An Edinburgh Speech Production Facility

Alice Turk[a], James M Scobbie[b], Christian Geng[a], Cedric Macmartin[a], Ellen Gurman Bard[a], Barry Campbell[a], Batoul Diab[a], Catherine Dickie[a,b], Eddie Dubourg[a], Bill Hardcastle[b], Phil Hoole[c], Evia Kainada[d], Simon King[a], Robin Lickley[b], Satsuki Nakai[a], Marianne Pouplier[c], Steve Renals[a], Korin Richmond[a], Sonja Schaeffler[b], Ronny Wiegand[a], Kevin White[a], Alan Wrench[e]

[a]University of Edinburgh, [b]Queen Margaret University, [c]LMU Munich, [d]University of Leeds, [e]Articulate Instruments, Ltd

1 Introduction

This unique facility is designed for the collection of articulatory and acoustic data from two synchronised dialogue participants, or single speakers. It will be open to the international research community for funded use as of September, 2010. Services will include data collection (preparation, sensor attachment, and recording), sensor position estimation at each sample point, head movement correction, synchronization (articulation to acoustics; speaker to speaker), and data archiving. The first product of the facility is a corpus of recorded dialogue, also available in September, 2010.

2 The Facility

The facility is built around two Carstens' AG500 electro-magnetic articulographs (EMA) and acoustic recording equipment (see Fig. 1). EMA recordings provide detailed information about speech movements. Each machine records 3D positions and rotations of 12 sensors every 5 ms. These sensors can be glued anywhere on the lips, tongue, jaw, and head. Acoustic recordings are made via an AKG CK98 hypercardoid mic, sampling rate: 32 kHz, bit rate 16. The EMA machines are positioned 8.5 m apart to avoid electromagnetic inter-machine interference. Communication among participants and experimenters is regulated via a talkback system (see Fig. 2).

Figure 1. *Carstens' new 3-D sensor 'cube'. The transparent plastic box surrounds a participant's head, creating a comfortable experimental experience which enables data collection sessions of 1 hour and more.*

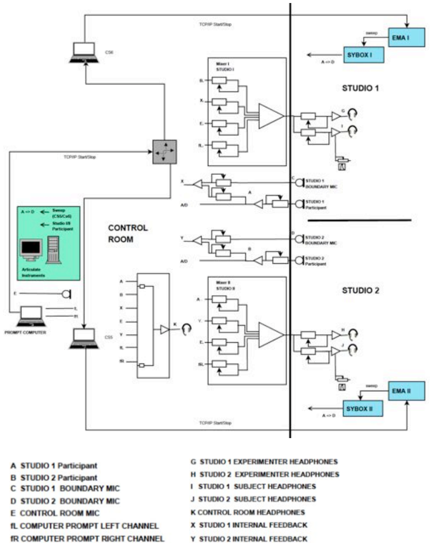


Figure 2. *Facility setup*

2.1 Synchronization

Synchronization of both EMA data sources and the acoustic waveforms is achieved by capturing (a) synch impulses of both machines and (b) the acoustic waveforms of both speakers by means of Articulate Instruments Ltd. hardware. After correcting for TCP-IP inter-machine communication delays, our inter-machine asynchrony is less than 1 ms. The hardware is also capable of synchronizing other time series data (EPG).

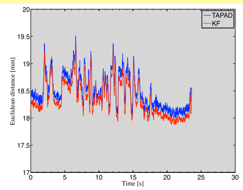
2.3 Data Analysis

Data analysis software (Articulate Assistant Advanced, EMA module) has been commissioned from Articulate Instruments Ltd. (2009). This software allows data visualisation, annotation and measurement extraction. It is user-friendly and does not require programming skills. The user interface provides a common platform for EMA, EPG and Ultrasound data. Analysts need only master one piece of software for all three techniques, and can transfer annotations between them.

2.2 Data Accuracy

Position-estimation procedures include those described in Hoole & Zierdt (in press) (TAPAD) and unscented Kalman filtering-based algorithms, developed by K. Richmond. Analyses for rigid body sensors suggest that accuracy is within 1 mm (see Fig. 3). Data accuracy for non-rigid body sensors is assured by comparing position results from TAPAD vs. Kalman filtering methods (cf. Fig. 3).

Figure 3. *Estimated distance between the central and lateral lower jaw sensors glued to a single participant during a stretch of speech. These rigid body sensors are always a fixed distance apart; our estimated distances suggest accuracy within ca. 1 mm.*



3 The Dialogue Corpus

So far, we have recorded 9 dual participant sessions primarily between Scottish and Southern British English speaking participants. Each session involves synchronized recordings of both EMA and acoustic data, and includes 30-60 minutes of speech. The corpus will be available in Sept. 2010 via a web-based, searchable archive system.

3.1 Sensor Positions

Sensors were attached behind the ears, to the bridge of the nose, to the upper jaw, lower jaw, upper lip, lower lip, tongue front, tongue mid and tongue back.

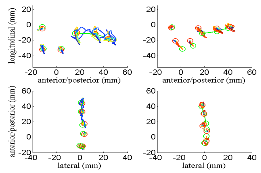


Figure 5. *Midsagittal and axial views of sensor positions for two participants during ca. 7 s. of dialogue speech. Time is represented by colour. Anterior is towards the left in the top panels, towards the bottom in the bottom panels.*

3.2 Speech Styles

Monologue
Story reading (Comma Gets a Cure, McCullough,Somerville & Honorof 2000), Wellsian lexical sets, spontaneous story telling, diadochokinetic tasks
Dialogue
Map tasks (Anderson et al. 1991), Spot the Difference picture tasks (Bradlow et al. 2007), Story-recall
Shadowing
One participant tells a familiar story, the other shadows.

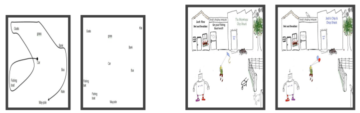


Figure 6. *Example prompting materials for map and spot the difference tasks*

3.3 Annotation

Annotation files include orthographic transcription and long pauses. Disfluency annotation is in preparation, and we are developing a guide for prosodic labeling (simplified ToBI).

3.4 Data Preview

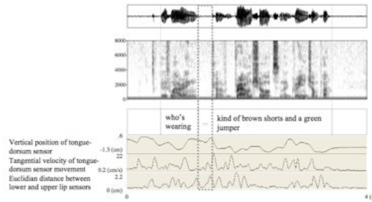


Figure 7. *From a 'Spot the difference' dialogue. Although the speaker could have held his tongue dorsum in position for /k/ in 'kind' following the velar coda in 'wearing', tongue dorsum movement traces suggest his tongue dorsum has moved downward during the hesitation pause. Lip movement traces suggest he has closed his mouth and opened it again during this interval.*

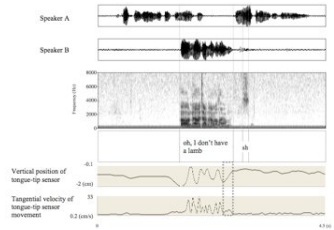


Figure 8. *From a 'Spot the difference' dialogue. Speaker B seems to have begun movement towards 'sh-' (of 'sheep'?) during the [m] closure of 'lamb' (see the box with gray dotted lines), before Speaker A interrupts. The duration between the onset of A's speech and the end of 'sh-' is ca. 300 ms, possibly the time it takes B to process that A is talking and to terminate his speech.*

References

- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34, pp. 351-366.
- Articulate Instruments Ltd (2009) Articulate Assistant Advanced: Version 2.09. Edinburgh, UK: Articulate Instruments Ltd.
- Bradlow, A.R., Baker, R.E., Choi,A., Kim, M. & van Engen, K.J. (2007). The Wildcat corpus of native- and foreign-accented English. Journal of the Acoustical Society of America, 121(5) Pt. 2, 3072.
- Carstens Medizinelektronik http://www.articulograph.de/
- Hoole, P & Zierdt, A. (in press). Five-dimensional articulography. In B. Maassen, van Lieshout, P.H..H..M. (Eds.) Speech Motor Control: New developments in basic and applied research., OUP, 331-349.
- McCullough, J., Somerville, B & Honorof (2000). Comma Gets a Cure.. http://web.ku.edu/~idea/readings/comma.htm
- Wells, J.C. (1982). Accents of English I: An Introduction. CUP.

Acknowledgements

This project was funded by EPSRC EP/E01609X/1 and EP/E016539.

FR - Other
Poster describing the facility