Major revisions:

* Generate a plot with articulator parameter time-courses and constriction degree time-course for an observed vowel-consonant-vowel sequence. (Can Fig. 5 substitute for this?)
* Generate synthetic data and estimate articulatory strategies for these data.
* Determine the relationship between articulatory strategy biomarkers and the weights of the pseudo-inverse of the jacobian matrix of the forward kinematic map.

Associate Editor’s comments:

* **page 16, equation 13. what does the plus sign (+) mean?** The plus sign denotes the Moore-Penrose pseudoinverse. This was mentioned on page 14. The revised manuscript adds the sentence “Superscript ``+’’ denotes the Moore-Penrose pseudoinverse” after Equation 13 on page 16. (TODO: line numbers, figure numbers)
* **line 429, 'velar and pharyngeal', do you mean 'pharyngeal'?** The biomarker values at the bilabial and palatal places did not significantly differ from those at the velar place (p=0.19, p=0.061, see lines 408 and 410) or from those at the pharyngeal place (p=0.96, p=0.5, see lines 412 and 414). The statement on line 429 of the original manuscript was correct. It is not amended in the revised manuscript.
* **Figure 5. I might have missed it, but figure 5 does not seem to be referred to in the text.** The revised manuscript now uses Figure 5 to explain Equation 26. (TODO: line numbers, figure numbers)
* **I noticed that no separate files were submitted for the figures included in the manuscript. If this is the case, please include such figure files in your resubmission.** The original manuscript included each figure as a separate file. These files were resubmitted with the revised manuscript.
* **Also, please make sure the supplementary materials are truly necessary. Reviewer #1 had a comment questioning the usefulness of some of the supplementary materials.** TODO

Reviewer #1’s comments:

Links to the theory:

* **The definition of articulator synergy as “a functional grouping of articulators such as the jaw, tongue, and lips whose coordinated movements produce constrictions during speech” is poorly motivated by concurrent definitions.**

We agree that the definition should mention that a reduction of degrees of freedom plays a central role in the definition of a synergy. The definition on page 3, lines 16-17 has been revised. It now reads as follows: “An articulator synergy is a functional grouping of articulators such as the jaw, tongue, and lips whose coordinated movements produce constrictions during speech, and which instantiates a reduction in the effective degrees of freedom from the set of articulators to the space of controlled variables.”

* **The proposed model of the speech production system underlying the imaged movements involves a reduction in dimension in two mappings: (a) the mapping from articulator shapes to articulator parameters and (b) the mapping from articulator parameters to constriction degrees. Do these two steps reflect two different levels of redundancy reduction in the motor system? If so, how are these two levels connected to available definitions of articulator synergies?**

The mapping from articulator shapes to articulator parameters is a function from the spatial coordinates of the articulator contour vertices in the scan plane to the parameters of an articulatory model. The forward kinematic map is a function from the parameters of an articulatory model to constriction degrees at the phonetic places of articulation. The reduction of dimension from the spatial coordinates of the articulator contour vertices to the parameters of the articulatory model is only necessary due to the specifications of the image segmentation outputs. (The articulator contours must be sampled at a finite set of points along their length.) The mapping from articulator shapes to the parameters of an articulatory model is only a reparameterization of the articulator contours. To the best of our knowledge, this step does not have scientific significance. Perhaps more scientifically important is the mapping from the parameters of an articulatory model to constriction degrees. TODO: continue this text

* **The definition should relate quantities obtained from real-time MRI to parameters of the task dynamics model of speech production.**

The revised manuscript now includes a graph of the articulator synergy biomarker as a function of the jaw weight parameter from Task Dynamics. (TODO: line numbers, figure numbers)

Links to top-down and bottom-up approaches to the study of motor synergies:

* **How is the proposed biomarker related to the uncontrolled manifold approach?**

The revised manuscript includes a paragraph in the discussion describing the relation to the uncontrolled manifold. (TODO: line numbers)

* **How is the proposed biomarker related to bottom-up methods of describing synergies?**

The revised manuscript includes a paragraph in the discussion describing the relation to the uncontrolled manifold. (TODO: line numbers)

Validation of the articulator synergy biomarker:

* **The method should be validated by analysis of simulated data. Alternatively, the method could be validated by analysis of real data on the basis of finding patterns that are expected because of available theoretical knowledge of mechanisms.**

Section VI of the revised manuscript estimates the measurement bias of the biomarker using synthetic data.

Potential limits of the study:

* **If the components obtained through guided PCA depend on the analyzed speech material, why is there such a similarity to the components reported in Toutios & Narayanan (2015), based on one speaker from the USC-TIMIT database?**

We have not quantified whether factors differ depending on speech material, and so we cannot confirm that there is similarity between the guided factor analysis results of Toutios & Narayanan (2015) and the guided factor analysis results of the present paper.

* **How does the assumption of linearity in the guided PCA approach affect the results?**

Work that we have presented at the LabPhon conference has shown that the error in approximating vocal tract contours as the linear combination of factors is smaller than the in-plane spatial resolution (see http://www-scf.usc.edu/~tsorense/pdf/LabPhon15\_Revised\_abstract\_315.pdf).

* **What happens when the spatio-temporal pattern of articulator parameters or task variables is not similar for different tokens of the same gesture? Should tokens be sorted into different groups on the basis of their spatio-temporal patterns and be submitted separately to the analysis?**

We respond to this question first by offering an assessment of how substantial the variability between different tokens of the same gesture may be in terms of spatio-temporal patterns.

Variability in the spatio-temporal pattern of articulator parameters is substantial in the sense that vocal tract shaping for a given gesture may be achieved using different factors depending on the participant. Since the factor model is estimated separately for each speaker, the factors of different speakers may have distinct spatial patterns, and the correspondence between factors of different speakers is not known a priori (e.g., the first tongue factor of Speaker A does not necessarily reflect the same deformation of the tongue as the first tongue factor of Speaker B).

Variability in the spatio-temporal pattern of constriction degrees for different tokens of the same gesture is minor in the sense that the same basic pattern is observed for all tokens of a particular gesture. Specifically, the constriction degree associated with the consonant closure decreases to a minimum at the point of maximum constriction and then increases again. Subtle variations in this spatio-temporal pattern may be present (e.g., peak velocity, time to peak velocity, etc.), but other techniques with finer temporal and spatial resolution are perhaps better suited to study fine details of the time-course of constriction degrees (e.g., electromagnetic articulography).

Whether or not it makes sense to group observations into different spatio-temporal patterns and to submit these patterns to the analysis separately depends on which stage of the analysis you are referring to. With regard to the analysis of biomarker precision, one could subset the data according to spatio-temporal pattern instead of by vowel-consonant-vowel sequence type, but different spatio-temporal patterns may contain different numbers of observations from each participant. As this would complicate the analysis, we did not pursue this approach. With regard to the cross-validation of the direct and differential kinematics, the data was divided into subsets by place of articulation. We believe that this is the most intuitive way to subset the data because the error being analyzed was the error in constriction degree at exactly these places of articulation. Without a sound reason for believing that different spatio-temporal patterns would involve errors of different magnitude, we did not pursue this approach. With regard to testing the main hypothesis (i.e., task-dependence of articulator synergies), our main interest was in determining differences between constrictions of the anterior vocal tract and constrictions of the posterior vocal tract. We subset the data by place of articulation (in addition to scan number and participant) because this straightforwardly reflected the anterior vs. posterior distinction, with bilabial being most anterior and pharyngeal being most posterior. As further subdivisions of the data on the basis of spatio-temporal patterning do not reflect this distinction, we have not pursued it here.

Although we do not believe it makes sense to subset the data on the basis of spatio-temporal patterns in the context of the present study, we believe that future studies might benefit from using the proposed articulator synergy biomarker to distinguish different spatio-temporal patterns.

Presentation of the method:

* **A tutorial style is adopted for some sections, whereas other sections use a terse, descriptive style. More explanation should be given where appropriate (e.g., Eq. 20).**

In the revised manuscript, Sections V.A. and VII have been edited to adopt a more tutorial style. (TODO: make sure these are the right section numbers)

Textual comments:

* **p.3, lines 16-17: "An articulator synergy is a functional grouping of articulators such as the jaw, tongue, and lips whose coordinated movements produce constrictions during speech (Turvey, 1977)." Motor synergies can be defined in different ways (see main comment I).**

The definition has been revised to the following: “An articulator synergy is a functional grouping of articulators such as the jaw, tongue, and lips whose coordinated movements produce constrictions during speech, and which instantiates a reduction in the effective degrees of freedom from the set of articulators to the space of controlled variables.”

* **P. 3, lines 24 -26: "healthy adult speakers of American English may use the jaw more for anterior constrictions at the bilabial, alveolar, and palatal places of articulation than for posterior constrictions at the velar and pharyngeal places of articulation". Please provide some explanation for this hypothesis.**

The revised manuscript adds the following sentence for explanation after the quoted text: “This may be due to the hinge-like motion of the jaw, whose rotational motion generates greater displacement in the anterior part vocal tract than in the posterior part of the vocal tract.”

* **p.4 lines 49-51: "The algorithm for computing the articulator synergy biomarker involves a computational model of the direct and differential kinematics of the vocal tract (Lammert et al., 2013a) based on the Task Dynamics model of speech production (Saltzman and Munhall, 1989)" I was a bit puzzled by this passage. As I understood Lammert et al. (2013), the authors of that paper presented a weighted linear regression model (as opposed to a model based on multilayer perceptron) permitting to predict vocal tract constriction degree from articulator positions. If this is the case, wouldn't be more appropriate referring to a "statistical model" rather than to a "computational model"? Moreover, the authors should be more specific about the features shared by the task dynamic model and the weighted linear regression model by Lammert et al (2013). Are the authors referring to the inventories of geometric articulator variables and task variables?**

In the revised manuscript, the quoted passage has been changed to the following: "The algorithm for computing the articulator synergy biomarker involves a statistical model of the direct and differential kinematics of the vocal tract (Lammert et al., 2013a), which uses MRI to estimate a forward kinematic map relating articulatory parameters and constriction degrees, as in the Task Dynamics model of speech production (Saltzman and Munhall, 1989)”. This change uses the term “statistical model” instead of “computational model” and clarifies that the statistical model of Lammert et al. (2013) is a model of the forward kinematic map and its jacobian matrix from the Task Dynamics model of speech production.

* **p.6 lines 66-75: Is it really necessary to summarize the results in the introduction?** According to the journal guidelines (https://asa.scitation.org/pb-assets/files/publications/jas/jasinfcon-1518211773377.pdf), it is not necessary. Accordingly, we removed the summary of results from the introduction.
* **p.6 line 85. The plan of the paper does not include the discussion and the conclusion sections. Moreover, it makes reference to modelling steps which have not been mentioned before. I strongly suggest to give a rapid overview of the method in the introduction in which the authors presents the pre-processing step (the manually assisted segmentation of the FMRI images) and the two main steps of the algorithm (the guided PCA analysis and the mapping between articulator parameters and the constriction degree).**

The plan of the paper now only refers to modelling steps that have been mentioned previously in the introduction (TODO: give line numbers for first mention of the following: constriction degree; articulatory modeling; direct and differential kinematics; articulator synergy biomarker). It reads as follows: “Section II describes the MRI experiment, scanner sequence, participant characteristics, and method for manually annotating the start and end time-points in the real-time MRI. Sections III and IV describe the segmentation of articulator contours in the images and use the segmentation results to estimate of constriction degrees and parameters of articulator shape and position, which are related by the direct and differential kinematics of the vocal tract. Section V estimates the direct and differential kinematics and evaluates the model through cross-validation. Section VI defines the articulator synergy biomarker and evaluates the precision of the articulator synergy biomarker in a test-retest experiment, which estimated the repeatability of the articulator synergy biomarker in two experiments on the same day under the same experimental conditions. Section VII uses the articulator synergy biomarker to test the task-dependence of articulator synergies. Sections VIII and IX offer discussion and conclusions.”

* **p. 10, line 148, "p is the number of contour vertices": If each vertex is a point on a plane, assuming that X is the vector of the stacked coordinates of the contour vertex, shouldn't p be the number of vertices divided by 2?**

The size of the matrix X is now written as n x 2p, where p is the number of contour vertices.

* **p.13, line 286: Was centering applied on the different dimension of X vectors before PCA?**

Yes. The revised manuscript states that the contour vertex coordinates were centered on zero before factor analysis.

* **p.16: Please add a multi-panel figure showing in each panel the trajectories of the q factor scores during the production of one different syllable. Each panel should also contain the trajectory of the relevant constriction degree. Such a figure will also be also useful when introducing the construction of the mapping between the factor scores and the constriction degree.**

TODO: ask co-authors whether we have the space

* **p.16, line 239: In order to introduce the construction of the forward map, please consider presenting it as a weighted linear regression problem as in Lammert et al. (2013).**

Section V indicates that the problem is solved using weighted least squares. The estimator is the weighted linear regression estimator (Equation 20), which optimizes the weighted sum of squared errors (Equation 17).

* **p.17, eq. 18: In Lammert et al. (2013) a regularization term was used. What motivates dropping that term?**

In Lammert et al. (2013), the regularization term was used to make the solution robust to the case where there are very few data points in the neighborhood of the query point. In our implementation, the number of data points in the neighborhood of the query point is set by the parameter (see answer to next question for details), and thus there are always a large number of data-points (specifically, data-points) in the neighborhood. Since the problematic case of few data-points does not occur, the proposed estimator does not use a regularization term.

* **p.19, line 245: Are h and f two independent free parameters or is h adjusted to obtain the desired number of neighbours? And by the way, is h constant over different dimensions?**

The revised manuscript clarifies that f is the free parameter, not h by adding the following text: “The parameter sets the percentage of data-points inside the spherical neighborhood within which the forward kinematic map is estimated. The parameter is the radius of the smallest sphere containing exactly data points. The parameter is found using the nearest neighbors algorithm, where .”

* **p.32, lines 454-456, "The study demonstrated low error in the estimator of the direct and differential kinematics and consistent estimation of the articulator synergy biomarker in a test-retest repeatability experiment.": If I am not wrong, the median values of the consistency index for the synergy biomarker were more often below 50% than above this percentage.**

The intra-class correlation coefficient measures the ratio of inter-subject variance to total (inter-subject plus intra-subject) variance. Thus, the reason for low values of the intra-class correlation coefficient may either be low inter-subject variability, high intra-subject variability, or both. Given that the voluntary movement of speech production can be quite variable even within a given subject, low intra-class correlation coefficient may be inevitable. TODO: mention this in discussion

**p.33, lines 460-462: I am not convinced that the results reported imply that the biomarker has adequate precision to characterize articulator synergies. See main comment III.**

Section VI of the revised manuscript estimates the measurement bias of the biomarker using synthetic data. No measurement bias was detected. The low precision may be an inevitable consequence of high intra-subject variability. TODO: mention this in discussion

* **p.33, lines 472-474: This is a cut and paste from p.4 lines 49-51. The same comment applies here.** The revised manuscript changes the text to “The algorithm for computing the articulator synergy biomarker involves a statistical model of the direct and differential kinematics of the vocal tract (Lammert et al., 2013a), which uses MRI to estimate a forward kinematic map relating articulatory parameters and constriction degrees, as in the Task Dynamics model of speech production (Saltzman and Munhall, 1989).”
* **p.35, lines 521-523: "The implication of the results is that articulator synergies are task-dependent in that they have different patterns of inter-articulator coordination depending on their place of articulation". Here I expected some comparison with other studies based on other approaches and supporting the idea that speech motor synergies are task dependent.** TODO
* **p.36, lines 534-535: Not all scripts are available as supplemental material. Strictly speaking, in the supplemental material there are wrappers that launch the core scripts residing on a server.** In the original submission, some scripts were obtained by cloning the Git repository “span\_contour\_processing” (https://github.com/usc-sail/span\_contour\_processing). This is no longer the case. All scripts are included in the main repository.

Reviewer #2’s comments:

* **The corpus … is strongly limited to the four isolated vowel-consonant-vowel utterances [apa], [ata], [aka], [aja]. This significantly hinders the validity of the conclusions, and should be clearly mentioned in the concluding section as well as in the abstract.**

TODO: The discussion section of the revised manuscript now states this point as a limitation of the present study.

* **It is thus important to quote some of this literature and to emphasize how the present approach is different.**
* **In fact, it would seem that these principal axes do capture the variance of tongue and lips that is linearly correlated with the jaw factors, as described in the literature. Could the authors clarify this issue?**
* **Note also that the upper lip might be influenced by jaw in some subjects: this is apparently not taken into account (e.g. Fig. 3, top left), though of minor acoustic importance.**
* **Finally, it would be very useful to conclude section IV with a comparison of the performances of the Guided Factor Analysis used and more standard Guided PCA.**
* **It would be useful to mention more explicitly in section IV.A which organs are really modelled (jaw, tongue, velum, lips?).**
* **Are the three degrees of freedom expected for a solid object in 2D enough to represent the jaw variance?**
* **In lines 52-53, the authors state that “direct kinematics relates the position and shape of articulators to the corresponding degree of constriction”. This definition clearly indicates that these maps are not kinematic, i.e. not related to time, but only represent relations between two partial representations of the geometry of articulators (articulator contours and construction areas). This expression is thus very confusing and should be avoided. Similarly, the authors state that “differential kinematics relates small increments of articulator movement to the resulting changes in the constriction degrees”. Here too, time is not involved either, which is also confusing.**
* **Finally, the authors write that “factor scores [that] characterize temporal variation in the position and shape of the articulators” (line 151). It is not obvious that these factor scores bear any temporal / kinematic information. The authors are thus invited to clarify this issue.**
* **Could the authors explain the relation/interaction between parameters “h” and “f”?**
* **In lines 46-48, a general definition of “imaging biomarker” is offered. In section VI.A however, no practical definition is proposed before the long technical development that leads to equation 27. The sentence (line 311-2) “The articulator synergy biomarker is the percent contribution of the jaw to narrowing and widening the vocal tract for a constriction.” suggests that a biomarker is a measurable consequence of articulator synergy, more related to latent variables that to visible geometrical parameters that may be deduced from the image. Clarifications about the nature of the biomarkers are thus strongly needed (note that Fig. 5 does not seem to be referred to, though it appears that it could contribute to this clarification).**
* **Note also that the notion of “elapsed change in constriction degree” needs clarification, in particular “elapsed” (between what instants?).**
* **The use of 10-fold cross-validation is indeed an interesting choice. However, RMSE tends to hide the largest errors. It could thus be very useful to supplement RMSE with another statistical parameter, for instance the 10th - 90th percentile range that gives a more accurate idea of the error extent.**
* **Statistical results presented at page 30 are a bit confusing. For instance, the authors state that “On average, the percent jaw contribution was 17% less at the velar place compared to the bilabial place (z = 1.9, p = 0.19)”. What is the point to offer a comparison and to mention indirectly that it is not statistically significant? This section would need more specific rewriting.**
* **Use of “study personnel”: if this expression refers to the authors of the manuscript and of the work, it sounds peculiar and should be rephrased. If it would refer to personnel who has participated to the work but is not listed as co-authors, it would not appear to comply with standard practice in the academic world.**
* **The notion of “constriction degree” is ambiguous. One might consider that a high degree of constriction refers to a very narrow constriction, but legend in Fig. 5 states “the jaw and tongue produce a narrowing at the palatal place, and constriction degree decreases to a minimum”, which seems to imply the opposite. More specific expressions such as “constriction area” or “constriction size” should be used.**
* **Line 148: Vertices in 2D are specified by two coordinates. This should reflect somehow in the number “p” that could be “2 x p”?**
* **Bottom of page 10: “Contour vertices x\_i” have apparently not been defined earlier (cf. remark above).**
* **Line 548: Please, correct to “sensitivity”.**
* **Figure 4. The tick marks on the right Y-axis are hardly visible.**
* **Figure 6. Lines and symbols are too thin to be neatly visible.**

Reviewer #3’s comments:

1. **First of all, one could say that it is not big news that the jaw contributes more to more anterior than more posterior constrictions. What would be more interesting to know is whether these differences can be explained by a simple view of jaw movement as pure rotation at the condyle - if so it would not be necessary to assume different synergies at different places of articulation, the differences would just fall out of the biomechanical arrangement of the articulators. In other words, do we need to assume, for example, different combinations of jaw rotation and translation at the condyle for different places of articulation?**
2. **Also the observation of interspeaker variability in jaw involvement does not, I feel, go as far as it might. That speakers differ is not surprising, but are they consistent within themselves across different places of articulation?**
3. **There is also a potential confound in the design of the experiment: the study does not present a pure comparison of place of articulation, but mixes it with manner of articulation (basically approximant for palatal and pharyngeal, but stop elsewhere). In fact, my first reaction to the paper was that it would have been more interesting if it had actually focussed on manner of articulation (this is not mentioned as a possible future direction at the end), since it is well known since e.g. Vatikiotis-Bateson & Ostry (JPhon, 1995) that manner of articulation can have a substantial impact on patterns of jaw movement. Concentrating on manner of articulation at a single place of articulation may also be easier to interpret directly as differences in synergy, given that the complicating factor of distance from the condyle is more or less absent.**
4. **I find the basic assumption making the current approach feasible, namely that the relevant relationships are locally linear, perfectly plausible. Nonetheless it would be interesting to have more discussion of the pre-requisites for this assumption, and also whether the present approach can also be used to identify potentially interesting cases where the assumption may start to break down. To take labial constriction as a specific example, it is presumably the case in the present approach that when the constriction size goes to zero then the movements of the lips, based on the external contours of the lips, also cease (so the relationship between constriction, and articulator movement is indeed quite straightforward). However, one could argue that in a more realistic view of the movements of the articulators, the lips continue to move even when the lip aperture has reached zero, and that perhaps other measurement techniques, or other approaches to extracting movement information from the MRI images, might capture this. Would such techniques then actually be less suitable for the present approach (since the relationship between change in articulator position and change in constriction size might then be decidedly non-linear)? Perhaps a similar question from a different point of view: How would the present approach mesh with a somewhat more abstract conceptualization of the constriction target, for example use of negative constriction size (as in TADA) as a target to achieve firm closure in stops?**
5. **It would be worth pointing out in the discussion that the present study only looks at one particular group of synergies, namely those involving the jaw. Many models assume that constrictions at the tongue-tip also involve synergistic movement of tip and dorsum. Would it be feasible to investigate this kind synergy too? And more specifically, can patterns of tip-jaw synergies be interpreted unambigously if at least one further synergy is involved?**
6. **The formulation of the hypotheses is somewhat ad hoc. The initial formulation (p. 5) aims to test for a difference between anterior (bilabial, alveolar, palatal) and posterior (velar, pharyngeal). In fact the motivation for this specific division is not very clear. Might it not be just as plausible to suggest a division based on main active articulator (lips, vs. tongue-tip vs. tongue-dorsum/root)? The formulation on p. 29 is different, since it now refers to the jaw not contributing to posterior constrictions (not just to being different from anterior constrictions). And the formulation of the null hypotheses at the bottom of p.30 is different again.**

* **p. 4 and elsewhere. I think phrases like "12ms temporal resolution" should be used cautiously. Is really the temporal resolution meant, or just something like frame-rate? Given the complex reconstruction algorithms used in real-time MRI the two may not be synonymous.** Temporal resolution is correct. As mentioned on p. 7, line 118, a single image is reconstructed from the MR signal acquired in two TRs. There is no view-sharing between consecutive images.
* **Fig.1 please spell out the colours for the different contours. For example, green is presumably jaw, but this should be mentioned explicitly since bone and teeth do not image directly in MRI, so the definition of "jaw" is not necessarily obvious. (What then is the yellowish contour around the chin? Is this actually used anywhere?)** Figure 1 now includes labels. Chin and mandible contours are analyzed together as jaw contours. This is now explicitly described in the “Jaw factors” section.
* **Fig. 2 The legend mentions "velopharyngeal port" in the list of "phonetic places of articulation", but it is not clear to me that the term "velopharyngeal port" has already been introduced in this connection. Also, I don't see the blue line that should apparently be associated to it in Fig. 2.** Mention of the velopharyngeal port has been removed from the revised manuscript.
* **p. 10 I think the work of Maeda is a sufficiently important part of the background that it should be referred to explicitly, rather than indirectly via the reference to Toutios & Narayanan (2015), since this could help many readers to quickly get to grips with some of the basic concepts of the current approach.** The revised manuscript now cites Maeda.
* **p. 21, l. 287. The definition of Time 0 and Time T is not completely clear. Presumably Time 0 means onset of movement towards the constriction, not the onset of an aerodynamically relevant consonantal constriction (referred to by some authors as target achievement). (Similarly the earlier description p. 7, l. 107 is rather vague. Perhaps a sketch of a typical closing-opening movement would be useful for defining terms)** The revised manuscript now refers to Figure 5 in order to explain these terms.
* **p. 19 . 250 typo: "is important parameter"** The typo “is important parameter” was corrected to “is an important parameter”.
* **p. 21, i. 275 Insert mm after 2.4** The units “mm” have been added after “2.4”.
* **p. 22 fig. 4 typo: "neighborood" I'm not sure the tiny tick marks in the right-most panels of this figure are useful. Why not just a single (visible) tick corresponding to the average over all speakers? Also, is the figure of 2.4mm really a realistic criterion for discussing error of the differential kinematics? Multiplying by the framerate this corresponds to an error of about 200mm/s, which is really huge (close to peak velocity of many articulator movements), so one would indeed hope that the error is nowhere near this.** The typo “neighborhood” has been corrected to “neighborhood”. The individual tick marks have been removed and an average over all speakers replaces them. The 2.4mm reference point has been removed from the figure.
* **p. 37, l. 548 typo "senativity"** The typo “senativity” has been corrected to “sensitivity”.