INTERSPEECH 2009 BRIGHTON

Using sensor orientation information for computational head stabilisation in 3D Electromagnetic Articulography (EMA)

Christian Kroos

MARCS Auditory Laboratories, University of Western Sydney, Australia

c.kroos@uws.edu.au

Abstract

We propose a new simple algorithm to make use of the sensor orientation information in 3D Electromagnetic Articulography (EMA) for computational head stabilisation. The algorithm also provides a well-defined procedure in the case where only two sensors are available for head motion tracking and allows for the combining of position coordinates and orientation angles for head stabilisation with an equal weighting of each kind of information. An evaluation showed that the method using the orientation angles produced the most reliable results.

Index Terms: Electromagnetic Articulography, head motion, head stabilisation, orientation angles

1. Introduction

Spoken language is produced by humans through the movements of the speech articulators (tongue, lips, velum, etc.) In order to investigate the properties of these movements they have to be accurately measured first. Among the currently used measurement methods such as Cineradiography, Electropalatography, structural Magnetic Resonance Imaging and ultrasound, only Electromagnetic Articulography (EMA) measures tongue movements in real time and returns data in the form of tongue surface flesh points. EMA began almost forty years ago with very limited 2D systems [1, 2] that were substantially improved and refined during the next decades [3, 4] until a 3D system was developed [5] in the late 1990s. Currently two commercial systems are available: the *AG500* (Carstens Medizinelektronik) and the *Wave Speech Research* system (Northern Digital). In this paper we will refer specifically to the Carstens AG500 though the considerations generally also apply to the Wave Speech Research system.

One of the many advantages of the 3D-EMA system over the 2D version is that the speaker's head movements are not constrained in any way during the recording. Head motion has been shown to accompany speech production in a meaningful way, and indeed conveys prosodic information about the utterance being spoken [6], making the extraction of head motion parameters from an articulatory recording a worthwhile undertaking. However, separating the contribution of head movements and articulatory movements from the combined signal returned by the EMA system is a far more important issue, since movements of the articulators are only meaningful within a head-centred coordinate system. To obtain true measures of articulator movements the head must be computationally stabilised post-hoc.

Head motion can be treated as rigid-body motion, that is, fully described by three rotational parameters, determining the rotation of the rigid body object around the major axes of a 3D Cartesian coordinate system (pitch, roll, and yaw), and three translational parameters, determining the translation along those axes. Computational head stabilisation is usually accomplished by applying a 3D pose estimation method sample-wise to the sensors that were designated to track head motions. The pose estimation algorithm finds the transformation parameters that would transform the head position of the sample to an arbitrary reference position which has to be chosen by the experimenter. Finding the transformation parameters is usually achieved by minimising the sum of the squared distances between the head tracking sensors in the actual position and the reference position with some kind of linear least squares approach. The resulting sequence of transformation parameter values can then be used to rotate and translate all sensors inversely and arrive at head-stabilised articulatory trajectories. Pose estimation algorithms have been employed in many different research fields, and are typically based on the Orthogonal Procrustes Rotation [7, 8].

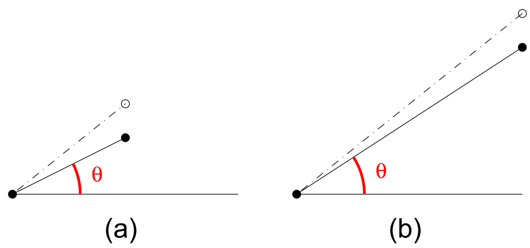

Figure 1: *Placing the head tracking sensors far from each other improves the estimation precision of rotational parameters (if the measurement accuracy is constant) shown here for a 2D example. The angle between the horizontal and the line through the true position of the second sensor (unfilled circle) is the same in (a) and (b),* 38.7°*, as is the vertical difference between true and measured position of the second sensors (filled circle). The measurement noise leads to an estimation of* θ *in (a) of* 26.7°*, while in (b), with the sensor twice as far away, of* 33.0°*.*

Relying on sensors that return only position coordinates - as most optical motion capture system do - requires at least three head tracking markers that must not lie on a straight line, i.e., they must form a triangle. For most applications (including head stabilisation) the centre of rotation (pivot point) can be assumed to be the centroid (centre of gravity) of the object. As uniform mass can be assumed, this is simply the mean of the sensor coordinates. Accordingly the translation parameters are computed as the difference between the position of the sensor centroids of the actual and the reference position *after* rotation. Importantly, the rotation parameters do not depend on the position of the pivot point (assumed or real), while the translation parameters always do. As a consequence, the translational head motion parameters obtained with the pivot point at the sensors' mean are not accurate, since it does not constitute the real pivot point. In fact, for head movements, a single point that acts as the centre of rotation does not exist, as movements are not accomplished by a single joint but by the combined action of eight joints of complex geometry of the upper and lower cervical spine (see [9], page 326-336). The contribution of any of these joints to the overall head-neck movement varies with the kind and amount of the executed movement. Note, however, that whole torso movements by the participant will always result in spurious head movement values, both translational and rotational. None of these issues affect head stabilisation as it does not require the 'true' centre of rotation, and henceforth the pivot point is assumed to coincide with the centroid.

Unlike most other motion capture systems, 3D-EMA also returns two of the three orientation angles of the sensors, given as azimuth (ϕ) and elevation (θ). The orientation information can be used to determine the rotational parameters (but not the translational ones) and there are good reasons to do so as becomes clear by examining the properties of the position coordinates-based (PCM) and the orientation angle-based (OAM) head rotation determination methods.

The accuracy with which the rotation parameters can be calculated using PCM depends on the distance of the head tracking sensors from each other in relation to the measurement noise of the system (see Figure 1), i.e., they need to be spread out. Since they are required to be attached to parts of the head that move as little as possible due to skin movements typical locations used consist of the upper jaw, the nasion and behind the ear lobes. 3D-EMA is known to be more accurate near the centre of the measurement field than the periphery; for the AG500 the manufacturer recommends to stay within a sphere of 15cm around the centre. Keeping the important tongue sensors close to the centre, means that the head tracking sensors approach the recommended limit and with more than negligible head motion can easily leave the 15cm sphere. In an analogous way, the accuracy with which the rotation parameters can be computed using OAM depends on the differences between the orientation angles among the head tracking sensors, i.e., they need to be spread out over the orientation angle space. Just as straight line configuration of all sensors would result in an error for PCM, if all orientation angles were the same, it would lead to a break down in OAM. However, they do not need to be positioned far from each other in terms of location on the participant's head, as the distance has no influence on the precision.

To our knowledge, there is currently only one approach that incorporates information from the orientation angles - the Matlab (The Mathworks, Inc) routine `rigidbodyana` which is part of the Matlab-based EMA processing software developed at the IPSK, Munich, Germany, by Phil Hoole. The routine is based on a pure PCM routine implemented by Christian Geng using the Gower Procrustes algorithm [8]. It was designed to allow the computation of head motion parameters from only two head tracking sensors, a situation that forces the use of orientation information. In `rigidbodyana` a virtual sensor is defined using the orientation angle of one of the head tracking sensors and a distance value set by the user. The virtual sensor is subsequently treated as a legitimate third sensor allowing the use of a PCM. The free parameter, the distance of the virtual sensor to its 'mother' sensor, amounts to a weight of the orientation information relative to the position information. It is, however, not clear what constitute an optimal value, leaving the user to rely on rules of thumb.

We propose a new algorithm consisting of three simple steps to obtain rotational head motion parameters from the orientation angles. Per extension it also allows the combined use of position coordinates and orientation angles in an optimal way. In addition, the special case of determining head motion parameters from two sensors only is optimally solved within the framework of the algorithm.

2. Algorithm

We assume that head movements were measured with three or more sensors.

1. Augment the orientation angles to full spherical coordinates by setting the unspecified radius to 1 for both the actual head pose and the reference pose.
2. Convert the spherical coordinates into Cartesian coordinates. (Matlab provides the function `sph2cart` for this purpose).
3. Determine head rotation with PCM using the origin as the centre of rotation (see below).

Note that no global centre of rotation needs to be specified, reflecting the fact mentioned above that the rotation parameters are independent of the choice for the pivot point. For the last step we implemented the least squares-based Orthogonal Procrustes algorithm for pose estimation [7, 8]. There is, however, one notable difference compared to ordinary pose estimation: For the orientation angles-turned-Cartesian coordinates no translation values are to be determined. As they lie on a unit sphere it is evident that rotation around the origin is sufficient to match the actual pose with the reference pose. The procedure returns the rotation and translation parameters (the latter, in this case, all zeros) in the form of a 4×4 transformation matrix with reference to homogeneous coordinates. The original position coordinates are then multiplied with the transformation matrix, rotating it to match the orientation of the reference pose. After that the centroid is calculated for both poses and the difference between the two centroids yields the translation parameters.

Should PCM and OAM combined (hybrid method, HM) to use all available information the centroid is computed for both poses in their raw form and subtracted from the position coordinates. Each sensor position is then converted to spherical coordinates yielding azimuth and elevation angles and a radius which is not further used. The angles are simply added to the measured sensor orientation angles and the OAM described above is applied. In the two-sensor case a third azimuth and elevation angle is derived in a similar way as in the HM by shifting the origin to the position of one of the sensors, and then converting the position of the other sensor to spherical coordinates and proceeding with the OAM.

3. Evaluation

We were specifically interested in how the three approaches (PCM, OAM, HM) would perform in the case of moderate to small distances between the sensors - if successful, this would enable us to avoid having sensor locations at the periphery of the measurement field.

3.1. Method

For the evaluation we used a data set recorded previously for a study to assess the measurement precision of the AG500 3D-

Figure 2: *Inside of the custom-built container showing the three EMA calibration cartridges*

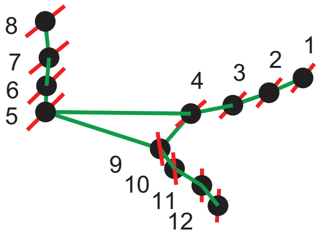

Figure 3: *Selected sensor triads: [4 5 9], [3 6 10], [2 7 11], and [1 8 12]. The 2-sensor sets consisted each of all combinations of the sensors in a triad, e.g, [4 5], [5 9], [4 9] formed the first 2-marker set. The results were averaged over all three combinations.*

EMA system [10]. As the data acquisition is described in detail in [10] we will give only a very brief overview here. A custom-built container (see Figure 2) fixed the positions of three EMA calibration cartridges with all 12 sensors inserted in its interior. In addition, 8 markers of the optical motions tracking system Vicon (Oxford Metrics) were placed on a protruding cross-shaped structure. A series of movements of the container within the EMA measurement field were recorded (each trial with 2500 samples at 200Hz) with both systems, consisting of predominantly translational (40 trials) or rotational movements (40 trials), or free movements without such constraints (20 trials). The movements were performed by the experimenter, with his hand extended into the cube.

In the three-sensor scenario we chose four different triads of sensors with the individual sensors coming from different cartridges (see Figure 3). Because of the arrangement of the cartridges, the orientation angles of the sensors were very similar from set to set, but the distances between the sensors changed substantially across the four sets. In the two-sensor scenario the sensor sets were chosen analogously (see Figure 3).

It may seem obvious to use the residual error of the pose estimation process (the remaining Euclidean distances between the sensor positions after head stabilisation and the sensor positions of the reference pose) as a measure of the quality of the head stabilisation. However, the residuals are not necessarily a good indicator of quality. As can be seen in Figure 4, during this interval of head movements OAM produces a larger residual error than the PCM, but the rotation parameter time sequence in the PCM shows clear discontinuities, which indicate a failure of the stabilisation. This seemingly counter-intuitive behaviour of the residual error is easy to explain. Mistracking of a single sensor can force the PCM to place the rigid-body object far away from its true path in order to minimise the residual error. Therefore one would also have to consider the less affected residual error of the sensor orientation angles. However, combining the two errors to arrive at a single number turns out to be difficult as they are scaled very differently. Accordingly we assessed the quality of the stabilisation by measuring the smoothness of the resulting motion parameters, using the integral on the squared second derivative which in our case of a sampled signal with a constant sampling rate can be computed as

$$R = \frac{1}{n-2} \sum_{t=2}^{n-1} (x_{t+1} - 2x_t + t_{t-1})^2 \quad \text{with } t \in \mathbb{Z} \qquad (1)$$

where x_t is the value of the motion parameters at sample t and n is the number of samples. The higher the value of R the lower is the smoothness being zero when applied to a constant signal. Since the values vary over a huge range and we preferred a measure of smoothness rather than roughness we used $S = -log(R)$ as the final measure. Because the stabilisation methods we compared are constrained by the least squares fitting process, unrealistic results that would nevertheless fare well with the smoothness measure, e.g., constant motion parameters, do not occur.

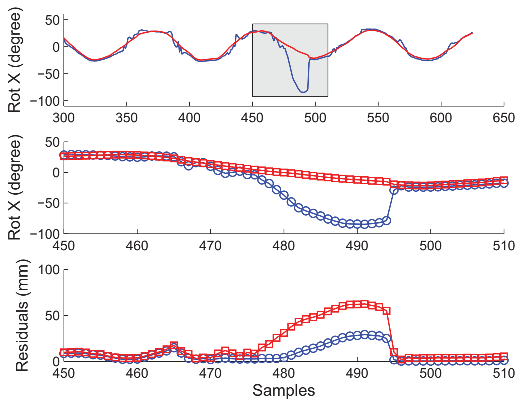

Figure 4: *Comparison of PCM (blue) and OAM (red) in the estimation of the rotation around the x-axis (top). Between sample 450 and 510 mistracking leads to an incorrect estimation of the motion parameter in the PCM (middle). However, this is inversely reflected in the magnitude of the residuals (bottom).*

We also calculated the RMS error between the resulting motion parameters from the stabilisation methods of the EMA data and the PCM of the Vicon data, the latter based in general on eight markers with a minimum threshold of four markers in the case of missing values.

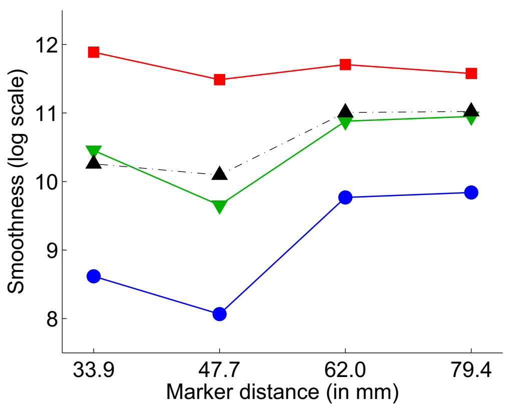

Figure 5: *Smoothness values for the three methods and four sensor distances averaged over all trials. PCM: blue solid line, circles; OAM: red solid line, squares; HM: green solid line, triangles; Two sensors: black dashed lines, triangles*

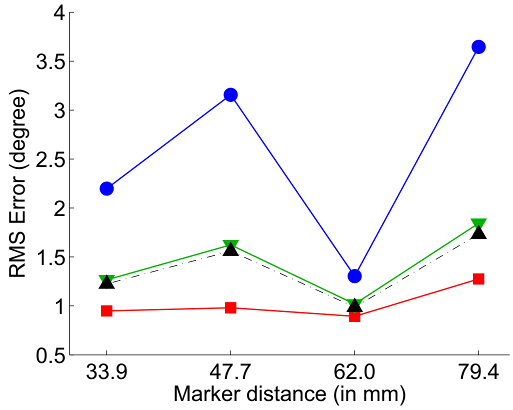

Figure 6: *RMS error for the three methods and four sensor distances averaged over all trials. PCM: blue solid line, circles; OAM: red solid line, squares; HM: green solid line, triangles; Two sensors: black dashed lines, triangles*

3.2. Results and Discussion

Figure 5 shows the smoothness values averaged over all trials for the three methods. Consistent with expectations, the OAM performance did not change over the different sensor distances, while the PCM and HM improved with larger distances. Within the range of sensor distances studied here, the OAM clearly outperformed the PCM at all distances, with the HM settling in between. Remarkably, there was no difference between the HM with three or two sensors. Figure 6 shows the RMS error of the EMA-based estimation with respect to the Vicon-based estimation. The results confirm the findings from the smoothness analysis though for the largest marker distance the expected improvement was not observed.

4. Conclusion and Outlook

We have proposed a simple algorithm to make use of the orientation angle information in 3D-EMA for computational head stabilisation. The algorithm also provides a well-defined procedure in the case where only two sensors are available for head motion tracking and allows for the combining of position coordinates and orientation angles for head stabilisation with an equal weighting of each kind of information. An evaluation showed that the method using the orientation angles produced the most reliable results when the head tracking sensors could or should not be placed far from each other.

The findings open up new possibilities for the placement of head tracking sensors in 3D-EMA when no longer required to spread them out in terms of spatial location. Even more promising is the use of the orientation angle information for accurately determining the movements of the jaw. Like the head, the jaw is a rigid-body object and the temporomandibular joint allows movements with all six degrees of freedom though not all of them are relevant in speech [11]. To investigate jaw-tongue coordination both the jaw movement parameters and the jaw-stabilised (intrinsic) tongue movements must be available. However, the appropriate locations for EMA sensors to track the jaw are limited and the sensors cannot be placed with a reasonable distance from each other. Here the orientation-based method allows recovery of all motion parameters from two sensors placed in different orientations on the lower gums. The position of the condyle can be easily factored in later as the centre of rotation.

Supported by NHMRC/ARC grant TS0669874.

5. References

[1] T. J. Hixon, "An electromagnetic method for transducing jaw movements during speech," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 603–606, 1971.

[2] G. van der Giet, "Computer-controlled method for measuring articulatory activities," *The Journal of the Acoustical Society of America*, vol. 61, no. 4, pp. 1072–1076, 1977.

[3] J. S. Perkell and D. Oka, "Use of an alternating magnetic field device to track midsagittal plane movements of multiple points inside the vocal tract," *The Journal of the Acoustical Society of America*, vol. 67, no. S1, pp. S92–S92, 1980.

[4] T. Kaburagi and M. Honda, "Calibration methods of voltage-to-distance function for an electromagnetic articulometer (EMA) system," *The Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2391–2394, 1997.

[5] A. Zierdt, P. Hoole, H. M., T. Kaburagi, and H.-G. Tillmann, "Extracting tongues from moving heads," in *5th Speech Production Seminar*, Kloster Seeon, Bavaria, Germany, 2000, pp. 313–316.

[6] H. C. Yehiaa, T. Kuratate, and E. Vatikiotis-Bateson, "Linking facial animation, head motion and speech acoustics," *Journal of Phonetics*, vol. 30, no. 3, pp. 555–568, 2002.

[7] P. Schönemann and R. M. Carroll, "Fitting one matrix to another under choice of a central dilation and a rigid motion," *Psychometrika*, vol. 35, no. 2, pp. 245–255, 1970.

[8] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1974.

[9] V. M. Zatsiorsky, *Kinematics of human motion*. Human Kinetics, 1998.

[10] C. Kroos, "Measurement accuracy in 3D Electromagnetic Articulography (Carstens AG500)," in *8th International Seminar on Speech Production*, R. Sock, S. Fuchs, and Y. Laprie, Eds., 2008, pp. 61–64.

[11] E. Vatikiotis-Bateson and D. J. Ostry, "An analysis of the dimensionality of jaw motion in speech," *Journal of Phonetics*, vol. 23, pp. 101–117, 1995.