# Reply to Coordinating Editor

**One Reviewer also notes that the term "kinematic" is not appropriate for the model of mapping between "the shape and position of the articulators" and "the corresponding constriction task" in which no time derivatives are involved.**

This comment was received in the first round of reviews. In the first review, Reviewer 2 writes, “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.”

Our first response to reviewers (JASA-03104R1) provided a justification for the use of the term "kinematics". This justification is repeated below for your reference:

*To the best of our knowledge, the terms “forward kinematics”, “direct kinematics”, “differential kinematics”, and “forward kinematic map” originated in the field of robotics. In the field of robotics, “[k]inematics is the science of motion which treats motion without regard to the forces which cause it. Within the science of kinematics, one studies the position, velocity, acceleration, and all higher-order derivatives of the position variables (with respect to time or any other variable(s)). Hence, the study of the kinematics of manipulators refers to all the geometrical and time-based properties of the motion” (Craig, 2005; emphasis my own). At least in the field of robotics, time is not necessarily involved in the study of kinematics.*

*The terms were subsequently adopted by the field of computational motor control. A definition of forward kinematics from a recent textbook on motor control reads as follows: “The kinematics maps relate the motions of the hand to the motions of the arm and come in two forms: direct kinematics, from arm’s joint angles to hand position, and inverse kinematics, from hand position to arm’s joint angles” (Shadmehr & Mussa-Ivaldi, 2012).*

*The use of the definition is not restricted to this one book, but appears frequently in academic papers on motor control. I sample two such papers below and indicate where in the paper these terms are defined (Todorov et al., 2005; see mathematical definition in Section 4.2; Atkeson, 1989; see Subsection “Kinematic Transformations” in Section “Motor Control involves Transformations”)*

*The terms are also used in Task Dynamics, the framework for motor control adopted in the present study. In the original Task Dynamics paper, these were simply called “kinematic relationships” rather than “kinematic maps” (Saltzman & Kelso, 1987; in particular, see Section “Joint variables and the task-dynamic network”, which defines functions relating “body-space variables” (i.e., controlled variables) as functions of the arm joint angles).*

*The kinematic maps were then defined for motor control of the vocal tract in speech production in the first application of Task Dynamics to speech production (Saltzman & Munhall, 1989; see in particular Appendix 2, which defines the “direct kinematic relationships” between articulator parameters and controlled “tract variable” parameters)*

*These terms are still used in more recent papers on Task Dynamics in speech (Ramanarayanan et al., 2016; see Fig. 2 for the graphical description of forward kinematics).*

*By comparison to the sources cited above in the fields of motor control and speech production, we believe that our usage of the terms “forward kinematics”, “direct kinematics”, “differential kinematics”, and “forward kinematic map” are consistent with common academic conventions.*

The second round of reviews states that the term “kinematics” is inappropriate because no time derivatives are involved. However, the text quoted from my first response to reviewers addresses this objection: according to convention, “[w]ithin the science of kinematics one studies the position, velocity, acceleration, and all higher-order derivatives of the position variables (with respect to time *or any other variable(s)*). Hence, the study of the kinematics of manipulators refers to all the geometrical and time-based properties of the motion” (Craig, 2005; emphasis my own). As outlined above, this convention carries over into the computational motor control literature. This convention provides a justification for the use of the term kinematics in the present manuscript.

Although we are reluctant to changing our use of the term “kinematics”, we would be willing to consider alternatives, given a suitable substitute. Absent an alternative, the revised manuscript continues to employ the term “kinematics” in a way that is consistent with its use in robotics and motor control.

Additionally, if the reviewers think it is necessary, the text of the manuscript could contain a paragraph or footnote describing the use of the term.

# Response to Reviewer 1

The revised manuscript addresses all points and revises the manuscript as Reviewer One suggests.

# Response to Reviewer 2

**I am wondering why the results of the replication experiment mentioned in Section X. are given in supplementary material #3 and not included in the text.**

Including the description of the data-set and results of the replication study would substantially increase the length of the manuscript and only provide information that is consistent/redundant with the primary study. In the interest of reducing redundancy in the main text, the replication study was provided as supplementary material for the reader who is interested in replication.

**The table of "Participant characteristics of the test-retest data-set" could also be integrated in the main text rather than in a separate supplementary material.**

The revised manuscript includes Table “Participant characteristics of the test-retest data-set” in the main text.

# Response to Reviewer 3

## Point 2

**Do speakers who show a low level of consistency for e.g. /p/ also show a low-level of consistency for /k/? Or conversely, do speakers exist who exhibit high consistency at one place of articulation but low consistency at another? Within a pattern of overall more jaw involvement for /t/ are speakers who show particularly strong jaw involvement for /t/ the ones who can be expected to be in the upper range of jaw involvement for /p/? These are interesting issues about which we currently know rather little. If I have misunderstood what messages can be pulled out of the current presentation of the data, please clarify.**

This is indeed an interesting issue. The revised manuscript takes it up in a new section, “Section VIII. Inter- and intra-participant variability”.

## Point 5

**Can patterns of tip-jaw synergies be interpreted unambiguously if at least one further synergy is involved?**

The revised manuscript now includes a new subsection “Section IX.D. Decomposing the tongue into multiple articulators” that explicitly describes an extension to the guided factor analysis that models the tongue body and tongue tip separately. The new subsection indicates how to interpret the tongue factors in the present study.

## Point 6

**This does not address my request for information about the motivation for the hypothesis, i.e. why expect a division according to place of articulation, and not rather a division based on active articulator? Why should the division between anterior and posterior be between palatal and velar?**

**Reviewer 1 also seems to find the hypothesis not very convincingly motivated (referred to under p. 3, lines 24 -26), and the response to this comment does not help much: It seems to me too that a hypothesis that is basically just a summary of previous findings is a very weakly motivated hypothesis.**

In revising the manuscript, the authors have come to see Research Goal 3 of the present study (cf. lines 60-64) to be “to use the articulator synergy biomarker to test the task-dependence of articulator synergies by determining whether the relative contribution of the jaw, tongue, and lips differs by constriction task”. This goal is more exploratory than the original Research Goal 3 we set at the beginning of the study, which was “to use the articulator synergy biomarker to test the task-dependence of articulator synergies by determining whether the jaw contributes 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”.

The manuscript has been revised to match the amended Research Goal 3. Within the context of the revised Research Goal 3, Section VII now uses the linear mixed effects model to test both the original null hypotheses as well as a set of supplementary hypotheses. Whereas the original null hypotheses were that contrasts between the “anterior” levels and “posterior” levels of the factor “constriction task” were zero, the revised set expands the original null hypotheses to include 10 contrasts consisting of all possible distinct combinations of levels of the factor “constriction task” (see Table II of the revised manuscript for enumeration). In addition, a new Section VIII now investigates inter- and intra-participant variability in the effect of factor “constriction task” on biomarker values (see below for the specific motivation for Section VIII).

Given the advances made in achieving Research Goals (i) and (ii) (cf. lines 60-64), the authors believe that the exploratory study of the articulator synergy biomarker that Sections VII and VIII provide is a reasonable next step towards confirmatory hypothesis testing with the articulator synergy biomarker. The patterns uncovered in Sections VII and VIII provide an indication of the size and direction of effects that will guide the design of experiments motivated by theory-driven confirmatory hypothesis testing.

**This is related to the Response to Point 1. No attempt seems to have been made to reflect this in the text. However, the results actually provide a good argument against a simple "distance from the condyle explanation", namely that jaw contribution does not simply increase from velar to alveolar to labial. It would have been easier to bring this out if the hypothesis had not been phrased with the, in my opinion, rather artificial division into anterior and posterior.**

Reviewer 3 entertains the possibility that differences in the jaw’s contribution to an articulator synergy may be explainable in terms of distance of the synergy’s place of articulation from the condyle. We view this as an interesting possibility. As Reviewer 3 points out, the results of the present study may suggest that distance between the place of articulation and the condyle does not explain inter-articulator coordination in synergies.

Our study focusses on dynamic imaging of the midsagittal plane, not on anatomical imaging of lateral structures such as the condyle. As the study does not include morphological measurements from anatomical MR images, the manuscript does attempt to explain differences in articulator synergies in terms of the distance between the place of articulation and the condyle. Instead, the study attempts to explain differences in articulator synergies in terms of the place of articulation along the anterior-posterior axis, which is visible in the midsagittal MRI scan plane.

**This is also related to the Response to Point 3 (possible confound in the design). I think the authors are letting themselves off too lightly by just briefly mentioning this in the conclusion as a direction for future work. I think the issue is serious enough that it should at least be "upgraded" to being dealt with in the discussion. Overall, I would have been much happier with the paper if it had just confined itself to a comparison of /p, t, k/. I don't see what useful conclusions can be drawn from any comparison of these consonants with a constriction for the vowel /a/. Could not any differences found also be due to the fact that synergies may differ for consonantal versus vocalic gestures?**

Sources of variance other than place of articulation influence variability in the biomarker value by level of the factor referred to as “place of articulation” in the original manuscript. For instance, manner of articulation and the consonant-vowel distinction influence variability between levels “bilabial/coronal/velar place” and “palatal/pharyngeal place”. In order to reflect this fact, the revised manuscript amends the name of the factor “place of articulation” to “constriction task” and refers to levels of the factor “constriction task” as “bilabial stop”, “coronal stop”, “palatal approximation”, “velar stop”, and “pharyngeal approximation”. The amended factor name is careful not to suggest that only place of articulation varies between its levels, and the amended names for the levels of the factor indicate both place and manner of articulation.

In addition to clarifying the nature of the factor “constriction task”, the revised manuscript upgrades the issue of having a possible confound in the design to the discussion. The relevant part of Section X is moved to a new paragraph in Section IX.A. This text addresses Reviewer 3’s concern that tasks can be distinguished not only by place of articulation, but also by active articulator (i.e., lips, tongue tip, tongue dorsum, tongue root) and manner of articulation (e.g., stop, approximation).

## Minor Points

**Are "alveolar" and "coronal" to be taken as synonymous in the abstract? If so, maybe just use one of the two terms.**

The revised manuscript replaces “alveolar” with “coronal” throughout.

**p. 5 l. 77-79 Insert "of" in "precision articulator". Also "evaluate the precision ..... with respect to .... precision" sounds strange.**

The revised manuscript removes the first occurrence of “precision” from the sentence.

**p. 6 Please provide information about whether elicitation of the speech items was randomized. If not randomized this could result in a further confound in the design, i.e. synergies could also be affected by position in the list of words.**

The order of the speech items was not randomized. It was a fixed list. The revised manuscript improves this information in Section II.A.

**p. 7 It only emerges somewhat indirectly that [a] is to be used to provide the material for the pharyngeal place of articulation. Also it didn't become clear to me whether V1 or V2 (or both) was used.**

The revised manuscript provides this information in the second paragraph of Section II.A.

**Also it should be made clear what is meant by the transcription symbol [a]. Did subjects really produce cardinal vowel 4? For an investigation of pharyngeal place a vowel near cardinal vowel 5 would probably be better.**

Although the study did not evaluate whether the speakers of American English recruited for the study produced the low front unrounded vowel [a] or the low back unrounded vowel [ɑ], these sounds are not generally distinguished in American English. In the original manuscript, the IPA symbol [a] was chosen to provide a phonetic transcription for the sound produced by participants reading the text “apa”, “ata”, “aka”, and “aia”. As you suggest, the IPA symbol [ɑ] may be more characteristic. (Supporting evidence that this is the convention may also come from the fact that ARPA transcribes IPA symbol [ɑ] as AA, but has no ARPABET equivalent for [a].) In order to minimize confusion, the revised manuscript switches the symbol [a] for the low front unrounded vowel to the symbol [ɑ] for the low back unrounded vowel, but notes that the speakers varied in exactly how the vowel was produced (see Section II.A.).

**p. 8, Fig.1 . Is [i] used as the template for [j]? If so, where did it come from?**

The original manuscript inconsistently used the symbols [i] and [j]. Speakers were instructed to produce [i]. However, speakers varied between [i] and [j] in producing the sound. The revised manuscript uses the vowel symbol [i] rather than the glide symbol [j], but notes that the speakers varied in exactly how the sound was produced (see Section II.A.). Given this clarification, it should be clear that the [i] template comes from the utterance [aia].

**p. 11. Is consideration of the velum necessary here? Even if part of the overall approach, could it be left out to streamline the presentation somewhat?**

The velum is an articulator that is used to estimate the forward kinematic map. Omitting it would increase the error of the forward kinematic map at the velar place of articulation. For this reason, the revised manuscript does not omit mention of the velum.

**p. 18, Fig. 4 Perhaps some additional notes on Fig. 4 would be useful (unless I missed something it is only referred to very briefly in the text). The amount of variance explained (for tongue and lip factors) looks quite low (if we assume that the most realistic number of jaw factor is 1), given that the corpus is very simple. Why are there no data points for e.g. 1, 2, 3 tongue factors and 1 lip factor? Are the results averaged over speakers?**

If the only variance component the statistical factors reflect is vocal tract movement, then the factor analysis model will not explain 100% of the variance in the data-set. This is because MR imaging and image processing introduce other sources of variance besides the imaged movement.

Nevertheless, depending on the parameterization, the percent variance explained is over 90% for the tongue and 80% for the lips. Given that the percent variance explained saturates as more factors are added to the model, adding still more factors to the model would yield diminishing returns. High-numbered factors would likely reflect nonlinguistic sources of variability. Including such factors may even add noise to the articulator synergy biomarker.

The reason that there are no data-points for 1, 2, 3, 5, and 7 tongue factors and for 1 lip factor is that these were not analyzed. The goal was to evaluate for a wide range of factor analysis parameterizations, not to be exhaustive.

Yes, results are averaged over speakers. The revised manuscript notes this in the caption.

**p. 22, Fig. 5. Maybe consider using either the same y-axis scaling for each of groups (a) and (b), or adjusting the scaling so that dashed line is at the same location graphically in all axes.**

**The last phase in the legend ("whenever ....") is not clear, as the dashed line actually seems to be present everywhere.**

The figure in the revised manuscript uses the same y-axis scaling for all plots.

**p. 23 Despite the information in Fig. 6, I still find the use of the terms "constriction onset" (and "release offset") somewhat confusing, e.g. I would find something like "onset of movement towards constriction" less ambiguous. Presumably also the second line of the legend of Fig. 6 could read "transition from vowel [a] to glide [j] and back to [a]" for completeness. And I may have missed it, but how was the start and end of movement to/from pharyngeal target defined?**

The revised manuscript adopts the suggested revisions. The figure caption refers to Section II.A. for definitions.

**p. 26/27 the reference to "open vocal tract posture of the vowel [a]" closely followed by "constriction at the pharyngeal place of articualtion of the vowel [a]" is confusing.**

The revised manuscript clarifies that the utterance has two [a] sounds (an initial and a final [a]).

**p. 27, line 357. Why is [i] involved?**

The revised manuscript standardizes reference to [i] and [j] by using [i] throughout and explaining that the speakers varied in exactly how the sound was produced (see Section II.A.).

**p. 31 Please recall to the reader what Scan1 and Scan2 refers to.**

The reader is reminded that “Study participants repeated the MRI experiment for a total of two MRI scans.”

**p.22 bottom and p.33 bottom. Given the complexity of the analyses, at the end of both these sections some kind of preliminary summary of the main points to be taken away from these sections would be helpful, before moving onto the next section.**

We added a short take-away to Section V.B. and expanded its segue into Section VI. In the case of Section VI.C., the section ends with a detailed analysis of the precision, which we believe provides useful guidance to the reader as they begin the section on task-specificity.

# Bibliography

Atkeson, C. G. (1989). Learning arm kinematics and dynamics. Annual review of neuroscience, 12(1), 157-183. DOI: ﻿10.1146/annurev.ne.12.030189.001105

Craig, J. J. (2005). Introduction to robotics: mechanics and control (Vol. 3, pp. 48-70). Upper Saddle River, NJ, USA:: Pearson/Prentice Hall. p. 6.

Ramanarayanan, V., Parrell, B., Goldstein, L., Nagarajan, S., Houde, J. (2016) A New Model of Speech Motor Control Based on Task Dynamics and State Feedback. Proc. Interspeech 2016, 3564-3568. DOI: 10.21437/Interspeech.2016-1499

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach.Psychological Review, 94(1), 84-106. doi:http://dx.doi.org.libproxy2.usc.edu/10.1037/0033-295X.94.1.84

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. Ecological psychology, 1(4), 333-382.

Shadmehr, Reza, and Sandro Mussa-Ivaldi. Biological Learning and Control: How the Brain Builds Representations, Predicts Events, and Makes Decisions, edited by Tomaso A. Poggio, and Terrence J. Sejnowski, MIT Press, 2012. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/socal/detail.action?docID=3339375. Created from socal on 2018-09-23 12:52:04.

J. E. Shoup, "Phonological Aspects of Speech Recognition," in Trends in Speech Recognition, W. A. LEA, Ed. Englewood Cliffs: Prentice Hall, 1980, pp. 125-138

Todorov, E., Li, W., & Pan, X. (2005). From task parameters to motor synergies: A hierarchical framework for approximately optimal control of redundant manipulators. Journal of robotic systems, 22(11), 691-710. DOI: 10.1002/rob.20093