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.
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.

2.1 Synchronization

Synchronization of both EMA data sources and the acoustic waveforms is achieved by capturing (a) the pose of both speakers 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).

2.2 Data Accuracy

Position-estimation procedures include those described in Hoole & Zekun (in press) (TAPAD) and unscant 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 computing position results from TAPAD vs. Kalman filtering methods (cf. Fig. 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

Dialogue

Story readings (Comma Gets a Cure, McCaugh, Somerville & Honfere 2000). Wellgün lexical sets, spontaneous story telling, diadochokinetic tasks

Shadowing

One participant tells a familiar story, the other shadows.

3.3 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 voiceless codas of ‘/km/’ (see the box with gray dotted lines), before Speaker A interrupts. The duration between the onset of A’s speech and the end of ‘km’ is ca. 100 ms, possibly the time it takes B to process that A is talking and to terminate his speech.

3.4 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 programming language, in this case C++. It can transfer annotations between them.

References


Acknowledgements

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