The University Of Sheffield.

CAST Clinical Applications of Speech Technology

Barnsley Hospital NHS NHS Foundation Trust

Voice-input voice-output communication aid (VIVOCA)

Simon Judge, Mark Hawley,
Fabian Cardinaux, Peter O'Neil,
Rebecca Palmer

Barnsley AT Team

- Assistive Technology (AT) team, covering 3 areas of S Yorkshire
- Assess for and provide a wide variety of AT
- Run training and provide support on AT
- Contribute to & run research and development projects…

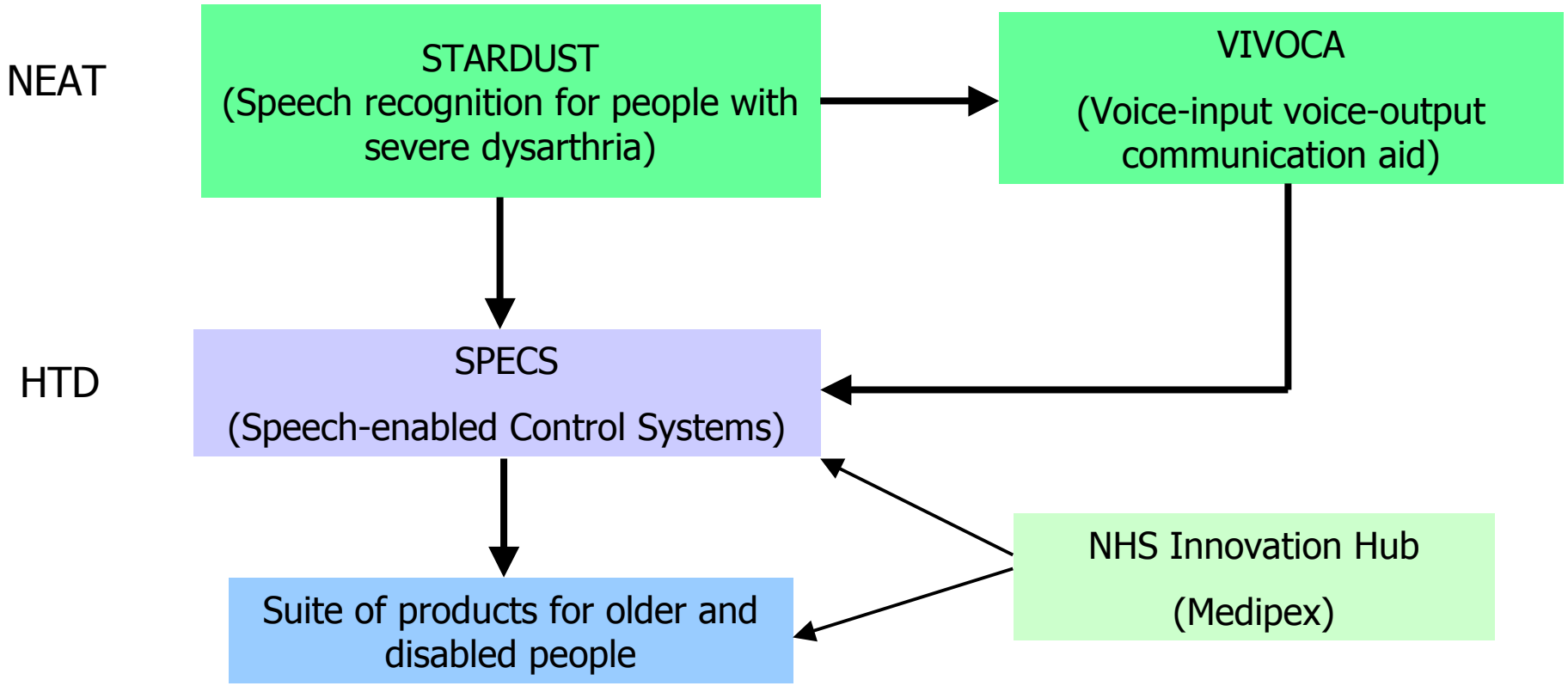
Research Groups

- Barnsley District General Hospital Foundation Trust – R&D Department, AT Team
- Sheffield University – Computer Science Dept, Health and Related Sciences School
- Collaboration & Track record on AT projects.
- New group forming involving AT
- CAST group: Clinical Applications of Speech Technology

Background

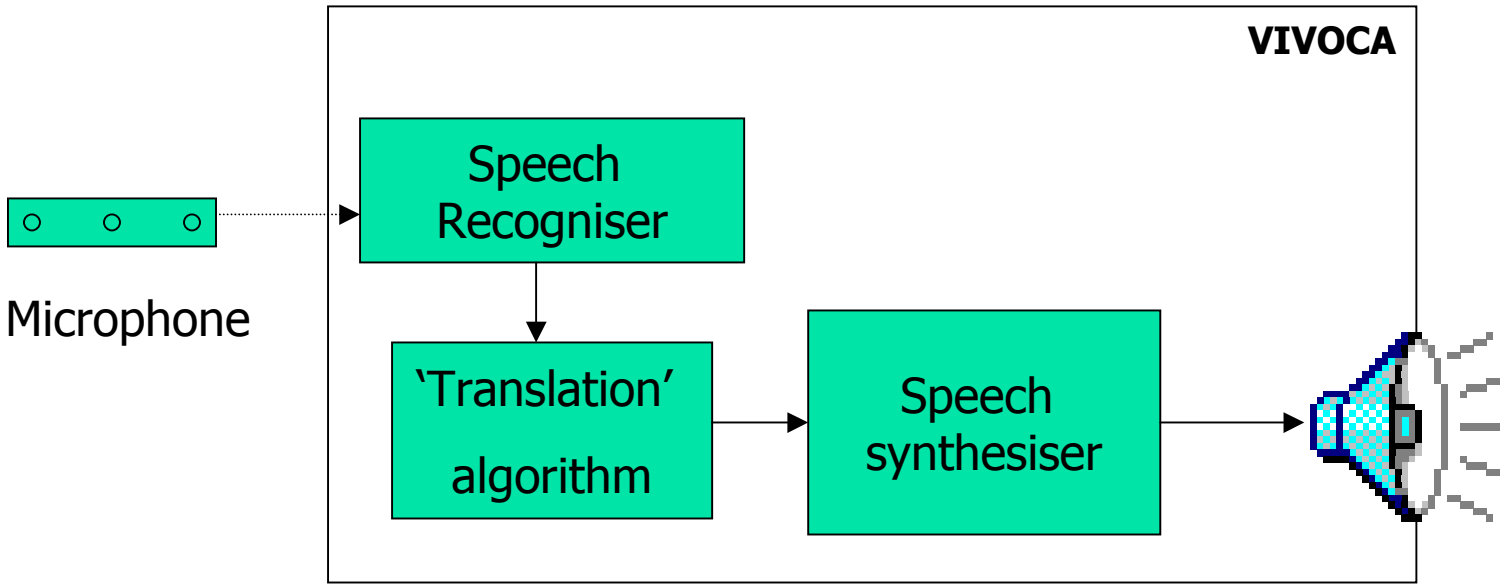
- Dysarthria is the most common acquired speech disorder (170 per 100,000)
- Many current communication aids (VOCAs) are slow and effortful to use
- Dysarthric speech can be an effective control input to assistive technology

CAST Projects

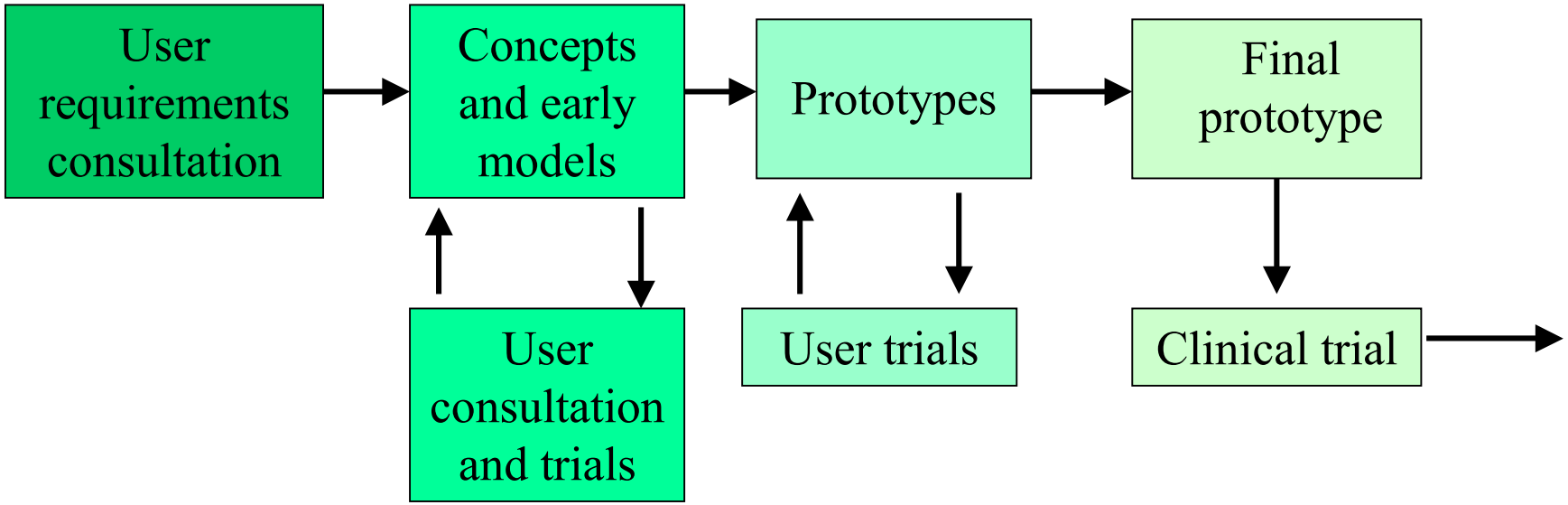


VIVOCA

Voice-input voice-output communication aid



User-centred design & development



User and professional consultation

Method

VOCA users and speech therapists

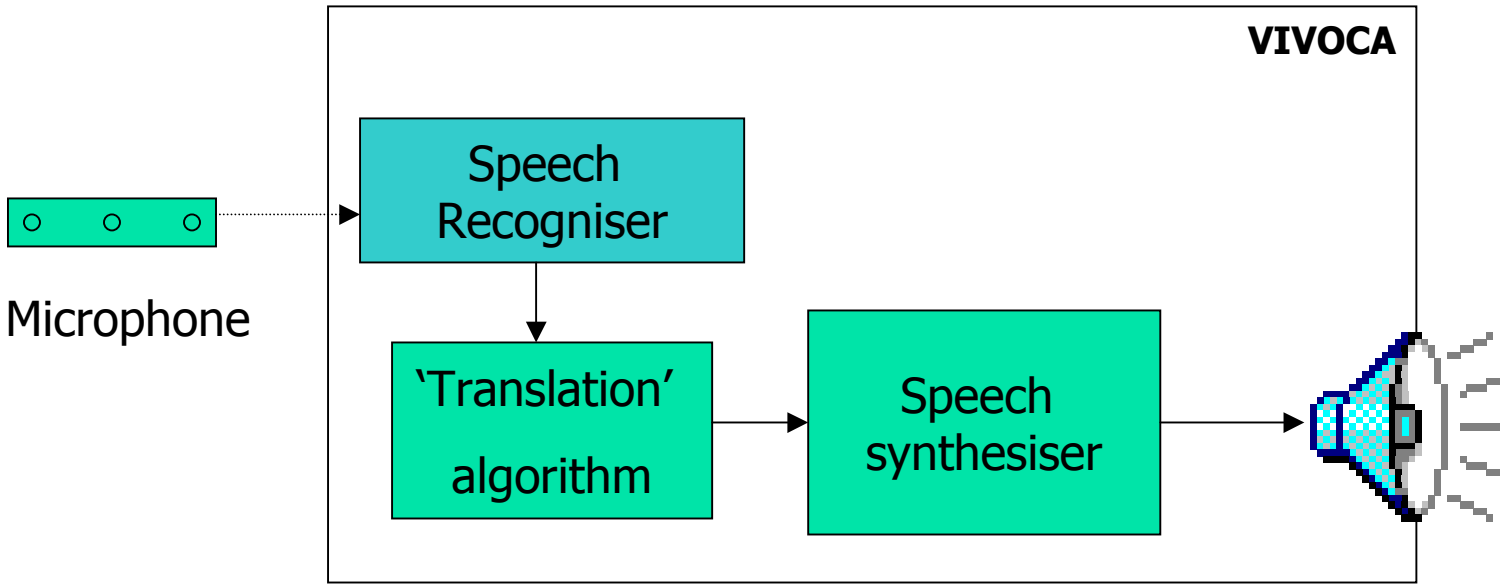
Semi-structured interviews and focus groups

Thematic analysis

Results

- Acceptable as a means of communication
- Potential advantages over conventional VOCA
  - *Quicker*
  - *Easier to use*
  - *Increased communication and independence*
- Useful where speed and intelligibility crucial
  - *Meeting new people*
  - *Telephone*
  - *Shopping*
- Range of requirements for hardware and software

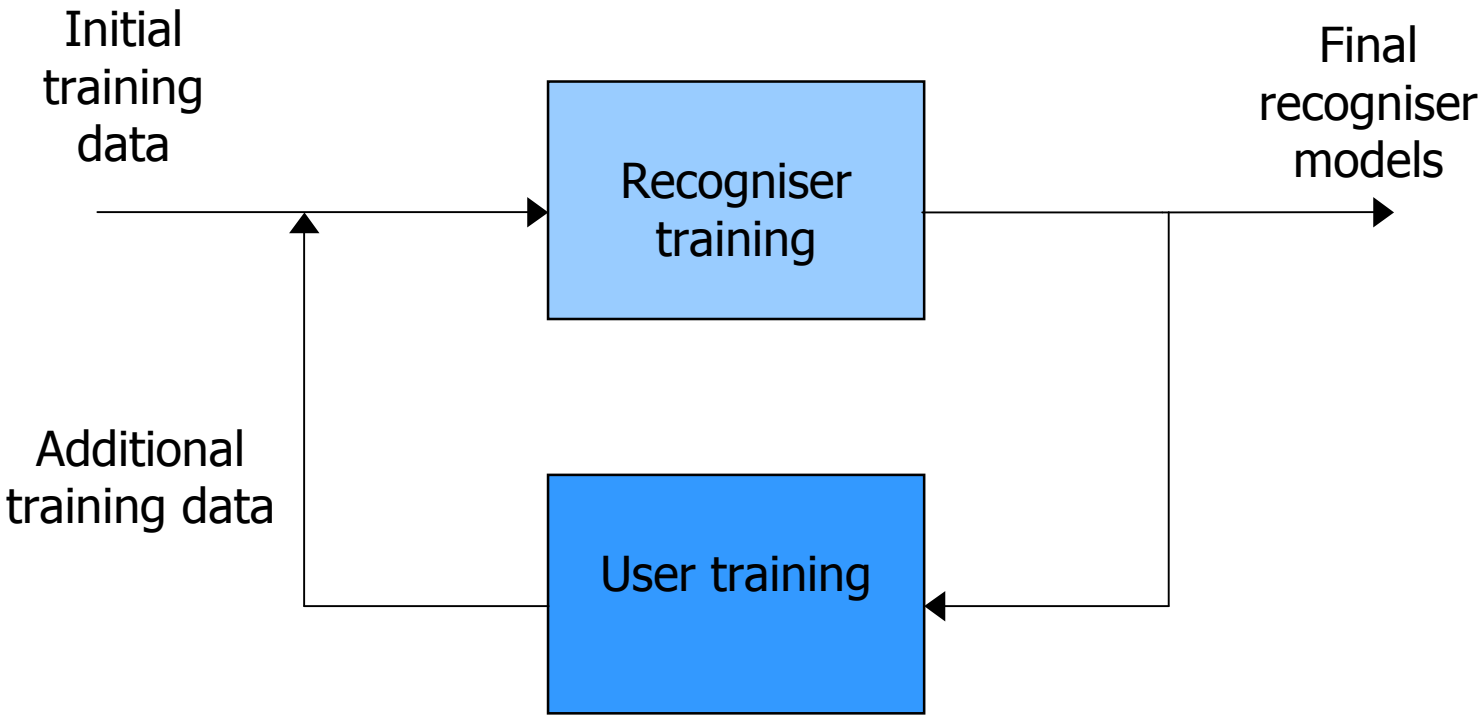
Speech Recogniser



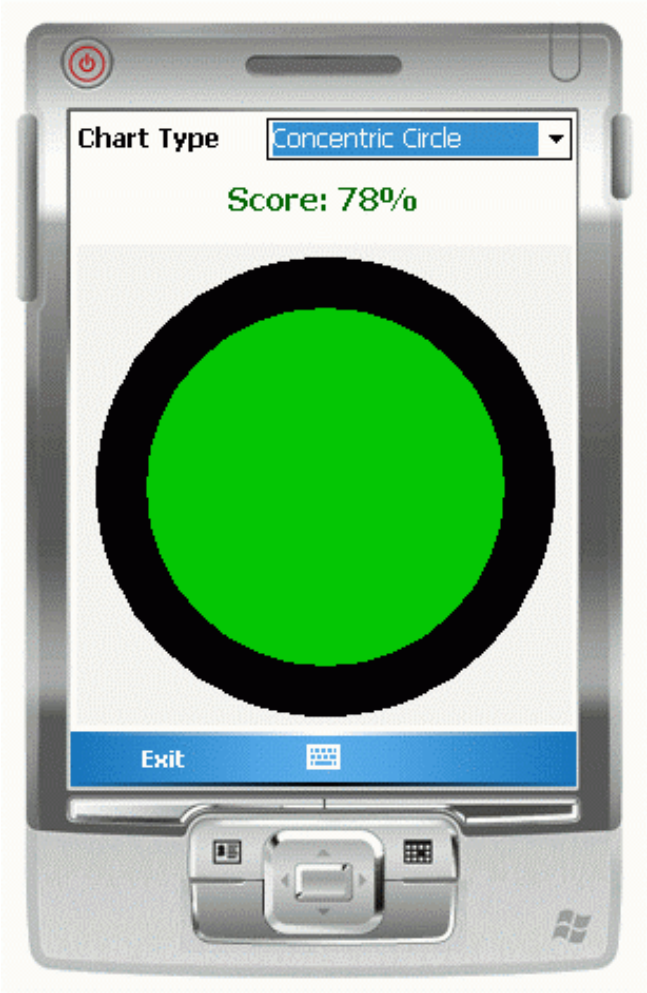
Speech recogniser for dysarthric speech

- Commercial speech recognisers do not work well for dysarthric speech
- User-centred approach – aim to make it work
- Speaker dependent recognition
- Vocabulary of discrete words tailored to speech capabilities of individual
- Closed loop between recogniser training and user training

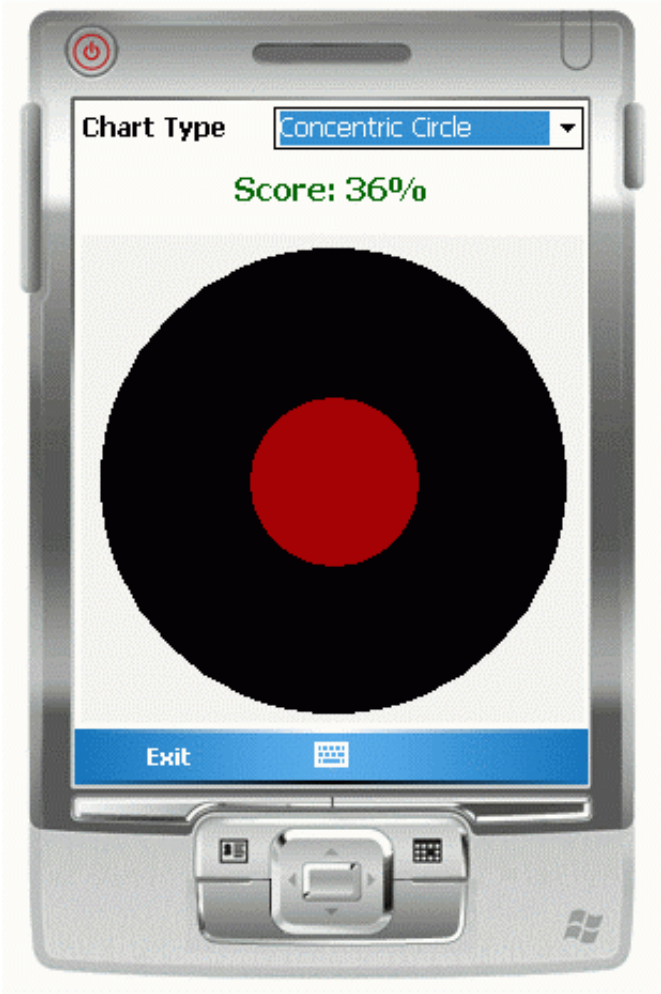
Speech recogniser for dysarthric speech



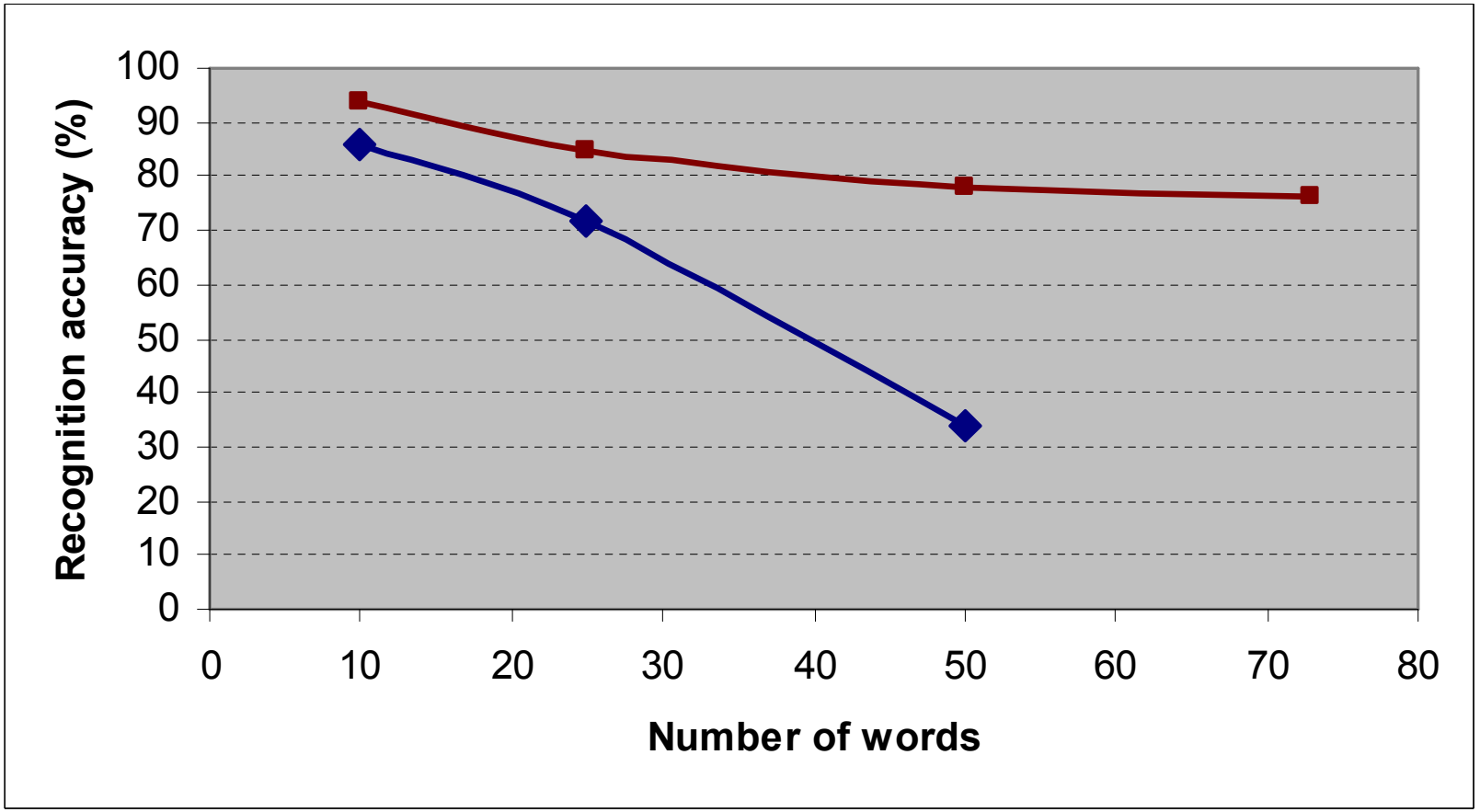
Training: User Feedback



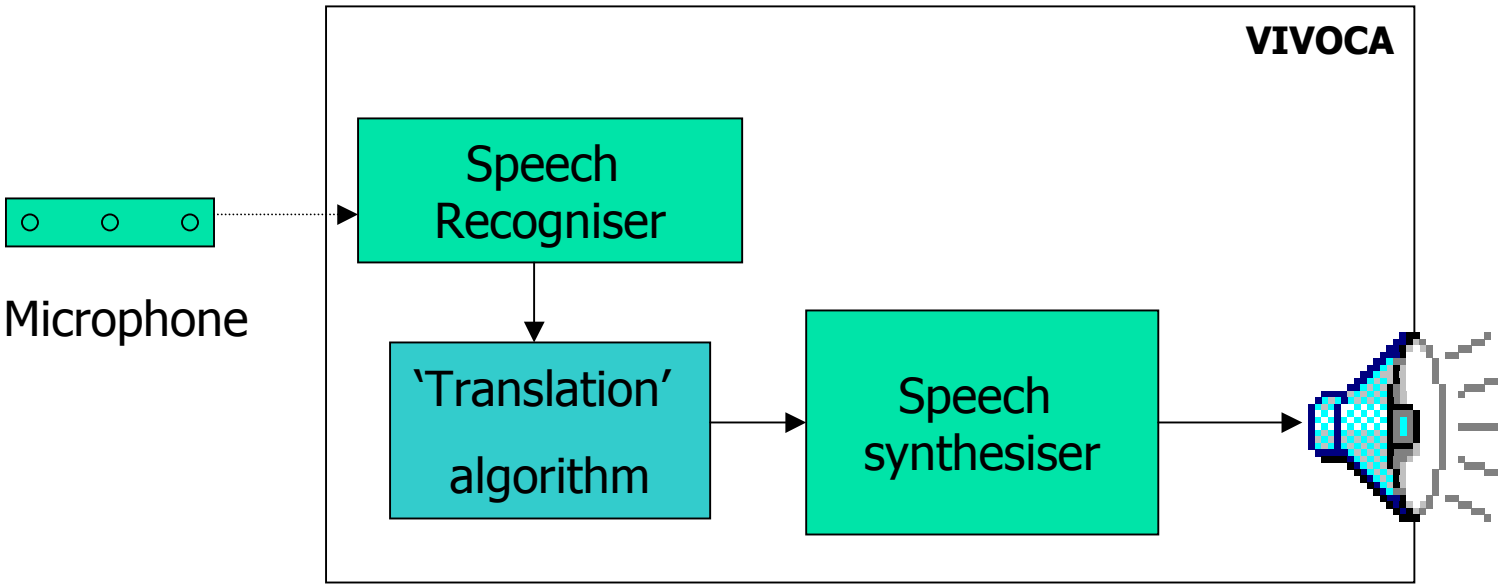
Training: User Feedback



Effect on Recognition Accuracy



Translation Algorithm



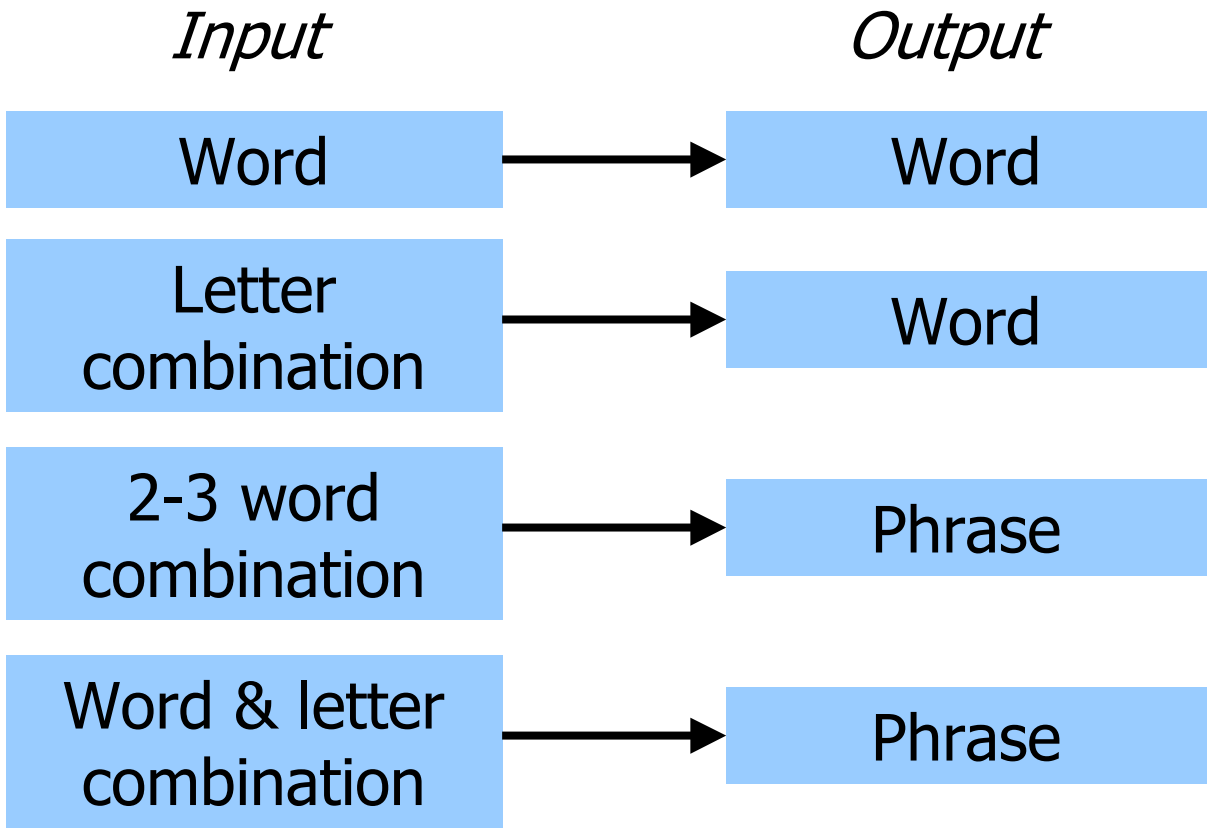
'Translation' methods

Small Number of inputs → *Large vocabularies*

1-5	~10	~30	~100	~1000
Switch scanning Morse code	Mobile phone/T9	Spelling Word prediction AAC	AAC fixed overlay AAC dynamic screen	Word for word translation

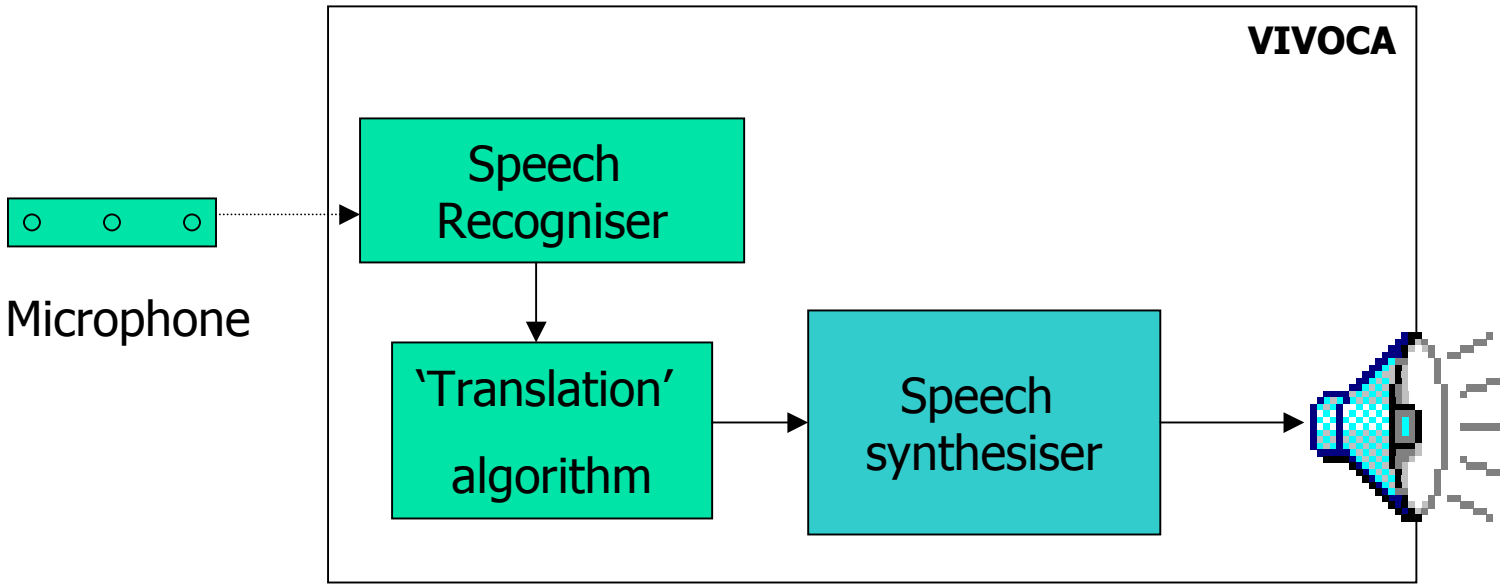
Coding → *Direct 'translation'*

Translation: input-output



"Want", "B", "R" *Can I have a beer please?*

Speech Synthesiser



Current speech synthesis: communication aids

- High quality voices available
- E.g.
  - DECtalk™ (Fonix) for American English 
  - Acapela for British English 
- Personalisation limited: age, gender, language

Personalisation

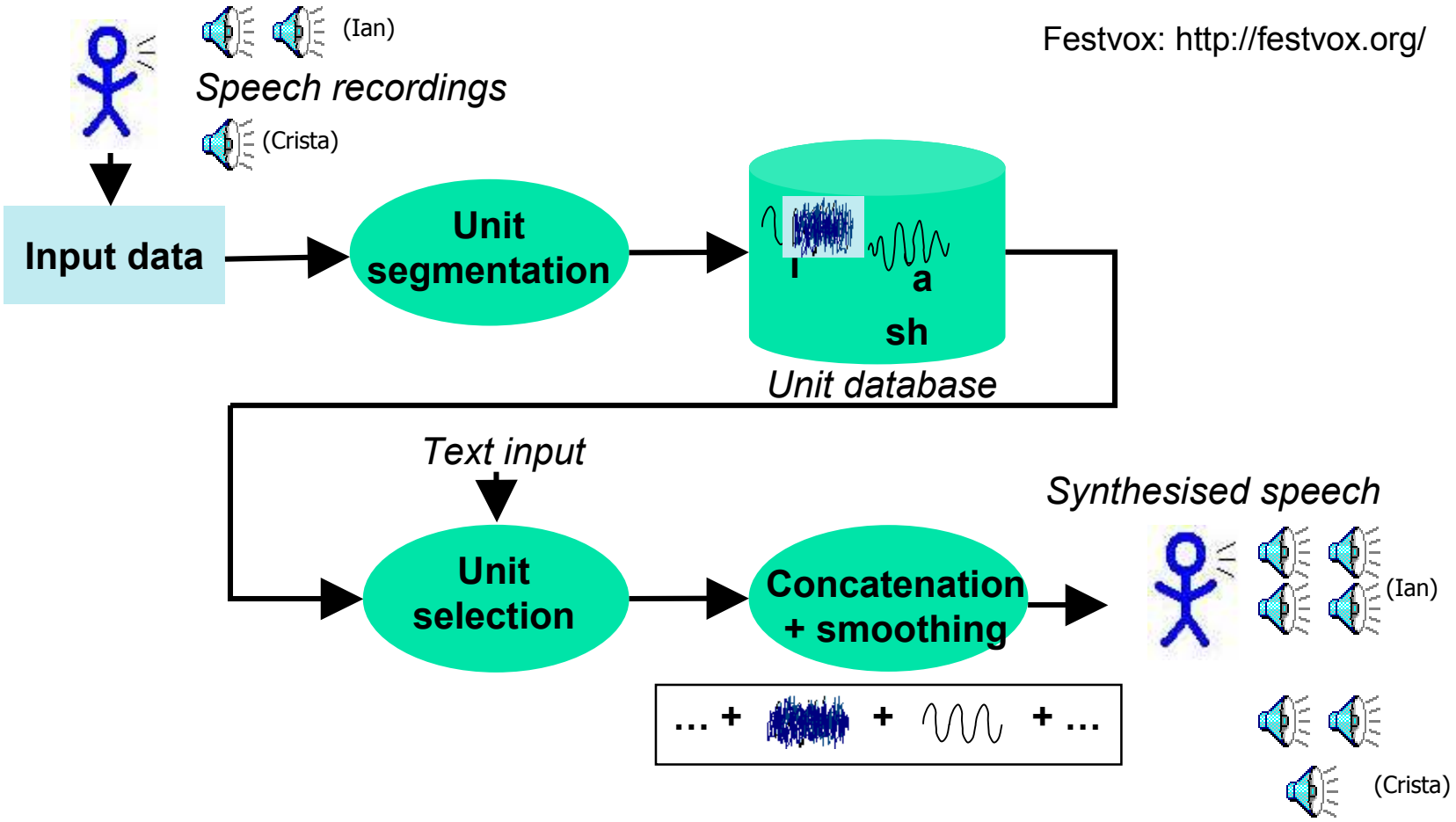
- Voice = identity
  - Gender
  - Age
  - Geographic background
  - Socio-economic background
  - Ethnic background
  - As that individual

- Maintains social relationships
- Maintains social closeness
- Sets group membership

VIVOCA: personalisation

- Sheffield/Barnsley user group
- Retain local accent
  - geographic identity
- Speaker database
  - Arctic database: 593 + 20 sentences
- Professional local speakers
  - Ian McMillan
  - Christa Ackroyd

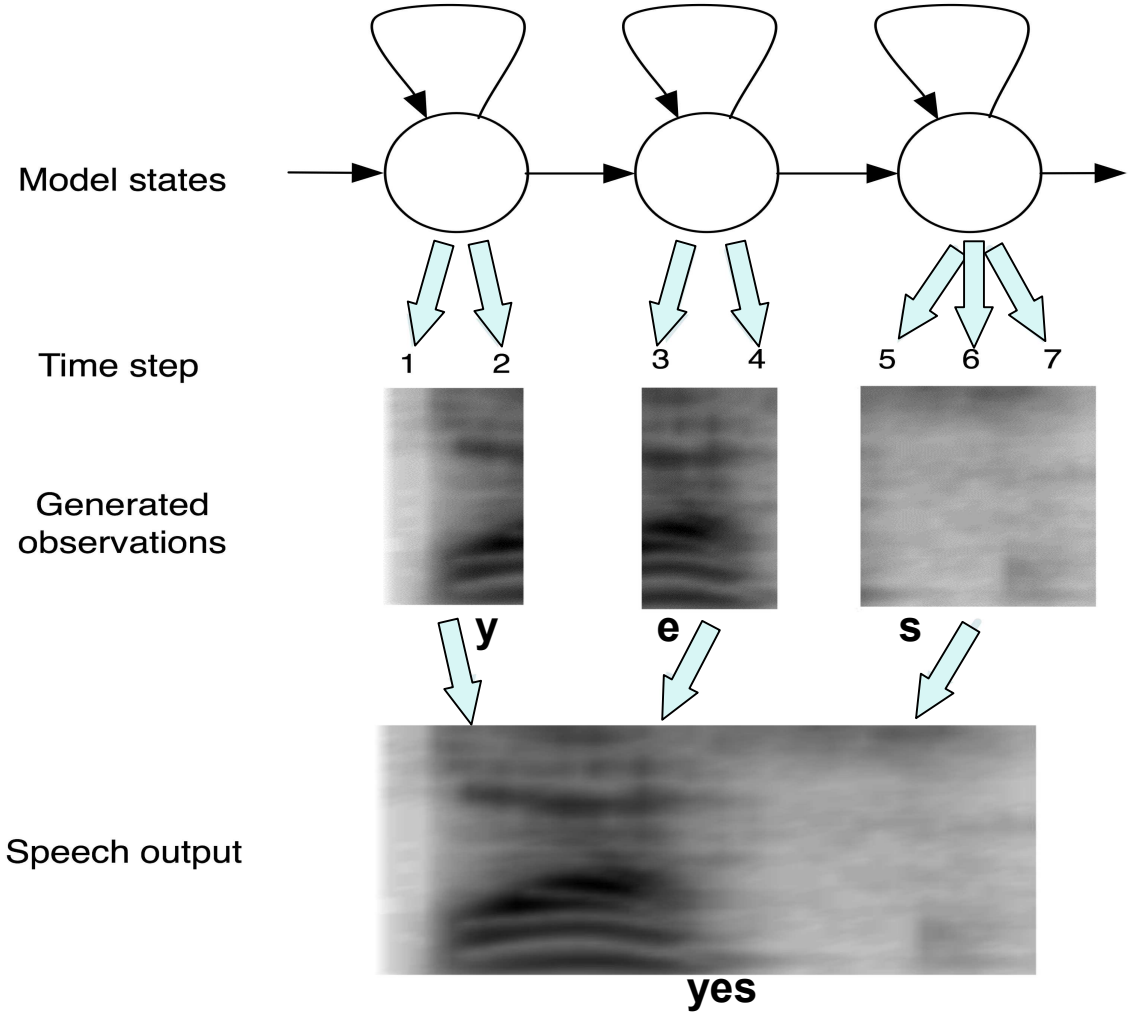
Concatenative synthesis



Concatenative synthesis

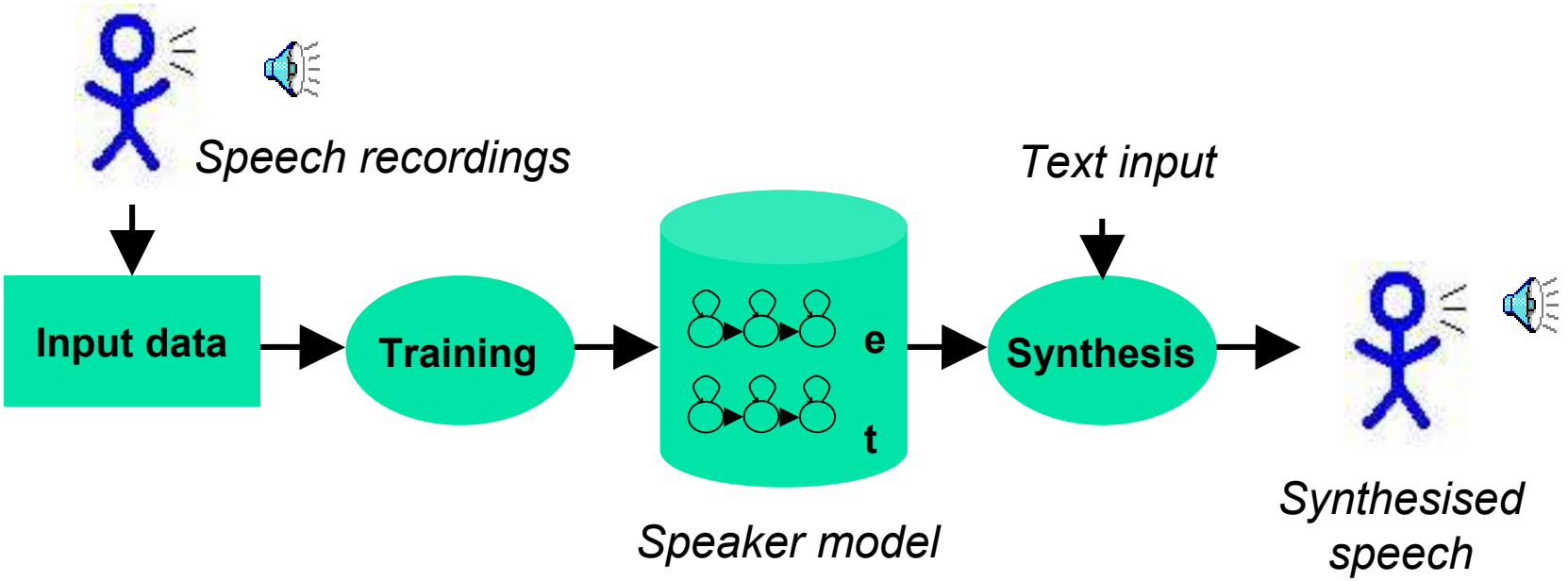
- ✓ High quality
- ✓ Natural sounding
- ✓ Sounds like original speaker
- ✗ Need a lot of data (~600 sentences)
- ✗ Can be inconsistent 
- ✗ Difficult to manipulate prosody

HMM synthesis



HMM synthesis procedure

HTS http://hts.sp.nitech.ac.jp/



HMM synthesis

- ✓ Consistent
- ✓ Intelligible
- ✓ Needs relatively little input (~20 mins)
- ✓ Can be adapted with small amount of data (>5 sentences)
- ✓ Easier to manipulate
- ✗ Buzzy quality
- ✗ Less natural than concatenative

Synthesis: Future research

- Further personalisation for individuals with progressive speech disorders
  - Capturing the essence of a voice
- Voice banking
  - Before deterioration
- Adaptation using HMM synthesis
  - Before or during deterioration

Summary

- Voice in-Voice out device based on a PDA
- Currently under development
- Recognising and improving discrete dysarthric words
- Regionalised, possibly personalised, speech synthesis

CAST Clinical Applications of Speech Technology

VIVOCA Team

Mark Hawley, Pam Enderby, Phil Green, Stuart Cunningham, Peter O'Neill, Rebecca Palmer, Siddharth Sehgal, Fabien Cardinaux, Andre Coy, Simon Judge

www.barnsleyrd.nhs.uk

www.shef.ac.uk/cast

simon.judge@nhs.net

A Plug!!

- RAATE 2007
- 26th and 27th November 2007

www.raate.org.uk

- (session on voice recognition!)