Ultrasound Tongue Imaging: some methodological protocols

James M. Scobbie*, Alan A. Wrench*†, Jane Stuart-Smith*, Marietta van der Linden* and Eleanor Lawson*
*Queen Margaret University, Edinburgh, †Articulate Instruments Ltd, *University of Glasgow
BAAP Sheffield 2008

Head-probe stabilisation

If the probe or head move during data collection relative to each other, the data is only useful for broad qualitative analysis. Two alternatives are

- Keep the head and probe immobilised for the whole experiment.
- Allow normal head movement.
  - Use headset to keep probe aligned to head.
  - Quantify and perhaps correct for [1] head-probe movement.

Method

In headset testing, Vicon motion capture was used to analyse position of the probe relative to normalised forehead/nose positions for three naïve subjects, and in 3D.

Long-term movement (Fig. 1)

Slippage of the headset on the head is small.

- 1.5mm - 3mm between rest periods.

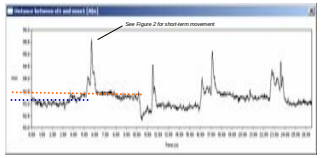


***Figure 1**. Linear distance between bridge of nose and probe. Orange and Blue lines show long term shift which occurred between two rest periods. The spikes are short-term movements caused during speech, e.g. in "ho-mo Maggie"*

Short-term movement (Fig. 2)

Jaw lowering pushes probe down and rotates it slightly in mid-sagittal plane only. A low vowel causes average peak translation errors of

- 3.5mm to 6mm backing.
- 1.1mm raising to 4.5mm lowering.
- Worst case: 10mm backer, 7.5mm lower.

The dynamics of short-term movement show some x-y rotation (two subjects were measured).

- 1.0º and 2.6º clockwise → a more anterior field of view. Worst case: 1.4º and 4.8º.

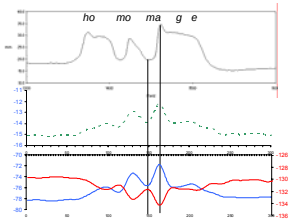


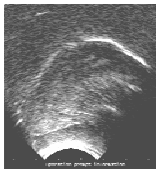
***Figure 2**. Dynamic changes in lip aperture (top) and associated mid-sagittal headset rotation in degrees (middle, green dashed line), and mid-sagittal translation (bottom, blue line posterior movement and red line superior movement, both in mm on colour-appropriate scales) during [ma] of "Maggie". Frames are synchronised and time is shown in the horizontal axis (mm)*

Conclusions

- Like temporal synchronisation and spatial resolution [2], stabilisation is crucial and should be quantified by each laboratory.
- The headset restricts all meaningful error to the mid-sagittal plane.
  - *The headset can very effectively reduce probe movement long-term throughout a session.*
  - *Dynamic translation and rotation are tolerable, especially for repeated measures of targets.*
  - *Speakers can rotate and move their head naturally, very important for some subject groups.*
- Error correction needs Vicon during all data collection plus post-processing.
- Error with the headset is small enough that the headset alone may be used.

Though error is correctable e.g. [1], no system can correct for error outside the mid-sagittal plane, and the data so affected, if detected, is discarded [1]. When the head is immobilised, error in very compliant subjects can be greatly reduced e.g. [3]. Non-mid-sagittal error still can occur – something the headset prevents.

[1] Whalen et. al (2005) The Haskins Optically Corrected Ultrasound System (HOCUS). Journal of Speech Language and Hearing Research 48: 543-553.
[2] Wrench, A.A. and Scobbie, J.M. (2006) Spatio-temporal inaccuracies of video-based ultrasound images of the tongue. *Proceedings of ISSP 06*, 451-458.
[3] Hau, C., Bressmann, T., Smallwood, R. and Wong, W. (2007) 3D ultrasound on a budget: Reconstruction of 3D tongue shapes from multiple coronal planes. Oral paper at Ultrafest IV, New York University.



Ecological validity of articulatory data

Laboratory articulatory data is taken at face value in phonetics research.

- Speakers tend to be colleagues or experienced subjects, style is formal, phenomena are only those found robustly in standard varieties.
- A long-standing problem [4]: How do we study vernacular articulation with naïve speakers and get results relevant for phoneticians *and* sociolinguists?

Method

We used sociolinguistic fieldwork methods to measure the extent to which vernacular speech is affected by the use of headset and Ultrasound Imaging

- 14 male subjects recorded in their school twice, first in audio-only condition, then on a different day with either UTI (n=10), or audio-only as control (n=4).
- Variables examined include easily transcribed TH-fronting, T-glottaling, and L- vocalisation.
- /r/ vocalisation also measured, and rates are reported elsewhere [5].
- Analysis of UTI in field vs. UTI in laboratory setting is ongoing.

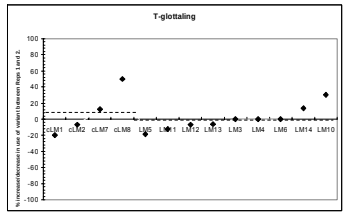


***Figure 3**. Change in use of variable T-glottaling showing individual results, ordered to show range of behaviour from greatest decrease to greatest increase. Controls are in the leftmost four columns.*

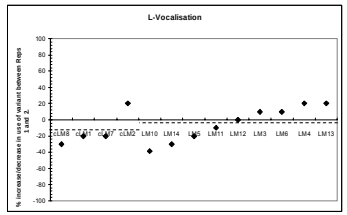


***Figure 4**. Change in use of variable L-vocalisation*

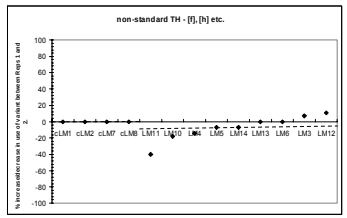


***Figure 5**. Change in use of variable TH-fronting*

Word-list Speech (Figs. 3, 4, 5)

In field-setting, there was little change when UTI was introduced, and any change was similar to audio-only retest control group.

Conclusions

- Field setting requires portable kit, is noisy, and is technically more complex.
- Field is fine for qualitative analysis, is quicker, and provides more subjects.
- Laboratory setting seems more suitable for quantitative sociophonetics.
  - *Friendship dyads in recording studio may help maintain vernacular speech style.*
  - *Having no experimenter present also helps but probably increases loss of data.*

Ultrasound Imaging data is very useful for sociolinguistic research, and vernacular speech need not to be so under-represented in experimental phonetic research.

In ongoing research into Scottish /r/ we are currently collecting an articulatory corpus and undertaking experimental work with a socially-stratified group of subjects in the laboratory.

[4] Kerswill, P. & Wright, S. (1990) On the limits of auditory transcription: a sociophonetic perspective. *Language Variation and Change* 2: 255-275.
[5] Lawson, E., Stuart-Smith, J. and Scobbie, J.M. (in press) Articulatory insights into language variation and change: preliminary findings from an ultrasound study of derhoticisation. *Selected papers from NWAV 36, Pennsylvania Working Papers in Linguistics* 14.2.