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, 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 electromagnetic 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 kfps, bit rate 16. The EMA systems 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 data from two synchronised dialogue participants, or single speakers. 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 all three techniques, and can transfer annotations between them.

2.2 Data Accuracy

Position-estimation procedures include those described in Houle & Zekun (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).

2.3 Data Analysis

Data analysis software (Articulate Assistant Advanced, EMA mode) 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 both EMA, EPG and ultrasound data. Analysis need only master one piece of software for all three techniques, and can transfer annotations between them.

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.

3.2 Speech Styles

Monologue
Story reading (Comma Gets a Cure, McCallough, Somerville & Honnold 2000). Wellman lexical sets, spontaneous story telling, dialectologically-annotated 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 TSB).

3.4 Data Preview

Figure 5: Mid axial and axoral 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.

Figure 8: From a ‘Spot the difference’ dialogue. Speaker B seems to have begun movement towards ‘sh’ (‘shoo!’) during the [f] closure of ‘Sam’ (see the bottom panel with the gray dotted lines), before Speaker A interrupts. The duration between the onset of A’s speech and the end of ‘sh’ is ca. 100 ms, possibly the time it takes B to process that A is talking and to terminate his speech.

4 Acknowledgements

This project was funded by EPSRC EP/E01690X/1 and EP/E016539.