CO-REGISTRATION OF SIMULTANEOUS HIGH-SPEED ULTRASOUND AND ELECTROMAGNETIC ARTICULOGRAPHY FOR SPEECH PRODUCTION RESEARCH

Sam Kirkham[1], Patrycja Strycharczuk[2], Emily Gorman[1], Takayuki Nagamine[1], Alan Wrench[3]

[1]Lancaster University, [2]University of Manchester, [3]Queen Margaret University
{s.kirkham, e.gorman, t.nagamine}@lancaster.ac.uk, patrycja.strycharczuk@manchester.ac.uk,
awrench@articulateinstruments.com

ABSTRACT

We outline a method for collecting simultaneous high-speed ultrasound and electromagnetic articulography (EMA) data using widely-available commercial hardware and software. Previous research demonstrates the utility of a simultaneous EMA-ultrasound set-up and reports the magnitude of probe rotation and displacement over an experimental session. We build upon this research by (1) combining systems with higher temporal and spatial accuracy than those previously used; (2) using a more principled method for temporal synchronisation; (3) reporting the effects of ultrasound on the accuracy of EMA tracking; (4) tracking probe movement using an alternative stabilisation method. Our system shows a high degree of temporal and spatial accuracy and can be easily implemented by other researchers.

Keywords: Electromagnetic articulography, ultrasound tongue imaging, speech articulation, co-registration

1. INTRODUCTION

In this study we report the design and evaluation of an approach to recording simultaneous ultrasound and electromagnetic articulography (EMA). Ultrasound provides a way of imaging the majority of the tongue surface from tip to root, and is a relatively non-invasive and cost-effective approach to imaging speech articulation. Ultrasound frame rates are lower with video-based capture (∼30 Hz), but modern digital systems are capable of 200+ Hz frame rates (although 80–100 Hz is more common given the settings used for most speech imaging). Ultrasound typically involves stabilisation of the probe relative to the head [1, 2] or tracking of head motion for post-hoc correction [3]. EMA differs from ultrasound in that it tracks sensors attached to specific fleshpoints such as the tongue and the lips, but it has very high temporal resolution (up to 1250 Hz but commonly downsampled to 250 Hz). A further advantage of EMA is that it can easily capture movement in three-dimensions and simultaneously capture movement in sagittal planes, e.g. by tracking lateral edges of the tongue [4, 5]. There exist a wide range of established timing measures for EMA data [6], while such measures for ultrasound are still largely in development [7].

Combining the high temporal accuracy and precise point-tracking of EMA with the rich spatial information of ultrasound is a promising avenue for a range of experimental paradigms in speech research. To this end, [8, 9] report the results of simultaneous EMA-ultrasound experiments based on a new probe stabilisation method. They find excellent stabilisation of the probe relative to the head, with transducer rotation limited to 1.25° and translation limited to 2.5 mm, which is within the tolerances of 2–4 mm translation and rotation of less than ∼5° outlined by [3]. In terms of empirical and theoretical research, [4] combine EMA and midsagittal ultrasound to analyse coda /l/ darkening and tongue lateralisation in New Zealand English.

There remain challenges, however, to conducting simultaneous EMA-ultrasound studies. These include accurate and principled temporal synchronisation between the two systems, ensuring no effects of the ultrasound hardware on EMA tracking accuracy, and minimising the effect of probe movement. In this study, we build upon previous work, outlining an approach to synchronisation between systems, quantifying the effect of ultrasound on EMA tracking, and reporting probe motion using an alternative probe stabilisation method.

2. EXPERIMENTAL SET-UP

2.1. Hardware

The hardware used in this study is a 16-channel Carstens AG501 electromagnetic articulograph and

a Telemed MicrUS scanner. The AG501 system records three positional coordinates and two angular coordinates at 1250 Hz for each sensor, and emits a synchronisation pulse every 4 ms for audio-EMA synchronisation. The Telemed ultrasound scanner emits nanosecond-level synchronisation pulses, which are converted to millisecond-level pulses using an Articulate Instruments Pulse Stretch unit. Each ultrasound frame is tagged with the corresponding sychronisation time, providing frame-level accuracy in temporal synchronisation between audio and ultrasound signals. The ultrasound system in this set-up uses a 64-element microconvex probe with 20 mm radius and 2 MHz frequency.

In our experiments, the acoustic speech signal is recorded using a Beyerdynamic Opus 55 omnidirectional microphone attached to the ultrasound headset, which is pre-amplified using a Grace m101 pre-amplifier. The signal is then split from the pre-amplifier's two outputs and recorded simultaneously onto both the ultrasound computer (via an Adlink DAQ-2213 analog-digital converter) and EMA computer (via an Alesis iO2 audio interface).

We stabilise the ultrasound probe relative to the speaker's head using the Ultrafit headset [10] produced by Articulate Instruments, which is entirely made out of plastic and non-metallic fabric (including replacement nylon bolts). The headset must be non-metallic for simultaneous EMA data collection, as metallic objects in or near the magnetic field significantly degrade the accuracy of EMA's position estimation.

2.2. Synchronisation

Synchronisation between EMA and ultrasound is achieved using synchronisation pulse signals from each system, which are already used for audio-ultrasound and audio-EMA synchronisation respectively. First of all, each system synchronises data separately at the time of recording, giving us a set of synchronised audio-ultrasound recordings and a set of synchronised audio-EMA recordings. Note that the audio-ultrasound and audio-EMA data are not synchronised with each other at this point, as there will be slight variation in the onset of recording and clock times between the two systems.

In order to precisely synchronise the EMA and ultrasound data, we record the AG501's TTL synchronisation pulse onto the computer recording ultrasound data as an extra audio track alongside the acoustic data and ultrasound sync signal. The EMA sync signal goes low 80 ms before recording begins and then sends a pulse every 4ms. This means we are able to align the edge of the synchronisation pulse that marks the onset of the EMA recording with the beginning of the EMA position file. As we describe below, all of the above procedures are carried out by the Articulate Assistant Advanced software using the Ultrasound and EMA modules, which only require the user to specify the relevant synchronisation channels and all synchronisation is then handled without further intervention from the user.

2.3. Software

This set-up uses the Articulate Assistant Advanced (AAA) software [11]. The software has a number of features that facilitate a simultaneous ultrasound-EMA experiment, which are as follows. First, it allows a network connection to the Carstens AG501's control server, which means that AAA can handle prompt presentation and remotely trigger the AG501 recording, meaning that the experimenter only has to operate one computer while recording data. Second, it automatically synchronises the ultrasound and audio data. Third, it allows EMA position data recorded using the AG501 to be imported and also keeps a record of which AAA recording corresponds to which EMA file. Finally, it automatically aligns the recorded EMA data with the audio-ultrasound, based on the EMA synchronisation pulse that is recorded during the experiment (as detailed in Section 2.2). This allows us to obtain synchronised audio, high-speed ultrasound and EMA data without any additional pre-processing steps.

2.4. Validation experiments

In order to examine the accuracy and validity of our system, we report the results of two controlled tests. The first examines the effects of the ultrasound hardware and headset on the accuracy of EMA tracking. The second aims to quantify the amount of ultrasound probe movement relative to the head over the course of a typical EMA session using the UltraFit headset. The data and specific approach used for each experiment are described in the respective sections that follow.

3. EFFECTS OF ULTRASOUND ON EMA ACCURACY

We first test the effects of recording simultaneous ultrasound and EMA on the accuracy of the EMA position calculation. While we anticipate that

the plastic ultrasound headset should not influence EMA tracking, we are nonetheless required to keep the ultrasound scanner placed 60 cm behind the centre of the tracking field in order appropriately position the probe on the speaker, so we test whether this nearby hardware has any effect on the accuracy of EMA tracking. In order to test this, we recorded six calibration sessions (three with ultrasound, three without ultrasound) in order to factor out any random differences due to the effect of calibration session, although calibration statistics were near identical within each set of three runs. The ultrasound condition involves the ultrasound probe and headset inside the EMA field, with the scanner hardware placed on the rear of the EMA stand, approximately 60 cm from the centre of the field. Calibration involves the rotation of EMA sensors placed on a purpose-built 'circal' device that moves the EMA sensors around a series of pre-determined rotations, allowing us to compare the ideal positions with the actual positions.

condition	Δz (sd)	σz (sd)
no ultrasound	0.661 (0.086)	0.148 (0.025)
with ultrasound	0.670 (0.090)	0.150 (0.026)

Table 1: Δz and σz statistics from calibration runs without ultrasound equipment and with ultrasound equipment placed in the EMA field. Values represent averages across 16 sensor channels for three different runs per condition, with standard deviations in brackets.

We evaluate the accuracy of calibration using two measures: (1) Δz, which is the peak-to-peak deviation of the z-coordinates from each channel during calibration; and (2) σz, the associated standard deviation. In both cases, smaller numbers represents more accurate calibration, and σz should remain below 0.25 mm. Table 1 shows only very minor differences between conditions, suggesting that the effect of the EMA hardware is near-zero and certainly too small to have a noticeable effect on measurements. In addition to this, all channels showed σz values between 0.11–0.21 mm, which are below the recommended threshold of 0.25 mm for accurate calibration.

4. TRACKING ULTRASOUND PROBE MOTION

We also evaluated translation and rotation of the ultrasound probe during a real-world experiment of vowels in British English. We recorded six female speakers of northern British English producing a large number of phrases of the form 'she says X', 'she says X dearly' and 'she says X eagerly', where X was a bV or bVd word. 58 phrases were produced in a block and each block was repeated 5 times, producing 1740 tokens across the 6 speakers. One speaker (F05) had to have the ultrasound probe re-positioned after 1 block due to improper fitting and subsequently recorded the full 5 blocks after the probe was moved; we only use the final 5 blocks for this speaker. A standard set of EMA sensors was used for each speaker and we corrected for head motion using two sensors behind the ears, one on the bridge of the nose, and one on the upper incisors.

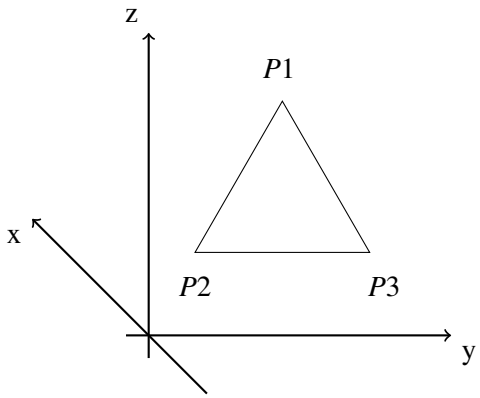

Figure 1: Schematic of rigid triangle attached to the ultrasound headset for tracking probe movement. $\{P1, P2, P3\}$ represent the corners at which EMA sensors were attached. The axes are oriented from the speaker's perspective, with $P2$ located on the speaker's left-hand side and $P3$ located on the speaker's right-hand side. In this coordinate system, x represents anterio-posterior motion, y represents lateral motion, and z represents vertical motion.

We tracked probe motion using a plastic triangle attached to the ultrasound headset, with EMA sensors fixed to each corner of the triangle. The triangle was custom designed and 3D printed to be retrofitted to the Articulate Assistant UltraFit headset. The STL file for the 3D printed triangle is available at the following link: https://github.com/samkirkham/ultrafit-ema. The triangle screws onto the headset using existing probe holder bolts and a small notch on the rear of the triangle prevents rotation around the bolt. We label the sensors on the triangle as $\{P1, P2, P3\}$, each of which has the (x, y, z) coordinates visualised in Figure 1. Note that any references to left-right orientation are from the perspective of the speaker.

As the plastic triangle is displaced from the actual ultrasound probe origin, we measured the distance in x and z dimensions between the top sensor on the headset triangle ($P1$) and the centre of the ultrasound

transducer array, and subtracted these values from the sensor coordinates. We then defined a virtual origin at the probe (reference object) and calculated the rotations required to align the reference object with the measured probe sensors. This was achieved by computing the rotation matrix that would align the position of the probe with the reference position for each data sample, which gives us the rotations around the respective x, y, z axes. We implemented this procedure in Python using the SciPy [12] spatial.transform sub-module, whereby we calculate the rotation between vectors using the algorithm in [13]. The final output is the displacement of the ultrasound probe in x, y, z dimensions, as well as the rotations around these dimensions: roll, pitch and yaw. For comparability with previous research [8, 9], we report the standard deviation of probe displacement and rotation across the experimental session for each speaker, as well as the 95% variance interval (1.98 x SD) for each measurement (note that for the rotations we calculated circular standard deviations). These values are shown in Table 2.

speaker	x	y	z	roll	pitch	yaw
F01	1.26	0.29	0.80	0.22	0.61	0.29
F02	1.39	0.82	0.86	0.68	0.93	0.39
F03	1.76	1.12	1.29	0.91	1.09	0.41
F04	1.21	0.36	0.70	0.38	0.71	1.00
F05	1.49	1.85	1.78	1.39	0.59	0.52
F06	0.66	0.73	0.80	0.52	0.34	0.89
mean	1.29	0.86	1.04	0.68	0.71	0.58
95%	2.56	1.71	2.06	1.35	1.41	1.15

Table 2: Standard deviation of displacement (mm) and circular standard deviation of rotation (degrees) of the ultrasound probe across all recordings for each speaker.

Table 2 shows mean translation of 0.86–1.29 mm and mean rotations of 0.58–0.71°, with maximum 95% variance values of 2.56 mm and 1.41°. This is comparable to the metal Articulate Instruments headset, which shows up to 1.8 mm translation and 1.3° rotation [2], and is well within the HOCUS tolerances of 2–4mm translation and 5–7° rotation [3]. Our statistics are also highly comparable to those in [8, 9], who report maximum values of 2.5 mm and 1.25°. We note, however, that our experimental materials used in this test are phonetically very similar to one another, so it would be worthwhile to conduct comparable tests using more phonetically diverse materials, which could potentially induce a wider range of probe movements.

5. CONCLUSIONS

We reported a method for simultaneous EMA-ultrasound data collection using widely-available hardware and software. Our system shows very high accuracy in temporal synchronisation, with minimal-to-no influence of the ultrasound scanner on EMA tracking accuracy, and ultrasound probe movement is well within accepted tolerances. This set-up can be implemented easily by other researchers who have access to the same commercial hardware and software, meaning that the experimental set-up can be replicated across different laboratories.

In future research, we aim to (1) use the known rotations and translations from this analysis to correct the ultrasound splines for probe movement, thereby locating both modalities in a common reference frame; (2) evaluate the comparability of articulatory timing measures based on EMA versus ultrasound data.

ACKNOWLEDGEMENTS

This research was funded by AHRC grant AH/S011900/1.

6. REFERENCES

[1] M. Stone and E. P. Davis, “A head and transducer support system for making ultrasound images of tongue/jaw movement,” *Journal of the Acoustical Society of America*, vol. 98, no. 6, pp. 3107–3112, 1995.

[2] J. M. Scobbie, A. A. Wrench, and M. L. van der Linden, “Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement,” *Proceedings of the 8th International Seminar on Speech Production*, pp. 373–376, 2008.

[3] D. Whalen, K. Iskarous, M. K. Tiede, D. J. Ostry, H. Lehnert-LeHouilier, and E. Vatikiotis-Bateson, “The Haskins Optically Corrected Ultrasound System (HOCUS),” *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 3, pp. 543–553, 2005.

[4] P. Strycharczuk, D. Derrick, and J. A. Shaw, “Locating de-lateralization in the pathway of sound changes affecting coda /l/,” *Laboratory Phonology*, vol. 11, no. 1, p. 21, 2020.

[5] J. Ying, J. A. Shaw, C. Carignan, M. Proctor, D. Derrick, and C. T. Best, “Evidence for active control of tongue lateralization in Australian English /l/,” *Journal of Phonetics*, vol. 86, no. 101039, pp. 1–22, 2021.

[6] S. Marin and M. Pouplier, “Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling

model," *Motor Control*, vol. 14, no. 3, pp. 380–407, 2010.

[7] P. Strycharczuk and J. M. Scobbie, "Velocity measures in ultrasound data: Gestural timing of post-vocalic/l/ in English," *Proceedings of the XVIII International Congress of Phonetic Sciences*, pp. 1–5, 2015.

[8] D. Derrick, C. T. Best, and R. Fiasson, "Non-metallic ultrasound probe holder for co-collection and co-registration with EMA," *Proceedings of the 19th International Congress of Phonetic Sciences*, pp. 1–5, 2015.

[9] D. Derrick, C. Carignan, W.-R. Chen, M. Shujau, and C. T. Best, "Three-dimensional printable ultrasound transducer stabilization system," *Journal of the Acoustical Society of America*, vol. 144, no. 5, pp. EL392–EL398, 2018.

[10] L. Spreafico, M. Pucher, and A. Matosova, "UltraFit: A speaker-friendly headset for ultrasound recordings in speech science," *Proceedings of Interspeech 2018*, pp. 1–4, 2018.

[11] A. Wrench, "Articulate Assistant Advanced, v. 220.04," 2022.

[12] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[13] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica*, vol. A32, no. 5, pp. 922–923, 1976.