This is a repository copy of *Voice-In, Voice-Out Communication Aids*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/10295/

---

**Conference or Workshop Item:**

---

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
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
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
Voice-input voice-output communication aid
User-centred design & development

- User requirements consultation
- Concepts and early models
  - User consultation and trials
  - User trials
- Prototypes
  - Final prototype
- Clinical trial

Barnsley District General Hospital
NHS Trust

The University of Sheffield

Assistive Technology Team
Barnsley
## 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 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

Initial training data → Recogniser training → Final recogniser models

Additional training data → User training →
Training: User Feedback

Score: 78%
Training: User Feedback
Effect on Recognition Accuracy

![Graph showing the effect on recognition accuracy. The x-axis represents the number of words, and the y-axis represents recognition accuracy in percentage. Two lines are plotted: one in blue and one in red, both indicating decreasing accuracy as the number of words increases.]
Translation Algorithm

Microphone

Speech Recogniser

‘Translation’ algorithm

Speech synthesiser

VIVOCA
### ‘Translation’ methods

<table>
<thead>
<tr>
<th>Small Number of inputs</th>
<th>Large vocabularies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>~10</td>
</tr>
<tr>
<td>~30</td>
<td>~100</td>
</tr>
<tr>
<td>~1000</td>
<td></td>
</tr>
</tbody>
</table>

- Switch scanning
- Mobile phone/T9
- Morse code
- Spelling
- Word prediction
- AAC
- AAC fixed overlay
- AAC dynamic screen
- Word for word translation

Coding

Direct ‘translation’
Translation: input-output

**Input**
- Word
- Letter combination
- 2-3 word combination
- Word & letter combination

**Output**
- Word
- Word
- Phrase
- Phrase

“Want”, “B”, “R”  
*Can I have a beer please?*
Speech Synthesiser

Microphone

Speech Recogniser

‘Translation’ algorithm

Speech synthesiser

VIVOCA
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
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

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

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

Speech recordings

Input data → Training → Speaker model

Text input

Synthesis → Synthesised speech
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
Voice in-Voice out device based on a PDA
Currently under development
Recognising and improving discrete dysarthric words
Regionalised, possibly personalised, speech synthesis
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
- RAATE 2007
- 26th and 27th November 2007

www.raate.org.uk

(session on voice recognition!)