Flinders
UNIVERSITY

inspiring achievement

The BigASC:

Designing, Collecting, Disseminating and Collaborating on a Big Australian Speech Corpus

Trent Lewis

Centre for Knowledge and Interaction Technology

Medical Devices and Research Institute

Artificial Intelligence and Language Technology Lab

Brain Signals Lab

Flinders UNIVERSITY

inspiring achievement

Speech Corpora

- AVSP requires large datasets
- Various corpora throughout world
  - including audio-visual

- ANDOSL (1990)
  - 200 Speakers, Audio-Only
- AVOZES, VidTIMIT, …
  - limited range, specific

Speech Corpora – The BigASC

https://austalk.edu.au/

Something for Everyone

- Phonetics
- Linguistics
- Cognitive Science
- Psycholinguistics
- Computer Science
- Speech Engineering
- Spoken Language Processing
- ASR & TTS
- Speech Pathology
- Forensic Speech Science
- …

Something from Everywhere

ARC LIEF, 2010: The Big Australian Speech Corpus:
An audio-visual speech corpus of Australian English
$650,000

Aims, Scope and Features

1. Design a functional heuristic speech database
   (a) Wide acceptability
   (b) Variability
   (c) Standardisation
2. Establish state-of the-art infrastructure to collect AV Australian English speech data

(a) Recording Equipment – black boxes

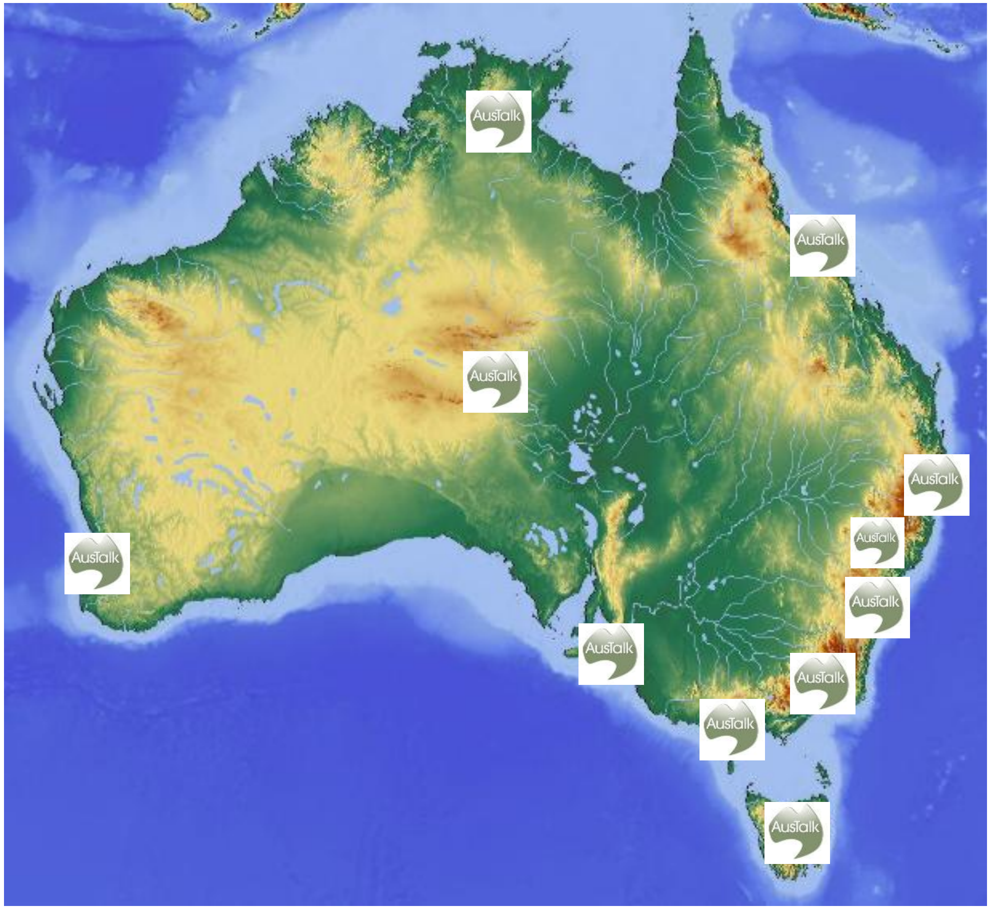
(b) Data Collection Protocol

(c) Public domain access to centralised storage facility

(d) Standardised Annotation

3. Collect large amount of speech data
   (a) Launch and advertising
   (b) Co-ordination and RA Training
4. Provide an extensible system for further data collection
5. Facilitate Australian/international speech science research

AusTalk


AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk
AusTalk

Aims, Scope and Features

1. Design a functional heuristic speech database
   - (a) Wide acceptability
   - (b) Variability
   - (c) Standardisation
2. Establish state-of the-art infrastructure to collect AV Australian English speech data

(a) Recording Equipment – black boxes

(b) Data Collection Protocol

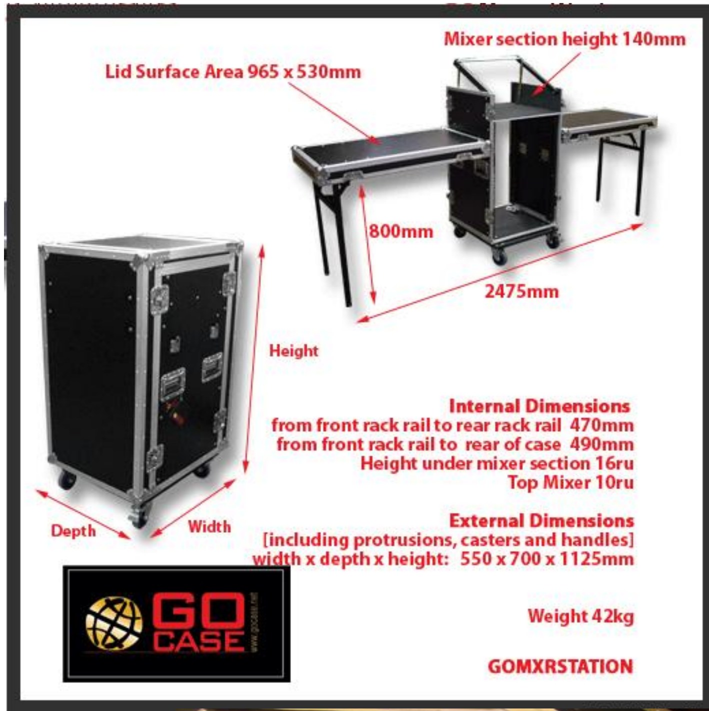
(c) Public domain access to centralised storage facility

(d) Standardised Annotation

3. Collect large amount of speech data
   - (a) Launch and advertising
   - (b) Co-ordination and RA Training
4. Provide an extensible system for further data collection
5. Facilitate Australian/international speech science research

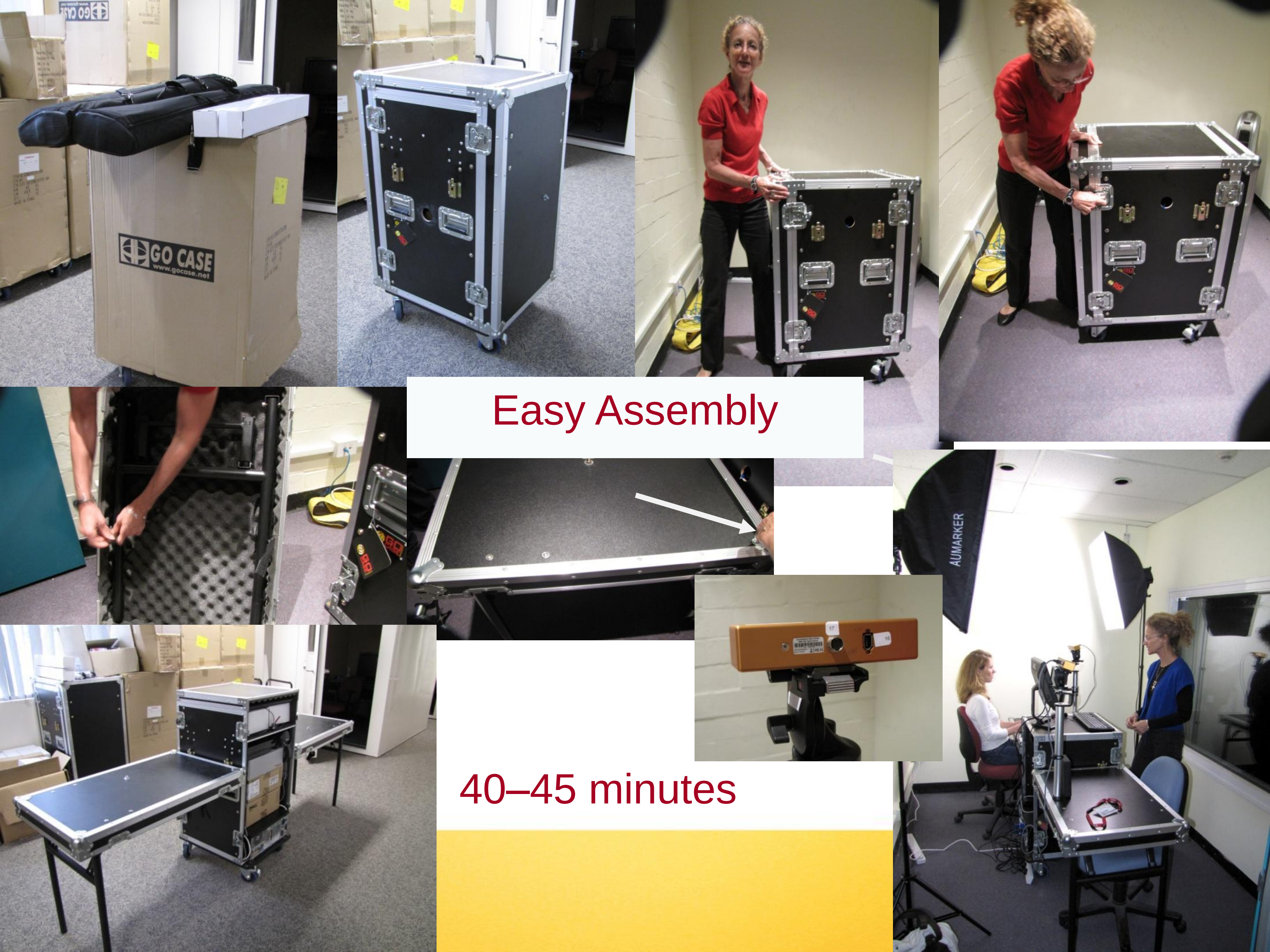
(a) Recording Equipment – Black Boxes

- Standard Speech Science Infrastructure Black Box
  - Standardised equipment, configuration, setup at all locations
  - Portability: Packed in reinforced box, folds out to a table + integrated shelving
  - Low cost: $AUD12K per unit
- Basic components
  - Computer, digital audio acquisition device ,desktop microphone, head-worn microphones, stereo cameras


Mixer section height 140mm
Lid Surface Area 965 x 530mm
800mm
2475mm
Height
Depth
Width
Internal Dimensions
from front rack rail to rear rack rail 470mm
from front rack rail to rear of case 490mm
Height under mixer section 16ru
Top Mixer 10ru
External Dimensions
[including protrusions, casters and handles]
width x depth x height: 550 x 700 x 1125mm
GO CASE
Weight 42kg
GOMXRSTATION

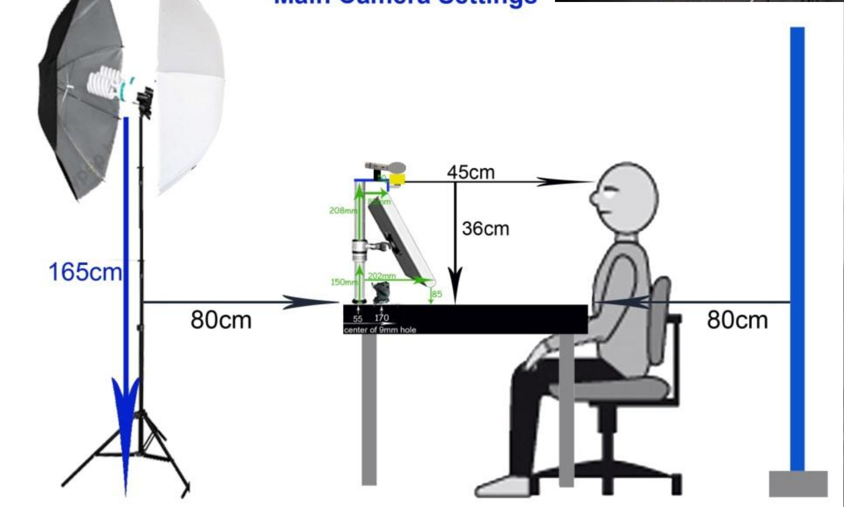
Recording Equipment – Black Boxes

- Black Box
  - Mixer Rack Workstation: the ‘Black Box’ for storing and transporting items; unpacks into 2 tables & computer rack
- Computing
  - Capture Computer: PC for protocol display and recording.
  - External hard drive: Samsung STORY Station 2TB.
- Audio recording:
  - M-Audio FastTrack Ultra8R.
- Microphones and Headphones
  - Head worn mic (x2): AudioTechnica AT892c.
  - AT8539 Phantom Power/XLR adapter to connect mic.
  - Far-Field mic: Shure MX391/O. On table, ~ 60cm from speaker.
  - Stereo mics (x2): Behringer C-2. On table, ~60 cm from speaker, to record hands-free voice interaction
  - Operator Head Phones: KOSS UR-20, for the RA.
- Cameras
  - Stereo Cameras BumbleBee2 (x2). Mounted ~50cm from speaker. Dual bus firewire card.
  - Tripod mount for camera (x3): Manfrotto 700RC2 tripod
- AV
  - Custom-made GPIO 2 audio Sync Cable. A/v synch: camera sends strobe signal →M-Audio DAQ to record waveform
- Monitors:
  - 17inch Monitors 4:3 (x2): Dell E170S 17 inch Flat Panel Monitor. To display prompts to speaker and for RA.
  - Monitor arm / stand: Atdec Visidec Focus MICRO LCD Single Arm, VF-M. To hold monitor and camera.
- Lighting
- 2 x (Soft Umbrella, Umbrella Reflector, Tripod, Dual lamp adapter, 2 x 65W lamps)
- Pull-up backdrop (x2) to provide uniform background.
- Chairs (x2) to ensure standardisation of video capture.



Easy Assembly

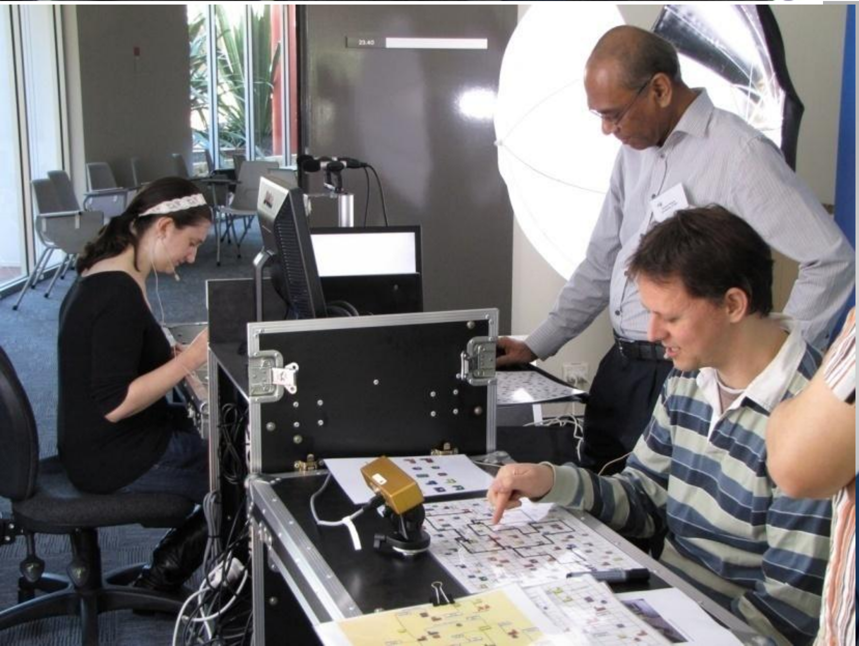
40–45 minutes

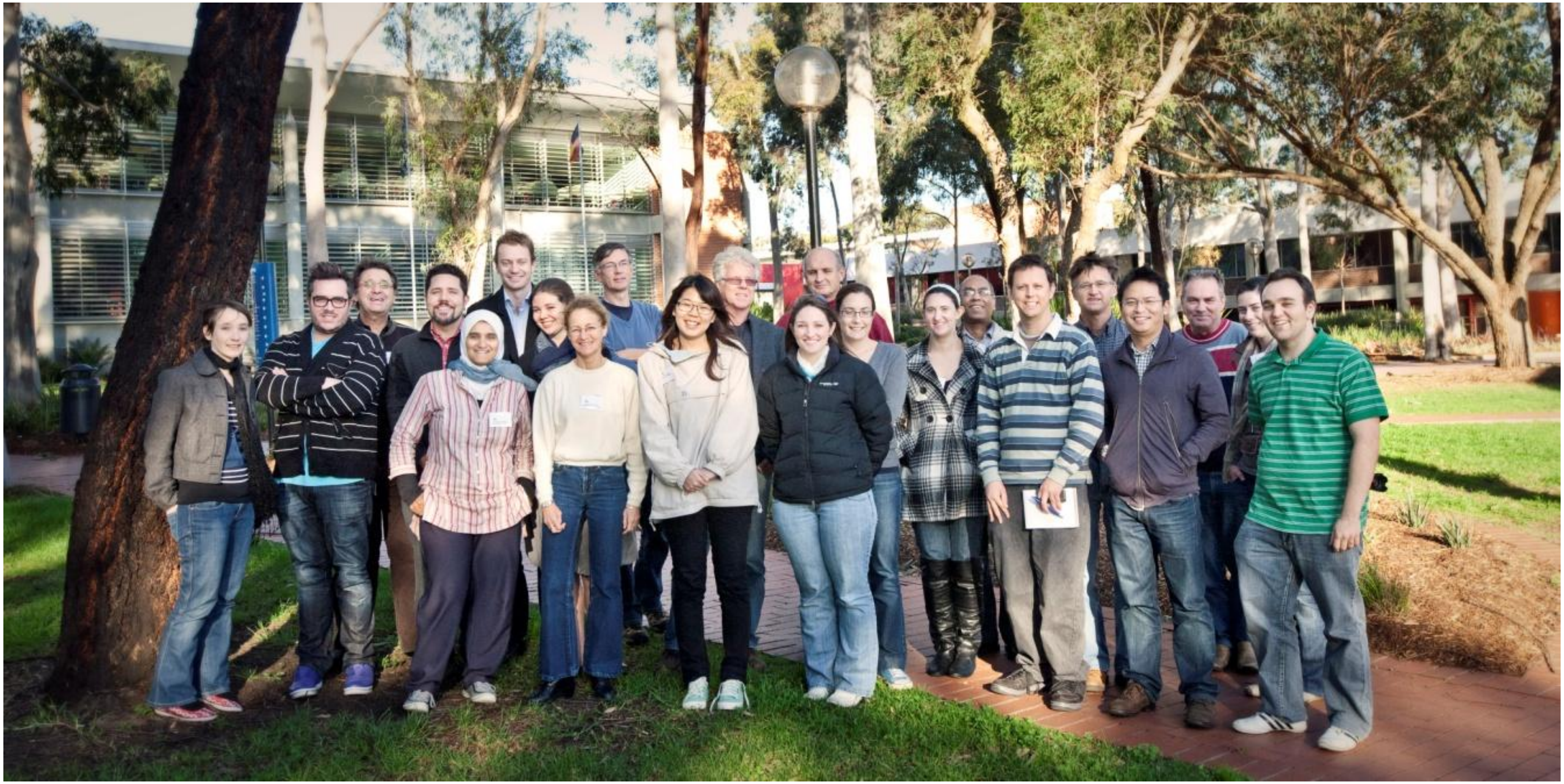

Main Camera Settings

Standardized Protocol

Session 1		Session 2		Session 3	
Task	**Time**	**Task**	**Time**	**Task**	**Time**
Calibration (+ 3D face)	10	Calibration	3	Calibration	3
Opening Yes/No	3	Opening Yes/No	2	Opening Yes/No	2
Words	10	Words	10	Words	10
Read Narrative	5	Interview	15	Map Task (First run)	20
Re-told Narrative	10			Switch Sp.A and Sp.B	5
Read Digits	5	Read Digits	5	Map Task (Second run)	20
		Read Sentences	8	Conversation	5
				Words	10
Closing Yes/No	2	Closing Yes/No	2	Closing Yes/No	2
	44		**45**		**77**

(b) 2-day Central Training Session





https://austalk.edu.au/

Aims, Scope and Features

1. Design a functional heuristic speech database
   (a) Wide acceptability
   (b) Variability
   (c) Standardisation
2. Establish state-of the-art infrastructure to collect AV Australian English speech data

(a) Recording Equipment – black boxes

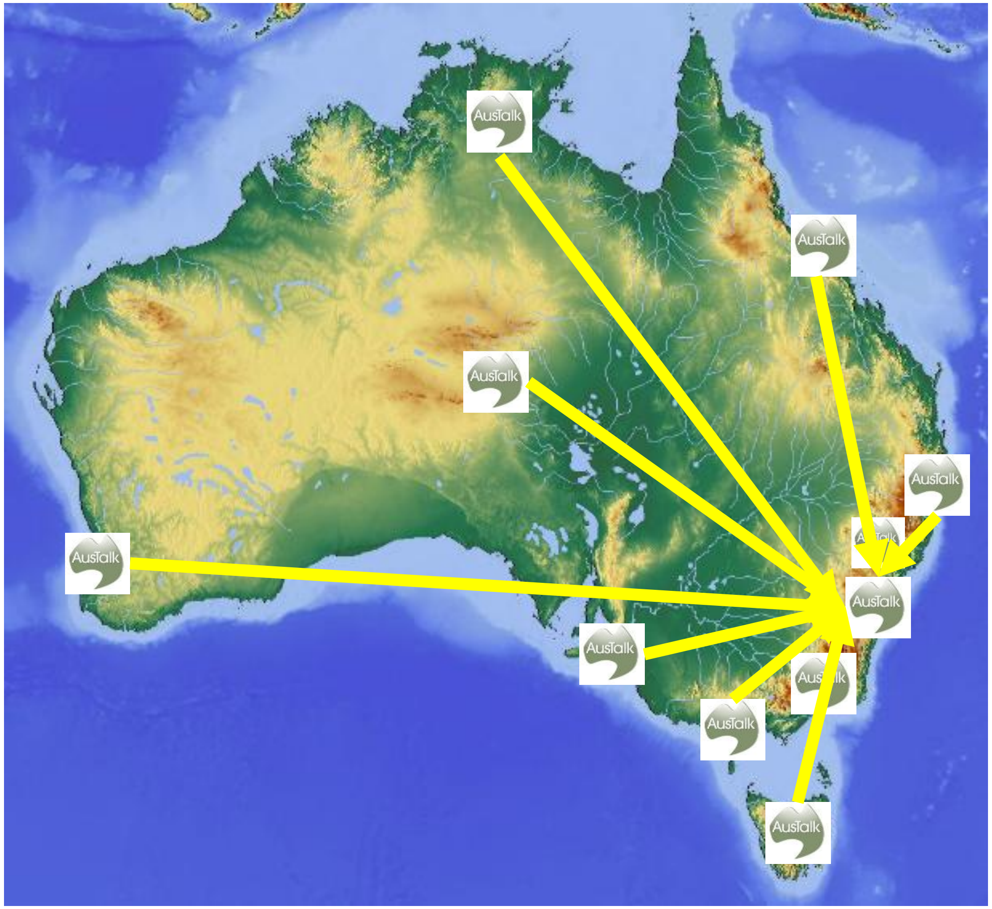
(b) Data Collection Protocol

(c) Public domain access to centralised storage facility

(d) Standardised Annotation

3. Collect large amount of speech data
   (a) Launch and advertising
   (b) Co-ordination and RA Training
4. Provide an extensible system for further data collection
5. Facilitate Australian/international speech science research

Centralised storage/annotation



1000 Speakers target
798 Speakers so far

Data upload

- 1 minute of recording = 1 Gb of data
- 3hrs*3 per speakers = 180GB / speakers
- 180,000GB/180TB in total
- **$1M for Tier 1 Storage**
- Conversion and Compression of vdo data on site (54:1)
- Typical 45 minute session, compressed = ~2Gb

Centralised storage/annotation

- Sharing Data?

Sharing Data?

http://hcsvlab.org.au/

NeCTAR Virtual Lab, 2012
Above and Beyond Speech, Language and Music: A Virtual Lab for Human Communication Science
Burnham (Lead), Powers (Flinders), Butcher (Flinders), Lewis (Flinders), et al.
$1.4m

NeCTAR

- The National eResearch Collaboration Tools and Resources project (NeCTAR)
- $47 million Australian Government, Super Science project.
- The University of Melbourne (UoM) is the lead agent
- $101 million to Australia's research infrastructure.

NeCTAR

- NeCTAR is building eResearch infrastructure in four areas:
  - Virtual Laboratories;
  - eResearch Tools;
  - Research Cloud;
  - A secure and robust hosting service (National Servers Program).

HCSvLab

- Human Communication Science
- Virtual Laboratory
- A platform for eResearch in HCS

http://hcsvlab.org.au/

HCSvLab

- Connects corpus data and tools
- Corpus data is:
  - normalised to standard formats
  - catalogued to enable search and browse
  - protected to respect licences on data
  - available for use by tools
- Tools are:
  - given (fast) access to data
  - integrated into the platform
  - made easy to use for non-technical users
  - connected together to enable pipelines

HCSvLab

- Corpora:
  - AusNC: ICE-AUS, ACE, COOEE, Mitchell & Delbridge, Braided Channels
  - PARADISEC
- Tools:
  - NLTK
  - Johnson Charniak Parser
  - Emu
- Environment
  - Web based browse/search of corpora (Blacklight)
  - Workflow/tool execution environment (Galaxy)

HCSvLab

- Connects corpus data and tools
- Sharing Data
- Sharing Tools
- Sharing Workflow
  - facilitate access of the Australian and international HCS communities
  - new tool–corpus combinations and new emergent research
  - allow analysis and annotation results to be stored and shared,
  - promoting collaboration between institutions and disciplines;
  - improve scientific replicability

The BigASC:

Designing, Collecting, Disseminating and Collaborating on a Big Australian Speech Corpus

Flinders
UNIVERSITY
inspiring achievement

Trent Lewis

Centre for Knowledge and Interaction Technology

Medical Devices and Research Institute

Artificial Intelligence and Language Technology Lab

Brain Signals Lab

Flinders
UNIVERSITY

inspiring achievement