[title page]

Accuracy Assessment of Two Electromagnetic Articulographs: NDI Wave and NDI Vox

Teja Rebernik[1*], Jidde Jacobi[1,2], Mark Tiede[3], Martijn Wieling[1,3]

[1]Center for Language and Cognition, University of Groningen, The Netherlands
[2]Department of Cognitive Science, Macquarie University, Australia
[3]Haskins Laboratories, New Haven, CT

*Corresponding author: t.rebernik@rug.nl

Abstract

Purpose: This study compares two electromagnetic articulographs (EMA) manufactured by Northern Digital, Inc.: the NDI Wave System (2008) and the NDI Vox-EMA System (2020).

Method: Four experiments were completed: (a) comparison of statically positioned sensors; (b) tracking dynamic movements of sensors manipulated using a motor-driven LEGO apparatus; (c) tracking small and large movements of sensors mounted in a rigid bar manipulated by hand; and (d) tracking movements of sensors rotated on a circular disc. We assessed spatial variability for statically positioned sensors, variability in the transduced Euclidean distances (EDs) between sensor pairs, and missing data rates. For sensors tracking circular movements, we compared the fit between fitted ideal circles and actual trajectories.

Results: The average sensor pair tracking error (i.e., the standard deviation of the EDs) was 1.37 mm for the WAVE and 0.12 mm for the VOX during automated trials at the fastest speed, and 0.35 mm for the WAVE and 0.14 mm for the VOX during the tracking of large manual movements. The average standard deviation of the fitted circle radii charted by manual circular disc movements was 0.72 mm for the WAVE sensors and 0.14 mm for the VOX sensors. There was no significant difference between the WAVE and the VOX in the number of missing frames.

Conclusions: In general, the VOX system significantly outperformed the WAVE on measures of both static precision and dynamic accuracy (automated and manual). For both systems, positional precision and spatial variability were influenced by the sensors' position relative to the field generator unit (FGU; worse when further away).

Introduction

Electromagnetic articulography (EMA) has been used for more than three decades in the study of speech and other processes involving movements of the articulators (e.g., mastication or deglutition; Schönle et al. 1987, Perkell et al. 1992, Hoole & Zierdt 2010). Commercial 5D (three spatial and two angular) dimensional articulographs are currently produced by two manufacturers, Carstens Medizinelektronik GmbH (Bovenden, Germany) and Northern Digital Inc. (NDI; Waterloo, Canada). NDI first entered flesh-point tracking technology market with its Aurora system (which was evaluated for the study of speech production by Kröger et al., 2008), followed by the release of two portable articulographs for speech tracking that are the focus of this paper. These two systems are the NDI Wave System (WAVE) which was released in 2008 and its successor, the NDI Vox-EMA System (VOX) which was released in 2020. Both systems have since then been discontinued.

This paper aims to compare the accuracy and reliability of the WAVE system to that of its successor, the VOX. We follow the general outline of previous researchers who have assessed available articulographs using various evaluation methods (see Table 1 for a brief overview). Two prior studies have evaluated the accuracy of the WAVE, namely Berry (2011) who focused exclusively on the WAVE, and Savariaux et al. (2017) who published a comparative evaluation of all then-available commercial articulographs.

Table 1: *Previous studies assessing the precision of EMA systems (from the most recent to the oldest).*

Authors	Device(s)	Evaluation method
Sigona et al. (2018) Stella et al. (2013) Stella et al. (2012)	AG500 AG501	Mounted rotating disk for automated controlled and regular rotations along the vertical axis; wooden stick (Stella et al., 2012) or plastic grid with holes (Stella et al., 2013) attached to the *Circal* disk; speech tracking
Savariaux et al. (2017)	NDI Wave Carstens AG200 Carstens AG500 Carstens AG501	*Mkal* device manually rotated in four conditions: static tracking with complete stops at 24 reference marks; slow, fast and very fast dynamic tracking

Hoole (2014)[1]	Carstens AG500 Carstens AG501	Calibration data; stability of head-movement correction of two systems
Kroos (2012; 2008)	Carstens AG500 Vicon	Rigid plastic container: static trials; manual movements of the container in all directions
Berry (2011)	NDI Wave	Static tracking with rigid body positioned at different locations; dynamic tracking of a rigid body made of LEGO building blocks; speech tracking
Yunusova et al. (2009)	Carstens AG500	Mounted rotating disk for controlled movements; dynamic tracking of a cartridge with sensors moved manually; speech tracking
Kröger et al. (2008)	NDI Aurora	Rigid object (ruler) tracking; speech tracking
Frantz et al. (2003)	NDI Aurora	Robot arm; hemispherical calibration device; handheld ceramic ball-bar

Berry (2011) was the first to evaluate the accuracy of the WAVE, using three experiments. First, he tested static tracking by attaching the sensors to a rigid body (namely, an engineer's scale) and positioning this rigid body at four different locations within the recording volume. Second, he tested dynamic tracking by building a rigid body with LEGO blocks. Sensors were attached to various structures that followed different trajectories and therefore moved at different speeds. The speed of movement was not manipulated across trials. Third, he tested speech tracking of jaw movement for 10 speakers. Berry's experiments showed that in optimal conditions, the WAVE has a tracking accuracy of 0.5 mm or better for 88% of tracked dynamic samples, and an accuracy of 0.5 mm or better for 95% of position samples during jaw-movement tracking in 9 out of 10 speakers. Extreme tracking errors (>2 mm) occurred in the 200-mm near field for <1% of the samples. When recording with the 500 mm cube field setting (as opposed to 300 mm cube), errors exceeded 1 mm when the sensors were further than 20 cm from the field generator.

The second study assessing the WAVE's accuracy was carried out by Savariaux et al. (2017). So far, this remains the only study to conduct a large-scale comparison of several systems from both Carstens and NDI, as they included the AG200, AG500, AG501 and NDI Wave. They

[1] The report from Hoole (2014) was based on recording experiences with the two systems rather than targeted accuracy tests.

embedded three pairs of sensors in an *mkal* device, which had originally been designed by Carstens GmbH for calibration of the AG100 and consists of a rotatable mechanical arm with a magazine that can hold five sensors. Savariaux et al. (2017) made recordings using the *mkal* device in four conditions. The first condition included static measures at 24 different reference marks, the remaining three conditions included dynamic manual rotations at three speeds (slow, fast, and very fast). They found that tracking accuracy was around 0.1 mm for all devices, but that it also depended on the position of the sensors within the operational measurement volume (300 mm cube in the case of the WAVE). For the WAVE, in particular, they found that the largest errors occurred on the positive side of the *x*-axis, with some errors on the positive side of the *y*-axis (see Fig. 5 below for axis orientations relative to the WAVE field generator unit; FGU). Consequently, the optimal recording region was associated with negative *x* and *y* values, leading Savariaux and colleagues to suggest that the participant should optimally be positioned with their right ear facing the text "WAVE" on the FGU plate. Furthermore, the WAVE's accuracy depended on how far the sensor was from the field generator unit (less accurate when further away) but not on the sensor velocity. Overall, the study showed that the AG500 and the WAVE performed similarly, but that the AG501 outperformed both. Noteworthy is that Savariaux and colleagues tested two NDI Wave devices in two different laboratories and obtained different results, suggesting that the data collection environment also plays a role (or that sensor and device performance may not be uniform).

In line with the previous studies, we carried out assessment in our lab to compare the accuracy of the WAVE and the VOX. In the following sections, we first describe the specifications, components, and software of both systems (see section *System components and specifications*). Then, we discuss four tracking experiments that were used to assess the precision of static sensors, and the tracking accuracy for automated (motor-driven) and manual movements (see the *Methods* section for a detailed description of the method and procedure). The first

experiment tested static performance of sensors at known offsets from the FGU. The second experiment tested dynamic tracking of sensors affixed to a LEGO apparatus driven by a motor at three speeds. The third experiment tested sensor tracking accuracy mounted in a rigid LEGO bar manipulated by hand for small (speech-scale) and large movements. The fourth and final experiment tracked movements of sensors rotated on a circular disc by hand.

System components and specifications

The WAVE and the VOX (Figs. 1 to 4) are portable EMA systems capable of tracking the position of sensors within the vocal tract. Both systems support up to 16 channels (for at most 16 sensors) and can record movements at 100, 200 or 400 Hz.[2] The two systems were released a decade apart, but the general functioning principles remain the same: a set of strobed transmitters mounted at different orientations within the field generator unit (FGU; see Fig. 1, above) induce current flow in the sensors. By measuring these flows, the spatial position and orientation of each sensor are determined with reference to the known characteristics of the electromagnetic field using proprietary software (raw data is not available to users). NDI considers sensors to be fully interchangeable and thus no calibration procedure is necessary. Note that descriptions that follow are based on our experience, combined with information found in the official Wave User Guide (2018, revision 12) and the official Vox-EMA System User Guide (2019, revision 4).

Both NDI systems have an FGU of similar size. The WAVE field generator is 200 mm x 200 mm x 80 mm, weighs 3.2 kg and has an operating frequency of 3.2 kHz. It has

[2] While the WAVE required a high-speed upgrade to achieve a recording speed of 400 Hz, the VOX supports this without an upgrade.

a measurement volume that is a 500 mm cube[3], offset from the front face of the field generator by 40 mm. The VOX field generator is 200 mm x 200 mm x 70 mm, weighs 2.2 kg and likewise has an operating frequency of 3.2 kHz. It has a dome-shaped measurement volume of 660 mm, which is offset from the field generator by 50 mm. When active, the VOX FGU emits a louder high-pitched sound than the WAVE FGU.

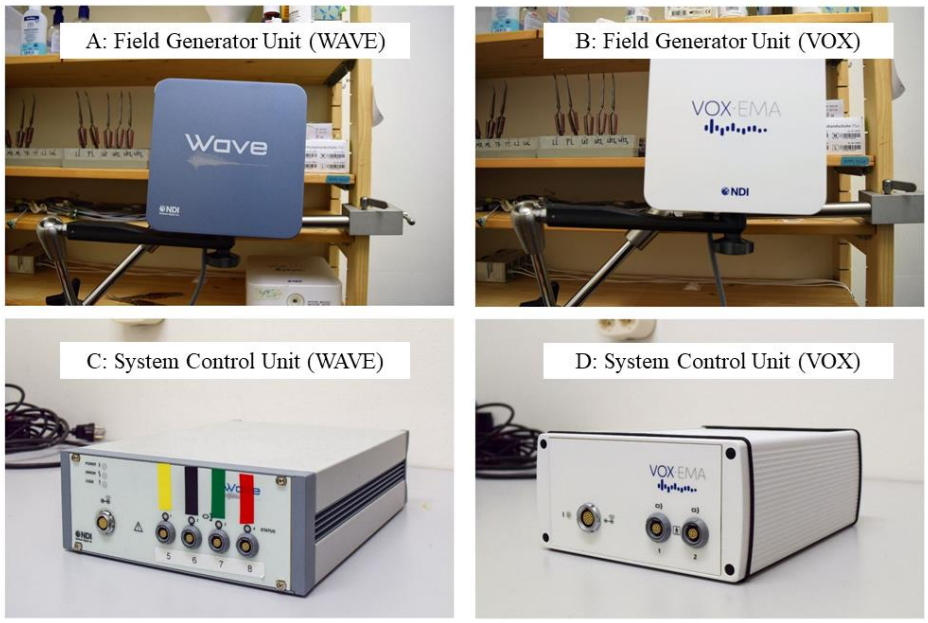

Figure 1: *Above: Field Generator (Transmitter) Units for the WAVE (1-A) and the VOX (1-B). Below: System Control Units (SCU) for the WAVE (1-C) and the VOX (1-D). Note that the WAVE does not come with coloured markers (these were added by our lab).*

The two devices come with the same field generator mounting arm (weighing ~5 kg; see Fig. 1, above). In both systems, the FGU is connected to a System Control Unit (SCU; Fig. 1, below), which is in turn connected to sensor interface units (SIUs). To use 16 sensors, the WAVE system needs two (connected) SCUs, which connect to eight SIU ports (i.e., four SIU ports per

[3] Earlier versions of the WAVE recording software (WaveFront) allowed the user to set the recording volume to either a 300 mm cube (used in the accuracy studies of both Berry, 2011, and Savariaux et al., 2017) or to a 500 mm cube. The final version of the software, which we used, records a volume of 500 mm by default.

SCU), while the VOX only needs one SCU with two SIU ports. The SCUs of the VOX are quieter, as they do not have fans. The SIUs of both systems are depicted in Fig. 2 (above).

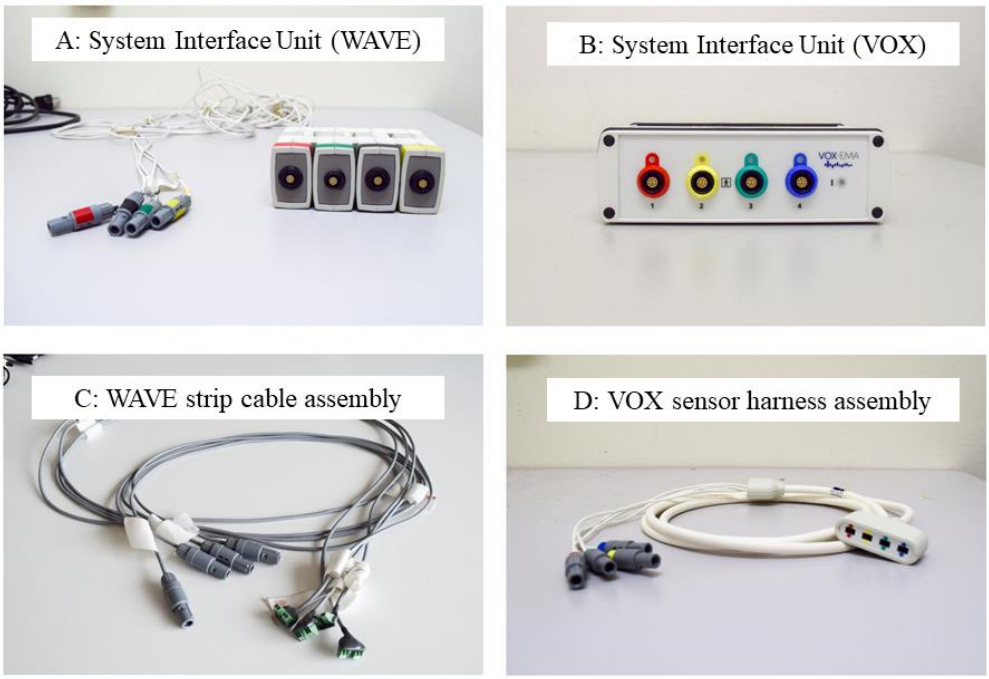

Figure 2: *Above: System Interface Units for the WAVE (1-A) and the VOX (1-B). Note that the WAVE does come with coloured markers (these were added by our lab). Below: WAVE strip cable assembly (2-C) and VOX sensor harness assembly (2-D).*

For the WAVE system, the eight SIUs connect to eight terminal strip cable assemblies (Fig. 2-C, below), with a SIU connector on one side and two sensor ports on the other. Each strip cable assembly thus enables the use of two sensors and eight strip cable assemblies are needed for the use of a complete set of 16 sensors. For the VOX, a sensor harness assembly (Fig. 2-D, below) has SIU connectors on one side and sensor ports on the other, and is thus plugged directly into each of the two SIUs. Two sensor harness assemblies are needed for a complete set of 16 sensors. Compared to the WAVE system, the VOX thus has fewer SIUs (two instead of eight), replaces terminal strip cable assemblies (which enable the connection of two sensors) with sensor harness assemblies (which enable the connection of eight sensors) and, consequently, has fewer parts (see schematic in Fig. 3). In addition, all connections of the VOX are clearly marked and colour-coded, which also makes it easier to set up. Both the WAVE and

the VOX are marketed as portable systems, but this is more easily achieved with the VOX due to fewer components and a lighter overall system weight.

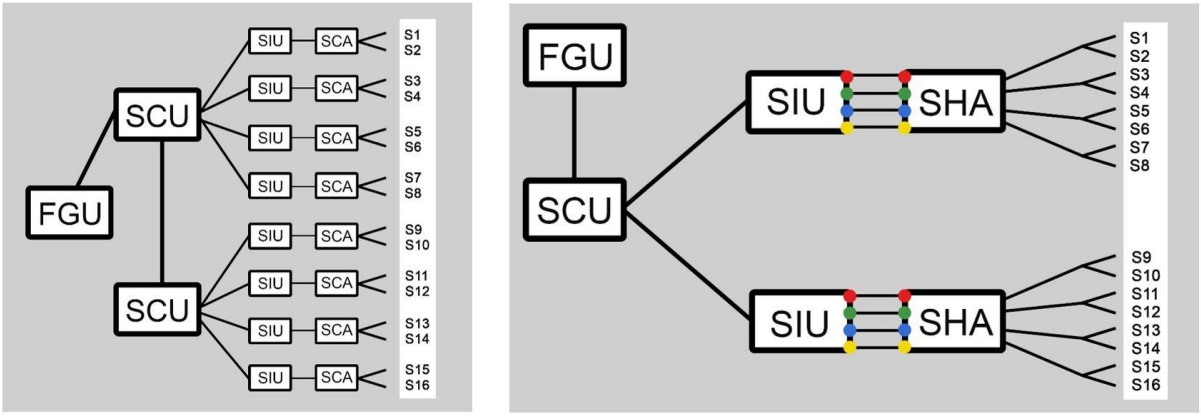

Figure 3: *Architecture of the WAVE (left) and the VOX (right): FGU (field generator unit), SCU (system control unit), SIU (system interface unit), SCA (strip cable assembly), SHA (sensor harness assembly), and S1-S16 (sensors).*

Significant changes have also been made to the sensors. There are two types of sensors for both devices: 5-degree-of-freedom sensors (5DOF), which track three dimensional position information (*x*-, *y*- and *z*-coordinates) and two angular coordinates (rotation around the *x*- and *y*-axis), and 6-degree-of-freedom sensors (6DOF), which additionally track a third angular coordinate (rotation around the *z*-axis). The 6DOF sensor (consisting of two 5DOF sensors mounted in a known mutual orientation) can be used to generate a new coordinate system such that the movement of the 5DOF sensors is tracked relative to the position and orientation of the 6DOF sensor. A 6DOF sensor used in this way attached to a speaker's head, for example, can be used to compensate for head movement in sensors tracking vocal tract articulators.[4] Both the

[4] However, it may be better to correct for head movement using at least three 5DOF sensors instead of one 6DOF sensor. Many labs (including ours; see Rebernik et al., forthcoming, 2021) use this approach because the effects of noise from individual sensors is reduced when the distance between them is increased. Furthermore, if four 5DOF sensors are used, head movement correction is still possible if one of the sensors malfunctions.

WAVE and the VOX can simultaneously accommodate 16 5DOF sensors or, alternatively, 14 5DOF sensors and 1 6DOF sensor (as the latter takes up two sensor ports).

WAVE 5DOF sensors have a square sensor head that is approximately 3 mm in length and 3 mm in diameter. Each sensor is attached to a wire pair whose ends are individually screwed into a strip cable assembly. Both the sensor head as well as most of the wire are encased in transparent plastic. VOX 5DOF sensors are oblong-shaped, with the sensor 7.3 mm long and 2.3 mm in diameter, and they come in pairs. Rather than being screwed in, a pair of sensors can be plugged into one slot of the sensor harness assembly, and the sensor wires are thicker, more flexible, and therefore less likely to break. The paired sensors have tips of a different colour, which enables quick recognition of the channels in the recording software. See Fig. 4 (left) for a side-by-side comparison of WAVE and VOX 5DOF sensors. The Wave 5DOF sensors are marketed as being disposable (although many labs reuse them after cleaning), whereas the VOX 5DOF sensors are explicitly marketed as being reusable, and cost about sixteen times as much.

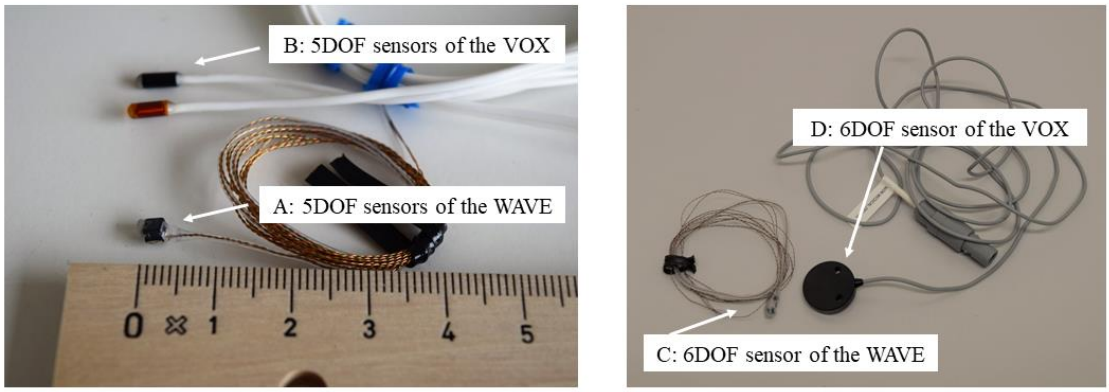

Figure 4: *Left: 5DOF sensors of the WAVE (4-A) and the VOX (4 B). Right: 6DOF sensors of the WAVE (4-C) and the VOX (4-D).*

Besides the differences in the size and shape of the sensor tips, there are also differences in the setup. The sensor wires are the same length for both systems (~1 m), but the wires of the VOX sensors are heavier (6g versus <1g), which means they need to be attached to the participant

more securely as the weight of the cable would otherwise put more strain on the attached sensor. The WAVE terminal strip cable assembly is shorter and lighter (1.8m, 133g) than the VOX sensor harness assembly (2.5m, 200g). However, whereas the WAVE SIUs have a long cable connecting them to the SCU, the VOX SIUs do not, which means they cannot be placed closer to the participant. Finally, while there are benefits to the paired sensors of the VOX, including an easier setup, a full pair needs to be replaced if a single sensor malfunctions.

There are also differences regarding the 6DOF sensor which can be used for automatic head movement correction during recording. The WAVE 6DOF sensor (Fig. 4-C) is oblong-shaped and screwed into the terminal strip cable assembly as two sensors. Consequently, the WAVE 6DOF sensor can function as a pair of 5DOF sensors (positioned close to each other). On the other hand, the VOX 6DOF sensor is a disc (Fig. 4-D), which is plugged directly into the SIU and is thus recognized by the system as a single sensor. Both systems also offer a separate 6D palate trace tool.

Movements can be viewed and recorded using NDI software WaveFront (2.0) and VOX-VRI. The VOX-VRI package allows the attribution of labels (names and colours) to sensors, which can be saved for future experiments (i.e., they remain stored even after the system or the laptop reboots). This functionality is not available in WaveFront. With VOX-VRI, the rotational coordinates (besides the three-dimensional positional information) can be recorded in either Euler angles or quaternions. The WAVE only allows recording of the rotational coordinates using quaternions. Both devices support a TCP/IP accessible server which enables researchers to control the system using third-party software (e.g., Matlab or Python).

Method

The purpose of the experiments conducted in this study was to compare the performance of the WAVE and VOX systems. For all tests, the field generator units (FGUs) of the WAVE and

VOX systems were positioned on a table in the same marked location and stabilized using duct tape (see Fig. 5 for positioning and marked axes). LEGO Technic parts were used to control repeated movements of the sensors (see description in subsection *Description of the LEGO apparatus*). Data could not be collected for both systems simultaneously due to their mutual electromagnetic interference. Both VOX and WAVE data were collected in the quaternion format, at 400 Hz, using VOX-VRI v3.0.61 and WaveFront v2.2.1 respectively. A Lenovo ThinkPad P52 laptop running Windows 10 and paired with an external USB sound digitizer (TASCAM US-600) was used in all experiments. The WAVE system we used for the experiments has been in use since 2013, while the VOX system has been in use since 2019.

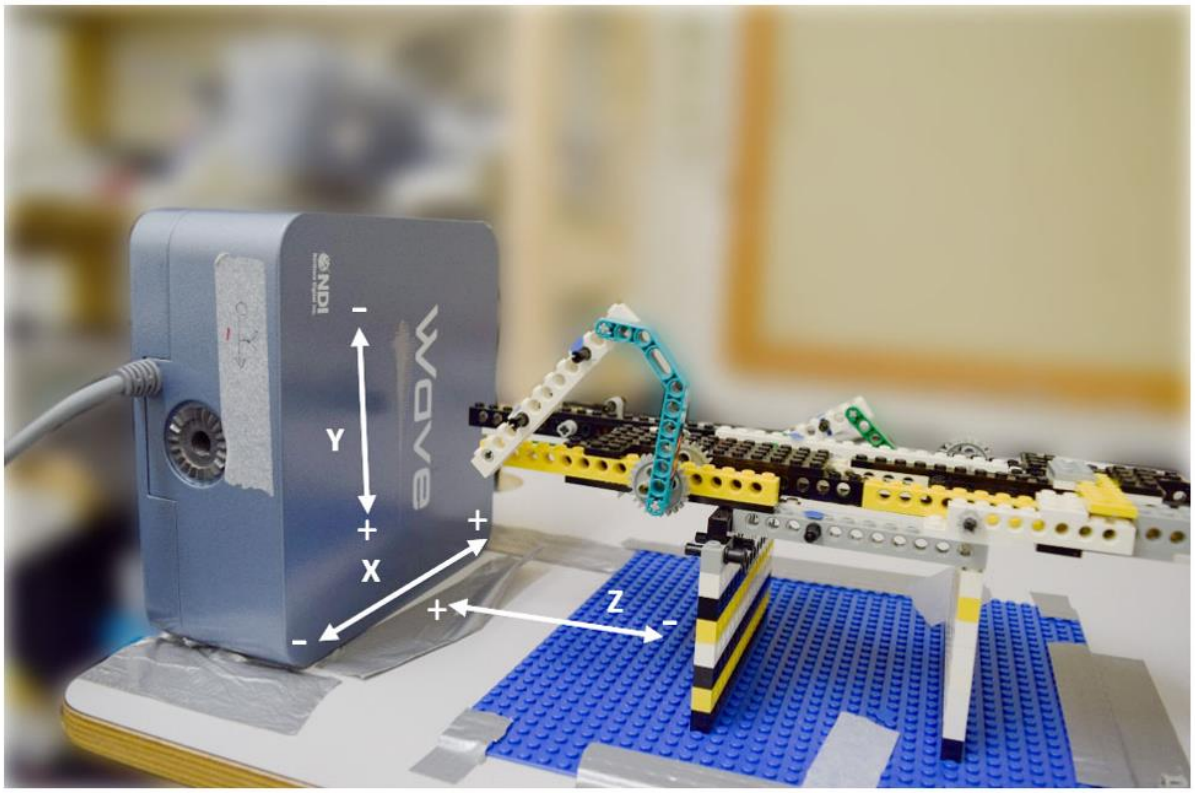

Figure 5: *Placement of FGU showing orientation of system axes (coordinate center is the center of the plate surface).*

For each of the four experiments, we collected four trials, using different sets of sensors. Specifically, we used two sets of sensors during the four trials of the static and manual dynamic tests (i.e., two trials recorded with each sensor set). Two different sensors sets were used for

the four trials of the automated tests (likewise, two trials were recorded per sensor set). This ensured that our results are more reliable and measurement errors across trials cannot be attributed to a defective sensor. As the four trials showed similar patterns, the values in the tables of this paper have been averaged across all trials. Likewise, the statistical analyses associated with the tables take into account all data.

Static tests

The static tests were performed to establish baseline noise levels in the absence of movement and to compare transduced with directly measured positions. Eight sensors were attached firmly to a LEGO bar (distance between sensors at least 10 mm) using LEGO Technic pins and placed on the LEGO baseplate at multiple distances and orientations relative to the FGU. Four sensors were attached on the left side of the LEGO bar and four sensors on the right side (during *z*-axis displacements, this meant that four sensor wires were pointing towards and four sensors away from the FGU; see Fig. 6). The LEGO bar with sensors was displaced 15 times along the *x*-, *y*- and *z*-axes. Twelve measures were made with the sensors positioned perpendicular to the FGU. These included six displacements along the *z*-axis, three displacements along the *y*-axis and three displacements along the *x*-axis. In addition, we included three measures along the *z*-axis made with the sensors positioned parallel to the FGU.[5]

Relative distance of the LEGO block from the FGU along the *z*-axis ranged from 87 mm to 335 mm for the WAVE and from 97 mm to 345 mm for the VOX. The difference of ~10 mm was to take into account the distance of the recording volume from the FGU for the VOX (50 mm) and WAVE (40 mm). The displacements along the *x*-axis and *z*-axis are additionally visualized

[5] When we refer to "parallel" and "perpendicular" trials, we refer to the position of the *sensors* relative to the FGU. Due to the way sensors were mounted in LEGO pins, this means that the LEGO bar itself was oriented parallel to the FGU in perpendicular trials, and perpendicular to the FGU in parallel trials (also see Fig. 6).

208 in Fig. 7 (for a single sensor). A video of the static tests has been made available online (Speech

209 Lab Groningen, 2020a).

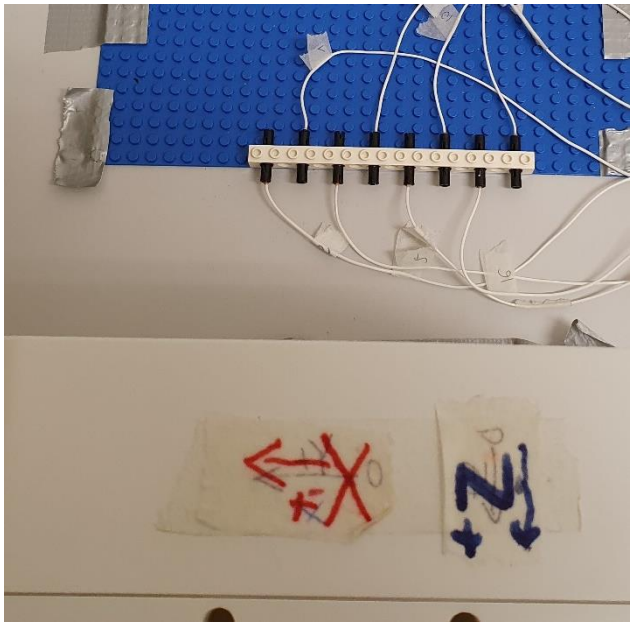

Figure 6: *LEGO block during static trials, in position 1 (the block is parallel to FGU; sensor heads are perpendicular to FGU).*

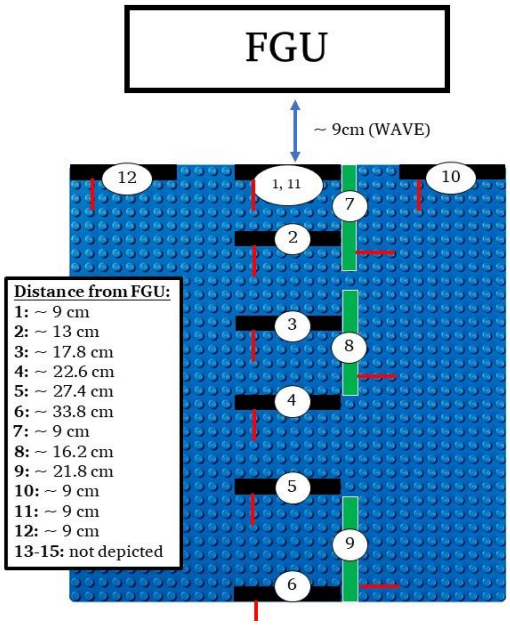

Figure 7: *Relative positions of static sensors (including distance from the FGU in the legend). The black square represents the LEGO block while the red line represents the wire of a single sensor (for illustration purposes). Positions 1-6 denote displacements along the z-axis (sensors perpendicular to the FGU), positions 7-9 denote displacements along the z-axis (sensors parallel to the FGU), and positions 10-12 denote displacements along the x axis (sensors perpendicular to the FGU). Positions 13-15 are not depicted but consisted of displacements along the y-axis when the block was in position 1 (the block was raised by two or five LEGO blocks).*

Two measurements were obtained: the standard deviation (SD) of transduced coordinates for each position per sensor, and comparisons of hand-measured distances of the LEGO block from the FGU plate vs. the transduced spatial positions of the sensors.

Dynamic tests

Dynamic performance was assessed in three conditions. In the first condition, a LEGO apparatus (see *Description of the LEGO apparatus*, below) was used to measure continuous movement of four pairs of sensors, each pair attached to one moving LEGO rigid body (see *Automated dynamic tests procedure*), plus one additional pair of sensors which was attached to the LEGO baseplate that did not move. In the second condition (see *Manual dynamic tests*), we recorded manual movements of a LEGO bar to which eight sensors were attached (same bar as used for static trials in Fig. 6). In the third condition, circular symmetry was assessed by recording two sensors on a compact disc rotated within its (handheld) jewel case (see Fig. 8).

Figure 8: *Sensors placed on CD (here depicted on the LEGO plate for easier visibility).*

Description of the LEGO apparatus

For the automated dynamic tests, we used LEGO Technic blocks to build an apparatus that could hold sensors in a fixed place, in line with the approaches of other researchers (see e.g., Berry (2011) who also used a custom-built LEGO apparatus, or Savariaux et al. (2017) who used the *Mkal* device). Our apparatus (see Fig. 9) was used to move four rigid bodies systematically within the tracking field, each geared to be driven by the same electric motor that was positioned approximately 0.75 meters away from the FGU to minimize field interference. Each rigid body held one pair of sensors.

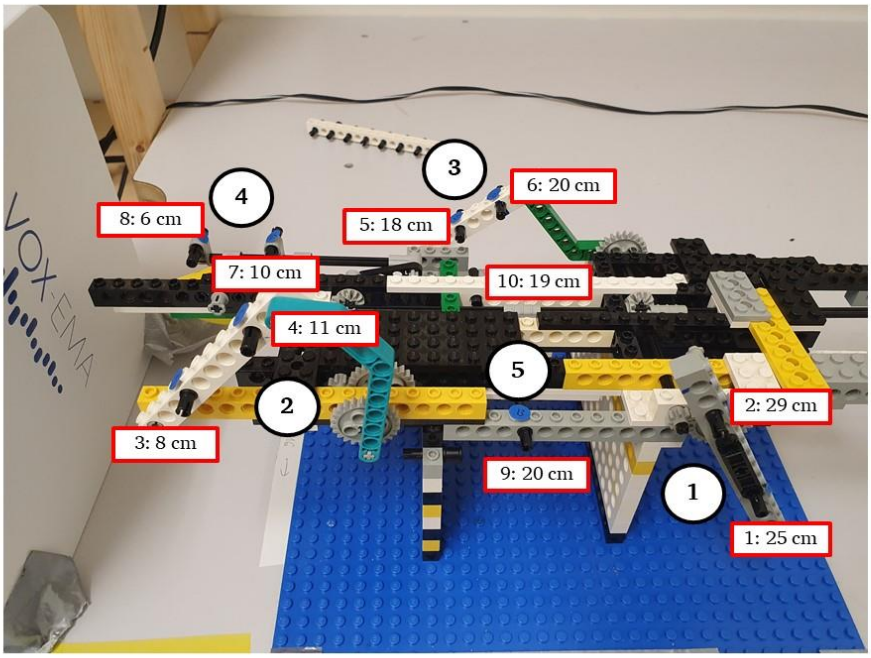

Figure 9: *LEGO apparatus for automated tests. 1: circularly rotating bar (sensors 1-2), 2: eccentric beam moving over a large range (sensors 3-4), 3: eccentric beam moving over a small range (sensors 5-6), 4: bar moving as a piston (sensors 7-8), and 5: static position (sensors 9-10). Approximate hand-measured distance between each sensor and the FGU are indicated.*

Sensors 1 and 2 were placed on an arm with a common axis rotating with circular motion (Fig. 9: number 1). Sensors 3 and 4 were placed on an elbow beam moving with eccentric motion over a large range (Fig. 9: number 2). Sensors 5 and 6 were placed on an elbow beam moving with eccentric motion over a small range (Fig. 9: number 3). Sensors 7 and 8 were attached to

a bar moving as a piston aligned with the long axis of the apparatus (Fig. 9: number 4). Finally, Sensors 9 and 10 were attached to the baseplate and served as static sensors (Fig. 9: number 5). A video of the LEGO apparatus in motion tests has been made available online (Speech Lab Groningen, 2020b). Fig. 9 additionally denotes the relative hand-measured distance of each sensor in the apparatus from the (front plate of the) FGU along the *z*-axis.

The sensors were held in place using LEGO Technic pins. These pins had an opening diameter of 3.4 mm, but due to the friction ridges, the effective internal diameter was smaller. The pins provided an immovable fit for both sensor types: specifically, VOX sensors fit snugly due to the cylindrical shape of the sensor head, while the WAVE sensors fit tightly in the pins due to the presence of two slots at the end of the pin, which enabled the two halves to extend slightly. The sensors of both systems were placed in the same LEGO Technic pin, but the exact position inside the pin varied depending on the sensor shape and size. All sensors were placed at least 1 cm away from each other, and the distances between adjacent sensors ranged between 23.9 mm and 89.9 mm (more details about these measurements can be found in the *Results* section).

As the sensors were attached to different (circularly or eccentrically rotating) structures, the relative velocities of their movement at each speed also differed depending on their exact placement (e.g., higher speed for sensors located further from a pivot point). The average sensor velocities, as measured by the VOX, ranged from 0.9 cm/s (a sensor attached to the large eccentric structure) to 14.8 cm/s (a sensor attached to the circular structure) in the slow condition, from 1.7 cm/s to 25.2 cm/s (same sensors) in the medium condition, and from 2.7 cm/s to 35.2 cm/s (same sensors) in the fast condition. While the speed of the circular structure was usually much higher than that of the other rigid bodies, the speed of the piston in the fast condition (18.9 cm/s) was slightly faster than the speed of the circular structure in the slow condition. In general, the sensor velocities were comparable to or faster than movement

of articulators during speech, which show peak velocities around 20 cm/s (e.g., Ostry & Munhall, 1985; Tasko & McClean, 2004).

Automated dynamic tests procedure

The goal of the automated dynamic tests was to evaluate the performance of both systems with sensors in motion over different orientations and positions within the field. Tests were performed with the LEGO apparatus positioned such that its long axis (and the movement of sensors) was oriented orthogonal to the FGU, as this allowed us to place the motor as far away from the FGU as possible. Sensors were positioned in such a way that all sensor cables were at least 3 cm away from the FGU (as recommended by NDI), but still within the limits of the recording volume. Trials were collected at three motor speeds: “slow”, “medium” and “fast”. We additionally collected a static recording in which no sensors were moving.

We recorded four trials of 20 seconds for each speed setting which at the 400 Hz sampling rate resulted in 8000 recorded samples per trial. Accuracy was assessed by determining the standard deviation (SD) of the Euclidean distances between sensor pairs on the same rigid body, and a comparison of hand-measured vs. transduced distances between sensors. We additionally evaluated the circle charted by the circularly rotating LEGO bar, by comparing the radius of the ideal circle fit to the recorded samples.

Manual dynamic tests

The goal of the second set of dynamic tests was to test tracking accuracy for hand-manipulated movements. A video of the manual dynamic tests has been made available online (Speech Lab Groningen, 2020a; from 2′30″). As described above, eight sensors attached to a LEGO bar were systematically moved along each of the three axes of the system. These hand-manipulated dynamic movements were either “small” (3-5cm) or “large” (10-15cm). For each system, we recorded four 20-second trials at 400 Hz for both types of movements. The same accuracy

assessment measure was used as above, namely the SD of Euclidean distances between sensor pairs.

An additional set of dynamic tests consisted of circle symmetry trials, recorded using five 5DOF sensors, of which three were reference sensors mounted on the corners of a compact disc (CD) jewel case and two more were movement sensors attached to a CD rotated in the jewel case. The CD was held perpendicular to the FGU plate and rotated by hand around the axis of the jewel case. The sensors were secured in LEGO Technic pins. We recorded four 20-second trials at 400 Hz.

Accuracy was assessed by first using the CD case reference sensors to correct all movement to one plane (i.e., similar to head correction), such that *x*- and *y*- dimensions described circular movement on the plane and the *z*-dimension deflection from the plane. The sampled (*x*, *y*) coordinates points were then fit to an ideal circle minimizing the error from that circle, comparing the ideal to the observed circle radius at each point.

Statistical methods

We assessed whether the two systems differed by fitting generalized additive (mixed-effects) regression models (Wood, 2017) in R using the *mgcv* package 1.8.33 (Wood, 2017; Wood, 2011). We used the *itsadug* package 2.4 (van Rij et al., 2020) for visualizing the models, as well as bean plots (Kampstra, 2008) which provide convenient visualizations of the distribution together with the mean (the widest) line.

As dependent variable we used the measurement error (usually in SD) and as single fixed-effect predictor the articulograph. Because the dependent variable was generally right-skewed, we fitted the models using a Gamma distribution with a logarithmic link function. Before fitting the model, we added a small amount to all values (i.e., 0.0001) to ensure no values equal to 0 remained. We included the optimal random-effects structure (including random intercepts and

slopes) including sensor (pair) as the random-effect factor. For numerical predictors, we assessed whether these had a non-linear relationship with the dependent variable (e.g., Wieling, 2018). Our annotated statistical analysis can be found in Appendix C.

Results

In the following section, the tables report mean values of all four trials as does the subsequent analysis. While the four trials showed some variability, the general pattern was consistent.

Static tests

We first examined spatial variability during static trials at different positions in the field. Table 2 reports the average standard deviation across all eight sensors in the LEGO block at 15 static positions (see schematic in Fig. 7 above). Standard deviations per sensor can be found in Appendix A (for Trial 1).

Table 2: *Standard deviations of positional coordinates (in mm) and rotational angles (in degrees), averaged across all eight sensors (the patterns for individual sensors of one example trial can be found in Appendix A). Tx: lateral movement (parallel to the FGU); Ty: inferior superior movement; Tz: anterior-posterior movement (away from the FGU). The values in the table have been averaged across all four trials. Please see Appendix B for the values of individual trials.*

	WAVE					**VOX**				
	Tx (mm)	**Ty (mm)**	**Tz (mm)**	**Pitch (deg)**	**Yaw (deg)**	**Tx (mm)**	**Ty (mm)**	**Tz (mm)**	**Pitch (deg)**	**Yaw (deg)**
***z*-axis (sensor perpendicular to FGU)**										
Position 1	0.02	0.03	0.02	<0.01	0.03	<0.01	<0.01	<0.01	<0.01	0.11
Position 2	0.03	0.04	0.03	<0.01	0.04	<0.01	<0.01	<0.01	<0.01	0.04
Position 3	0.07	0.09	0.08	0.01	0.07	0.01	0.01	<0.01	<0.01	0.06
Position 4	0.14	0.18	0.16	0.02	0.14	0.02	0.03	0.02	0.01	0.11
Position 5	0.29	0.36	0.33	0.04	0.23	0.04	0.05	0.03	0.02	0.18
Position 6	0.68	0.80	0.75	0.06	0.43	0.08	0.09	0.06	0.03	0.27
***z*-axis (sensor parallel to FGU)**										
Position 7	0.08	0.07	0.10	0.05	10.74	<0.01	0.01	<0.01	<0.01	0.03
Position 8	0.25	0.18	0.34	0.12	0.80	0.02	0.02	0.02	<0.01	0.06
Position 9	0.45	0.43	0.61	0.17	1.61	0.04	0.05	0.04	<0.01	0.09
***x*-axis (sensor perpendicular to FGU)**										
Position 10	0.02	0.03	0.03	0.01	0.04	0.01	<0.01	<0.01	0.01	0.16
Position 11	0.02	0.02	0.02	<0.01	0.03	<0.01	<0.01	<0.01	<0.01	0.07
Position 12	0.02	0.03	0.02	<0.01	0.05	<0.01	<0.01	<0.01	<0.01	0.03
***y*-axis**										
Position 13	0.02	0.02	0.02	<0.01	0.03	<0.01	<0.01	<0.01	<0.01	0.08
Position 14	0.02	0.02	0.02	<0.01	0.03	<0.01	<0.01	<0.01	<0.01	0.12
Position 15	<0.01	0.01	0.01	<0.01	0.02	<0.01	<0.01	<0.01	<0.01	0.14

Table 2 shows that the SD's for the WAVE are generally higher than for the VOX. While the different trials varied to some extent, these differences were not significant (all *p*-values > 0.6). The statistical analysis across all trials (visualized in the bean plot in Fig. 10) found significantly lower standard deviations for the VOX in comparison to the WAVE across all measures (β_{log} = 1.7, *SE* = 0.46, $p < 0.001$; effectively the WAVE had an SD which was about 0.1 mm or degree higher than the VOX on the non-transformed scale).[6] We also assessed whether distance from the FGU (only including positions 1 to 6, see Fig. 7) affected system accuracy. This appeared to be the case for both systems (both *p*'s < 0.001), with the WAVE showing a significantly (β_{log} = 0.03, *SE* = 0.006, $p < 0.001$) more detrimental effect of distance from the FGU than the VOX (whose effect was estimated to be somewhat non-linear). Figure 11 visualizes the estimated effects of the model for both articulographs, including the bean plots for both devices per position (cf. Fig. 7). Note that the greater variability for lower values is expected as on the log-scale large negative values represent much smaller values than small negative values.

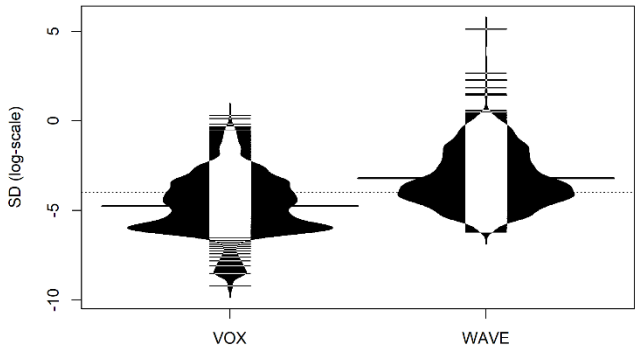

Figure 10: *Bean plot visualizing average standard deviations of the positional coordinates for both devices during the static tests.*

[6] While the units of measurement differed for positional vs. angular coordinates (mm. vs. degrees), we conducted a single analysis as the patterns were similar for both measures, and to limit the number of models fitted.

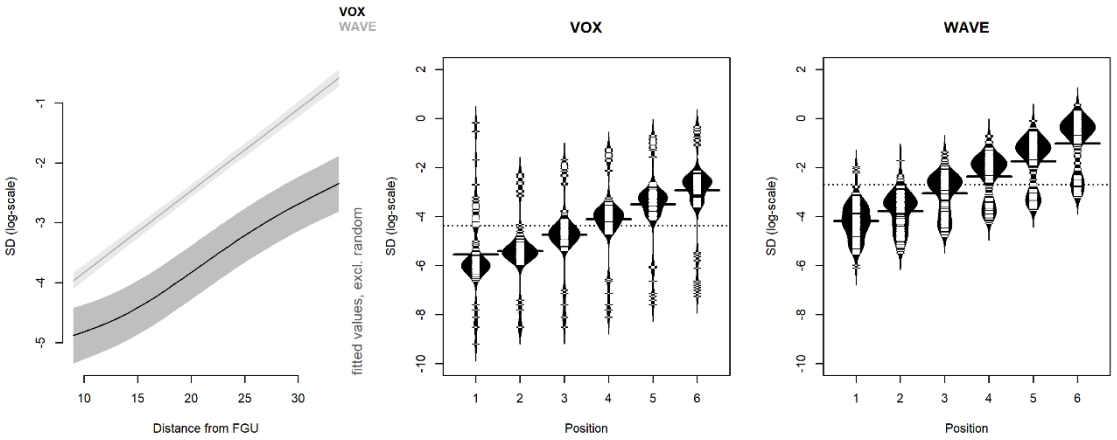

Figure 11: *A. Model predictions for the effect of the distance from the FGU on the SD (log-transformed); B. Bean plot visualizing the relationship between average static SD and position for the VOX; C. Bean plot visualizing the relationship between average static SD and position for the WAVE.*

Table 3 compares the actual (i.e., manually measured) distance between subsequent positions to the distance determined using the reported sensor coordinates.[7] The statistical analysis shows that the VOX and WAVE did not differ significantly regarding the inferred differences in positions (β_{log} = 0.26, *SE* = 0.38, p = 0.5), despite the WAVE showing more extreme errors. This is visualized in Figure 12, which also shows the difference in actual (non-transformed) errors.

Table 3: *Difference between the actual distance between positions (averaged across two sensors; one on each side of the LEGO block) and the inferred distance based on the WAVE and VOX during static trials. Movements were along the z-axis (sensors perpendicular; positions 1-6), along the z-axis (sensors parallel; positions 7-9), along the x-axis (sensors perpendicular; positions 10-12), and along the y-axis (sensors perpendicular; positions 13-15). Positions 6:7 and 9:10 are not included because the movement was not along the same axis. The values in the table have been averaged across all four trials.*

Position pair	**Movement** (axis)	**Actual distance** (mm)	**WAVE** (absolute error)	**VOX** (absolute error)
1:2	z-axis (perp.)	40.0	0.06	0.07
2:3	z-axis (perp.)	48.0	0.16	0.07
3:4	z-axis (perp.)	48.0	0.25	0.13
4:5	z-axis (perp.)	48.0	0.49	0.22
5:6	z-axis (perp.)	64.0	0.98	0.35
7:8	z-axis (par.)	72.0	1.80	0.13
8:9	z-axis (par.)	56.0	1.05	0.34
10:11	x-axis	64.0	0.16	0.06
11:12	x-axis	64.0	0.12	0.58
13:14	y-axis	19.2	0.06	0.19
14:15	y-axis	28.8	0.05	0.43

[7]As we used a LEGO grid (with submillimetre precision), we can estimate the actual movements of the sensors with submillimetre precision, and compare them to the distance measured on the basis of the WAVE and VOX sensor positions.

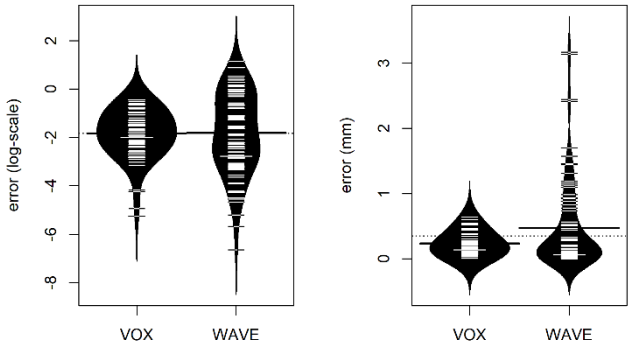

Figure 12: *Bean plot visualizing average error (A: log-transformed, B: non-transformed) of the inferred distances per articulograph.*

Dynamic tests

For the dynamic tests, we examined the *variability* in the Euclidean distances (ED) between all pairs of sensors attached to the same rigid body. Table 4 reports these results. Specifically, we calculated these values for all three dynamic tests: automated dynamic (static and at three increasing speeds), manual dynamic (small and large movement patterns), and circle tests. Besides the standard deviation, we report the range (the difference between the maximum and minimum value), and the 95% range (the difference between the 2.5^{th} percentile and 97.5^{th} percentile) to limit the influence of incidental outliers. In the ideal case, these measures should be close to zero (since the EDs should remain constant during movement, as the sensors are not moving relative to each other). Note that as two sensors are involved in the calculation of EDs, the inaccuracy of a single sensor will be lower than the combined inaccuracy of the ED between sensor pairs. For the circle tests we report the difference between the calculated ideal and observed radius.

Table 4 shows that the VOX consistently outperforms the WAVE. The subsequent figures supplement the table. Figures 13 and 14 visualize the difference between the WAVE and the VOX regarding the stability of the Euclidean distances between several pairs of sensors

352 (specifically for fast automated movements and manual large movements). Figures 15 and 16
353 visualize how well the circles were fit during the automated (Fig. 15; radius ranging from 23.8
354 mm to 32.0 mm) and manual circle rotations (Fig. 16; radius ranging from 29.9 mm to
355 31.7 mm). As Figures 13 to 16 show trial-specific visualizations, these have been based on
356 Trial 1.

Table 4: *The range, standard deviation (SD) and 95% range (all in mm) of Euclidean distances (ED) and radii between pairs of sensors for the static block tests, all rigid bodies of the LEGO apparatus, and the manual block and CD tests. The values for the static block have been averaged across all eight sensors at Position 2 (see Fig. 7; Position 2 was chosen because it was approximately closest to the position of the block during manual movements). The values in the table have been averaged across all four trials. See Appendix B for the values of individual trials.*

		WAVE			VOX		
Rigid body	**Condition**	**SD (mm)**	**95% (mm)**	**Range (mm)**	**SD (mm)**	**95% (mm)**	**Range (mm)**
Dynamic tests (automated)							
Large eccentric ED	*static*	0.03	0.11	0.2	0.01	0.03	< 0.1
	slow	0.09	0.37	0.8	0.04	0.17	0.6
	medium	0.12	0.49	1.1	0.08	0.33	0.9
	fast	0.19	0.76	1.9	0.12	0.50	1.4
Small eccentric ED	*static*	0.10	0.37	0.6	0.02	0.07	0.1
	slow	0.24	0.94	2.4	0.05	0.18	0.3
	medium	0.30	1.09	3.5	0.06	0.22	0.5
	fast	0.36	1.36	4.5	0.07	0.28	0.6
Piston ED	*static*	0.01	0.05	0.1	0.01	0.05	0.1
	slow	0.11	0.41	2.0	0.04	0.17	0.4
	medium	0.21	0.78	3.4	0.07	0.26	0.6
	fast	0.32	1.30	4.3	0.10	0.42	1.0
Circle ED	*static*	0.58	2.26	4.0	0.05	0.21	0.4
	slow	3.28	12.80	60.3	0.21	0.82	1.5
	medium	4.80	17.83	77.5	0.25	0.97	2.0
	fast	5.85	22.63	88.6	0.30	1.16	2.5
Circle radii	*slow*	2.46	9.26	45.7	0.18	0.68	1.8
	medium	3.50	12.61	61.8	0.26	1.04	3.0
	fast	4.28	15.36	73.8	0.31	1.23	3.8
Static ED	*static*	0.10	0.40	0.6	0.01	0.05	0.1
	slow	0.11	0.41	0.7	0.01	0.05	0.1
	medium	0.11	0.44	0.8	0.01	0.06	0.1
	fast	0.11	0.44	0.8	0.02	0.07	0.1
Average variability in ED (all bodies)	*static*	0.16	0.64	1.1	0.02	0.08	0.2
	slow	0.77	2.97	13.4	0.07	0.28	0.6
	medium	1.11	4.13	17.3	0.09	0.37	0.8
	fast	1.37	5.30	20.0	0.12	0.49	1.1
Average variability in ED (all bodies, except circle)	*static*	0.06	0.23	0.4	0.01	0.05	0.1
	slow	0.14	0.53	1.5	0.04	0.14	0.4
	medium	0.19	0.70	2.2	0.06	0.22	0.5
	fast	0.25	0.97	2.9	0.08	0.32	0.8
Dynamic tests (manual)							
Block ED	*static*	0.05	0.17	0.3	< 0.01	0.03	< 0.1
	small	0.12	0.41	3.3	0.06	0.22	0.4
	large	0.35	1.19	10.3	0.14	0.52	1.0

CD rotation - ED	*case*	1.05	4.41	10.6	0.13	0.48	1.3
	CD	0.72	2.59	14.1	0.14	0.51	1.1
CD rotation - radii	*sensor 1*	0.65	2.52	12.6	0.18	0.71	1.5
	sensor 2	0.49	1.90	11.6	0.19	0.72	1.3
	average	0.57	2.21	12.1	0.19	0.72	1.4

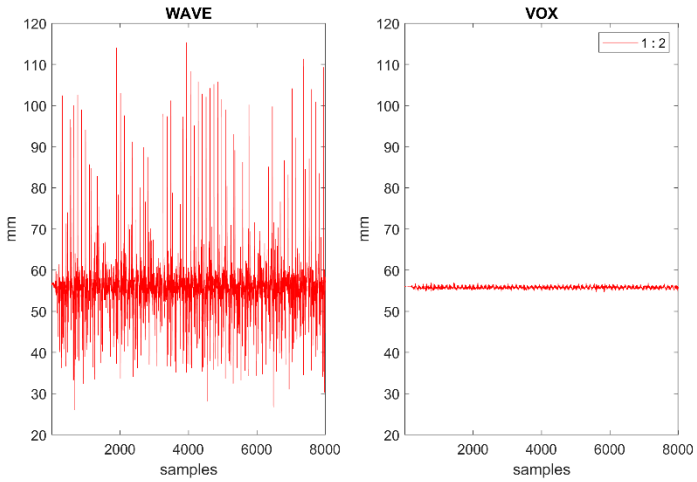

Figure 13: *Euclidean distances between sensor pairs 1:2 on the circularly rotating bar for the WAVE (left) and the VOX (right) during fast automated movements.*

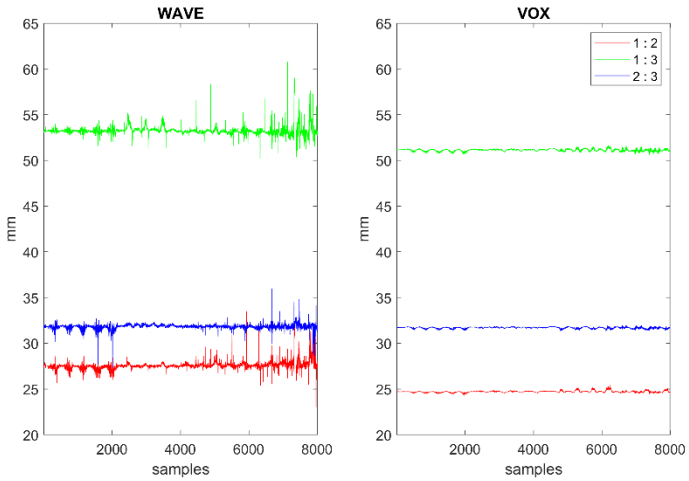

Figure 14: *Euclidean distances between sensor pairs 1:2, 1:3 and 2:3 for the WAVE (left) and the VOX (right) during large manual movements.*

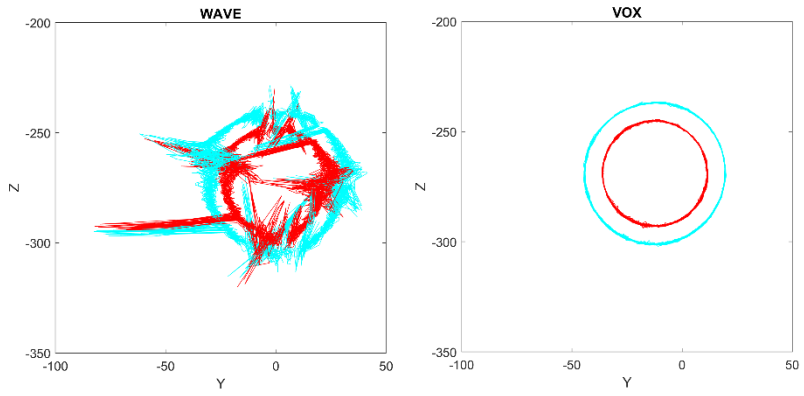

Figure 15: *Fitted circles for fast automated trials based on WAVE (left) and VOX (right) sensor movements.*

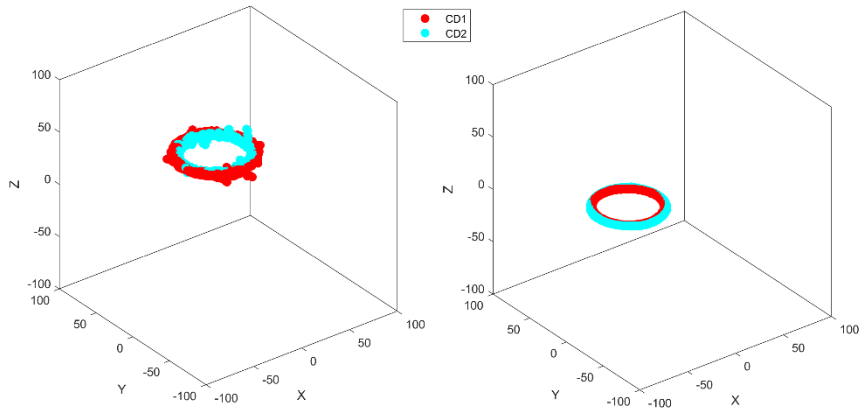

Figure 16: *Fitted circles for manually rotated CD trials based on WAVE (left) and VOX (right) sensors. Only sensors on the rotating CD (CD1, CD2) are depicted.*

The statistical analysis (including all trials, but excluding the manual dynamic movements as these are likely not entirely consistent across trials and articulographs) shows a similar pattern to the one shown in Table 4. The difference between the two articulographs was significant, with the WAVE showing a higher SD than the VOX (β_{log} = 1.7, *SE* = 0.69, p = 0.02). Figure 17 visualizes the general difference, but also the difference per rigid body. Effectively the difference on a non-transformed scale is about 0.1 mm for all rigid bodies, except for the circle, and 3.4 mm for the circle (i.e., at the largest distance from the FGU).

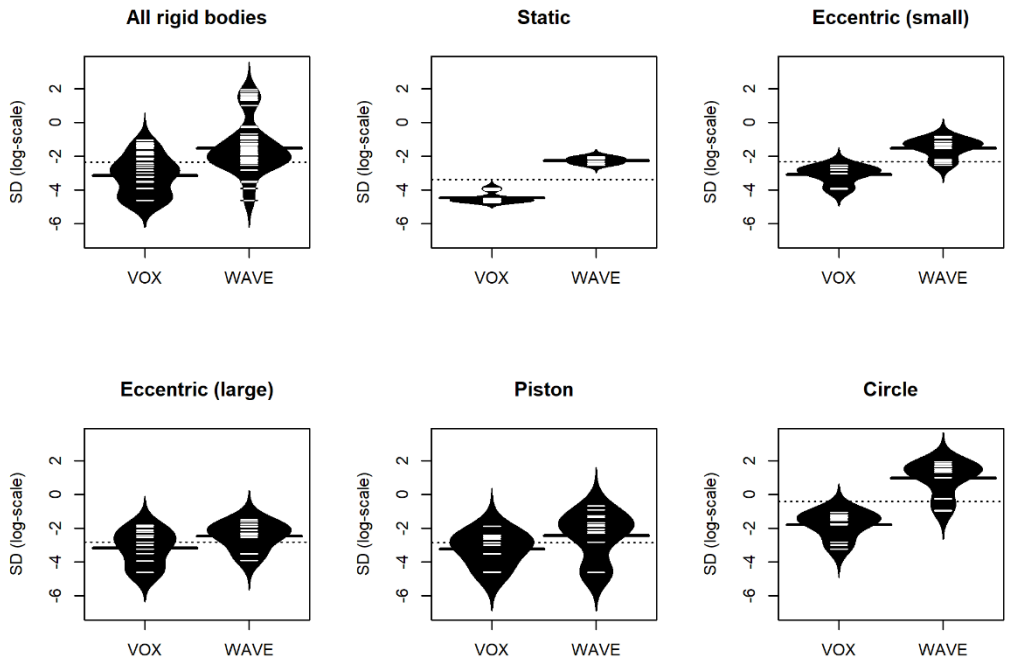

Figure 17: *Bean plot visualizing average SD per articulograph for the automatic dynamic trials for all rigid bodies (A) and the five separate rigid bodies (B-F).*

364 Additionally, we assessed the effect of movement speed on the SDs. For both articulographs,
365 increasing speed resulted in a significant ($p < 0.001$) (non-linear) increase in errors, but the
366 increasing speed was not more detrimental for one system compared to the other ($p = 0.9$).
367 Figure 18 visualizes the result of the model as well as includes bean plots for the different
368 speeds per articulograph.

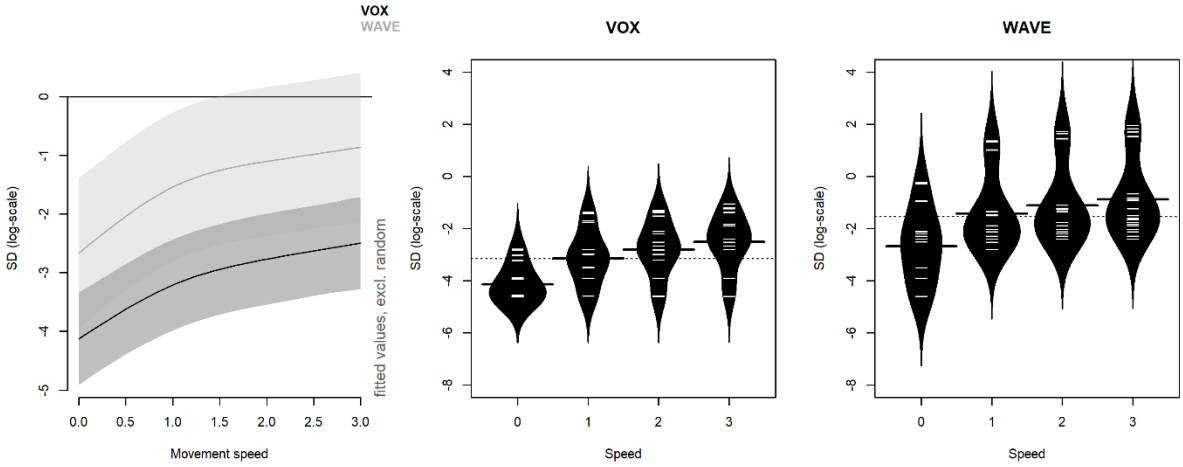

Figure 18: *A. Model predictions for the effect of movement speed on the SD (log-transformed); B. Bean plot visualizing the relationship between average dynamic SD and speed for the VOX; C. Bean plot visualizing the relationship between average dynamic SD and speed for the WAVE.*

Finally, we report missing data percentages for the moving sensors to assess sensor dropouts in Table 5. Here, likewise, the VOX outperforms the WAVE. During fast automated tests, across all four trials, the WAVE had up to 3277 missing frames for a single sensor (6532 missing frames in total for three sensors), while the VOX had 26 missing frames for one sensor (and in total). Averaged across all moving sensors, during all dynamic tests, the total percentage of missing values was 0.85% for the WAVE and 0.04% for the VOX. The statistical analysis (using logistic mixed-effects regression, with the number of missing frames vs. the number of non-missing frames as the dependent variable), however, did not reveal a significant difference between the two (β_{logit} = 1.8, SE = 1.26, p = 0.16). Figure 19 visualizes the (non-significant) difference.

Table 5: *Missing data during dynamic tests (automated and manual), averaged across all four trials. The table lists the total number of frames per sensor (in brackets next to each condition), the number of missing frames, and the percentage of missing frames. A zero indicates that there were no missing frames for this sensor in any trial. The last line (Dynamic tests: all conditions, all sensors) summarizes the number of total missing frames.*

	WAVE			**VOX**		
Dynamic tests (automated) – trial 1						
Sensor	**Condition**			**Condition**		
	Slow (32,024)	**Mid** (32,024)	**High** (32,024)	**Slow** (32,012)	**Mid** (31,997)	**High** (31,980)
1 (circle)	531 (1.7%)	1897 (5.9%)	3277 (10.2%)	0	0	0
2 (circle)	816 (2.5%)	1775 (5.5%)	3196 (10.0%)	0	0	0
3 (eccentric L)	0	0	0	0	0	0
4 (eccentric L)	0	0	0	0	0	0
5 (eccentric S)	0	0	0	0	0	0
6 (eccentric S)	0	0	0	0	0	0
7 (piston)	0	0	0	0	0	0
8 (piston)	71 (0.2%)	25 (0.1%)	59 (0.2%)	0	73 (0.2%)	26 (0.1%)
Dynamic tests: manual – block						
Sensor	**Condition WAVE**			**Condition VOX**		
	small (32,024)	**large** (32,024)		**small** (31,981)	**large** (31,932)	
1	4 (<0.1%)	174 (0.5%)		0	56 (0.2%)	
2	0	12 (<0.1%)		0	0	
3	0	0		0	14 (<0.1%)	
4	0	60 (0.2%)		0	0	
5	0	15 (<0.1%)		0	28 (<0.1%)	
6	0	2 (<0.1%)		0	0	
7	0	20 (<0.1%)		0	14 (<0.1%)	
8	1 (<0.1%)	70 (0.2%)		0	0	
Dynamic tests: manual – CD						
	WAVE (32,024)			**VOX** (32,029)		
1 (CD)	38 (0.1%)			189 (0.6%)		
2 (CD)	66 (0.2%)			189 (0.6%)		
3 (case)	0			49 (0.2%)		
4 (case)	0			0		
5 (case)	0			0		
Dynamic tests: all conditions, all sensors						
total frames	1,441,080			1,439,361		

missing frames	12,208	638
% missing	0.85%	0.04%

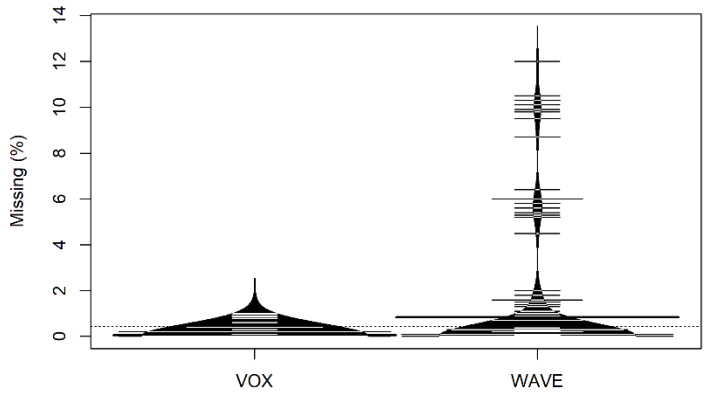

Figure 19: *Bean plot visualizing missing rates (in percentages) per articulograph.*

Discussion

The focus of our study was to compare the tracking accuracy of two NDI electromagnetic articulographs: the NDI Wave System, which was released in 2008, and its successor, the NDI Vox-EMA System, which was released in early 2020.

We evaluated the accuracy using static, automated, and manual movements. The VOX generally outperformed the WAVE according to our evaluation criteria (i.e., having lower variability of Euclidean distances between sensors and less variable fitted-circle radii). When assessing static precision, the WAVE showed significant higher variability than the VOX. In addition, variability significantly increased when sensors were placed further away from the FGU, with a more detrimental effect for the WAVE than for the VOX. For positional coordinates averaged across all sensors (Table 2), we observed an average SD of 0.02 mm for the WAVE and less than < 0.01 mm for the VOX when the LEGO bar was closest to the FGU (about 9 cm), and an average SD of 0.7 mm for the WAVE and 0.08 mm for the VOX when the LEGO bar was furthest away (about 30 cm). Neither positional (*x*, *y*, *z*-coordinates) nor rotational (pitch, yaw) precision was entirely uniform across sensors (see Appendix A for static precision of individual sensors during Trial 1).

Next, we compared the hand-measured vs. transduced positional information. Examining the position in the field during static trials (Table 3), we found that the expected vs. transduced displacements for two sensors along the *z*-axis differed up to 1.80 mm for the WAVE and up to 0.58 mm for the VOX (the further away the sensors were, the less accurate the detected change in position). These differences were not significant, however.

During dynamic trials, the VOX showed ED standard deviations of up to 0.3 mm (see Table 4, rows of *automated dynamic tests*), which is about ten times lower than those associated with the WAVE measurements. Circular movements were found to be less precise than piston and eccentric movements, with a variability range of up to 3.8 mm for the circular movements tracked with the VOX but increasing to 73.8 mm when tracked with the WAVE (Table 4, rows of *Circle ED*). For the piston and eccentric movement, the range for the VOX was at most 1.4 mm (95%-range: 0.5 mm), whereas the range for the WAVE was at most 4.5 mm (95% range: 1.36 mm). The low performance of WAVE sensors on the circularly rotating bar was likely caused by the distance from the FGU (which was highest of all rigid bodies), and not due to the circular movement pattern (see also Fig. 11 which shows the detrimental effect of distance on positional precision). In version 1 of the WaveFront recording software, the default recording volume had been limited to a field of 300 mm, as performance in the 500 mm field was lower (see also Berry et al., 2011). The accuracy of the VOX decreased slightly for higher movement speeds, whereas this effect seemed to be more pronounced for the WAVE. However, this difference between the two articulographs was not significant.

For the WAVE, data was missing more often than for the VOX, but this difference was also not significant. Averaged across all moving sensors, during all dynamic tests (namely automated, LEGO block, and CD), the total percentage of missing samples was 0.85% for the WAVE and

418 0.04% for the VOX. Sensor dropouts occurred especially during automated trials for sensors
419 on the circularly rotating bar (Table 5).[8]

420 When comparing our results to those of Berry (2011), who previously evaluated the WAVE,
421 we see that our WAVE system performed comparably. First, regarding static precision, Berry
422 (2011: Table 1 on p. 1299) found that the 95% range for static sensors recorded 10 cm and 35
423 cm away from the FGU (using the 500 mm cube recording setting) was 0.88 mm and 5.11 mm,
424 respectively. Accuracy was better for the 300 mm cube recording field (specifically 0.46 mm
425 and 3.27 mm for the same distances). In our case, the LEGO block positioned at approximately
426 the same distances (at 9 cm and 34 cm) resulted in a 95% range of 0.08 mm and 3.6 mm,
427 respectively.

428 Second, regarding dynamic test results, Berry (2011: Table 2 on p. 1299) found that the 95%
429 range for moving sensors placed at 5 cm, 10 cm and 30 cm from the FGU (using the 500 mm
430 cube recording setting) was 1.16 mm, 1.36 mm and 7.33 mm. In our study, the rigid bodies
431 positioned at approximately those distances during fast trials resulted in a 95% range of 0.8 mm,
432 1.3 mm, and 22.6 mm.[9] Differences between the results from our study and those of Berry
433 (2011) could be due to the difference in tasks, the speed during dynamic movements, the exact
434 positioning of the sensors (e.g., Berry moved the device whereas we placed different sensors at
435 different positions), as well as a difference in the recording software (i.e., we used a newer
436 version of WaveFront), and system variability (Savariaux et al., 2017). Nevertheless, also when
437 taking into account the results from Berry (2011), the VOX appears to be a substantial
438 improvement over the WAVE.

[8] In earlier trials, with a slightly different setup, we experienced a large amount of sensor dropouts for sensors on the piston rigid body.

[9] The 95% range found for moving sensors at approximately 30 cm from the FGU was much higher in our study (22.6 mm) than in that of Berry (2011; 7.33 mm). The difference might have occurred because our sensors were not positioned to the center of the recording field but more towards the side (along the x-axis). In addition, these sensors were embedded in the circularly rotating bar, which showed worst performance of all moving rigid bodies.

Sigona et al. (2018) showed that the precision of the Carstens AG501 was 0.3 mm within the optimal region in the recording volume, which decreased to about 0.8 mm outside of this volume. As we tried to position our device optimally with respect to the field generator, it is likely that the precision of the VOX is lower than that of the AG501 (as the standard deviations for the radii were generally somewhat higher than 0.3). However, a comparison of both devices would be required to ascertain this. Savariaux et al. (2017) suggest that a precision of 0.5 mm is acceptable for speech tracking, which makes both the VOX and the WAVE suitable for this purpose.

Besides improved precision and accuracy, the VOX has some additional practical benefits in comparison to the WAVE. In particular, the VOX sensors can be attached to the system faster (in contrast to the WAVE sensors for which a screwdriver is necessary), the new SCUs are smaller, quieter, and do not have fans, and the revised FGU is less heavy (although emitting a slightly louder high-pitched sound during operation). Given these benefits, there is no reason to use the WAVE when a VOX device is available.

There are several limitations to our study. First, we tested static sensor tracking at different points in the field and dynamic sensor tracking accuracy at different speeds. We did not, however, test all sensor orientations. Second, as mentioned before, the WAVE in our lab has been in use since 2013 whereas the VOX has been in use since 2019. It is possible that some system degradation has occurred over time that caused the WAVE to perform significantly worse than the VOX. However, considering that our WAVE results are at least partially comparable to those of Berry (2011), we believe that our results have not been unduly influenced by the age of the system. Third, we did not yet systematically test the VOX during speech data collection[10] and therefore cannot compare the two systems for their designed

[10] However, in our experience with the VOX so far, data collection is similar to the WAVE.

purpose: tracking speech articulation. However, our study does show that the VOX system is accurate, and further experiments should evaluate the dynamic tracking performance of the VOX during speech production.

Funding Statement

This study was supported by funding from the Dutch Research Organisation (NWO) to Martijn Wieling (grants no. 019.2011.3.110.016, 016.144.049 and PGW.19.034) and by an International Macquarie University Research Excellence Scholarship (iMQRES) grant to Jidde Jacobi.

Conflict of Interest Statement

There are no conflicts of interest. Speech Lab Groningen (SLG) was a seed testing site for Northern Digital Inc. and received an NDI Vox-EMA for testing before its release in January 2020. However, NDI was not involved in the setup, the execution, nor the reporting of the experiments in this paper.

Acknowledgments

We would like to thank the editor and two anonymous reviewers who helped us improve the paper substantially.

References

Berry, J. J. (2011). Accuracy of the NDI Wave Speech Research System. *Journal of Speech, Language, and Hearing Research, 54*(5), 1295–1301. https://doi.org/10.1044/1092-4388(2011/10-0226)

Frantz, D. D., Wiles, A. D., Leis, S. E., & Kirsch, S. R. (2003). Accuracy assessment protocols for electromagnetic tracking systems. *Physics in Medicine and Biology, 48*(14), 2241–2251. https://doi.org/10.1088/0031-9155/48/14/314

Hoole, P. (2014). *Recent work on EMA methods at IPS Munich*. Retrieved from https://www.phonetik.uni-muenchen.de/~hoole/articmanual/ag501/carstens_workshop_summary_issp2014.pdf

Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. In B. Maassen & P. van Lieshout (Eds.): *Speech motor control: New developments in basic and applied research* (pp. 331-349).

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software, 28*. https://doi.org/10.18637/jss.v028.c01

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. *Journal of Speech, Language, and Hearing Research, 51*(4), 914–921. https://doi.org/10.1044/1092-4388(2008/067)

Kroos, C. (2008). Measurement accuracy in 3D electromagnetic articulography (Carstens AG500). *Proceedings of ISSP 2008 - 8th International Seminar on Speech Production*, 61–64.

Kroos, C. (2012). Evaluation of the measurement precision in three-dimensional Electromagnetic Articulography (Carstens AG500). *Journal of Phonetics, 40*(3), 453–465. https://doi.org/10.1016/j.wocn.2012.03.002

Northern Digital Inc. (2018, revision 12). *Wave User Guide*. Retrieved from http://support.ndigital.com

Northern Digital Inc. (2019, revision 4). *Vox-EMA System User Guide*. Retrieved from http://support.ndigital.com

Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America, 77*, 640-648. https://doi.org/10.1121/1.391882

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America, 92*(6), 3078-3096. https://doi.org/10.1121/1.404204

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (forthcoming, 2021). Retrieved from http://www.martijnwieling.nl/files/Rebernik-forthcoming.pdf.

Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017). A Comparative Study of the Precision of Carstens and Northern Digital Instruments Electromagnetic Articulographs. *Journal of Speech, Language, and Hearing Research, 60*(2), 322-340. https://doi.org/10.1044/2016_JSLHR-S-15-0223

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language, 31*(1), 26-35. https://doi.org/10.1016/0093-934x(87)90058-7

Sigona, F., Stella, M., Stella, A., Bernardini, P., Gili Fivela, B., & Grimaldi, M. (2018). Assessing the position tracking reliability of Carstens' AG500 and AG501 electromagnetic articulographs during constrained movements and speech tasks. *Speech Communication, 104*, 73–88. https://doi.org/10.1016/j.specom.2018.10.001

Speech Lab Groningen. (2020a). *EMA testing (static and manual dynamic tests)* [video]. YouTube. **https://youtu.be/2oZekNcGAoU**

Speech Lab Groningen. (2020b). *EMA testing (LEGO apparatus)* [video]. YouTube. **https://youtu.be/vifxgBCI8kY**

Stella, M., Bernardini, P., Sigona, A., Grimaldi, M., & Fivela, B. G. (2012). Numerical instabilities and three-dimensional electromagnetic articulography. *The Journal of the Acoustical Society of America, 132*(6), 3941-3949. https://doi.org/10.1044/2016

Stella, M., Stella, A., Sigona, F., Bernardini, P., Grimaldi, M., & Fivela, B. G. (2013). Electromagnetic articulography with AG500 and AG501. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1316–1320.

Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47*(1), 85-100. https://doi.org/10.1044/1092-4388(2004/008)

van Rij, J., Wieling, M., Baayen, H., & van Rijn, H. (2020). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.3.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86-116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3-36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. (2017). *Generalized Additive Models: an introduction with R.* 2nd edition. Boca Raton: CRC press.

Yunusova, Y., Green, J. R., & Mefferd, A. (2009). Accuracy assessment for AG500, electromagnetic articulograph. *Journal of Speech, Language, and Hearing Research, 52*(2), 547–555. https:///doi.org/10.1044/1092-4388(2008/07-0218)

Figure captions

Figure 1: Above: Field Generator (Transmitter) Units for the WAVE (1-A) and the VOX (1-B). Below: System Control Units (SCU) for the WAVE (1-C) and the VOX (1-D). Note that the WAVE does not come with coloured markers (these were added by our lab).

Figure 2: Above: System Interface Units for the WAVE (1-A) and the VOX (1-B). Note that the WAVE does come with coloured markers (these were added by our lab). Below: WAVE strip cable assembly (2-C) and VOX sensor harness assembly (2-D).

Figure 3: Architecture of the WAVE (left) and the VOX (right): FGU (field generator unit), SCU (system control unit), SIU (system interface unit), SCA (strip cable assembly), SHA (sensor harness assembly), and S1-S16 (sensors).

Figure 4: Left: 5DOF sensors of the WAVE (4-A) and the VOX (4 B). Right: 6DOF sensors of the WAVE (4-C) and the VOX (4-D).

Figure 5: Placement of FGU showing orientation of system axes (coordinate center is the center of the plate surface).

Figure 6: LEGO block during static trials, in position 1 (the block is parallel to FGU; sensor heads are perpendicular to FGU).

Figure 7: Relative positions of static sensors (including distance from the FGU in the legend). The black square represents the LEGO block while the red line represents the wire of a single sensor (for illustration purposes). Positions 1-6 denote displacements along the z-axis (sensors perpendicular to the FGU), positions 7-9 denote displacements along the z-axis (sensors parallel to the FGU), and positions 10-12 denote displacements along the x axis (sensors perpendicular to the FGU). Positions 13-15 are not depicted but consisted of displacements along the y-axis when the block was in position 1 (the block was raised by two or five LEGO blocks).

Figure 8: Sensors placed on CD (here depicted on the LEGO plate for easier visibility).

Figure 9: LEGO apparatus for automated tests. 1: circularly rotating bar (sensors 1-2), 2: eccentric beam moving over a large range (sensors 3-4), 3: eccentric beam moving over a small range (sensors 5-6), 4: bar moving as a piston (sensors 7-8), and 5: static position (sensors 9-10). Approximate hand-measured distance between each sensor and the FGU are indicated.

Figure 10: Bean plot visualizing average standard deviations of the positional coordinates for both devices during the static tests.

Figure 11: A. Model predictions for the effect of the distance from the FGU on the SD (log-transformed); B. Bean plot visualizing the relationship between average static SD and position for the VOX; C. Bean plot visualizing the relationship between average static SD and position for the WAVE.

Figure 12: Bean plot visualizing average error (A: log-transformed, B: non-transformed) of the inferred distances per articulograph.

Figure 13: Euclidean distances between sensor pairs 1:2 on the circularly rotating bar for the WAVE (left) and the VOX (right) during fast automated movements.

Figure 14: Euclidean distances between sensor pairs 1:2, 1:3 and 2:3 for the WAVE (left) and the VOX (right) during large manual movements.

Figure 15: Fitted circles for fast automated trials based on WAVE (left) and VOX (right) sensor movements.

Figure 16: Fitted circles for manually rotated CD trials based on WAVE (left) and VOX (right) sensors. Only sensors on the rotating CD (CD1, CD2) are depicted.

Figure 17: Bean plot visualizing average SD per articulograph for the automatic dynamic trials for all rigid bodies (A) and the five separate rigid bodies (B-F).

Figure 18: A. Model predictions for the effect of movement speed on the SD (log-transformed); B. Bean plot visualizing the relationship between average dynamic SD and speed for the VOX; C. Bean plot visualizing the relationship between average dynamic SD and speed for the WAVE.

Figure 19: Bean plot visualizing missing rates (in percentages) per articulograph.