

A comparative study of the precision of Carstens and NDI electromagnetic articulographs

Christophe Savariaux^(1, 2), Pierre Badin^(1, 2), Adeline Samson^(3, 4), and Silvain Gerber^(1, 2)

(1) Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France

(2) CNRS, GIPSA-Lab, F-38000 Grenoble, France

(3) Univ. Grenoble Alpes, Laboratoire Jean Kuntzmann, F-38000 Grenoble, France

(4) CNRS, Laboratoire Jean Kuntzmann, F-38000 Grenoble, France

{christophe.savariaux, pierre.badin}@gipsa-lab.grenoble-inp.fr

Research Article

Abstract

Purpose: The study compares the precision of the electromagnetic articulographs used in speech research: NDI's WAVE and Carstens' AG200, AG500 and AG501 systems.

Method: The fluctuation of distances between three pairs of sensors attached to a manually rotated device that can position them inside the measurement volumes were determined. For each device, two precision estimates based on the 95% quantile range of these distances (QR95) were defined: the local QR95 was computed for bins around specific rotation angles, while the global QR95 was computed for all angles pooled.

Results: For all devices, while the local precision lies around 0.1 cm, the global precision is much more worrisome, ranging from 0.03 cm to 2.18 cm and displays large variations as a function of the position of the sensors in the measurement volume. No influence of the rotational speed was found. The AG501 produced – by far – the lowest errors, in particular concerning the global precision.

Conclusions: The local precision can be considered suitable for speech articulatory measurements, but the variations of the global precision need to be taken into account by the knowledge of the spatial distribution of errors. A guideline for good practice in EMA recording is proposed for each system.

Keywords: electromagnetic articulograph, articulatory measurements, precision, NDI WAVE, Carstens AG200, AG500, AG501

1 Introduction

The study of speech production mechanisms at peripheral levels, *i.e.* the articulatory and acoustic levels, requires a detailed knowledge of vocal tract shape and movement. This type of studies involves therefore recording movements of articulators such as the jaw, the tongue, the lips or the velum. As the majority of these articulators are not directly visible, various measurement techniques are used, ranging from anatomical Magnetic Resonance Imaging to electromagnetic articulography or ultrasound imaging.

Since very long, electromagnetic articulographs have been widely used for this purpose in the speech community, from Schönle, Gräbe *et al.* (1987), Perkell, Cohen *et al.* (1992), Zierdt (1993), Hoole & Nguyen (1997) to Kaburagi, Wakamiya *et al.* (2005), or Hummel, Figl *et al.* (2006), without mentioning numerous more recent studies.

The operating principle of these devices has remained substantially the same: alternating electromagnetic fields generated and radiated by multiple emitting coils at different frequencies induce currents of varying intensity in sensors (small coils) placed inside the measurement volume. The current intensity in each coil depends on the distance and the orientation of the sensor in relation to each of the emitting coils. The location, orientation and tilt of the sensors can therefore be obtained from the intensities of the induced currents at the different emitting frequencies by solving equations with multiple unknowns (for details *cf.* Perkell *et al.* (1992); Kaburagi *et al.* (2005); Kröger, Pouplier *et al.* (2008); Hoole & Zierdt (2010)).

The electromagnetic articulographs allow the simultaneous measurement of 8 to 24 sensors at sampling

frequencies ranging from 100 to 1250 Hz. The previous 2D systems suffered from fairly severe drawbacks: (1) the sensors' movements had to be restricted to the measurement plane of the system, as well as their receiving coil axes; this implied that the subject's midsagittal plane is aligned with this symmetry plane, which is often tricky to ensure; (2) one of the main components of these systems was a helmet that was attached to the subject's head by means of straps and that could become quite uncomfortable when recording long corpuses.

A new generation of articulographs has appeared in the 1990s (Zierdt (1993), Kaburagi *et al.* (2005), Hummel *et al.* (2006)). The main drawback has been eliminated: the positions of the receiving sensors are not limited any longer to the midsagittal plane, and are obtained in 3D: 5 degrees of freedom are provided for each sensor (three coordinates x , y and z and two angular orientations corresponding to the azimuth and elevation of the sensor coil main axis). This offers the supplementary advantage of preventing errors from increasing when the sensors are located outside the midsagittal plane or oriented in any particular direction. Moreover, the head of the subject is allowed to move fairly freely in a small region of the measurement space.

A number of electromagnetic articulographs are routinely used in speech research all over the world: the WAVE system (Northern Digital Instruments, Canada), the AG200 and AG500 systems (Carstens Medizinelektronik GmbH, Germany), and more recently the AG501 system aiming to replace the AG500.

The major difference between the systems lies in their designs and their dimensions. The AG200 is a 2D midsagittal system that uses three emitting coils affixed to a Plexiglas helmet worn by the speaker: the receiving coils (reference HS220) that must stay in the midsagittal plane of the subject aligned with the symmetry plane of the helmet are tracked at a frequency of 200 Hz.

The AG500 is a 3D system that uses six emitting coils, distributed in a spherical space, affixed to a Plexiglas cubic structure of approximately 50×50×50 cm. Each coil radiates at frequencies ranging between 7.5 and 13.75 kHz. The 3D coordinates and orientations of at most 12 receiving coil sensors (reference HQ220-L165-S) can be obtained at a frequency of 200 Hz inside a spherical registration volume with a radius of 150 mm. The AG501 is still different from the AG500: nine transmitter coils stand above the registration volume, which is not any longer encumbered by a cube structure; up to 24 sensors (reference HQ220-L120-B) can be tracked at 1250 Hz.

The WAVE system uses an unknown number of transmitter coils embedded in a parallelepiped field generator connected to the system control unit. The field generator of dimension 20×20×7 cm generates an electromagnetic field in a volume set by software with two size options: 50×50×50 cm and 30×30×30 cm (chosen for this study). This system can track the 3D coordinates and orientations of up to 16 sensors (reference 5DOF) at frequencies from 100 to 400 Hz.

Note that the dimensions and shapes of all the coils sensors are slightly different, as illustrated in Figure 1c by a picture of a magazine equipped with one Carstens HS220 and one Wave 5DOF coils.

Considering the diversity of systems presently used, it was felt of high potential interest to the speech community to perform a comparative study of these systems' performances. A number of publications is indeed available in the literature and are summarized in Table 1 (Kröger *et al.* (2008); Yunusova, Green *et al.* (2009); Berry (2011), Stella, Bernardini *et al.* (2012); Kroos (2012); Stella, Stella *et al.* (2013)).

However, these studies are not always comparable one with each other because they use different methodologies for the precision assessment.

For the Aurora, the preceding version of the current WAVE system, Kröger *et al.* (2008) found a standard deviation (SD) of the Euclidean distance between two sensors glued on the lower jaw ranging from 0.054 to 0.103 cm, depending on the corpus (from the syllable /ba/ to a nodding head movement for "No", including a text reading task). They also studied the impact of the location and of the velocity on the distance between two sensors attached to a ruler and showed that the measurement error, evaluated as the SD of the mean Euclidean distance between the two sensors, was greater when the ruler was moved away from the field generator than when it was displaced in direction parallel to the field generator. Kröger *et al.* observed that measurement error increased with increasing velocity.

For the AG500, Yunusova *et al.* (2009) proposed various evaluation experiments using four different metrics: Median, Inter Quartile Range (IQR), 95% quantile and maximum error of the *absolute mean-corrected Euclidean distance* (calculated as the absolute difference between the measured and the estimated X and Y coordinates). For two sensors attached to a cartridge on a rotating disk provided by Carstens (*Circal*), the median errors were under 0.05 cm in each dimensions: 0.024 cm for X, 0.022 cm for Y and 0.038 cm for Z. Then, they glued two sensors on the jaw of one speaker: they observed that the median of the absolute mean-corrected Euclidean distance between two sensors ranged from 0.009 to 0.022 cm depending on the speech corpus used (respectively vowel /a/ vs. reading a paragraph). Further, errors in three spatial orientations (Coronal, Sagittal and Transverse) were estimated as the absolute mean-corrected Euclidean distance between two sensors fixed on a magazine for manual movements of the magazine inside the entire measurement field. They found that the errors were greater than 0.05 cm everywhere except in the central region. Finally, Yunusova *et al.* (2009) recommended to position the sensors attached to the subject as close to the cube center as possible.

Berry (2011) measured the precision of the WAVE system thanks to three experiments: tracking (1) static positions and (2) dynamic movements of a rigid-body across different locations of the electromagnetic volume, and (3) tracking 2 coils glued to the jaw of one subject uttering speech. These experiments were performed for two field volume settings (50×50×50 cm and 30×30×30 cm). The positional tracking errors were estimated as the Root Mean Square of five distances measured for adjacent sensors affixed to the rigid body with reference to the five known distances (evaluated from the average of 30 seconds of recording in static condition). These distance were evaluated using different metrics (Median, IQR, 95% quantile and maximum error). A grid drawn on a table was used to guide the manual positioning of a ruler for the static conditions, while Lego building blocks mounted on a gear system were driven by a precision drill motor for the dynamic movements. Berry (2011) found that static tracking errors stayed below 0.05 cm within a 20 cm near field for both field volume settings. For dynamic tracking with the 50×50×50 cm field setting, they reported errors exceeding 0.1 cm for locations further than 20 cm from the field generator. For both field volume settings the dynamic tracking error was greater than the static one. The human jaw movement tracking error was obtained as the difference between the measured distance between two sensors glued on the jaw, for 10 subjects reading a paragraph, and a reference value estimated as the average registered distance between these two sensors over 10 seconds of stationary position of the jaw. The tracking errors obtained were always lower than 0.05 cm for all subjects. Berry found that the WAVE system has no vertical effect on the tracking errors. He concluded finally that the WAVE system should be used with the 30×30×30 cm volume setting and that subjects should be positioned so that sensors stay within 20 cm of the field generator surface and with their midsagittal profile parallel to this surface.

In extension of the pilot study made by Kroos (2008), Kroos (2012) evaluated the measurement accuracy of the AG500 system using data acquired simultaneously with the Vicon optical motion tracking system (OPT). 12 EMA sensors and 8 OPT markers were attached to a single rigid object to be able to predict the location and orientation of the EMA sensors from the OPT motion tracking data. The rigid object was moved by both hands by the experimenter with translational, rotational and unconstrained movements. The error between OPT and EMA was computed as the root mean square (RMS) error for the two different calculation routines: Calcpos (CP) proposed by Carstens, and TAPAD (TA) proposed by (Zierdt (2007)). Globally, for all the experiments, Kroos (2012) found that the maximum deviations from constant inter-sensor distances (relative error) were 0.088 cm for CP and 0.115 cm for TA; moreover, the difference between the measured and estimated positions (absolute error) were 0.0138 cm (for CP) and 0.195 cm (for TP). Kroos also showed that sensor velocity appeared to have little impact.

Stella *et al.* (2012) tested whether some of the large anomalies observed for the AG500 were due to the calculation method or originated from other causes. They used an experimental setup consisting of a wooden stick affixed to the *Circal* rotating disk device provided by Carstens. The rotational motion of the *Circal* was constant in time and determined circular trajectories lying on the X-Y plane at a fixed height Z. The X, Y and Z coordinates of nine sensors attached to wooden stick (spaced by 1.3 cm) were then compared with the reference ones estimated as those of the theoretical circular trajectories. Stella *et al.* showed that there were large perturbations in different zones inside the measurement field in which the SD of errors varied from 0.02 to 0.39 cm for the Z height. They also confirmed these high errors regions while looking at trajectories of 3

sensors glued on the tongue of a subject during the repetitions of pseudo words. For some repetitions, the deviations of the trajectory were observed to be not compatible with the actual movement of the tongue even though the sensors were placed near the center of the measurement volume. Thus they concluded that the optimal position for the head of the subject was the bottom of the left side of the measurement field and that errors were essentially due to instabilities of the numerical algorithm rather than to any physical or external source of noise. Note that this recommendation appears not to be consistent with those of Yunusova *et al.* (2009).

More recently, Stella *et al.* (2013) analyzed the precision of the recent AG501. They used the same experimental setup than Stella *et al.* (2012) with 16 sensors affixed to the rotating *Circal*. They showed that the standard deviation for the receiver coils of the radius and the height coordinates was always inferior to 0.03 cm when sensors were moving inside a spherical volume with a radius of 11 cm. They concluded that the accuracy of the AG501 was by far better than that of the AG500. This accuracy was confirmed by the recording of sensors trajectories during repetition of words and for which no difference in reliability were found in either displacement or velocity dimensions.

Note also that Geng, Turk *et al.* (2013) evaluated a synchronized double electromagnetic articulograph setup for which they studied the electromagnetic interference between two AG500 systems (the inter device distance was varied between 1 m and 8.5 m). As the authors have not directly evaluated the accuracy of the AG500s in such conditions, this work, though very interesting, is unfortunately not relevant to the present question.

**** Insert Table 1 here ****

Though all these studies provide very useful information concerning the precision of various articulographs, this summary shows that these evaluations are not performed in very comparable ways (*cf.* also Table 1). This convinced us to perform our own evaluation of these systems in similar conditions to be able to compare them in a consistent way. Formally, the evaluation of a measurement system involves concepts of *bias*, *precision* and *accuracy*, as suggested by Walther & Moore (2005). For the *bias* they propose Bainbridge (1985)'s definition: "difference between a population mean of the measurements or test results and an accepted reference or true value". Following West (1999), they consider *precision* as a measure of "the statistical variance of an estimation procedure". Finally, they define *accuracy* as "the overall distance between estimated (or observed) values and the true value".

Ideally, for the present evaluation study, we would like to assess the accuracy and the bias of the articulographs. This would however necessitate the knowledge of the ground truth values, and thus to position the coils with a high precision, *i.e.* with an error at least less than 0.01 cm. As no suitable mean of precisely positioning coils was available to us, we resorted to a less conservative approach: we analyzed the *fluctuations* of the distances between pairs of coils attached to a common solid object for various positions of this object, as already done by Kröger *et al.* (2008), Yunusova *et al.* (2009), Berry (2011) and also Kroos (2012).

We have finally tested six devices, with the help of many colleagues: (1) the AG200 available at GIPSA-lab; (2) the AG500 used by the LORIA/PAROLE team in Nancy, France; (3) the AG500 located at Laboratoire Parole et Langage in Aix-en-Provence, France; (4) one AG501 of Carstens brought to the ISSP10 meeting in Cologne, Germany; (5) the WAVE used by the LORIA/MAGRIT in Nancy, France; (6) the WAVE available at GIPSA-lab (see dates and names in Table 2). For both WAVE devices, the electromagnetic field volume was set to 30×30×30 cm.

Note that we are aware that the AG200 presents some limitations compared to the 3D systems, namely the fact that errors increase considerably when the coils depart from the midsagittal plane or get a too strong oblique tilt. However, an estimation of its precision is of potential interest since there are still AG200s in use and since it could help interpreting past publications based on it.

2 Method and protocol

The purpose of the measurement protocol was threefold: (1) to determine the fluctuation of distances between

pairs of sensors for each system for series or circular movements, (2) to estimate the potential influence of the location, velocity and orientation of the sensors relative to the electromagnetic field on these fluctuations, and (3) to compare the performance of the six devices. We have therefore manually applied movements to sets of sensors attached to a solid object, using the *mkal* mechanical device designed by the Carstens Company.

2.1 The *mkal* device

The *mkal* device was designed and built by the Carstens Company with the initial purpose of calibrating its former AG100 articulograph system. It was designed to receive the helmet of the AG100 or AG200 so that the center of the device was maintained in the center of the helmet in a very stable way. As illustrated in Figure 1a, the *mkal* device is based on a *fixed horizontal base frame*, which maintains the helmet. A *circular plate* can be rotated without limits around a horizontal axis borne by a *vertical frame* firmly affixed to the base frame. The rotation performed manually can be approximately monitored thanks to 24 marks spaced by 15 degrees drawn on the circular plate. The plate is equipped with a container designed to hold up to three magazines positioned at distances respectively of 0 (magazine #1), 4 (magazine #2), and 8 (magazine #3) cm from the rotation center (Figure 1b). Each magazine contains 5 slots (notches) that can receive the sensors (Figure 1c). Altogether, this setup allows moving sensors continuously along circles of different radii according to the position of the sensors in the magazines (*cf.* Hoole (1996) for more details). This setup ensures that the sensors cover at least an area of about 9×9 cm typical of the space of human vocal tract articulators' movements.

For this study two sensors per magazine only were used, thus six sensors altogether. One sensor was attached at each end of each magazine, ensuring a distance of approximately 4 cm between them. Note that the notches carved in the *mkal* magazines are not specifically adapted to receive any given type of coils: they are not more suitable for AG200 coils than for AG501 or NDI coils (see Figure 1c). Therefore, in every case, the coils were attached in the slots using plasticine and were subsequently covered with adhesive tape, in order to ensure that the sensors were firmly affixed to the *mkal* plate during manual rotation, whatever the type of coil used. The slots were designed so that the magnetic axes of the sensor coils were always perpendicular to the plane of the *mkal* plate. The three magazines were positioned respectively at 0, 4 and 8 cm from the center. They were positioned eccentrically from the center of the *mkal* device to ensure that the trajectories of the two sensors on the same magazine were not on the same circle. Note that this setup results in the coil main axes being maintained in the same direction with reference to the *mkal* plate. However, changing the *mkal* orientation does change the orientation of the coil axes; thus it is impossible to disentangle the sources of errors between the position of the coils in the measurement volume and their orientation. This constitutes a clear limitation of the study, though it would not be a problem in practice when dealing with articulatory measurements.

**** Insert Figure 1 here ****

2.2 Measurement protocol

The measurement protocol was nearly identical for all systems. Note that the Carstens devices were first calibrated using the procedures proposed by the maker; the WAVE does not need any calibration. Once the three magazines and the six sensors were in place on the *mkal* plate, the movements manually operated were recorded in four sessions designed to generate a large range of rotational velocities:

- static condition: the *mkal* plate was rotated with complete stops at approximately each of the 24 reference marks (for at least one second); a complete revolution was performed,
- slow dynamic condition: the *mkal* plate was slowly but continuously rotated in one direction (180°) and then in the opposite direction, making between 1 and 3 full revolutions,
- fast dynamic condition: the *mkal* plate was rotated by quarters of a turn four times in each direction, more quickly than in the previous condition, making at least 3 full revolutions.
- very fast dynamic condition: the *mkal* plate was rotated in the same way as in the previous condition, but with very fast movements.

2.3 Common coordinate system

The coordinate systems used by Carstens and NDI are different. Our assessments of the performance of the

various devices are based on the analysis of the coordinates of the six sensors measured for the different conditions. In order to be able to compare the data obtained from the different devices, it was needed to swap some of the X, Y and Z coordinates provided by each device as well as to change some of their signs to recast them into a common coordinate system.

The *common coordinate system* attached to the subject's head has been arbitrarily defined as X running from neck to nose, Y running from feet to head, and Z running from left to right.

Next, a *reference head position* was decided for each system. The top of the head is always up.

For the WAVE, the head midsagittal plane is parallel to the field generator, positioned such as to have the connector for the supporting arm underneath; the right ear is close to the plate on the side painted with the "WAVE" logo.

For the AG200, the nose is pointing toward the open side of the helmet.

For the AG500, the right ear is facing the open side of the Plexiglas cube.

Finally, for the AG501, the neck is close to the supporting column while the nose is pointing outwards from the column.

2.4 Positioning the *mkal* device

An important point was to define the *reference location* of the *mkal* plate with reference to the articulographs: the *mkal* base frame is standing on the supporting table, *i.e.* oriented towards the feet; the plane of the *mkal* plate is aligned with the midsagittal plane of the reference head; the vertical frame supporting the rotating plate is either on the left or on the right (this will be specified whenever needed in the following).

For the AG200, the helmet is positioned in a fixed way on the *mkal* device in such manner to place its coordinate center at the center of the helmet, as designed by Carstens.

For the AG500, the Carstens Company suggests placing the midsagittal plane of the subject in the symmetry plane of the Plexiglas cube. The *mkal* plate was thus set in the X-Y midsagittal plane of the cube, its coordinate center being in the center of the cube.

For the AG501, no instruction is given by the manufacturer; the plane of the *mkal* plate was thus arbitrarily aligned with the reference head position mentioned above.

For the WAVE, the *mkal* plate was placed parallel to the field generator with its center at a distance of 13.5 cm.

For all devices, except for the AG200_G, it was decided to study the effect of the orientation of the *mkal* plate compared to the reference location described above. Thus, the entire protocol was performed for 3 *mkal* orientations, namely 0° corresponding to the reference location, and orientations of approximately 45° and 90°. Note that for the WAVE, in these latter conditions, the center of the *mkal* plate had to be positioned further away from the field generator (17 cm) to allow room for the base frame to be oriented as needed.

The different experimental conditions are summarized in Table 2.

**** Insert Table 2 here ****

Note that, due to the variety of experimental conditions (four different labs or locations), some exceptions to these procedures inadvertently occurred. They are actually two possible ways to plug the magazines on the circular plate: for five devices, the magazines were plugged on the left, but for the AG501_K, it was discovered later that they had been inadvertently plugged on the right. This means that the trajectories of the coils for the AG501_K are different from those of the other devices (*cf.* Figure 3).

Similarly, the *mkal* frame was not always turned in the same direction. The orientation angle, as seen from above the device, is considered positive if the base frame was turned clockwise, and negative otherwise, as indicated in Table 2. This will be taken care of in the statistical analyses.

3 Overview of data

3.1 Preprocessing of the raw data

The reference sensor of the WAVE system (6 degrees of freedom) was not used. The coordinates, referenced to the field generator, were directly obtained from the acquisition software provided by the NDI Company without low-pass filtering.

For the AG500 devices, the two calculation steps of the *Calcpos* algorithm (provided by the Carstens Company) were applied: the first step with the forward/backward option for calculating trajectories and a second run from the initial position obtained in the first step (Hoole & Zierdt (2010)). For the AG501 device, the data were generated by the new *Calcpos* software provided by the manufacturer.

In order to evaluate the influence of the position and velocity of the sensors on the precision of the measurements, the rotation angle of the *mkal* plate and the rotational velocity of the plate had to be estimated, since they were not a priori known. For each experimental setup (device, *mkal* orientation, rotational velocity condition), the *plate rotation axis* was estimated as the vector normal to the plane defined by three coils using the cross product of the vectors from one coil to the other two coils. The relative coordinates in this plane were then computed by projecting the 3D coordinates on this plane. Low-pass filtering at 20 Hz was applied to the resulting coordinates; the *rotation angle* of the *mkal* plate θ was then computed from the coordinate of the coils of magazine #1, taking the initial low vertical position of the magazine container as reference ($\theta = 0$ corresponds thus to a low vertical position of the container – and thus to an horizontal position of the magazines –, while $\theta = \pm 180^\circ$ corresponds to its high vertical position). Finally, time derivation was applied to yield *the rotational velocity* of the plate θ_{vel} .

A total of 15 pair wise distances were obtained from the six sensors. Figure 2 exemplifies the boxplot of all these distances in the case of the WAV_N, for the *mkal* device at 0° orientation, all rotation angles and rotational velocities pooled together. Similar plots were obtained for the other systems. We observe that the dispersion of the distances lies in the same range whatever the combination of coils and therefore of distances. In order to analyze specifically the influence of the proximity of the coils to the center of the measurement volume we have finally considered only the pair wise distances between two sensors on each magazine, *i.e.* 3 distances. For each time sample, each distance was computed as a 3D Euclidian distance according to the following formula:

$$D_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

where $(x, y, z)_{(1, 2)}$ represent the coordinates of the pairs of sensors on magazine #1. Similarly D_2 and D_3 are obtained for magazines #2 and #3.

**** Insert Figure 2 here ****

3.2 Trajectories of the sensors

The spatial traces of the sensors for the various devices are illustrated in Figure 3 by the circular trajectories of the six coils projected in the plane of the *mkal* plate (coordinates X_{2D} / Y_{2D}). Note that for the WAV_N, AG500_A and AG200_G, the center of rotation of the *mkal* plate is almost aligned with the coordinate center. For the other experiments, we see that the center offset does not go beyond 2 cm. This ensures that all devices are approximately operated in their optimal regions. It can be observed that the sensors cover an area typical of the space of human vocal tract articulators' movements.

According the several studies, the maximum articulators' velocity for speech lies around 20 – 25 cm / sec (Ostry & Munhall (1985) : 20 cm / sec, Payan & Perrier (1997) : 22 cm / sec, Westbury, Severson *et al.* (2000): 20 cm / sec). The rotational velocities for each of our experiments was found to range between 0 and $800^\circ / \text{sec}$. The largest percentage of data (95 %) ranged between 0 and $200^\circ / \text{sec}$. The corresponding maximal tangential velocity for the fastest sensors, *i.e.* those of magazine #3 on circles of about 8 cm radius, is thus about 28 cm / sec, which is fairly higher than the maximum reported for speech: this ensures that our experiments encompass well speech conditions in terms of velocities.

**** Insert Figure 3 here ****

3.3 Distance vs. rotation of the *mkal* plate

The first result of interest is the fluctuation of distances $D_{i=1:3}$ as a function of the *mkal* plate rotation angle θ . Figures 3 to 5 display the distances D_i for each magazine (all rotational velocities pooled together) and for each device studied. Figure 4 presents the data for the six devices, for a 0° *mkal* orientation, while Figure 5 presents the data for four devices for $\pm 45^\circ$ *mkal* orientations, and Figure 6 the data for four devices for $\pm 90^\circ$ *mkal* orientations. Note that in these figures, the samples have been sorted according to their θ : the time series of measurements are thus not directly visible.

The first obvious observation is that for all devices, except the AG501, the mean distances are fairly much fluctuating as a function of plate rotation angle θ , especially for magazine #3. This attests to a systematic distortion or bias of the measurements related to the position of the sensors in the magnetic field of the devices. Another important observation is the fairly large fluctuation of the noise around these mean distances. Both types of fluctuations are also globally seen to increase when the *mkal* orientation departs from the 0° . More specifically, in the case of the WAVE devices, this behavior could be ascribed to the fact that for 45° and 90° orientations, the coil sensors are further away from the field generator, and thus in a region of weaker field and presumably of worse precision. Note that NDI does not issue any recommendation concerning the optimal region of use. In the case of the AG5xx devices, similar fluctuations can be observed. The settings of the emitting coils seem more homogeneous as the emitting coils are distributed all around the region of measurement; however Carstens mentions that measurements should be made in the midsagittal plane for the AG500.

Note that, when the articulograph software cannot find a suitable solution to the optimization problem that determines the coordinates, NDI delivers an empty data point, according to an internal threshold (not provided to the user), whereas Carstens delivers a data, whatever reliable or not. This explains why large errors are visible in some regions for the AG500 in Figures 3-5.

Globally, two different precision issues related to these two types of fluctuations are worth being considered:

- (1) the *local precision*, characterized by the fluctuation estimated *locally* for a given range of the *mkal* plate rotation angle θ ;
- (2) a *global precision* defined as the fluctuation estimated *globally* over all θ 's. This notion reflects the extent of the large and slow variations of the mean fluctuations of the distances as a function of the *mkal* plate rotation globally visible in Figures 3 to 5: it is thus related to the concept of bias, as the true values of distances cannot be expected to depend on the location of the sensors.

The exact calculation procedures will be described below.

**** Insert Figure 4, Figure 5, and Figure 6 here ****

4 Statistical analysis

4.1 General approach

The previous sections have qualitatively shown that the distances fluctuations depend on the *mkal* orientation, on the *mkal* plate rotation, on the devices and on the magazine number. The present section offers a quantitative statistical analysis of the influence of a number of factors on these fluctuations.

First of all, we needed to define a metrics to evaluate the dispersion of the measured distances which reflects the measurement errors. A preliminary analysis of the data has shown that the statistical distributions were not quite Gaussian. Therefore, rather than using the standard RMS error, we decided to follow Berry (2011) or Yunusova *et al.* (2009) and to use the 95% quantile range, *i.e.* the difference between the limits of the histogram that contains 95% of the values, which is less sensitive to the distributions asymmetries. This parameter will be denoted by QR95 in the following.

The aim of the study is therefore to estimate the influence of the four following factors on QR95's:

- (1) The device (AG500_N, ...), which is a categorical variable,
- (2) The *mkal* orientation ($0^\circ, \pm 45^\circ, \pm 90^\circ$), which is an ordinal variable,
- (3) The magazine number (#1, #2, #3), which is an ordinal variable,
- (4) The *mkal* plate rotational velocity θ_{vel} . As 95% of the θ_{vel} values ranged between 0 and $200^\circ / \text{sec}$; this parameter was divided in 3 bins of equal width between 0 and $200^\circ / \text{sec}$, providing an ordinal factor with three levels.

Note that ideally, a continuous covariate θ_{vel} should have been used; this was however not possible since QR95's can only be computed over bins of data.

As mentioned above, we were interested in both local and global issues. Therefore, for each set of distance data corresponding to a combination of these four factors, we defined precision estimates as follows.

- (1) To measure the **local precision**, the 360° range of the *mkal* plate rotation angle θ was divided into 10 equally large bins of 36° , and the distance data were accordingly distributed in these bins. QR95 values were then computed over each bin and averaged over the 10 bins to provide the estimation of the local precision for the set of data (named QR95_loc in the following).
- (2) To estimate the **global precision**, a global QR95 (named QR95_gbl in the following) was simply computed over each set of data, all θ pooled.

Finally, ANOVA linear models have been adjusted to the QR95 data with four factors (device, *mkal* orientation, magazine number, rotational velocity). Linear statistical models must take into account all significant main effects of factors as well as all their interactions. Therefore, we have determined the statistically relevant combinations of factors through a procedure of variable selection based on a step by step descending method with a Fisher test. This ensures that the ANOVA models are appropriately fitted to the data. Note that as we are not interested in the interactions between factors, these interactions will however not be analyzed in this study.

In addition to the ANOVA, contrast analyses have also been performed. For each of the models for which it was needed, the analysis of contrasts was performed using the method described by Hothorn, Bretz *et al.* (2008). Thereby, the *glht* function of the *multcomp* package of the R software (R_Development_Core_Team (2008)) was used. This method can perform multiple comparisons of means after building appropriate contrasts matrices. Importantly, it ensures that the overall *type I* error associated with the simultaneous decisions does not exceed the pre-specified significance level (0.05 in this study) by adjusting p-values.

4.2 Various analyses

Ideally, though one global ANOVA model should have been used for all data with all four factors and their interactions, it was however not possible, due exceptions mentioned above.

Recall that the *mkal* orientation had not been fully controlled for all devices: as summarized in Table 2, it had been turned in either $\pm 45^\circ$ or $\pm 90^\circ$. In the case of $\pm 45^\circ$, the trajectories of the sensors cannot be superposed for all devices; the case of $\pm 90^\circ$ is less problematic since this difference could be compensated by flipping the X axis direction.

Consequently, all six devices have been compared for the 3 magazines, the 0° *mkal* orientation, and 3 θ_{vel} values of the factors (analysis A).

Next, the four devices for which different *mkal* orientations are available (WAV_N, AG501_K, WAV_G, AG500_A) have been separately compared for the 3 magazines, 3 *mkal* orientation and 3 θ_{vel} values (analysis B).

Before presenting the various results, it is important to define a threshold under which differences cannot be considered as physically meaningful. A threshold of 0.05 cm may be considered as acceptable for speech, referring to various articulatory modelling studies with RMS reconstruction errors in the range of 0.10-0.15 cm (*cf.* Serrurier, Badin *et al.* (2012) for midsagittal EMA data, Beutemps, Badin *et al.* (2001) for cineradiographic images, or Badin, Valdés Vargas *et al.* (2013) for midsagittal MRI images).

4.2.1 A- Analysis for six devices for the 0° *mkal* orientation

In this analysis, we observe the influence of device, of magazine, and of θ_{vel} on the global and local QR95s for a 0° *mkal* orientation.

The results are illustrated in Figure 7 to 9 that displays the box plots of QR95_gbl and QR95_loc. For the global precision the box plots show differences between devices (Figure 7) and between magazines (Figure 8), but not between θ_{vel} 's (Figure 9). On the other hand, the differences of the QR95_loc between devices, magazines and θ_{vel} 's are all very weak.

In order to test the statistical significance of these differences, the data have been submitted to ANOVA. The estimates of the pairwise mean differences defined in a contrast matrix have been obtained using the *glt* function. Factors for which all mean differences estimates are below the threshold of 0.05 cm have been considered non physically meaningful, and were thus not taken into account in the multiple comparisons and in the estimation of the p-values.

For QR95_gbl, the linear mixed ANOVA model obtained after the step by step descending method (Bolker, Brooks *et al.* (2009)) takes into account all three factors and the interactions between device and magazine. All mean difference estimates for θ_{vel} were below the threshold: the velocity of the sensors appears thus to have no influence on the global precision, all devices and magazines pooled. For all pair wise comparisons between devices, a significant difference between their means estimates (all magazine and θ_{vel} pooled) was obtained ($p < 0.001$) except between device AG200 and WAV_G ($p > 0.05$). Significant differences between means estimates were also obtained for all pair wise magazine comparisons ($p < 0.001$), all devices and θ_{vel} pooled.

The model obtained for QR95_loc takes into account all three factors and also the interactions between device and magazine and between device and θ_{vel} . All mean difference estimates for the three factors were below the threshold: they have thus no influence on the local precision.

**** Insert Figure 7, Figure8, Figure 9 here ****

4.2.2 B- Analyses for *mkal* orientations from 0° to 90° for 4 devices

We present here a detailed analysis of the effects of magazine, *mkal* orientation, and θ_{vel} on the precision of the four devices for which different *mkal* orientations are available. As observed in the previous section, these three factors have no influence on the local QR95_loc for 0° *mkal* orientation; we have observed the same results for the other *mkal* orientations. Therefore we focus on QR95_gbl offers a summary of the results. We see that the QR95_gbl are higher for the 45° and 90° *mkal* orientations than for 0° for the AG500_A and the two WAV devices, but not the AG501_K. Similarly, the QR95_gbl are higher for magazine positions further from the measurement volume center for the AG500_A, the WAV_N and the WAV_G, but not for the AG501_K.

For the AG500_A, the linear mixed ANOVA model obtained takes into account the magazine and *mkal* orientation factors as well as their interaction. Note that θ_{vel} was not retained in the model, which means that it has no effect. Significant mean difference estimates were observed between magazine 1 and 3 ($p < 0.05$), and between the 45° *mkal* orientation and the others ($p < 0.01$). We have also found that the QR95_gbl are particularly high for the 45° and 90° *mkal* orientation of this device (*cf.* Figure 10). We can conclude that for the AG500, the recommendation made by Carstens of placing the subject only in the 0° plane (referred to as midsagittal plane by Carstens) is of great importance.

For the AG501_K, the model takes into account all three factors and all their interactions. As the estimates of all mean differences for each pairwise combination of each factor are below the threshold of 0.05 cm, we considered that these factors have no effect on the global precision. Moreover, we observed that all QR95_gbl values remain below 0.1 cm, whatever the magazine position, the *mkal* orientation and the velocity of the sensors. This is also reflected in Figure 10 for the mean values.

For the WAV_N, the model takes into account all three factors and only the interaction between magazine and *mkal* orientation. As mean difference estimates for all θ_{vel} were below the 0.05 cm threshold, this factor was considered having no influence on the global precision. On the other hand, we observed significant differences between the 45° *mkal* orientation and the two others ($p < 0.001$), and between the 0° and 90° *mkal* orientation ($p < 0.05$). Significant pairwise differences were also observed between all magazine values ($p < 0.001$).

For the WAV_G, the model is the same as for WAV_N, but supplemented with the interaction between *mkal* orientation and θ_{vel} . As for the WAV_N, there is no influence of θ_{vel} on the global precision. On the other hand, significant differences were observed between all magazine ($p < 0.001$) and between all *mkal* orientation values ($p < 0.001$).

In summary, these analyses confirm that the sensor velocity has no influence on the global precision, all devices and magazines pooled.

**** Insert Figure 10 here ****

4.3 Summary of statistical analyses

The analysis of the fluctuation of pair wise distances $D_{i=1:3}$ between sensors can be summarized as follows. First, we have found that no factors have any meaningful influence on the local precision QR95_loc, and that this precision stays globally around 0.1 cm, which can be considered as negligible for speech experiments.

The situation is quite different for global precision QR95_gbl: it ranges from an average of 0.03 cm over all θ_{vel} for magazine #1 and 0° *mkal* orientation for the AG501_K to an average of 2.18 cm over all θ_{vel} and all magazines for 45° *mkal* orientation for the AG500_A, which is more problematic for speech. We have found that θ_{vel} has no influence on QR95_gbl. Oppositely, except for the AG501_K, there is a clear general tendency for QR95_gbl to increase with magazine distance from the center of the measurement volume and when *mkal* orientation is departing from 0°. For a 0° *mkal* orientation, the QR95_gbl is significantly influenced by the device and by the magazine position.

The AG501_K has – by far – the lowest errors, in particular for the global precision. It should also be remarked that two devices from the same manufacturer (AG500 and WAV) have notably different results in different laboratories (*cf.* the discussion in the last section of the article).

5 Localization of highest errors

Recall that the figures in section 3.3 testify to large variations of global precision as a function of θ . This means that the amount of error depends on the localization in the measurement volume. To complement the previous statistical analysis, it is thus interesting to establish maps of these errors in the measurement volume of each device. This will help drawing practical conclusions from this study and establishing the optimal positioning of the speaker's head for each device. Since it has been shown above that errors are globally smaller for the 0° *mkal* orientation, the analysis is restricted to this orientation.

So far, we have analyzed the influence of the position of the sensors in terms of the angular rotation of the *mkal* plate θ , since this was the most direct way to apply statistical analysis to our data. We have thus attempted to establish maps of the highest errors for the different devices in the X_{2D} / Y_{2D} coordinates in the plane of the *mkal* plate. We have pooled all rotational speeds together since the influence of speed was no found relevant. We have ensured that the X_{2D} / Y_{2D} coordinates correspond to the same reference head position for all devices, *i.e.* X_{2D} horizontal from neck to nose, and Y_{2D} vertical from feet to head.

Note that the global precision discussed above cannot be related to any specific time instance (none of the four factors reflects time information) and thus not to geometric location either. The error associated with each measurement sample was thus computed as the difference between the $D_{i=1:3}$ coil distances for each magazine and their average obtained over all data pooled for the 0° *mkal* orientation.

Figure 11 displays the X_{2D} / Y_{2D} trajectories of the six coils; the points associated with differences higher than a threshold of 0.1 cm are marked with black circles. Figure 11 shows a tendency for the WAVE devices to have

their largest errors in the rightmost region of the measurement volume (the neck region for our reference head position), with some errors also in the upper region. For the AG500, the largest errors are located in the leftmost region (the nose for our reference head position). The region where AG200 has the largest errors is quite different: it is concentrated in the upper region, likely closer to the emitting coil near the subject's forehead. Note finally that the AG501 has no points with errors larger than 0.1 cm.

**** Insert Figure 11 here ****

Plots of pair wise sensor distances as a function of *mkal* plate rotation angle have uncovered rather large variations related to the localization in the measurement volume. We have used this knowledge of the spatial distribution of errors to produce guidelines for good practice for EMA recording for each of the four systems (AG200, AG500, AG501, WAVE), as has already proposed in the literature. For the AG500, Yunusova *et al.* (2009) recommend “that the sensors attached to the tongue and lips be positioned as close to the cube center as possible”, while Stella *et al.* (2012) state that “it has often been necessary to ask the subject to move his/her head to the bottom of the left side of the measurement field, a portion of the cube in which the resulting trajectories were more stable”. For the WAVE, Berry (2011) mentions that “on the basis of the model data, dynamic tracking within the near field (< 200 mm) can be characterized by positional tracking errors that tend to be < 0.5 mm.”

From our own experiments, for the AG500, the region to avoid is associated with negative X coordinates and corresponds to the area around the position of nose in our *reference head position* (cf. 2.3). Therefore, it is advisable to position the speaker in the opposite direction, as illustrated in Figure 12. This is consistent with Yunusova *et al.* (2009) recommendation, but not with Stella *et al.* (2012) statement.

For the WAVE, the optimal region is associated with negative X and Y coordinates. Thus, the speaker should be positioned in the same way as the *reference head position* (right ear close to the plate on the “WAVE” logo side), as illustrated in Figure 12b.

For the AG200, by design, the speaker's nose must be pointing toward the open side of the helmet by design, which happens to be the optimal position with regards to measured errors.

Finally, the AG501 has been found little sensitive to the speaker's head orientation, which leaves some freedom for positioning the speaker. However, using the *reference head position* (neck close to the supporting column, nose is pointing outwards from the column) leads to still smaller errors.

**** Insert Figure 12 here ****

6 Discussion and conclusion

We have presented a very extensive comparative study of the precision of various electromagnetic articulographs, from the ancient Carstens AG200 2D system (released in the beginning of the 2000's) to the most recent AG501 3D system, including the Carstens AG500 and the NDI WAVE. The influence of several factors on both local precision and global precision has been analyzed.

Globally, for all devices, all magazines and for the 0° *mkal* orientation, we have found that the local precision (QR95_loc) lies around 0.01 cm. This type of error is rather low, and acceptable for speech articulatory measurements.

The values obtained for the global precision (QR95_gbl) are larger, ranging from 0.03 cm to 2.18 cm. This type of error is actually much more worrisome. No influence of the rotational speed has been found, which is in line with Kroos (2012) who found a very little influence of the sensors' velocity on the precision. Oppositely, except for the AG501_K, the *mkal* orientation has an important influence on the global precision, which is better at 0°. In the same way, the global precision increases with distance from the measurement volume center (magazine from #1 to #3) for all devices. Altogether, the AG501_K is the more accurate device whatever factors, with QR95_gbl values always below 0.1 cm.

The comparison of similar devices (AG500_A and AG500_N, and WAVE_N and WAVE_G) has revealed

significantly different errors, which is surprising at first sight. A first explanation could be that the devices are not strictly identical, in particular due to the use of different coil sensors. Another explanation, not exclusive of the first one, would be that the devices may be sensitive to the experimental environment (background magnetic field in the experimental room, possible reflection of electromagnetic waves on the walls...). However, we observed similar patterns of errors localization for the similar devices, which are likely due to identical emitting coil arrangements. As done by Yunusova *et al.* (2009), it could be envisaged to determine the optimal location for each room by moving the device through various locations in the room and monitoring changes in the sensor positions.

In summary, the present study has evaluated six different electromagnetic articulographs devices (two manufacturers, four systems) in the same systematic way, which can be considered as a major contribution, as the literature in this domain is rather heterogeneous.

We could also note that, faced to the impossibility to access ground truth coordinate values as reference to compute absolute measurement errors, a number of studies have attempted to model these “expected” or “mean-corrected” values. In order to avoid the risk to introduce systematic biases due erroneous models, we have followed a more conservative approach, and determined a global precision that reflects the fluctuations related to sensor location, without reference to expected values.

Recall finally that our experimental setup did not allow to disentangle the sources of errors between the position of the coils in the measurement volume and their normals’ orientation, which somehow constitutes a limitation of the study. In the future, we could also consider evaluating the influence of the sensors normals’ orientation in the measurement volume: indeed, the absolute orientation of the sensors’ axis might be highly variable considering both the general orientation of the organs (*e.g.* incisors *vs.* tongue dorsum) and large deformations of some organs such as the tongue (*e.g.* the reversal of the tongue tip sensor in retroflex consonants).

7 Acknowledgements

We are very thankful to all colleagues who helped us collecting data at various places : Slim Ouni and Yves Laprie (LORIA/PAROLE, Nancy, France), Brigitte Wrobel-Dautcourt (LORIA/MAGRIT, Nancy, France), Thierry Legou and Noël Nguyen (Laboratoire Parole & Langage, Aix-en-Provence, France), and Ulrich Szagun (Carstens, Germany). We thank also the companies NDI (Germany) and Carstens (Germany) for their help in this study.

We are also much indebted to Jean-Luc Schwartz (GIPSA-lab/PCMD, Grenoble, France) and Rémy Drouilhet (Laboratoire Jean Kuntzman, Grenoble, France) for useful advice on statistical analysis.

This work has been partially funded by the project ANR-13-TECS-0011-06 “e-SwallHome – Swallowing & Respiration: Modelling & e-Health at Home”, by the ANR-11-LABX-0025-01 “LabEx Persyval-LAB” funded by the French ANR agency, and by the PEPS program funded by the Univ. Grenoble Alpes COMUE and the French CNRS.

8 References

- Badin, P., Valdés Vargas, J.A., Koncki, A., Lamalle, L. & Savariaux, C. (2013). Development and implementation of fiducial markers for vocal tract MRI imaging and speech articulatory modelling. In *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, vol., pp. 1321-1325. Lyon, France.
- Bainbridge, T.R. (1985). The Committee on Standards: precision and bias. *ASTM Standardization News* **13**, 44-46.
- Beautemps, D., Badin, P. & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *Journal of the Acoustical Society of America*, **109**(5), 2165-2180.
- Berry, J.J. (2011). Accuracy of the NDI Wave speech research system. *J Speech Lang Hear Res*, **54**(5), 1295-1301.

- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**(3), 127-135.
- Geng, C., Turk, A., Scobbie, J.M., Macmartin, C., Hoole, P., Richmond, K., Wrench, A., Pouplier, M., Gurman Bard, E., Campbell, Z., Dickie, C., Dubourg, E., Hardcastle, W., Kainada, E., King, S., Lickley, R., Nakai, S., Renals, S., White, K. & Wiegand, R. (2013). Recording speech articulation in dialogue: Evaluating a synchronized double electromagnetic articulography setup. *Journal of Phonetics*, **41**(6), 421-431.
- Hoole, P. (1996). Issues in the acquisition, processing, reduction and parameterization of articulographic data. *Forschungsberichte des Instituts für Phonetik und Spachliche Kommunikation der Universität München, FIPKM*, **34**, 158-173.
- Hoole, P. & Nguyen, N. (1997). Electromagnetic Articulography in coarticulation research. *Forschungsberichte des Instituts für Phonetik und Spachliche Kommunikation der Universität München, FIPKM*, **35**, 177-184.
- Hoole, P. & Zierdt, A. (2010). Five-dimensional articulography. In *Speech Motor Control: New developments in basic and applied research* (B. Maassen & P.H.H.M. Van Lieshout, editors), pp. 331-349.
- Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, **50**(3), 346-363.
- Hummel, J., Figl, M., Birkfellner, W., Bax, M.R., Shahidi, R., Maurer Jr, C.R. & Bergmann, H. (2006). Evaluation of a new electromagnetic tracking system using a standardized assessment protocol *Physics in Medicine and Biology*, **51**(10), N205-N220.
- Kaburagi, T., Wakamiya, K. & Honda, M. (2005). Three-dimensional electromagnetic articulography: A measurement principle. *The Journal of the Acoustical Society of America*, **118**(1), 428-443.
- Kröger, B.J., Pouplier, M. & Tiede, M.K. (2008). An evaluation of the Aurora system as a flesh-point tracking tool for speech production research. *J Speech Lang Hear Res*, **51**(4), 914-921.
- Kroos, C. (2008). Measurement accuracy in 3D electromagnetic articulography (Carstens AG500). In *8th International Seminar on Speech Production, ISSP8* (R. Sock, S. Fuchs & Y. Laprie, editors), vol., pp. 61-64. Strasbourg, France.
- Kroos, C. (2012). Evaluation of the measurement precision in three-dimensional Electromagnetic Articulography (Carstens AG500). *Journal of Phonetics*, **40**(3), 453-465.
- Ostry, D.J. & Munhall, K.G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America*, **77**(2), 640-648.
- Payan, Y. & Perrier, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, **22**, 185-205.
- Perkell, J.S., Cohen, M.M., Svirsky, M.A., Matthies, M.L., Garabieta, I. & Jackson, M.T.T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, **92**, 3078-3096.
- R_Development_Core_Team. (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, <http://www.R-project.org>.
- 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 and Language*, **31**(1), 26-35.
- Serrurier, A., Badin, P., Barney, A., Boë, L.-J. & Savariaux, C. (2012). The tongue in speech and feeding: Comparative articulatory modelling. *Journal of Phonetics*, **40**, 745-763.
- Stella, M., Bernardini, P., Sigona, F., Stella, A., Grimaldi, M. & Gili Fivela, B. (2012). Numerical instabilities and three-dimensional electromagnetic articulography. *The Journal of the Acoustical Society of America*, **132**(6), 3941-3949.
- Stella, M., Stella, A., Sigona, F., Bernardini, P., Grimaldi, M. & Gili Fivela, B. (2013). Electromagnetic Articulography with AG500 and AG501. In *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, vol., pp. 1316-1320. Lyon, France.
- Walther, B.A. & Moore, J.L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**(6), 815-829.

- West, M.J. (1999). Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias. *Trends in Neurosciences*, **22**(2), 51-61.
- Westbury, J.R., Sevenson, E.J. & Lindstrom, M.J. (2000). Kinematic event patterns in speech: Special problems. *Language and Speech*, **43**(4), 403-428.
- Yunusova, Y., Green, J.R. & Mefferd, A. (2009). Accuracy Assessment for AG500, Electromagnetic Articulograph. *J Speech Lang Hear Res*, **52**(2), 547-555.
- Zierdt, A. (1993). Problems of electromagnetic position transduction for a three-dimensional articulographic measurement system. *FIPKM (Proceedings of the Accor Workshop on Electromagnetic Articulography in Phonetic Research)*, **31**, 137-142.
- Zierdt, A. (2007). EMA and the crux of calibration. In *XVIth International Congress of Phonetic Sciences* (J. Trouvain & W.J. Barry, editors), vol., pp. 593-596. Saarbrücken, Germany.

9 List of Figures

Figure 1. (a) Photo of the <i>mkal</i> device equipped with three magazines; (b) Zoom on the circular plate, the container (vertical) and the 3 magazines (horizontal); for each magazine, the small circles denote the positions of the sensors closest to the magazine container, while the stars correspond to the sensors furthest from the container; (c) Zoom on a magazine equipped with two sensors (AG200, ref. HS220: left; WAVE (ref. 5DOF): right).....	17
Figure 2. Box plot of all 15 pair wise distances (in cm) for the WAV_N, with the <i>mkal</i> device at 0° orientation, all rotation angles and rotational velocities pooled together.	18
Figure 3. Spatial trajectories of the six sensors, in the reference midsagittal plane for the six devices (X_{2D} / Y_{2D} coordinates in the plane of the <i>mkal</i> plate, in cm), for the 0° <i>mkal</i> orientation, all θ_{vel} pooled. For each magazine, the small circles denote the positions of the sensors closest to the magazine container considered at the first instant of the recordings in static conditions. The stars correspond to the sensors furthest from the container. Note that if the plate in the initial position for the WAV_N is rotated by 180°, all initial positions are identical for all devices, except for the AG501_K, for which the magazines were unfortunately plugged in an opposite direction in the container.	19
Figure 4. Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ (°) for the six devices, all conditions pooled (see §0) for 0° <i>mkal</i> orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.	20
Figure 5: Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ (°) for four devices, all conditions pooled (see §2.2) for $\pm 45^\circ$ <i>mkal</i> orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.	21
Figure 6: Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ (°) for four devices, all conditions pooled (see §2.2) for $\pm 90^\circ$ <i>mkal</i> orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.	22
Figure 7. Box plots of QR95_gbl (left) and QR95_loc (right) in cm for the six devices, for all magazine and all θ_{vel} pooled, for 0° <i>mkal</i> orientation. In these plots, the central red line is the median, the edges of the box are the 25 th and 75 th percentiles, and the whiskers extend to the most extreme data points not considered outliers.....	23
Figure 8. Box plots of the QR95_gbl (left) and QR95_loc (right) in cm for the three magazines, for all devices and all θ_{vel} pooled, for 0° <i>mkal</i> orientation. In these plots, the central red line is the median, the edges of the box are the 25 th and 75 th percentiles, and the whiskers extend to the most extreme data points not considered outliers.....	24
Figure 9. Box plots of the QR95_gbl (left) and QR95_loc (right) in cm for the three θ_{vel} , for all devices and all magazine pooled, for 0° <i>mkal</i> orientation. In these plots, the central red line is the median, the edges of the box are the 25 th and 75 th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as red “+”.	25
Figure 10: Means of QR95_gbl in cm for the three <i>mkal</i> orientations, all magazines and θ_{vel} pooled (left), and for the three magazines, all <i>mkal</i> orientations and θ_{vel} pooled (right). Note that large values have been clipped for AG500_A (values are indicated on the graph).....	26
Figure 11. Spatial trajectories of the six sensors (as in Figure 3), where points for which the distance departs from the average distance (computed over all data pooled for the 0° <i>mkal</i> orientation) by more than 0.1 cm have been marked with black circles.	27
Figure 12. Illustration of the optimal placement of the speaker’s head: outlines of the articulator’s contours (red) and example of sensor positions (filled red circles) superposed to the same elements as in Figure 11 for the AG500_N and the WAVE_N.	28

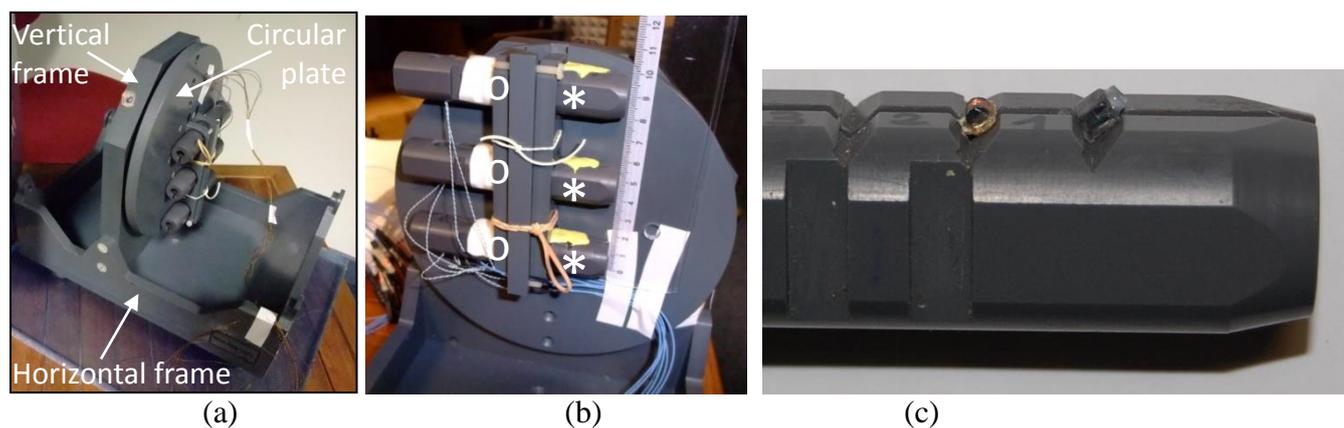


Figure 1. (a) Photo of the *mkal* device equipped with three magazines; (b) Zoom on the circular plate, the container (vertical) and the 3 magazines (horizontal); for each magazine, the small circles denote the positions of the sensors closest to the magazine container, while the stars correspond to the sensors furthest from the container; (c) Zoom on a magazine equipped with two sensors (AG200, ref. HS220: left; WAVE (ref. 5DOF): right).

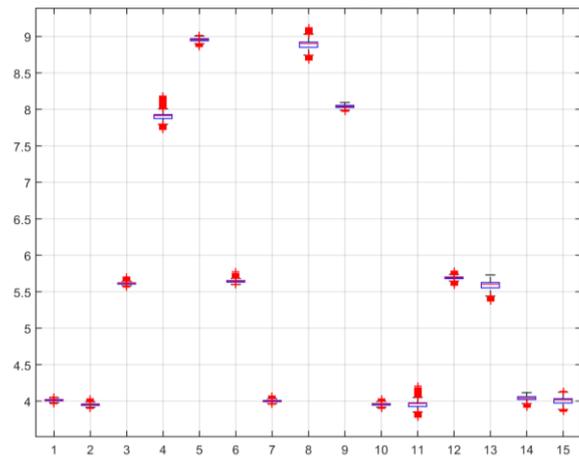


Figure 2. Box plot of all 15 pair wise distances (in cm) for the WAV_N, with the *mkal* device at 0° orientation, all rotation angles and rotational velocities pooled together.

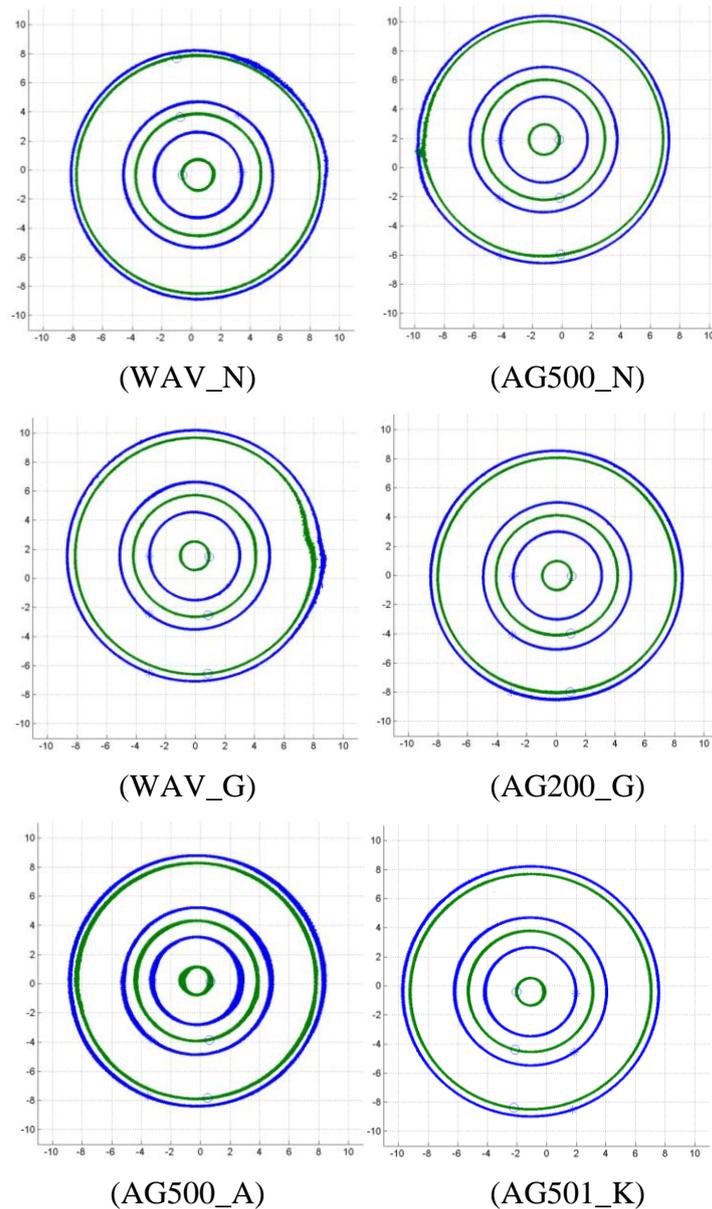


Figure 3. Spatial trajectories of the six sensors, in the reference midsagittal plane for the six devices (X_{2D} / Y_{2D} coordinates in the plane of the *mkal* plate, in cm), for the 0° *mkal* orientation, all θ_{vel} pooled. For each magazine, the small circles denote the positions of the sensors closest to the magazine container considered at the first instant of the recordings in static conditions. The stars correspond to the sensors furthest from the container. Note that if the plate in the initial position for the WAV_N is rotated by 180° , all initial positions are identical for all devices, except for the AG501_K, for which the magazines were unfortunately plugged in an opposite direction in the container.

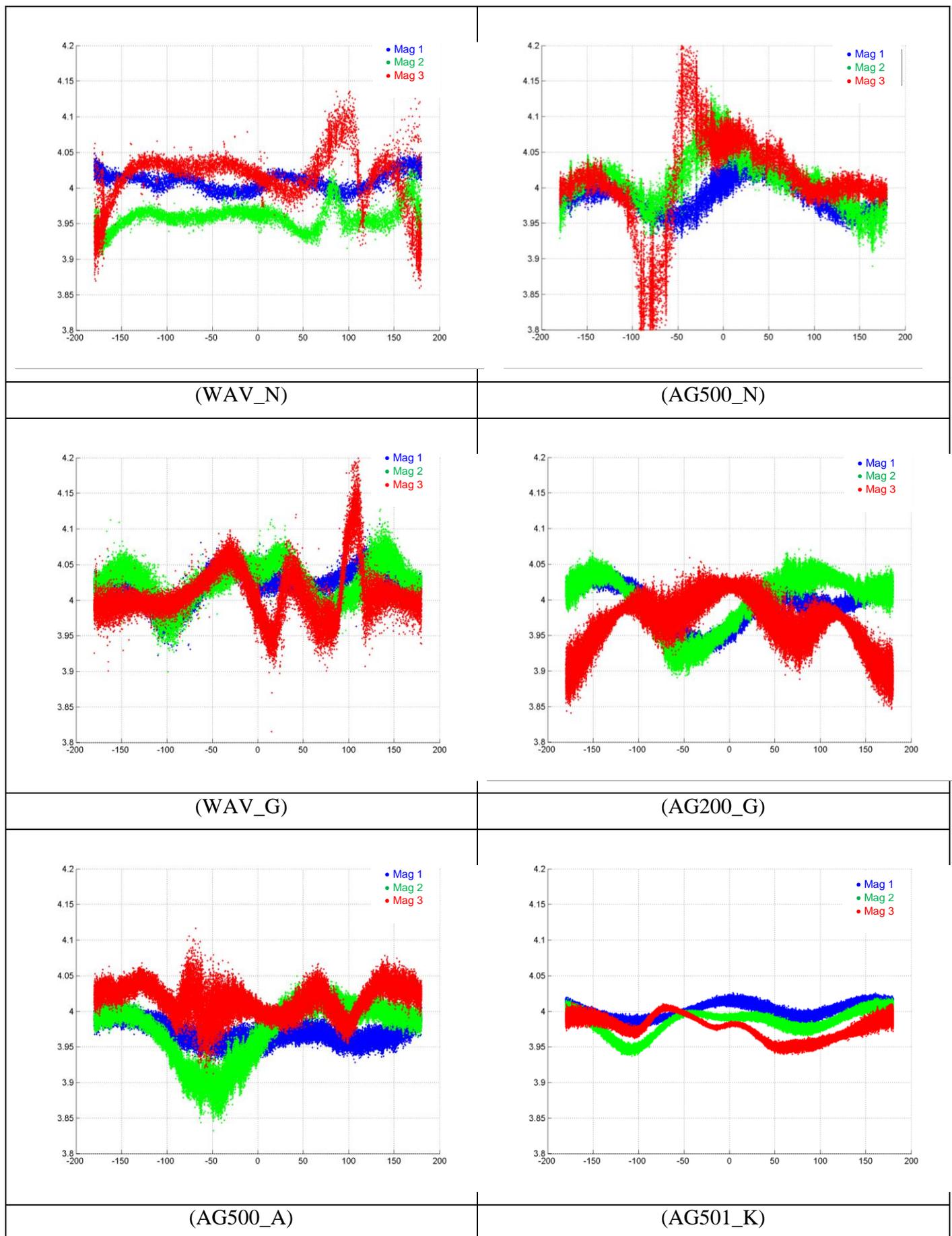


Figure 4. Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ ($^{\circ}$) for the six devices, all conditions pooled (see §0) for 0° *mkal* orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.

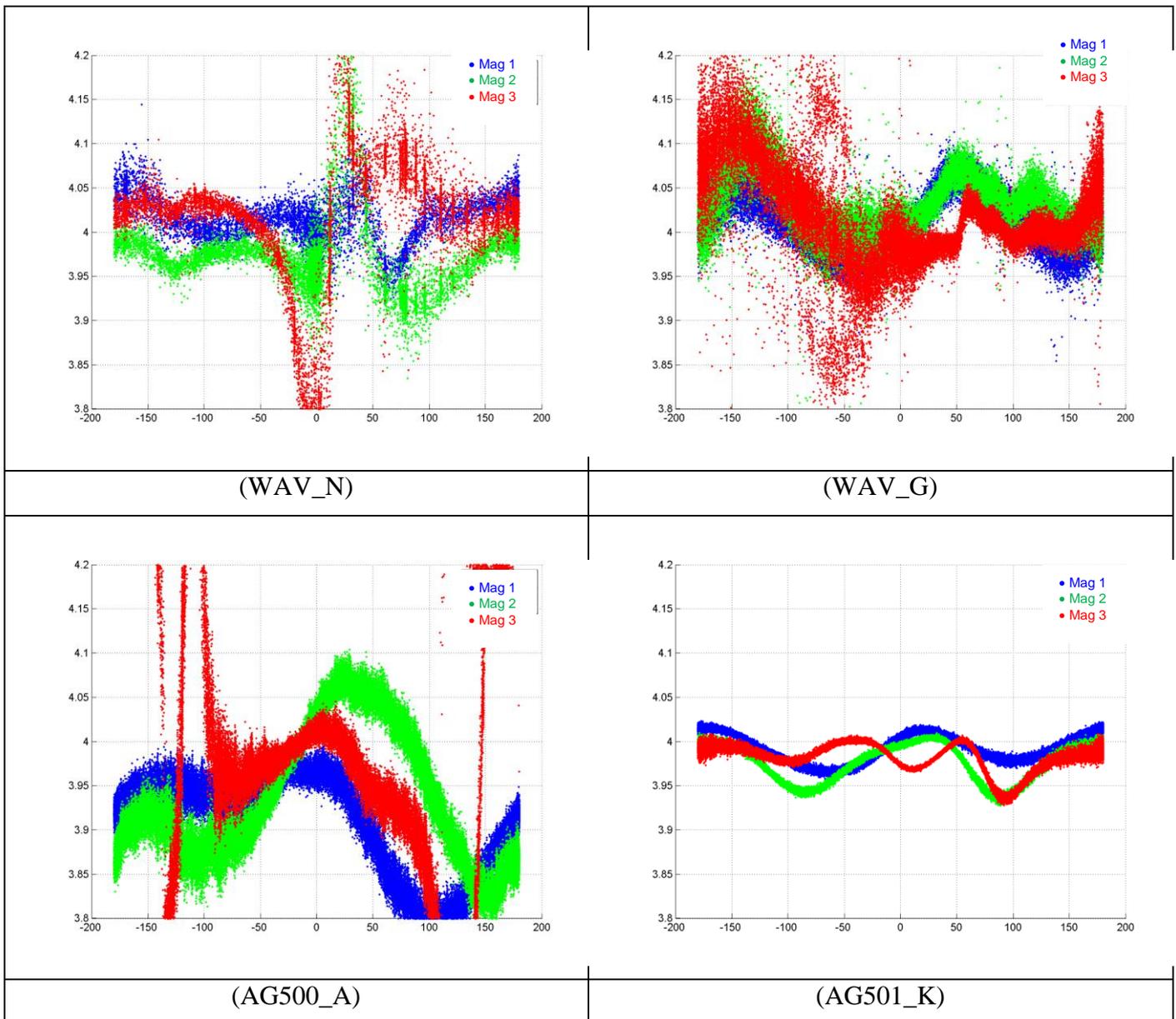


Figure 5: Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ ($^{\circ}$) for four devices, all conditions pooled (see §2.2) for $\pm 45^{\circ} mka$ orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.

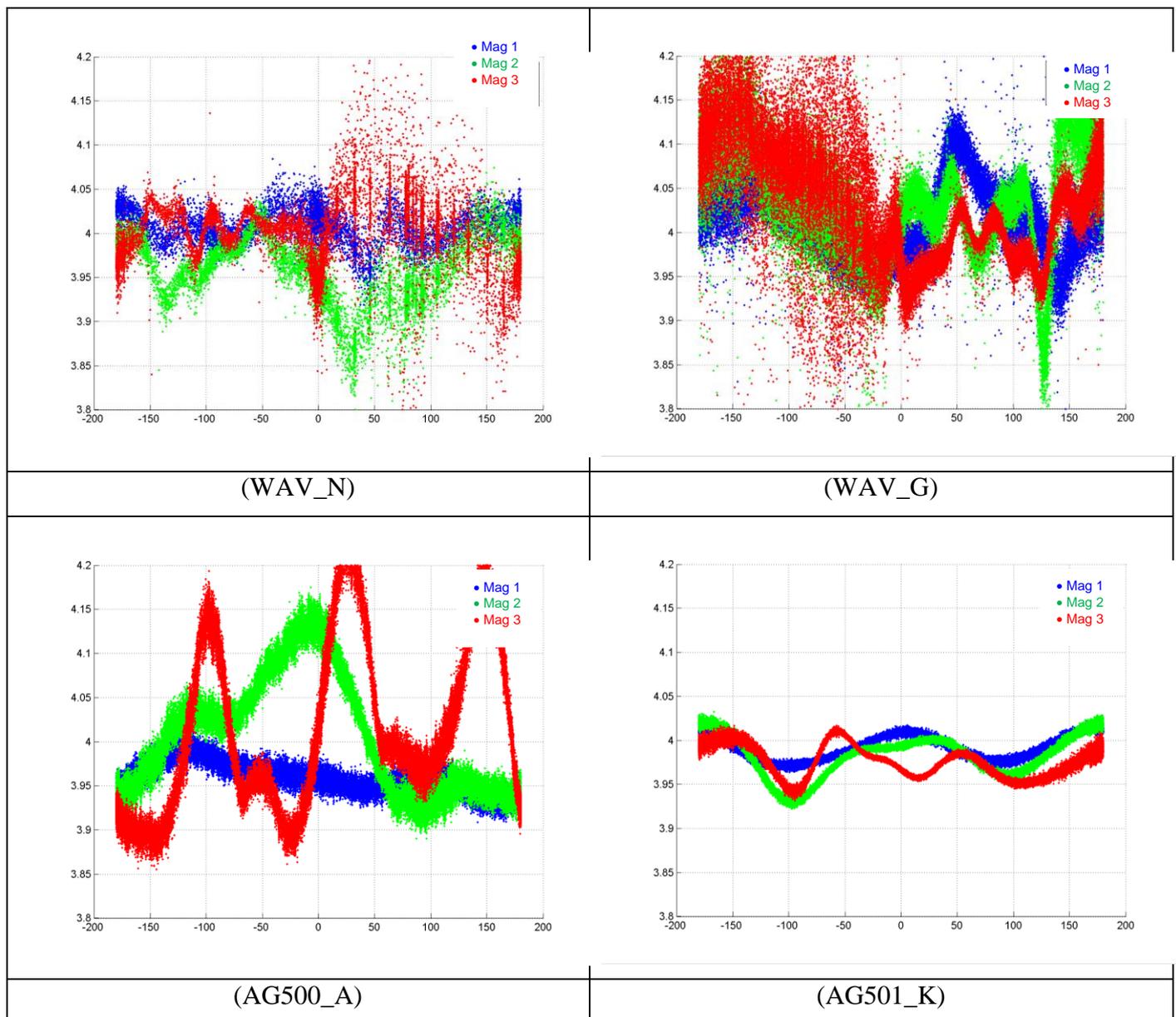


Figure 6: Distances $D_{i=1:3}$ (cm) as a function of plate rotation θ ($^{\circ}$) for four devices, all conditions pooled (see §2.2) for $\pm 90^{\circ}$ *mkal* orientation. Magazine #1 (blue) is positioned at 0 cm from the center, magazine #2 (green) at 4 cm and magazine #3 (red) at 8 cm.

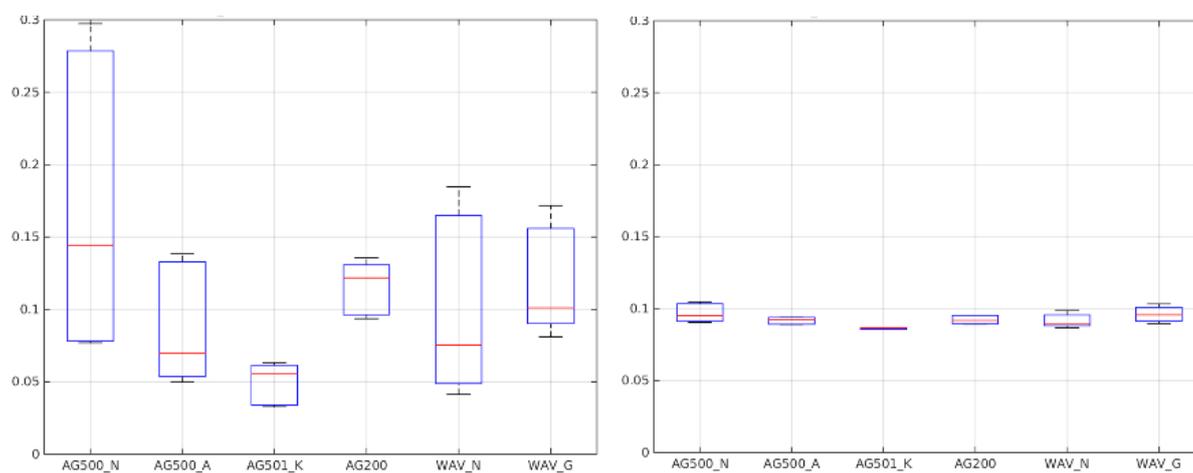


Figure 7. Box plots of QR95_gbl (left) and QR95_loc (right) in cm for the six devices, for all magazine and all θ_{vel} pooled, for 0° *mkal* orientation. In these plots, the central red line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers.

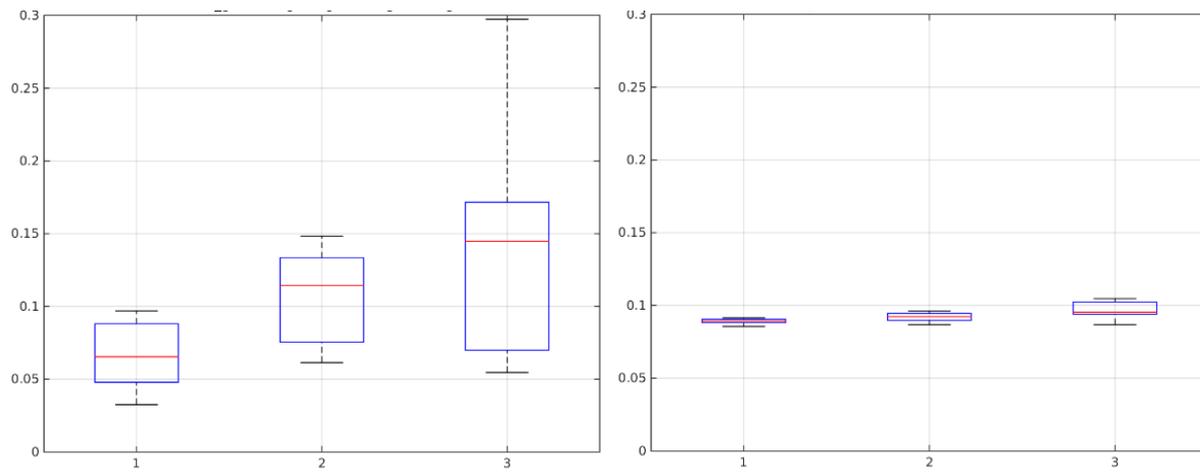


Figure 8. Box plots of the QR95_gbl (left) and QR95_loc (right) in cm for the three magazines, for all devices and all θ_{vel} pooled, for 0° *mka* orientation. In these plots, the central red line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers.

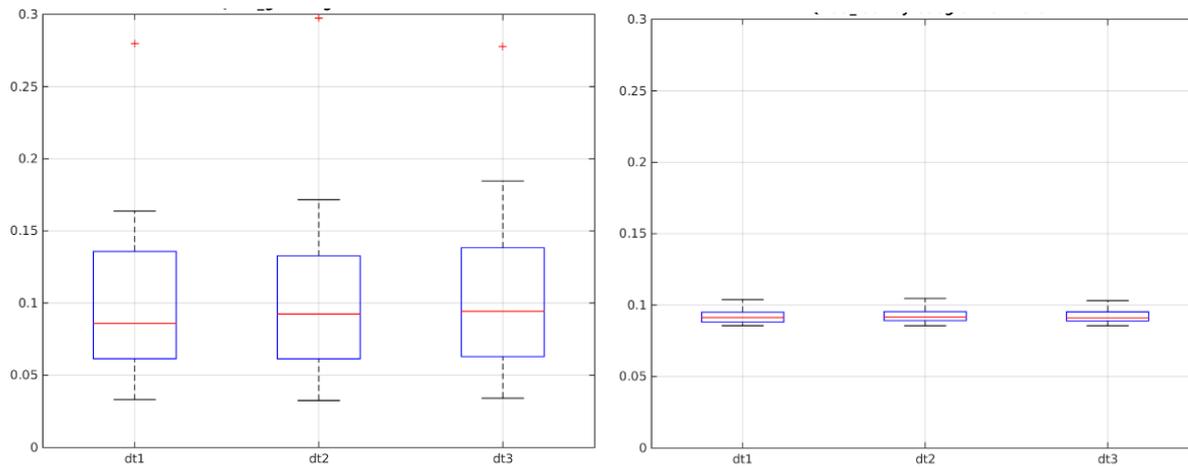


Figure 9. Box plots of the QR95_gbl (left) and QR95_loc (right) in cm for the three θ_{vel} , for all devices and all magazine pooled, for 0° *mkal* orientation. In these plots, the central red line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as red “+”.

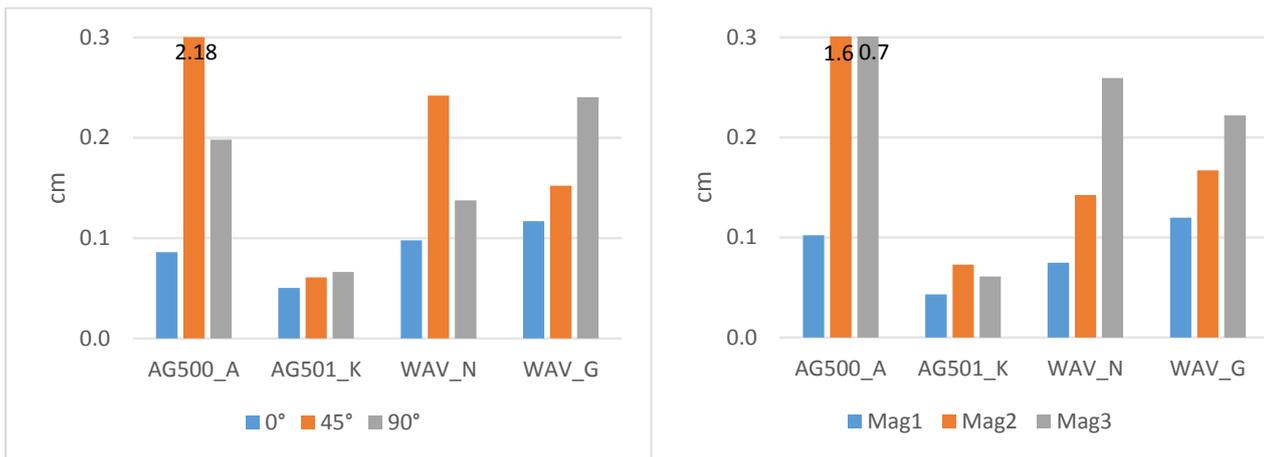


Figure 10: Means of QR95_gbl in cm for the three *mkal* orientations, all magazines and θ_{vel} pooled (left), and for the three magazines, all *mkal* orientations and θ_{vel} pooled (right). Note that large values have been clipped for AG500_A (values are indicated on the graph).

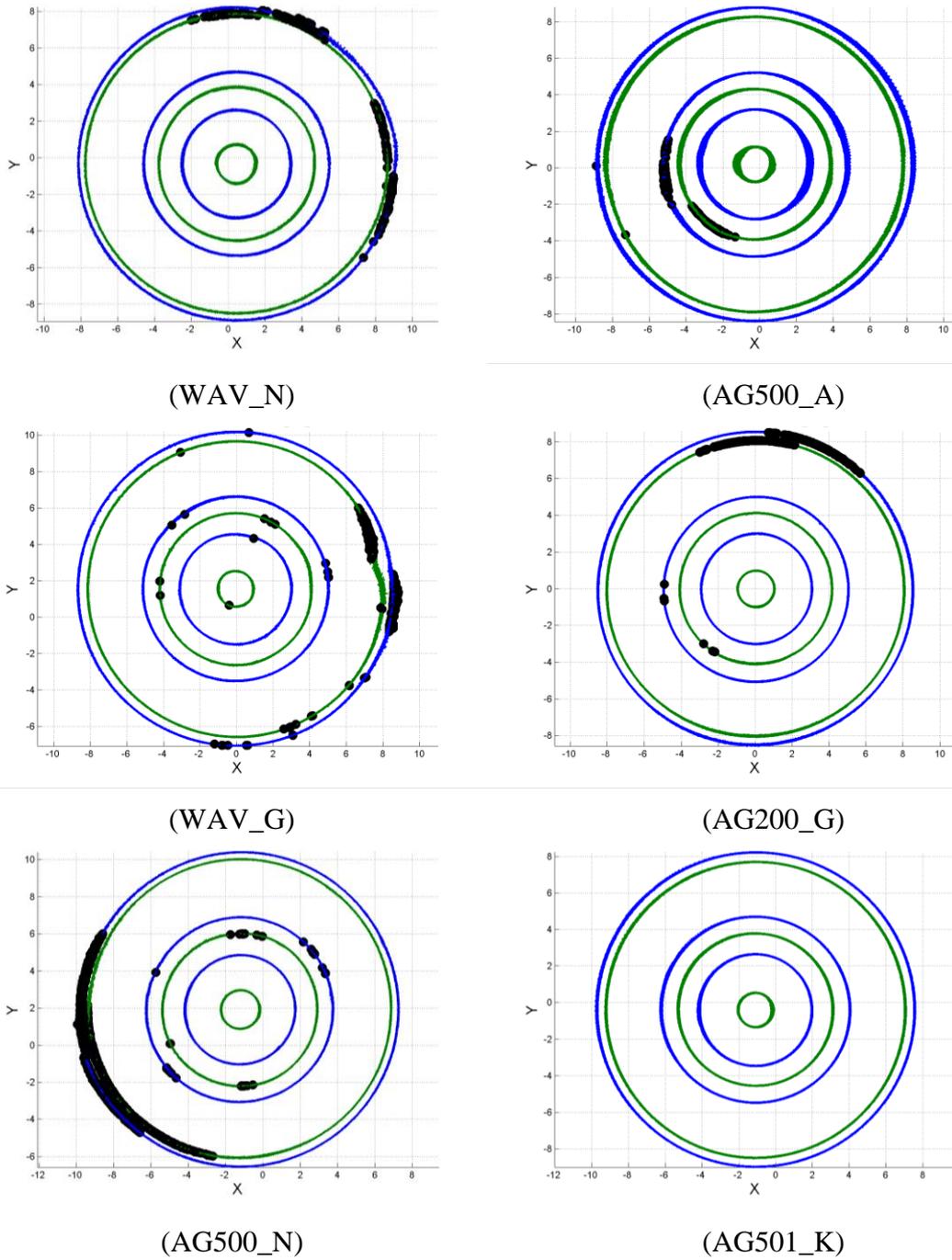


Figure 11. Spatial trajectories of the six sensors (as in Figure 3), where points for which the distance departs from the average distance (computed over all data pooled for the 0° *mkal* orientation) by more than 0.1 cm have been marked with black circles.

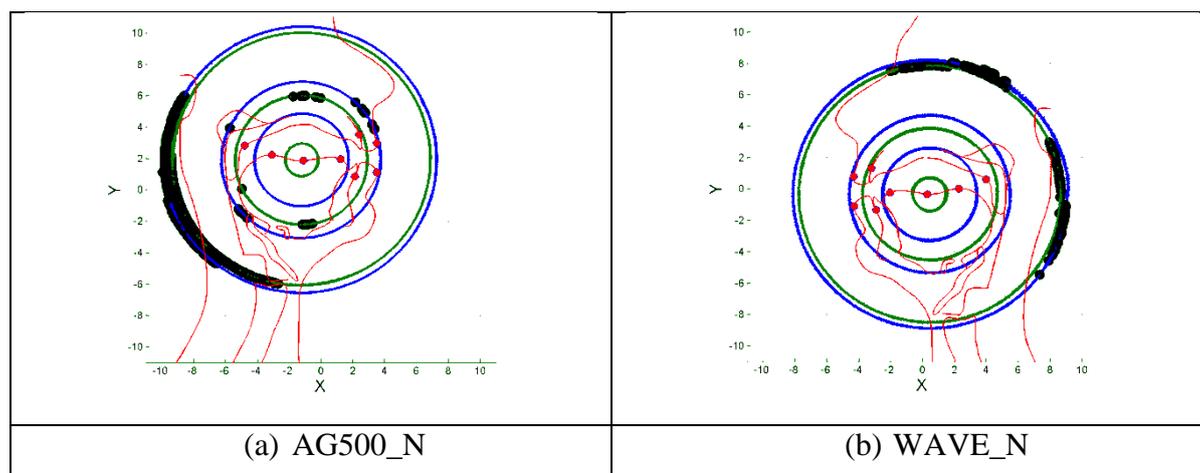


Figure 12. Illustration of the optimal placement of the speaker's head: outlines of the articulator's contours (red) and example of sensor positions (filled red circles) superposed to the same elements as in Figure 11 for the AG500_N and the WAVE_N.

10 List of Tables

Table 1: Comparison of different evaluation studies.....	31
Table 2. Summary of experimental and analysis conditions.	32

Author	Device	Distance between sensors attached to the jaw	Distances between sensors attached to a solid object	Volume exploration	Guidelines	Absolute error	Metrics
Kröger <i>et al.</i> (2008)	NDI Aurora	2 coils nodding movements, /ba/ syllable, one sentence, reading task) One speaker	2 sensors (2 cm apart) on a ruler Manual movements 3 speeds (metronome)	One plane			Standard deviation of Euclidean mean distance between sensors
Yunusova <i>et al.</i> (2009)	Carstens AG500	2 coils (/a/ and /ba/ syllables, one sentence, reading task) One speaker	2 sensors (1.25 cm apart) on a cartridge manual movement without speed control	3 planes (coronal, sagittal, transverse)	Center of measurement volume	Reference: perfect circular trajectories with constant Z, R, and θ fitted to measurements	Absolute mean-corrected Euclidean distance for <ul style="list-style-type: none"> • Median • IQR • 95% • Maximum error
Berry (2011)	NDI WAVE	2 coils (reading task) 10 speakers	<i>Static</i> : 6 sensors on a ruler positioned at various locations in a 3D cube <i>Dynamic</i> : 6 coils on Lego building block Dynamic movements different speeds	3D measurements in a cube volume	No effect in vertical direction Distance from field generator \in [5 - 20 cm] Midsagittal speaker face profile // field generator	Reference: distance between sensors averaged across samples recorded during 30 seconds for static conditions and 10 seconds for jaw recordings.	RMS for <ul style="list-style-type: none"> • Median • IQR • 95% quantile • Maximum error
Stella <i>et al.</i> (2012)	Carstens AG500		9 sensors (spaced by 1.3 cm) on a static support (<i>Circal</i>) Circular movements at one slow constant speed	9 circular trajectories	Bottom, left side	Reference: expected perfect circular trajectories	

Kroos (2012)	Carstens AG500		12 sensors on a rigid object Manual movements (translational, rotational, unconstrained)			Reference: positions provided by an Optical Vicon device.	RMS error of the EMA location coordinates relative to the locations predicted using the OPT system
Stella et al. (2013)	Carstens AG501		16 sensors on a static support (<i>Circal</i>) Circular movements at one slow constant speed	16 circular trajectories			RMS between the measured and the estimated binary induced amplitudes

Table 1: Comparison of different evaluation studies.

Device, place, date	Name	A	B			
WAVE, Nancy, June 2011	WAV_N	X	X			
AG500, Nancy, June 2011	AG500_N	X				
AG200, Grenoble, April 2014	AG200_G	X				
WAVE, Grenoble, April 2014	WAV_G	X				X
AG501, Cologne, May 2014	AG501_K	X		X		
AG500, Aix, June 2014	AG500_A	X			X	
Mkal orientation						
0°		X	X	X	X	X
45° (clockwise)			X	X		
45° (anti-clockwise)					X	X
90°			X	X	X	X

Table 2. Summary of experimental and analysis conditions.