3D EMA: Examples of speech movement patterns

Phil Hoole & Andreas Zierdt
Institut für Phonetik
Munich University
hoole|andi@phonetik.uni-muenchen.de

Work supported by German Research Council grant TI 69/30 to Hans G. Tillmann

1. Introduction

This report presents selected examples of simple speech movements recorded with the prototype AG500 system (Carstens Medizinelektronik). They were chosen to illustrate in particular what may well prove to be the most useful new features in the final system, namely (1) information on the orientation of the sensors and (2) freedom of head movement of the subject.

The first group of examples is taken from a recording in which five sensors were used. They are referred to as ***head_l, head_r, upper_incisors, dorsum, tip.***

head_l and ***head_r*** were located on the left and right side of the head, respectively, adjacent to the ears.

upper_incisors was located (roughly) on the midline of the gums of the maxillary incisors.

tip was actually located about 2cm posterior to the tongue tip. As we will see, "tip" is thus not a very good designation for this sensor.

dorsum was located about 3.5cm behind ***tip***.

Both tongue sensors were located with the main axis of the sensor (roughly) aligned with the midline of the tongue. This meant that the orientation information could be used to help reconstruct the midline contour of the tongue (see below).

The data has been mapped from the raw coordinate system of the transmitter assembly to a skull-based coordinate system in which the naming conventions for the three spatial coordinates are:

X = Lateral (increase to left)
Y = Anterior-posterior (increase from front to back; labelled "a_p" in the figures)
Z = Longitudinal (increase from foot to head)

The mapping to the skull-based system was defined as follows:

- The line joining ***head_l*** to ***head_r*** is horizontal (parallel to the x/y plane), and perpendicular to the y/z plane (i.e ***head_l*** and ***head_r*** have the same anterior-posterior and longitudinal location).

- The line joining ***upper_incisors*** to the midpoint of the two head sensors was oriented at 15 deg. This was estimated to correspond to a normal upright position of the head (other normalization procedures could easily be implemented depending on what sensors are available and what reference tasks were performed).
- The origin of the coordinate system is located at the upper incisors.

Unless specified otherwise the data has been corrected for head-movement on a sample-by-sample basis.

It will be recalled that each sensor provides five coordinates, consisting of three positions (x/y/z) and two rotations (the latter can be thought of as the azimuth and elevation of the sensor in a spherical coordinate system). If the sensor is regarded as a rigid body this leaves one of the six degrees of freedom not accounted for (corresponding to rotation about the main axis of the receiver coil, since this results in no change in the induced field).

2. Tongue movements: Sensor orientation as additional information

The first movement examples show tongue movements as captured by the tip and dorsum sensors.

The figures are intended to be "read" as follows[1]:
The trajectories of the tip and dorsum sensors are given by traces that change colour from dark blue (at the beginning of the movement) through green and yellow to red (at the end of the movement). At selected time points on the trajectories the position and orientation of the sensors has been marked by coloured bars (the colour is determined by the time-point on the trajectory to which they correspond). The position of the sensor is marked by the large circle at the midpoint of the bar. For most examples the length of the bars has been arbitrarily chosen to be 2cm from end to end (it has nothing to do with the size of the sensors themselves that are mounted on the tongue). Most of the examples are of simple VCV sequences in which sensor positions and orientations have been displayed roughly at the midpoint of each the three sounds in the sequence. In these cases, a colour coded transcription is included in the figure to indicate which time point corresponds to which sounds.

The four examples in Figs. 1-4, showing the sequences /ita/, /itu/, /isa/, /isu/ in a "traditional" sagittal view, all make essentially the same point. As just indicated, the time points shown by bars are in the first vowel (/i/), in the medial consonant (/s/ or /t/), and the final vowel (/a/ or /u/).

[1]They are an attempt to capture in a static figure movement features that may be easier to appreciate as animations; some are available as QuickTime movies at www.phonetik.uni-muenchen.de/~hoole/5d-examples.html

The main point to note is that there is often very little movement of the sensors (especially the tip sensor) from V1 to the consonant (i.e the large circles at the midpoint of the bars are close together). However, there is a substantial change in the *orientation* of the sensor, consistent with the change from a bunched tongue configuration with lowered tongue tip for /i/ to raised tongue-tip for the consonant (recall that the tip of the tongue is actually about 2cm anterior to the sensor we refer to as 'tip').

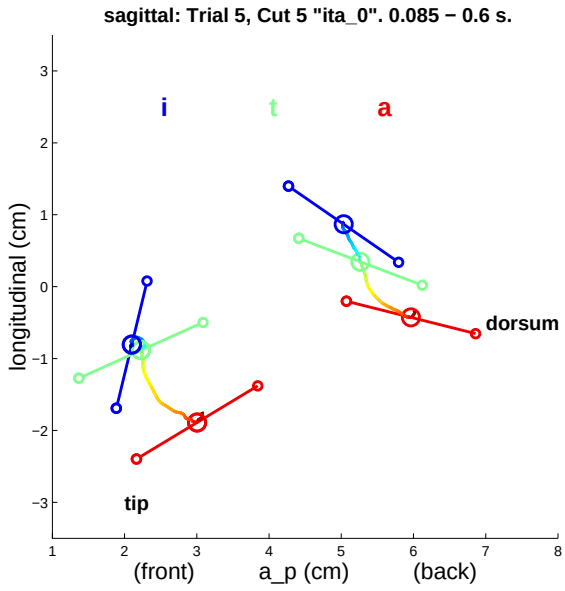

***Fig. 1**:Tongue movements for /ita/*

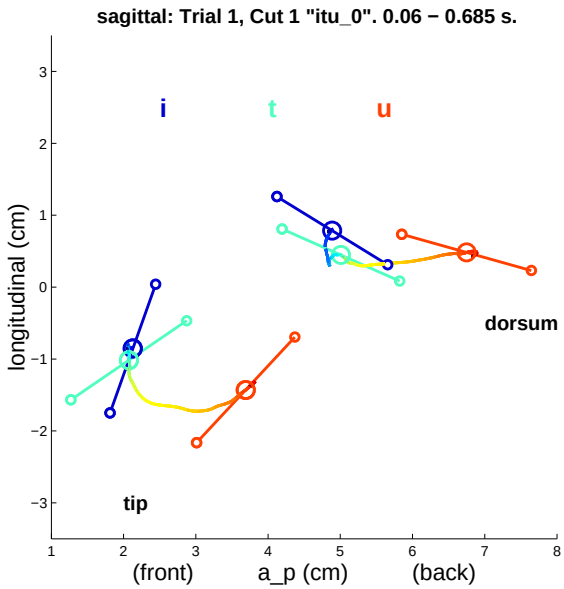

***Fig. 2**:Tongue movements for /itu/*

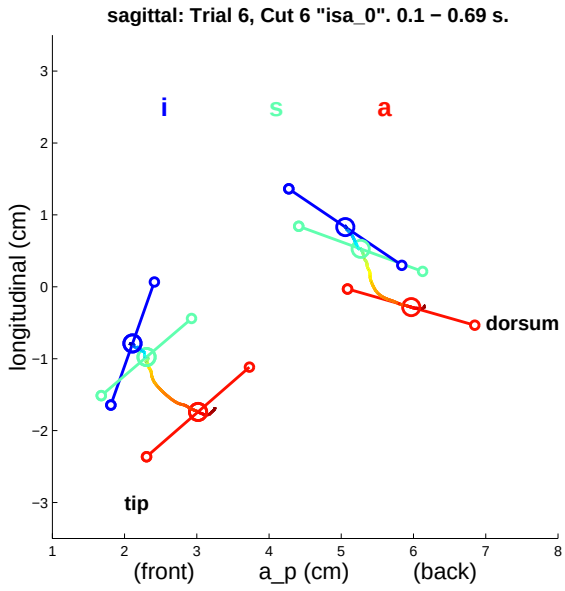

***Fig. 3**:Tongue movements for /isa/*

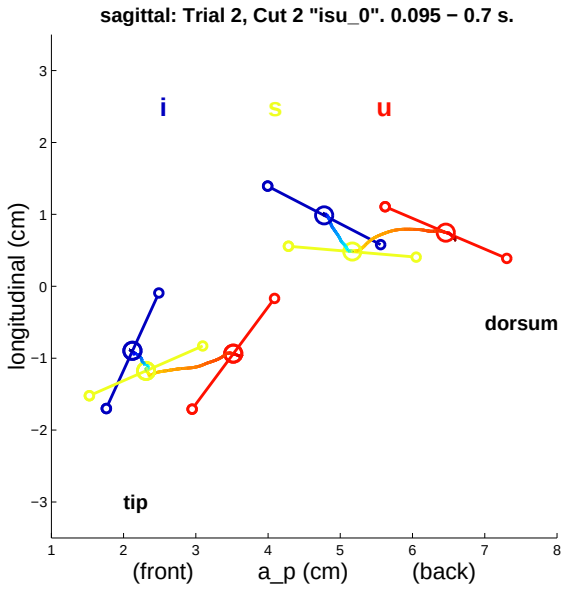

***Fig. 4**:Tongue movements for /isu/*

With only the *positional* information from the two tongue sensors it would be impossible to realistically assess the configuration of the tongue. With the orientation information we would appear to get about as much information on the configuration of the tongue from only two sensors in the new system as we would have got from four sensors in the old 2D system (4 sensors has proved to be in practice the maximum number of sensors that can easily be attached to the midline of the tongue). The other way of looking at this is, of course, that with four sensors in the new system we would expect to obtain a more reliable and detailed picture of the the shape of the tongue than would be possible in the old system. In particular, it is worth noting that it is not possible to locate a sensor right on the tongue tip (because of disturbance of articulation) so the orientation information will help to give information about locations on the tongue that cannot be monitored directly (this should apply to the tongue root as well as to the tongue tip).

3. Separating tongue and head movement

In the above examples, head movement was not restricted, but the subject generally moved his head very little while speaking . We also recorded trials with the same sound sequences but with subject asked to nod his head up and down (a “yes”-movement) while speaking.

The two pairs of figures below (Figs. 5+6; Figs. 7+8) show on the left the tongue movements after correcting for the accompanying head movement, and on the right the head movement itself as monitored at the sensors on the upper incisors, left and right temples. The time segments are the same in the corresponding left and right views (i.e the colour-coding of time in the trajectories is the same; however, to avoid making the right-hand panels too complicated only two selected time instants are shown with the orientation bars). The upper incisor traces in particular show that about 2cm of head movement occurred during the corresponding speech utterances on the left. Nevertheless, the trajectories of the tongue sensors are virtually identical to the tongue movements spoken with negligible head movement (compare Fig. 5 to Fig.2, and Fig. 7 to Fig.4). Thus the system seems potentially able to measure head-movement itself, as well as allow this movement to be factored out from tongue movements, leaving the tongue movements anchored in a skull-based frame of reference.

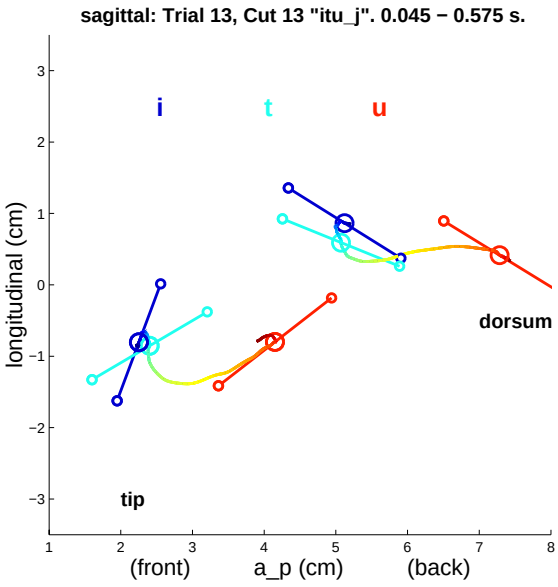

***Fig. 5**:Tongue movements for /itu/ after correction for head movement shown in right panel*

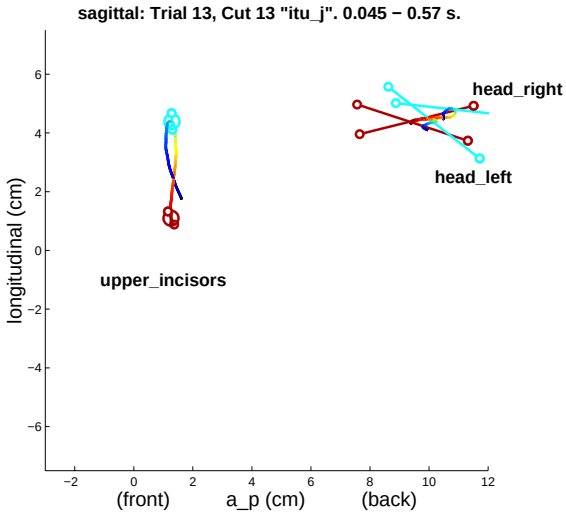

***Fig. 6**:Head movments during utterance of /itu/ shown in left panel*

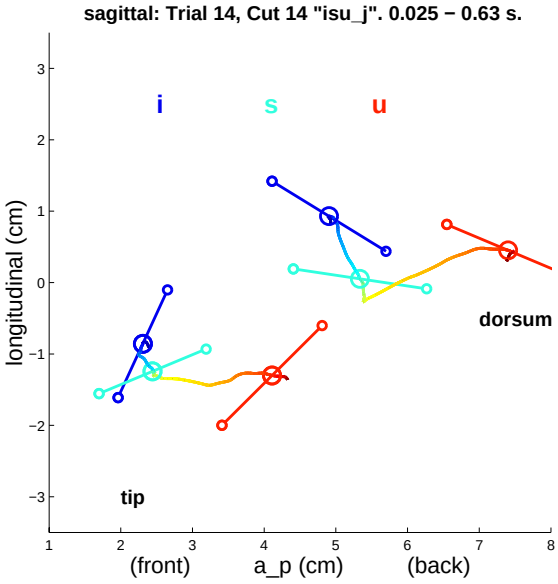

***Fig. 7**:Tongue movements for /isu/ after correction for head movement shown in right panel*

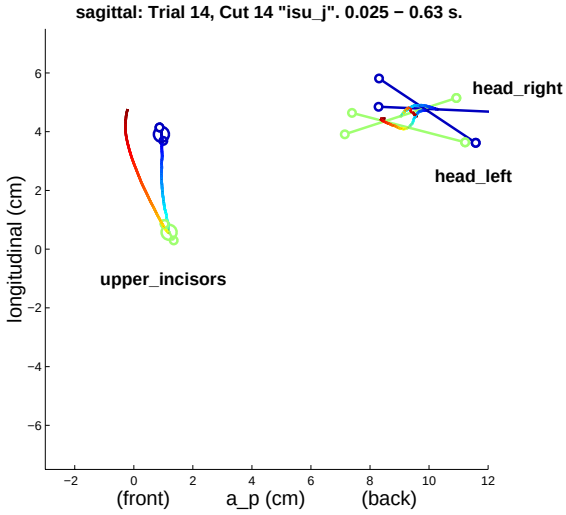

***Fig. 8**:Head movments during utterance of /isu/ shown in left panel*

4. Additional views of head movements

The head movements shown in Figs. 6 and 8 above may be easier to understand after referring to Fig. 9. This shows the upper-incisor and the two head sensors in both sagittal and transversal views (the latter as if viewing from below). Again, a nodding movement is displayed, but with orientation bars only at one time instant. The reason why no (or very little) bar is seen for the upper-incisor sensor in the sagittal view is that the main orientation of the sensor is in a lateral direction. Thus very little of the total length of the bar is projected onto the sagittal plane.

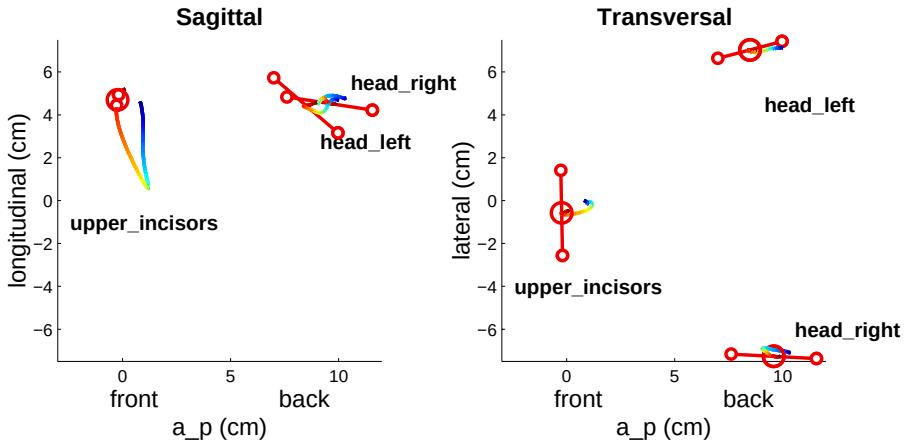

***Fig. 9**: Illustration of reference sensors on head and upper incisors in both sagittal and transversal views*

The view of the head_left and head_right sensors in the sagittal plane is rather confusing, because in this plane, of course, they overlay each other. (Note: The bar for the head_left sensor is shorter in the transversal view because, as seen in the sagittal view, it is inclined upward more than the head_right sensor.) Nevertheless, it is interesting to note that the two head-sensors do not show much translational motion during the nodding movements in this and the previous examples. Presumably they are quite close to the centre of rotation of the head. But the changes in *orientation* of these two head sensors in Figs. 6 and 8, where two time instants are displayed, are clearly consistent with the changes in *position* (up-down) of the upper-incisor sensor.

A further example of sensor trajectories and orientations in a simple head-movement task is shown in the next figure. The subject was required to shake his head (a "no"-movement) without speaking. The subject and recording session are different from the above examples. The figure shows the movements of two *tongue* sensors during this task. The dorsum coil was mounted on the tongue in a similar way to the previous examples, however the tip coil was mounted with the main axis in a lateral orientation (i.e roughly parallel to the orientation of the upper-incisor sensor

in the above examples)[2]. The movement is shown in a transversal view. This example serves again to illustrate the point that consistent orientation information can be extracted. Even though the two sensors are mounted in very different ways on the tongue, they both show changes in orientation during the largely rotational movement of the head that are consistent with each other and with the rotational movement suggested by the positional changes of the sensors. And once again it becomes clear how much more information a sensor with both position and orientation information provides: For this particular movement a single sensor with both types of information would suffice to clearly reveal the nature of the movement of the underlying rigid body (the skull), whereas a single sensor with only position information would not distinguish between largely rotational motion, and translational motion along a curved trajectory.

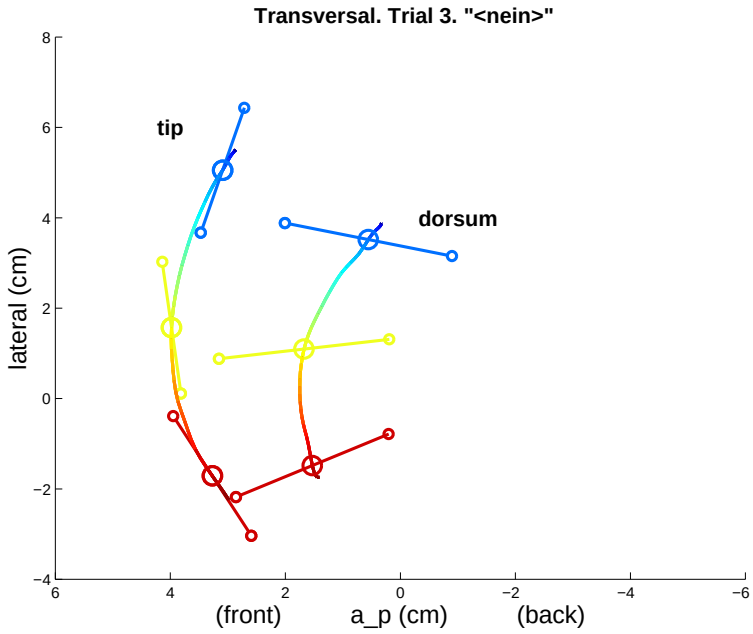

***Fig. 10**: Transversal view of movement of two sensors on tongue during shaking movement of head. Note that orientation of tip sensor is different from that used in the previous examples*

We will return to this sensor arrangment again in Fig. 12 below. Before this, Fig. 11 shows a final example using the original sensor arrangement on the tongue.

[2]In this early recording the direction of increasing values on the anterior-posterior axis had the opposite definition to that currently used. For this reason the axis direction has been reversed in order to orient all examples in the same way.

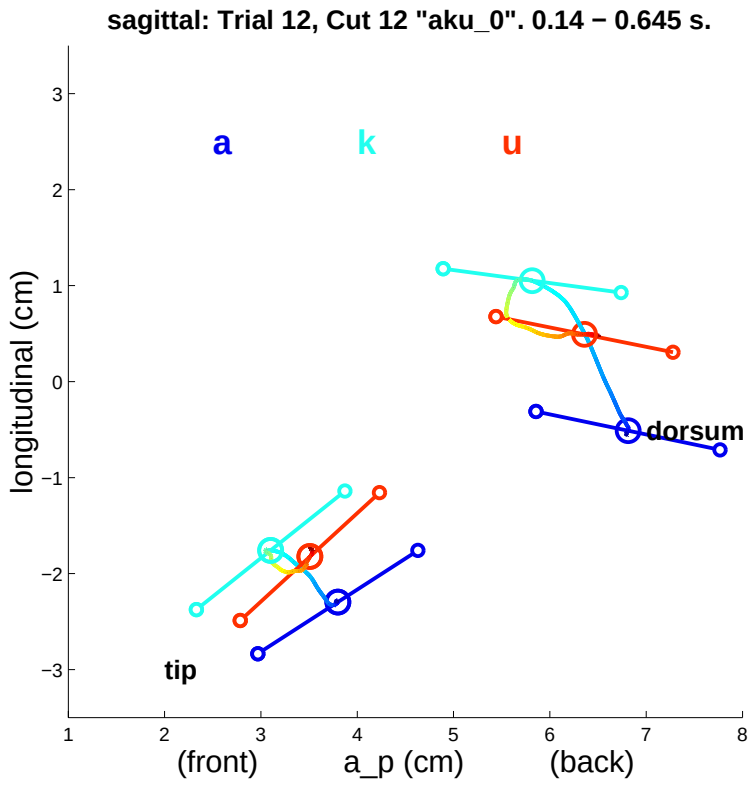

***Fig. 11**: Tongue movements for /aku/*

5. Tongue loops

The example in Fig. 11 simply serves to confirm that subtle but consistent features well-known from investigations with the old 2D system are also readily reproducible in the new system. The example is of movement from /a/ throught /k/ to /u/. It is well-known that especially in the context of back vowels the tongue tends to move forward during the closure for velar consonants, even though this may seem to be taking it further away from the following vowel /u/. (cf. Mooshammer, C., Hoole, P. & Kühnert, B. (1995). ***On loops***. J. Phonetics, 23: 3-21). In a sense, then, this simply documents the minimal, but nonetheless necessary requirement for the new system that it can reproduce established findings. In the next example (Fig. 12), we turn to a more forward-looking example which shows that the new system can be used successfully in cases where the old system would have probably failed to generate reliable data.

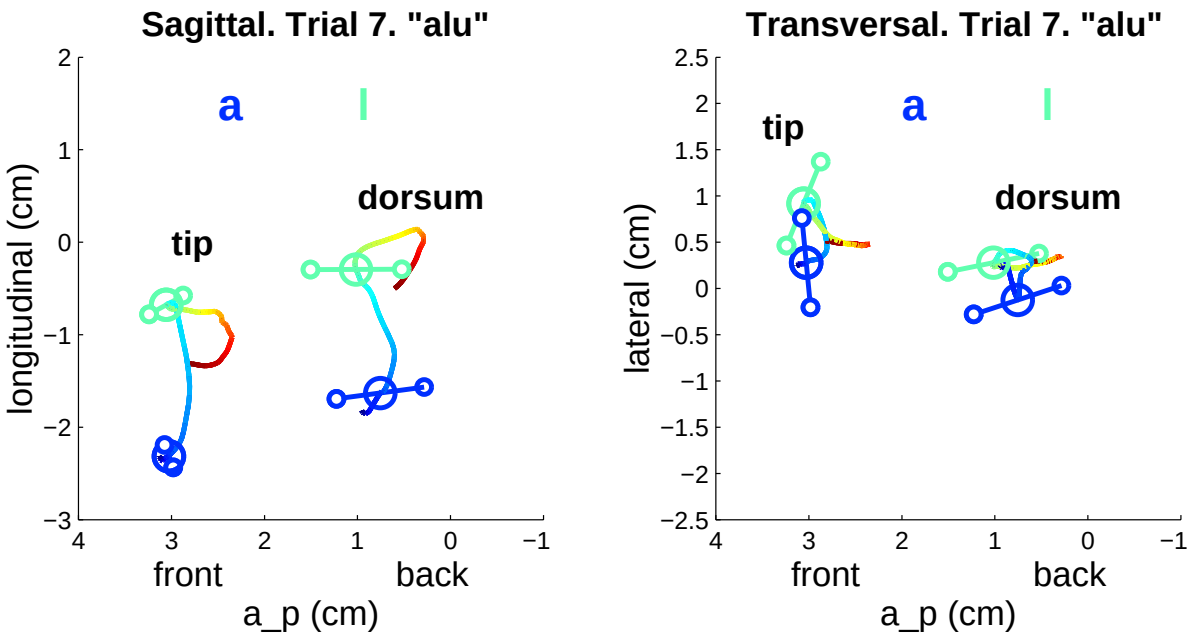

***Fig. 12**: Sagittal and transversal views of tongue movment for /alu/*

6. Lateral tongue movements

This example (Fig. 12) returns to the configuration used in the head-shaking example in Fig. 10, i.e to the recording setup in which tip and and dorsum sensor are oriented roughly at right angles to each other.

It shows in both transversal and sagittal views the trajectories of tip and dorsum for the sequence /alu/. Orientation bars are shown at about the midpoint of the /a/ and the /l/. By looking simultaneously at both views it is possible to observe that as the tongue tip raises from /a/ to /l/ (i.e as the colour changes from dark blue to light blue) it also moves laterally, and changes its orientation. Movement patterns of this kind can be expected to be quite common with apical sounds like /l/, but hitherto it has not been possible to measure them. Indeed, such movements (several millimetres of movement away from the midsagittal plane coupled with 10 or 20 degrees of rotation), have been shown in bench tests to lead to errors in the old 2D EMMA system (Honda, M. & Kaburagi, T. (1993) "Comparison of electromagnetic and ultrasonic techniques for monitoring tongue motion", FIPKM 31, 121-136; see also Kaburagi & Honda, "Determination of sagittal tongue shape from the positions of points on the tongue surface", JASA 96(3), 1356-1367). And in actual experiments we have heard several reports from experienced EMMA users of data having to be discarded (particularly for apical sounds) because the amount of rotational misalignment estimated by the system indicated that the data was unreliable.

7. Detection of unreliable data

Although the new system resolves the above-mentioned limitation of the 2D EMMA system, the reliability of the new system nonetheless needs to be assessed.

As discussed elsewhere, at the time these recordings were made it was clear that several problems remained to be resolved regarding the robustness and accuracy of the calibration procedures. Thus, in the experiments from which the above speech examples are taken, trials in particular regions of the measurement space frequently gave highly anomalous movement patterns. However, even though we were aware of some clear imperfections, at the same time we are quite confident about the reliability of the examples shown above, because the algorithms have been designed to provide feedback to the user that allows the the likely reliability of the data to be assessed. Currently the TAPAD algorithm that solves the non-linear equations to estimate position and orientation for each sensor from the raw field strengths also returns the following three diagnostic parameters on a sample-by-sample basis for each sensor (they are discussed in more detail elsewhere):

n_iter:	Number of iterations. Stable solutions are usually found within a couple of iterations (<10)
n_diverge:	Number of divergences. Non-zero value may indicate that the algorithm had difficulty finding a solution
residue:	A low value indicates a good match between measured amplitude values and the values that would be predicted for the chosen solution from the calibration data. Unstable solutions may be marked by a sudden increase in the residue.

Below (Fig. 13) we show an example of time signals from a trial similar to those shown above in which two short “glitches” are apparent in the the position values. The bottom panel in the figure shows the “residue” parameter. Clearly, there is a temporary sharp increase in the residue in the vicinity of these glitches, giving an independent indication that these short portions of the data might better be discarded. As the calibration improves, the frequency and magnitude of problems of this kind can be confidently expected to decrease. For this reason we have not yet attempted to provide hard and fast criteria for detecting unreliable data. The more crucial point at the moment is to note that diagnostic features of this kind are an integral part of the system; even in a perfectly tuned system it will probably never be possible to completely rule out the possibility of anomalous behaviour triggered by some unexpected combination of circumstances. This was probably one of the most important lessons that was learnt from the 2D EMMA system, where the estimated amount of rotational mislignment acted as a kind of quality control procedure: however frustrating, it is vastly preferable to be forced to discard data occasionally than to unwittingly retain data that may lead to unwarranted conclusions.

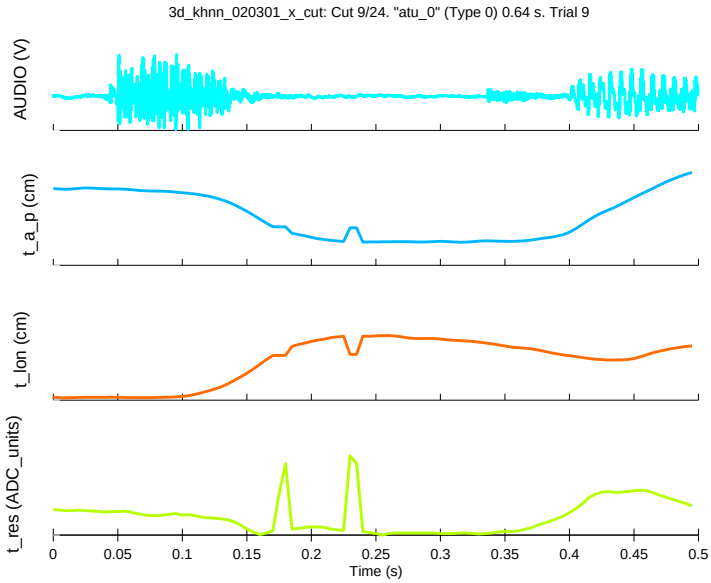

Fig. 13

Another good quality-control procedure in the old system was to to monitor the distance between reference sensors (e.g on upper incisors and bridge of nose), since ideally this should remain constant. Clearly, measures of this kind are also possible in the new system, and will be looked at in detail over the next development phase. At the moment, there is certainly room for improvement. Even for the trials shown above, where the data was considered to be basically reliable, it was possible to observe several millimetres of change in the apparent distance between the reference sensors (upper_incisors, head_left, head_right) during the nodding movements of the head. These changes were generally very systematic during these cyclical movements, suggesting some distortion - but of a systematic nature - in the calibration functions.