An Edinburgh Speech Production Facility

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
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, synchronisation (articulation to acoustics and 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 EMA 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 CM9 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).

2.1 Synchronization

Synchronization of both EMA data sources and the acoustic waveforms of both speakers by means of Articulate Instruments Ltd. hardware. After correcting for TCP/IP machine and (b) the acoustic waveforms of both speakers by means of Articulate Instruments Ltd. hardware. After correcting for TCP/IP machine and inter-machine interference.

2.2 Data Accuracy

Position-estimation procedures include those described in Hausle & Zünk (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 assumed by comparing position results from TAPAD vs. Kalman filtering methods (cf. Fig. 3).

3 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.

3.2 Speech Styles

Monologue

Story reading (Connu Gets a Cure, McCaugh, Somerville & Honnold 2000). Well trained actors, spontaneous story telling, diadochokinetic tasks

Dialogue


Shadowing

One participant tells a familiar story, the other shadows.

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).

4 Data Preview

Figure 7. From a ‘Spot the difference’ dialogue. Although the speaker could have held his tongue dorsal in position for /k/ in ‘kind’ following the vowel code in ‘quantity’, tongue dorsal movement traces suggest his tongue dorsal has moved downward during the hesitation pause. Lip movement traces suggest he has closed his mouth and opened it again during this interval.

Figure 7. From a ‘Spot the difference’ dialogue. Although the speaker could have held his tongue dorsal in position for /k/ in ‘kind’ following the vowel code in ‘quantity’, tongue dorsal movement traces suggest his tongue dorsal has moved downward during the hesitation pause. Lip movement traces suggest he has closed his mouth and opened it again during this interval.

Acknowledgements

This project was funded by EP/S010690X/1 and EP/ E016559.

References

Korin Richmond, Satsuki Nakai, Barry Campbell, Evia Kainada, Konrad Richmond, Sonja Schaeffer, Kevin White, Alan Wrench, Bill Hardcastle, Eddie Dubourg, Ellen Guzman Bard, William Hardcastle, Mariam Hartinger, Simon King, Robin Hickley, Cedric Maccartain, Satsuki Nakai, Steve Renals, Korin Richmond, Sonja Schaeffer, Kevin White, Alan Wrench

Address: a, EDDIE DUBOURG, b, BILL HARDCASTLE, c, PRINCE, c, KONRAD RICHMOND, c, Robin Hickley, d, ARTICULATE INSTRUMENTS, LTD. Articulate Instruments, Ltd

University of Edinburgh, d, Queen Margaret University, e, LMU Munich, f, University of Leeds, g, Articulate Instruments, Ltd

E016559.