

# Watch your Tongue: A point-tracking visualisation system in Blender

Kristy James<sup>1,2</sup>, Ingmar Steiner<sup>1,3</sup>, Martijn Wieling<sup>2</sup>, Alexander Hewer<sup>1,3</sup>, and Angelika Braun<sup>4</sup>

<sup>1</sup>*Computational Linguistics & Phonetics and MMCI, Saarland University*

<sup>2</sup>*Information Science, University of Groningen*

<sup>3</sup>*Language Technology Lab, DFKI*

<sup>4</sup>*Phonetics, University of Trier*

**Abstract** We present a new software tool for the visualisation of EMA data, using 3D animation in a game engine. This tool displays the movement of articulators in real-time, extrapolating from point-tracking data to a basic representation of tongue and lip movement, with plans to include more accurate tongue and palate models in the future. The tool is written in Python and reads data into Blender, an animation and game engine, in real-time. In addition, Blender game-like resources have been developed, so that a basic mouth ‘scene’ is provided, of which the user can fully customise the appearance and behaviour to their own needs.

The tool features functionality to display pre-recorded data from various data formats, which may be of use in demonstrating recorded data from different speakers, as well as rendering a live data stream from an articulograph, which could be adapted to provide online feedback for pronunciation training. In both modes, game controls allow the user to choose their preferred viewpoint and set game parameters, whilst the researcher can set other parameters before the streaming commences.

Furthermore, effort has been taken to incorporate several modalities: in static data mode, simultaneously recorded ultrasound videos can be overlaid on the image (motivated by the data described in Steiner et al., 2014), and in live mode a webcam recording can also be shown.

Competing systems utilising proprietary software already demonstrate that some degree of placement feedback can be given using 3D animation in real-time (Katz et al., 2014), and feedback for the acquisition of foreign phones (Levitt and Katz, 2010) and phonetic correction in L2 (Badin et al., 2010) have been investigated. The degree to which EMA data can describe differences in L2 accent and the differences between native and foreign speech are an active area of research as well (Wieling et al., 2015).

The open-source nature of both this package, as well as the ease of scripting with Python mean that this package would be ideally adapted for experimenting with visual feedback and feedback timing, as well as individualising placement feedback to a subject’s articulator location.

## References

- Badin, Pierre et al. (2010). “Visual articulatory feedback for phonetic correction in second language learning”. In: *Proc. Second Language Studies: Acquisition, Learning, Education and Technology*, P1–10. URL: [http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010\\_P1-10.pdf](http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_P1-10.pdf).
- Katz, William et al. (2014). “Opti-Speech: A Real-Time, 3D Visual Feedback System for Speech Training”. In: *Proc. Interspeech*. URL: [http://www.isca-speech.org/archive/interspeech\\_2014/i14\\_1174.html](http://www.isca-speech.org/archive/interspeech_2014/i14_1174.html).
- Levitt, June S. and William F. Katz (2010). “The effects of EMA-based augmented visual feedback on the English speakers’ acquisition of the Japanese flap: a perceptual study”. In: *Proc. Interspeech*, pp. 1862–1865. URL: [http://www.isca-speech.org/archive/interspeech\\_2010/i10\\_1862.html](http://www.isca-speech.org/archive/interspeech_2010/i10_1862.html).
- Steiner, Ingmar et al. (2014). “Investigating the effects of posture and noise on speech production”. In: *10th International Seminar on Speech Production (ISSP)*.
- Wieling, Martijn et al. (2015). “Comparing L1 and L2 speakers using articulography”. In: *Proc. 18th International Congress of Phonetic Sciences*. URL: <http://www.icphs2015.info/pdfs/Papers/ICPHS0551.pdf>.