# The Edinburgh Speech Production Facility

# DoubleTalk Corpus

*James M Scobbie*
*Alice Turk*
*Christian Geng*
*Simon King*
*Robin Lickley*
*Korin Richmond*

↑ *Data Archive*

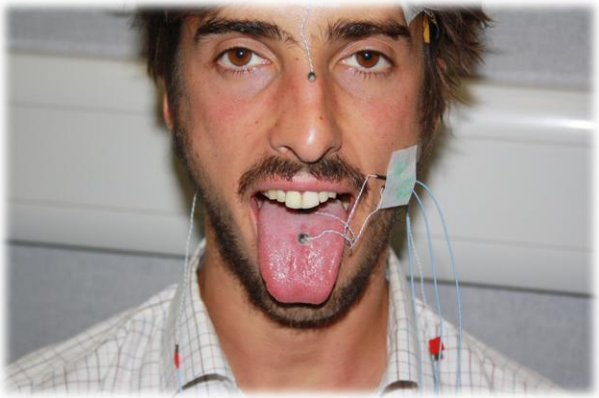
## speakers

5 naïve stranger pairs:
Southern British English
& Scottish English
1 pilot pair: Northern
English & Gen American

## all connected speech

~39,000 word tokens
~2,000 word types
~168 talk minutes
~4,600 utterances
2.2 seconds / utterance
3.9 words / second
8.5 words / utterance

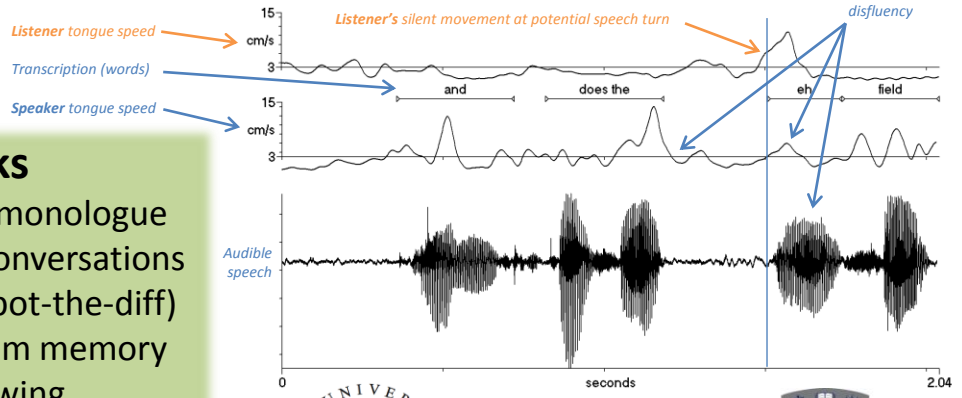
| All c.s. | and | yeah | so | all | made | threw |
|---|---|---|---|---|---|---|
| Rank | 2 | 9 | 15 | 42 | 121 | 210 |
| Cum % | 10% | 25% | 33% | 50% | 66% | 75% |
| Tokens | 1592 | 637 | 445 | 155 | 48 | 27 |



## just the map-task & spot-the-difference

~17,000 word tokens
~1,000 word types
~71 talk minutes
~2,450 utterances

| Conv. | yeah | ok | uhm | uh | uh-huh | ehm | eh |
|---|---|---|---|---|---|---|---|
| Rank | 2 | 7 | 11 | 17 | 29 | 32 | 39 |
| Tokens | 587 | 438 | 305 | 256 | 116 | 107 | 76 |




## data

dual synchronized
Carstens AG-500 EMA
AAA™ software module
free data download

## annotation

PRAAT textgrid
utterance-level
orthographic
includes dysfluencies

## tasks

spontaneous monologue
spontaneous conversations
(map-task & spot-the-diff)
repetition from memory
shadowing
reading passage
+ word list & DDK

Geng, C., Turk, A, Scobbie, J.M.,. et al. Recording speech articulation in dialogue: Evaluating a synchronized double Electromagnetic Articulography setup. (In Press, 2013) *Journal of Phonetics*.
Data download from ESPF at http://espf.ppls.ed.ac.uk/ This Interspeech 2013 poster at http://eresearch.qmu.ac.uk/3219/

THE UNIVERSITY OF EDINBURGH
EPSRC Pioneering research and skills
Queen Margaret University EDINBURGH