDA:**
 - Interspeech 2011, Florence / Firenze

Perspectives

- **COCOSDA goals have taken up by other initiatives, projects, conference series, in other domains**
- **De facto re-definition and merging of fields:**
 - **multimodality, cooperation with NLP, signing, ...**
 - **multilinguality**
 - **technologies for under-resourced languages**
 - **self-bootstrapping learning systems**
- **Standardisation of resources**
 - **compatibility/extendability of metadata conventions**
 - **corpus formats**
 - **lexicon formats**
 - **affordable, interoperable tool types & toolkits (cf. 'BLARK')**

And now for the details ...

COCOSDA reports: Regions

- **Asia:**
 - **Shyam Agrawal**
- **Americas:**
 - **Chris Cieri**
- **Europe:**
 - **Khalid Choukri**
- **Africa:**
 - **Justus Roux**
- **Australia:**
 - **n.a.**

COCOSDA reports: Topics

- **Multimodal resources & assessment**
 - **Nick Campbell**
- **Multilingual multimodal resources – creating a levelling field in India**
 - **Girish Nath Jha**
- **Resource infrastructures for endangered and other minority languages**
 - **Peter Wittenburg**
- **Speech technology developments for non-funded languages**
 - **Dafydd Gibbon**