

UC Berkeley

UC Berkeley PhonLab Annual Report

Title

Head correction of point tracking data

Permalink

<https://escholarship.org/uc/item/28b3c9fx>

Journal

UC Berkeley PhonLab Annual Report, 15(1)

Authors

Johnson, Keith
Sprouse, Ronald L.

Publication Date

2019

DOI

10.5070/P7151050341

Copyright Information

Copyright 2019 by the author(s). All rights reserved unless otherwise indicated. Contact the author(s) for any necessary permissions. Learn more at <https://escholarship.org/terms>

Peer reviewed

Head correction of point tracking data

Keith Johnson & Ronald L. Sprouse

Department of Linguistics
University of California Berkeley
USA

keithjohnson@berkeley.edu
ronald@berkeley.edu

Abstract

This is a short paper comparing two approaches to head correction for Electro-Magnetic Articulography (EMA) data collected with the Northern Digital Instruments “Wave” system. In both of these approaches, it is necessary to translate and rotate the sensor locations to the occlusal coordinate system. We found that point tracking error is greater by as much as double with the built-in NDI head correction method, compared to a three-sensor head correction algorithm. However, we conclude that the data are comparable, and that the two-sensor NDI method is acceptable for phonetic research. A Python library for head correction was developed for this work, and is available on github.com.

Key Words: Articulatory Phonetics, EMA, ElectroMagnetic Articulography, head correction

1. Introduction

In electromagnetic articulography (EMA; Perkell, et al., 1992; Schönle et al., 1987), the Northern Digital Instruments (NDI) Wave System can use a 6 degree of freedom (x,y,z, pitch, roll, yaw) sensor to reference other sensors relative to the location of the head. The 6D reference sensor is actually two sensors, using two data channels, fused in a known configuration with each other so the location and orientation of the head can be fully specified. This built-in referencing system allows head movement to be removed from the articulatory point-tracking data as a part of the recording process (Ji et al., 2014; Wieling et al., 2016). With a proper translation and rotation, the sensor data can be transformed to the maxillary occlusal coordinate system that was recommended by Westbury (1991, 1994).

Other researchers (Shaw & Kawahara, 2018; Hoole & Zierdt, 2009; Tilsen, Das & McKee, 2014) use a configuration of three sensors (e.g. nasion, left and right mastoid) to register the current position of the head in each frame of data, and then translate and rotate the data from each frame so that the position of the head matches an ideal head position, again placing the sensor data onto the maxillary occlusal coordinate system.

This paper reports a small study comparing these two methods for head correction in EMA data.

2. Method

Five sensors were attached with medical tape to the skin of a speaker: One on the chin to track jaw movement, one on each of the left and right mastoid processes behind the ears, and two on the nasion (a 6D reference sensor, and also a standard 5D sensor). In addition, to register the occlusal plane, the subject bit into warmed dental impression wax and then two sensors were placed on the mid-sagittal plane in the wax - one just anterior to the incisors, and one on the mid-sagittal plane between the second molars.

Figure 1 illustrates the placement of the sensors. The three dots outlined in green mark the approximate locations of the nasion, and left and right mastoid sensors. The two dots outlined in red mark the approximate locations of the wax bite plate sensors. The dot outlined in blue marks the location of the jaw movement sensor. The 6D reference sensor was located just below the nasion sensor.

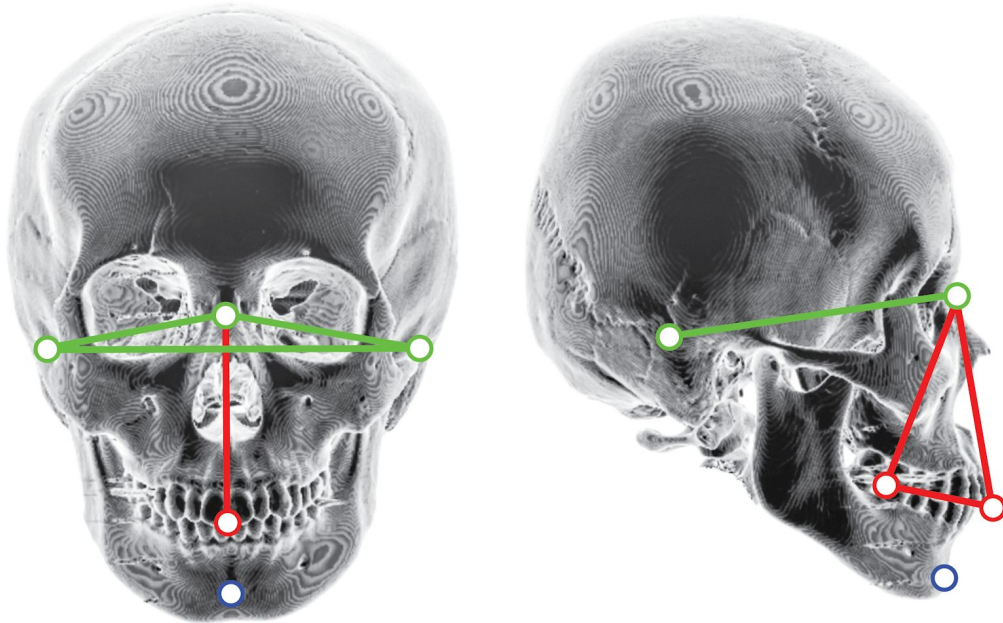


Figure 1. *White dots mark the approximate locations of the sensors, and the lines connecting the dots indicate two triangles that were used in head correction. Red lines mark a calibration triangle in the mid-sagittal plane, the base of which defines the occlusal plane. Green lines mark head location during data collection in the three-point correction method (after an image by Puwadol Jaturawutthichai/shutterstock.com).*

Three recordings were taken with each type of head correction (NDI-referenced, and 3-point-referenced): (a) a recording with the wax bite plate held in the mouth to register the maxillary occlusal plane, (b) a mid-sagittal trace of the palate from the front incisors to the soft palate, and (c) a longer recording in which the subject simultaneously moved his head and opened and shut his jaw. During this last recording, three kinds of large head movements were made - up and down 'head nodding', side to side 'head turning', and a combination of nodding and turning forming a circular form of 'neck roll'.

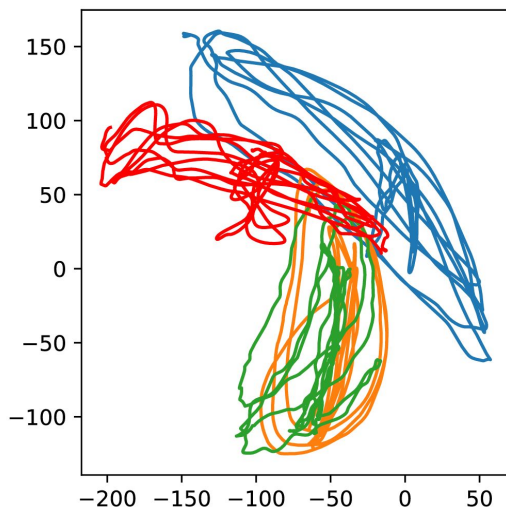


Figure 2. Raw data during a head movement trial with jaw movement. Blue = nasion sensor; Red = jaw sensor, Green = left mastoid, Gold = right mastoid sensor.

3. Transformation

From the triangle of three mid-sagittal sensors during the ‘bite-plate’ recording (nasion, and two bite plate sensors; the red triangle in Figure 1), a rotation and translation matrix was defined to transform the data onto the occlusal coordinate system (see Ji et al., 2013 for an alternative, quaternion-based approach). We can call the sensor at the front incisors *OS*, the sensor on the wax plate between the molars *MS*, and the nasion sensor *NS*. The average location of these sensors in 3D space $[x,y,z]$ was taken over a duration of about five seconds.

Subtracting the location of the origin sensor *OS* from all sensor data points translates the data into a coordinate system in which *OS* is at the origin $[0,0,0]$. The position vector of *OS* is therefore the translation matrix for all sensors that we wish to put on the maxillary occlusal coordinate system. The translation of the molar and nasion sensors into *M* and *N* is shown in (1) and (2).

The rotation matrix is created by finding a set of axes that rotate the coordinate system into a desired reference orientation. We start by selecting the vectors *M* and *N* which meet at the origin and define a plane, which we take to be the mid-sagittal (*xy*) plane in the reference space. The *z* axis must be perpendicular to these vectors and is found by taking the normalized cross-product of them (3). The vectors *M* and *z* define the maxillary occlusal (*xz*) plane, and the *y* axis is found by taking the normalized cross-product of them (4). Finally, the *x* axis is perpendicular to the *z* and *y* axes and is also found by taking the cross-product (5). These vectors $[x, y, z]$ define a rotation matrix (6) that, together with the translation matrix *OS*, will put the data (*d*) onto the maxillary occlusal coordinate system (7). Formula (7) is the “occlusal transformation”.

- 1) $M = MS - OS$
- 2) $N = NS - OS$
- 3) $z = (N \times M) / \|N \times M\|$
- 4) $y = (M \times z) / \|M \times z\|$
- 5) $x = (y \times z) / \|y \times z\|$
- 6) $m = [x, y, z]$
- 7) $d_o = (d - OS) \cdot m^T$

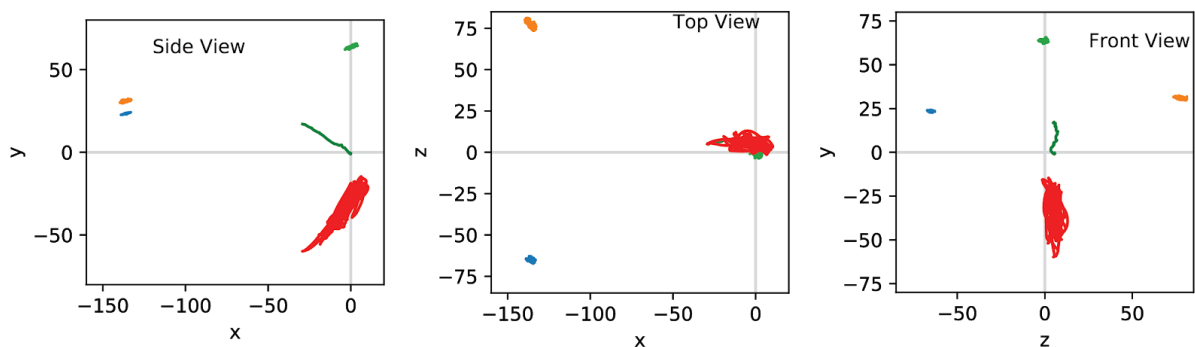
In this coordinate system, the tip of the incisors is located at the origin, the mid-sagittal plane is at $z=0$, and the maxillary occlusal plane is at $y=0$. Occlusal transformation matrices were derived from both the NDI-referenced and 3-point-referenced bite plate recordings. For the NDI-referenced data, the occlusal transformation was applied to the palate trace and head movement recordings. For the 3-point-referenced data, the locations of the head position sensors (a triangle formed of the nasion and left and right mastoid; the green triangle in Figure 1) during the bite plate recording were rotated to define an ‘ideal’ head position - the position of the three head sensors in the occlusal coordinate system. Then for the palate and head movement recordings the frame-by-frame quaternions that rotate the locations of the head position sensors to that ideal position were found with procrustean fitting using the Davenport Q method (Keat, 1977; Horn, 1987). These quaternions and the translation defined by OS were applied to all sensor data in each frame to transform into the desired reference orientation.

Trajectories were smoothed with the Garcia (2010) robust smoothing algorithm.

4. Results

Figure 3 shows a comparison of the corrected data for two separate head movement recordings. The raw data from the first of these recordings was shown earlier in figure 2. If head movement is perfectly removed then the head location sensors (nasion, left and right mastoid) should be completely stationary. Clearly both head correction methods remove head movement while showing movement of the jaw. But there is also an apparent difference between the two methods, such that head movement was more completely removed using the 3-point head correction method (notice the size difference in the green, gold and blue clouds).

(a) 3-point head correction



(b) NDI built-in head correction

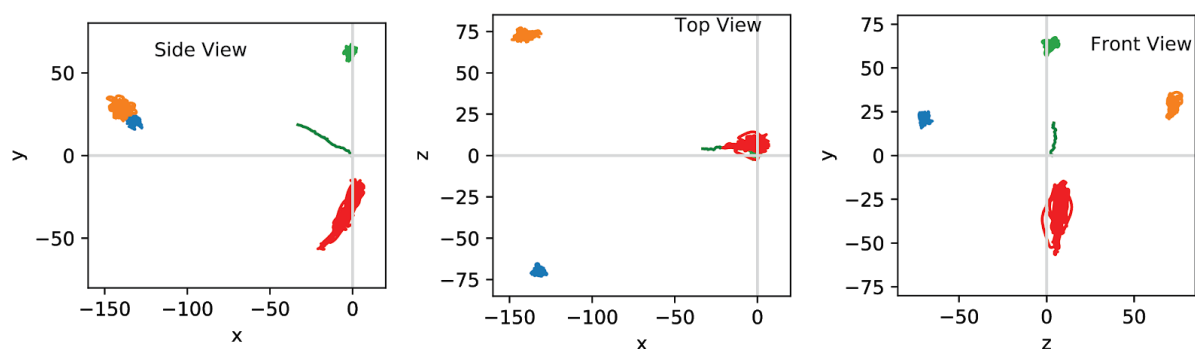


Figure 3. Head-corrected data from two separate recordings with extreme head movement, and simultaneous jaw wagging. In (a), the top row, the data from one recording were corrected using the 3-point head correction methods. In (b), the bottom row, the data from a different recording were corrected by the built-in NDI 6D reference sensor correction system. In each row the left panel shows a sagittal view, the middle panel shows an axial view, and the right panel shows a coronal view.

The amount of sensor movement in the ‘head movement’ recording was measured as the square root of the largest eigenvalue of the 3D location data for each sensor. This is the standard deviation of the first principal component of sensor location. If the head correction algorithm was perfect, then the eigenvalues of the nasion, and right and left mastoid sensors would be equal to zero. Table 1 shows the magnitude of sensor movement for the three head sensors and the jaw in two separate head movement recordings. The first recording (the first two columns in the table) was done without the built-in NDI head correction. The raw data were shown in figure 2, and the corrected data were shown in figure 3a. The two rightmost columns of the table report measurements from a recording that was done with the built-in NDI referencing system. The “raw” data, after translation and rotation, were shown in figure 3b. These data were then additionally submitted to the 3-point head correction algorithm and the resulting error is shown in the last column of the table. The right mastoid sensor is furthest from the EM sending coils, which explains the slightly larger error for this sensor. In general, the magnitude of the errors found here are comparable with the findings of previous studies of the NDI Wave system (Berry, 2011; Savariaux et al., 2017).

Table 1. Standard deviation of the first principal component (in mm) of sensor position with two methods of head correction. Measurements in the first two columns were made from a recording that was made without the NDI global referencing. Measurements in the last two columns are from a recording that was made with the built-in NDI head correction.

	no built-in referencing		with built-in NDI referencing	
	raw data	3-point correction	raw data	3-point correction
nasion	82.1	1.33	1.46	1.18
right mastoid	47.2	1.56	2.85	1.37
left mastoid	55.7	0.75	1.72	0.75
jaw	57.7	11.6	13.2	13.4

Figure 4 shows two separately recorded traces of the speaker’s hard palate, again with the two methods of head correction. It is reassuring to see that two recordings of the same physical structure, using the two different methods of head correction, give a highly comparable result. The small discrepancy in these recordings is attributable to a slightly more lateral placement of the palate trace probe in the NDI global referenced recording.

In a further comparison of the phonetic utility of the two methods of head correction, figure 5 shows jaw movement from a single recording, where the data were corrected using the two head correction methods. The two processing streams (translation and rotation of data collected with built-in NDI head referencing, and head position correction using the 3-point referencing method) produce almost identical results. Only at a couple of points during the very large side-to-side head movements do the two traces deviate from each other.

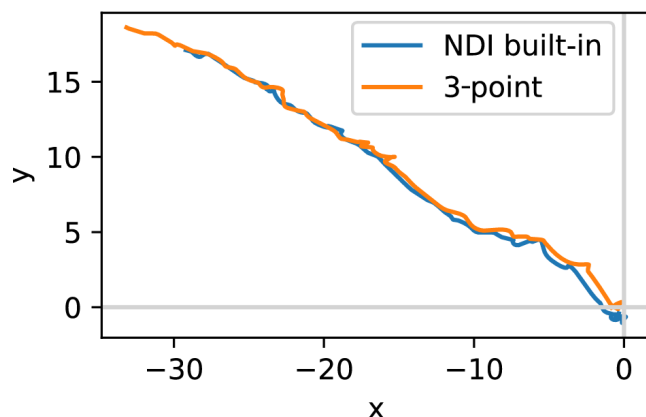


Figure 4. Two mid-sagittal traces of the hard palate. Separate recordings with two methods of head correction.

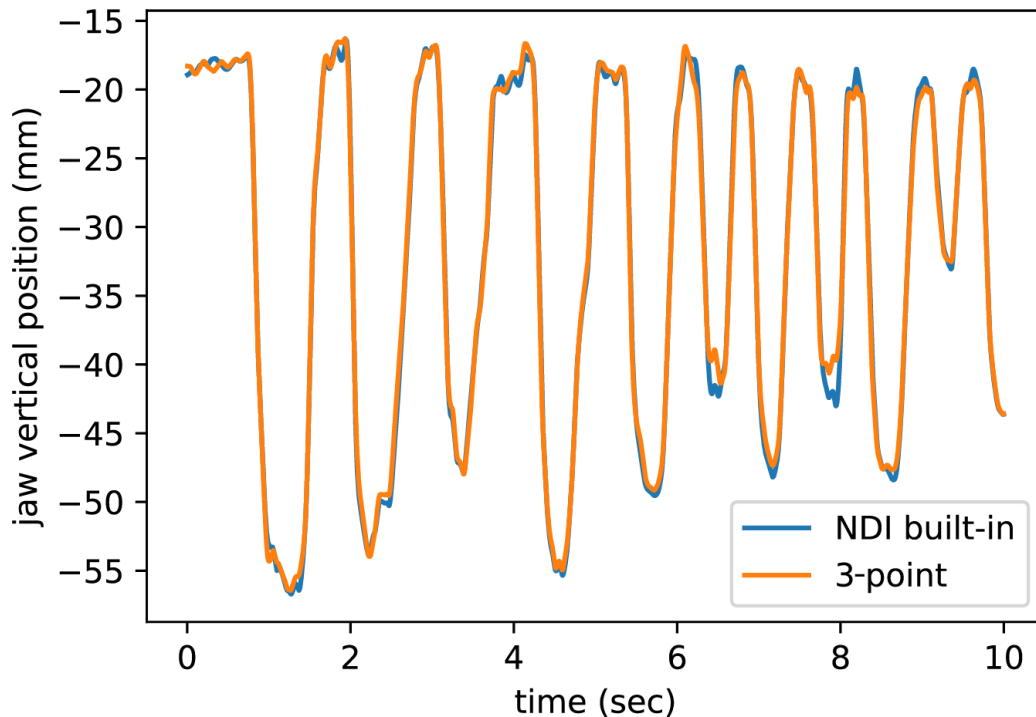


Figure 5. Jaw vertical (y) movement from a single recording that was head-corrected in two different ways.

5. Conclusion

For labs who have enough sensor channels to spend three of them referencing the location of the head with three sensors located at different non-articulating points, head movement is more completely removed by 3-point head correction. However, the two channel NDI-referenced head correction is quite good, and labs with an 8 channel system who do not want to spend an extra channel on 3-point head correction will get accurate, phonetically useful data from the built-in method. A Python library to remove head movement and put data into the occlusal coordinate system is available at <https://github.com/rsprouse/ematools>.

Acknowledgements

Thanks to Jason Shaw for fruitful discussion of EMA methods and giving the initial push on this project. We learned and benefited from software published by Andrew Kolb: <https://github.com/andrewkolb08>. Thanks to GENG Puyang and Cassandra Serrano for their help in the lab. Both authors worked on the analysis code. The data analysis and first draft of the paper were done by KJ, and both authors revised and corrected the manuscript.

References

Berry, J.J. (2011) Accuracy of the NDI wave speech research system. *J Speech Lang Hear Res.* 54(5):1295–1301. doi: 10.1044/1092-4388(2011/10-0226)

- Garcia, D. (2010) Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis* **54**, 1167-1178. doi: 10.1016/j.csda.2009.09.020
- Hoole, P & Zierdt, A. (2009) Five-dimensional articulography. In Maassen, B. & van Lieshout, P. (Eds.), *Speech Motor Control: New Developments in Basic and Applied Research*(pp.331-350). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199235797.003.0020
- Horn, B.K.P. (1987) Closed-form solution of absolute orientation using unit quaternions, *Journal of the Optical Society of America A* **4**(4) 629–642 (1987). doi: 10.1364/JOSAA.4.000629
- Ji, A., Berry, J.J. & Johnson, M.T. (2014) The Electromagnetic Articulography Mandarin Accented English (EMA-MAE) corpus of acoustic and 3D articulatory kinematic data," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, pp. 7719-7723. doi: 10.1109/ICASSP.2014.6855102
- Ji, A., Johnson, M.T., & Berry, J.J. (2013) Articulatory space calibration in 3D electro-magnetic articulography. In *China SIP International Conference on Signal and Image Processing*.
- Keat, J. (1977) *Analysis of Least-Squares Attitude Determination Routine DOAOP*, CSC Report CSC/TM-77/6034.
- Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I., & Jackson, M.T. (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *J. Acoust. Soc. Am.* **92**(6):3078–3096. doi: 10.1121/1.404204
- Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017) A Comparative Study of the Precision of Carstens and Northern Digital Instruments electromagnetic articulographs. *J. Speech. Lang. & Hear. Res.* **60**(2), 322-340. doi: 10.1044/2016_JSLHR-S-15-0223
- Schönle, P.W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987) Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain Lang.* **31**(1):26–35. doi: 10.1016/0093-934X(87)90058-7
- Shaw, J.A., & Kawahara, S. (2018) The lingual articulation of devoiced /u/ in Tokyo Japanese. *J Phon* **66**, 100-119. doi: /10.1016/j.wocn.2017.09.007
- Tilsen, S., Das, D., & McKee, B. (2014). Real-time articulatory biofeedback with electromagnetic articulography. *Linguistics Vanguard*. doi: 10.1515/lingvan-2014-1006.
- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Broker, F., Thiele, S., Wood, S.N., & Baayen, R.H. (2016) Investigating dialectal differences using articulography. *J Phon* **59**, 122-143. doi: 10.1016/j.wocn.2016.09.004
- Westbury, J.R. (1991) The significance and measurement of head position during speech production experiments using the x-ray microbeam system. *J. Acoust. Soc. Am.* **89**, 1782. doi: 10.1121/1.401012
- Westbury, J.R. (1994) On coordinate systems and the representation of articulatory movements. *J Acoust Soc Am.* **95**(4):2271–2273. doi: 10.1121/1.408638