Tools
for work that uses
acoustic-phonetic analysis

Sarah Hawkins and Richard Ogden
University of Cambridge, University of York

with thanks to
Suzanne Boyce, Mirjam Ernestus and Brechtje Post

sh110@cam.ac.uk rao1@york.ac.uk

Overview

- S2S overview
- Corpus use:
  - focus on phonetic-linguistic requirements
- Tools for speech scientists:
  - workshop proposed to discuss what speech scientists want/need
  - emphasis on interdisciplinary collaborations with ‘non-computational scientists’ e.g.
    - ‘other professionals’: neuroscientists, audiologists
    - mid-career and older teaching academics with no time to learn yet another system
    - undergraduates etc (even if they can program)

Sound to Sense: EC Marie Curie Research Training Network (RTN)

A Marie Curie RTN:

- trains young researchers
- interdisciplinary methods
- to further scientific progress

S2S: 2007-11 €2.8 million

- 14 institutions, 11 countries
- coordinated from Cambridge

Sound to Sense: EC Marie Curie Research Training Network (RTN)

- engineers, computer scientists, psychologists, phoneticians
  - 7 postdocs
  - 11 PhD students + 3 transcribers
  - 46 “senior scientists”
- “paradigm shift”?
  reassess what information is available in the signal, especially in conversational speech
- explore new ideas about how humans understand speech
- and apply to machine speech recognition (ASR) and some synthesis (TTS)

Sound to Sense: EC Marie Curie Research Training Network (RTN)

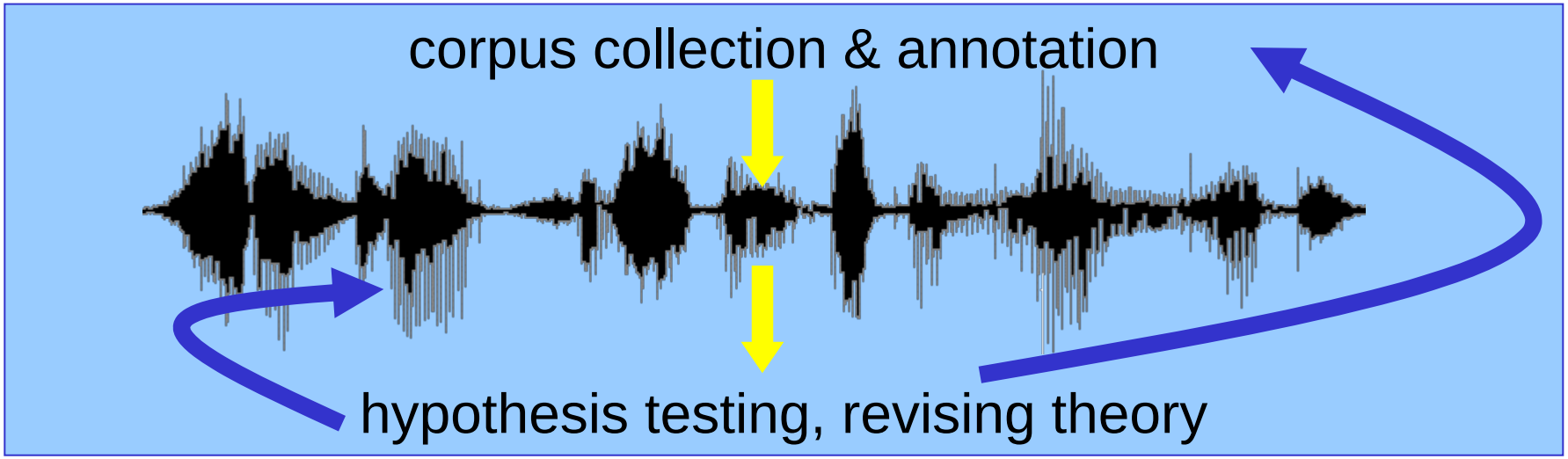
Some relevant focuses:

- multilingual
- conversational speech
- unit identification:
  - short and longer units
- feature extraction
- word segmentation
- discourse and prosody
- exemplars and abstraction
- Bayesian modelling
- adverse listening conditions
  - type of noise interference
  - native vs foreign language listeners
- signal streaming (audio, audio-visual)
- automatic labelling

Corpus use: What everyone wants

- Coordinated…integrated…instant access
- S2S has not solved this problem

- Crucial:
  - the time to research available methods
  - wise planning to allow continued development



Some “special” types of of speech

- natural conversational (hardly special, but certainly challenging within current theory and common practice)
- under adverse conditions
  - foreign language learners (speaking and listening)
  - multiple speakers
  - in noise
- medical conditions (for diagnosis, for assessing change)
- natural ‘atypical’ states: fatigue, fear etc

Conversational speech

- How to label unclear sequences:
  - for what the underlying ‘words’ probably are?
  - for what’s in the signal? (then, what type of label?)
  - for how to retrieve?
  - for its context? (what type of context?)

- Tension:
  - clarity, tradition, general exchange ½ phonemes, words, parts of speech etc
  - information in new theoretical frameworks ½ more

Automatic Annotation

- Types of annotation e.g.
  - orthographic transcription
  - phonetic transcription
  - lexical information (includes morphemic structure)
  - syntactic function (direct object, subject, etc)
  - conversational function; pragmatic function
  - complete conversational turn
- what may be unusual in S2S:
  - type of phonetic transcription (value of detail)
  - annotation for conversational speech & ‘adverse conditions’
  - emphasis on relationships between units
  - long units, long contexts
- Any system needs agreement on guiding theory: novel needs may need new theoretical approaches

What theory?

- usual categories of linguistic theory
- longer contexts:

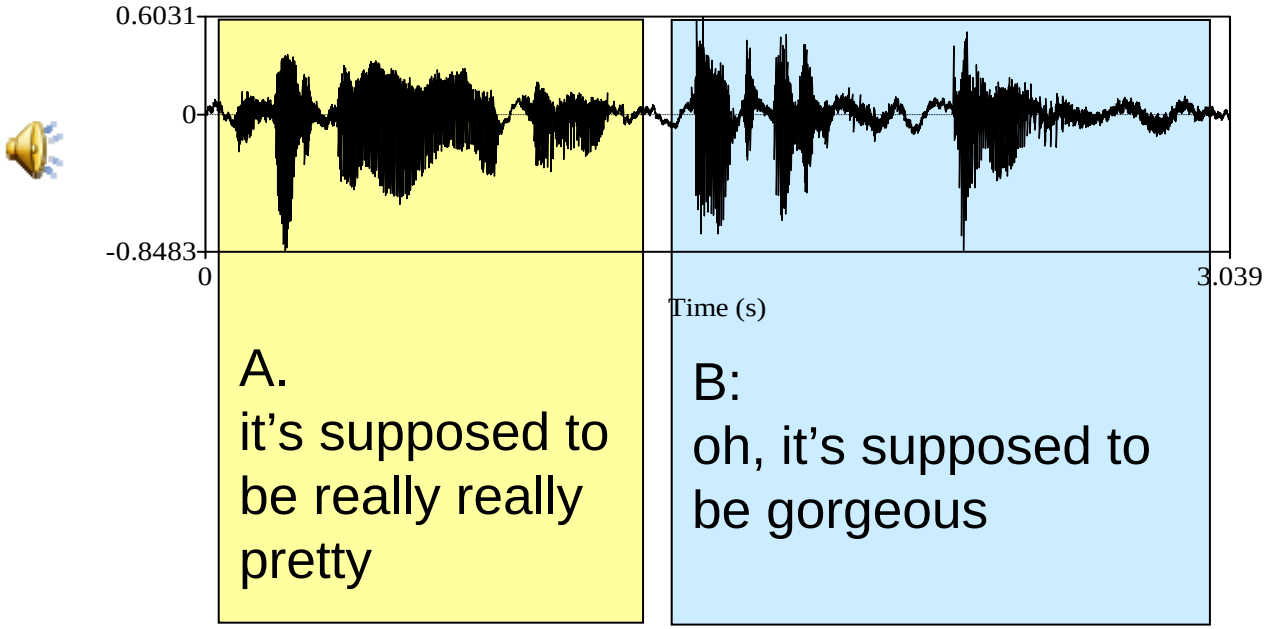
  S2S uses **prosody** & **CA** (**C**onversation **A**nalysis)
- thus:-

and

» structural categories

» functional categories

Agreement sequence



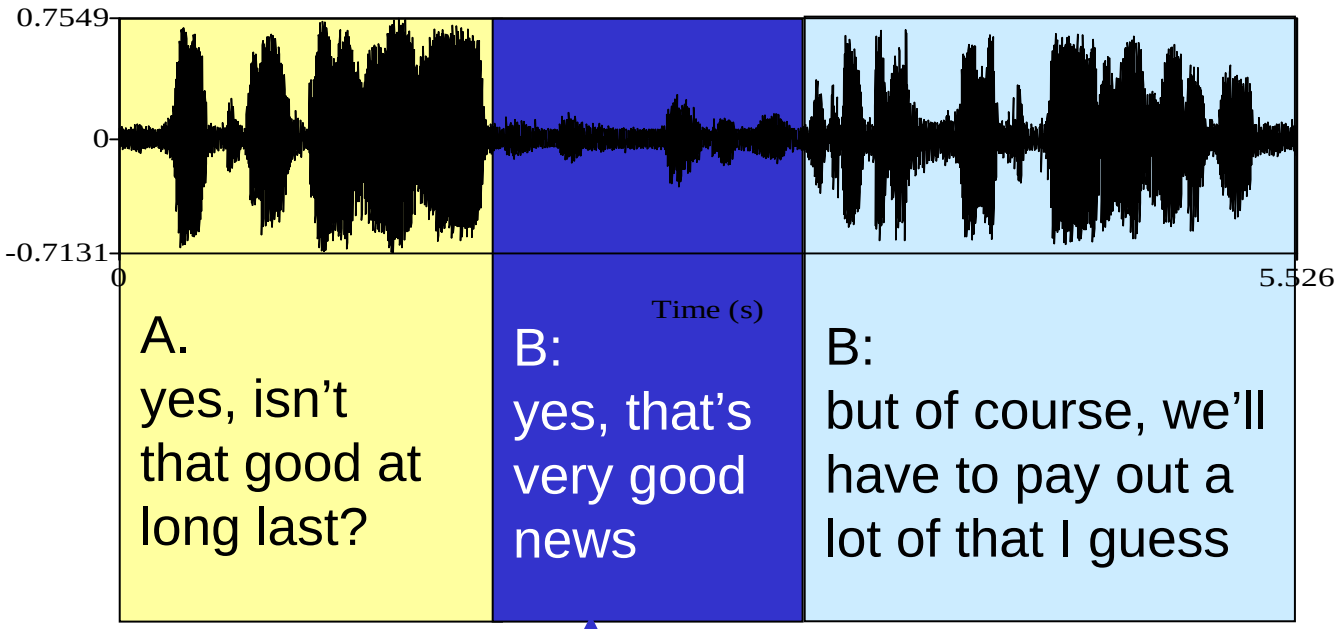
A's turn:

- syntactically complete
- prosodically complete
- recognisable action (in CA framework)
- allows B to respond

B's turn relative to A:

- syntactically parallel
- lexically parallel but 'upgraded'
- phonetically **upgraded:** slower, louder, expanded semitone range (7.3 vs 5.7 st)

Agreement + disagreement sequence



A's turn:

- syntactically complete
- prosodically complete
- recognisable action (in CA framework)
- allows B to respond

B's turn relative to A:

- syntactically matched (parallel to Q)
- lexically parallel
- phonetically **downgraded:** softer, faster, reduced semitone range

Efficient corpus search (1)

- currently, programs lack (for linguists)
  - word finding unless word boundaries are marked (e.g. by spaces)
  - simultaneous search of different annotation tiers (essential for anything phonetically sophisticated)
  - user friendliness

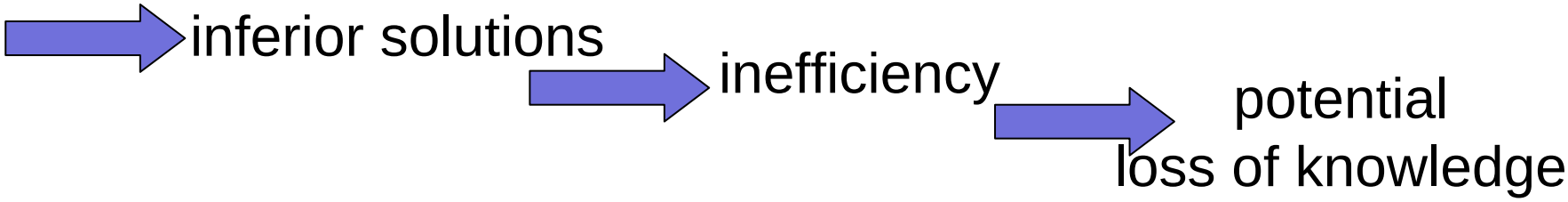
- essential to mark boundaries….of all types
  - capability to label new boundaries (add tiers?) to accommodate new theoretical frameworks

Efficient corpus search (2)

- results saved to separate sound files (+ labels)
- with options for amount of material on either side, specified in a choice of units (range 0-max) e.g.
  - ms
  - phrases
  - words
  - where people start and stop speaking
  - kind of function (action)
  - ??
    (basically anything that has been annotated) and…
  - a ‘free’ entry that could combine such terms, with standard logical (if, >, etc) & arithmetic operators

Tools for speech scientists

- Powerful PCs + much good freeware, but:
  - too many sources of software, each with strengths and limitations
  - Steep learning curve
    before you can do something interesting (as always)
  - Interdisciplinary exchange/remote collaboration
    often in small centres (replace large centres)



- Need:
  - intuitively **usable gui** for short-learning curve (e.g. wavesurfer)
  - many **in-built options** (e.g. praat)
  - common **general-purpose language** (easy-to-learn e.g. python)
  - powerful, flexible **modular** computing capability (cf. Matlab)

Tools for speech scientists

- intuitive gui
- in-built functions
- general language
- modular

- Ideal aims:
  - new users can learn system easily
  - experienced users of other systems can use their own plug-ins via specific interfaces/filters
- Requires:
  - significant research into needs, and then into preferable and possible solutions
  - envisaged lifespan: 15 years?
    - aim for flexibility to allow future functionality
- Proposal: international workshop

Tools for speech scientists

- intuitive gui
- in-built functions
- general language
- modular

- fast to learn
- plug-ins easy
- small- & large-scale analysis
- manual & automatic

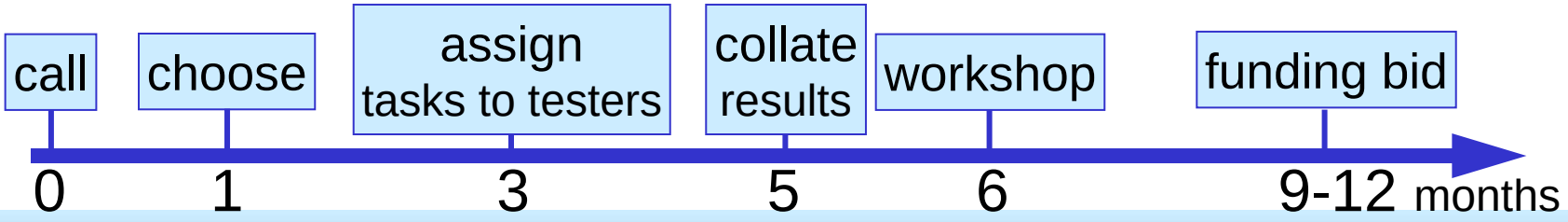
- Ideal aims:
  - new users can learn system relatively easily
  - experienced users of other systems can use their own plug-ins via specific interfaces/filters
- Requires:
  - significant research into needs, and then into preferable and possible solutions
  - envisaged lifespan: 15 years?
    - aim for future functionality
- Proposal: international workshop

Plans for international tools workshop

- intuitive gui
- in-built functions
- general language
- modular

- fast to learn
- plug-ins easy
- small- & large-scale analysis
- manual & automatic

- Scientific committee
  - N. Am: Boyce, Black, Espy-Wilson, Liberman + LDC
  - Europe: S2S (Beskow (+ KTH), Hawkins, Cutugno, Van Compernolle); others?
  - Asia: Yegnarayana, (Kawahara, (Australasia?))
  - S. America: Barbosa
- Invite: representatives of main freeware: emu, praat, sfs, wavesurfer….
- + Open call: existing systems; desired tasks; testers
- 2-stage data collection → funding bid



Summary

- corpus annotation and search:
  - multiple representations
  - hierarchical and other relationships (contexts)
  - long domains
  - communicative function, not just structural categories
- tools:
  - common platform and language
  - accessible to non-computational people
  - appealing to computational people
  - to enhance cross-disciplinary fertilization