Ref: PHON15-82  
Title: Investigating dialectal differences using articulography  
Authors: Martijn Wieling, Dr.; Fabian Tomaschek, Dr.; Denis Arnold, Dr.; Mark Tiede; Franziska Bröker; Samuel Thiele; Simon N Wood, Prof.dr.; R. Harald Baayen, Prof.dr.  
Manuscript Type: Research Article  
  
Dear Mr. Wieling,  
  
Thank you for submitting the above paper to the Journal of Phonetics. My apology for not having communicated my decision with you earlier. I have long waited for a third reviewer's comments in vain, and, as I promised, I decided to make a decision with two reviewers' comments that are already very thorough and constructive. Both reviewers agreed that the paper provides very interesting and useful methods to quantify articulatory differences between speaker groups (this time, of different dialectal populations), having potentially innovative contribution to the field of phonetics. Both reviewers, however, also expressed some reservations due to the paper's failure of contextualizing the study, and discussing the results in linguistically significant ways, which makes the ms quite far from being publishable in its current form.  
  
I read through the whole paper closely before I looked at these reviewers' comments, and my own assessment of the paper was indeed in line with that of these reviewers', especially regarding the necessity for the ms to be fully expanded in terms of what specific linguistic questions can be directly tested and answered using this very interesting and innovative method.  
  
Here are some highlights from reviewers' comments.  
Both R1 and R2 provided helpful pointers to how to address the biggest question on the linguistic relevance of the present study (except for showing that the groups of different dialects have different articulatory postures). You need to select some subsets of data regarding specific phonological differences (e.g., certain vowel quality differences that are assumed to be phonemically the same) between the two dialects and discuss the results in linguistically relevant ways.

You also need to divide the data into different subsets, so that, for example, you should strengthen the study regarding regional/dialectal pronunciation vs. standard pronunciation; and words vs. non-word sequences for CVC.

We have created a single model allowing the direct comparison between the dialect vs. standard results.

In fact, as R1 rightly pointed out, you need to include the list of the test words and their phonetic transcriptions used in the present study, which will certainly illuminate the nature of the differential differences observed in the present study.

OK. We’ve included the list and their associated (dialect) transcriptions.

Both reviewers also noted that  
something needs to be done regarding the possible confound that comes from other factors, particularly the age factor, given that there was a huge age gap in the data.

OK. We’ve reanalyzed the data excluding the older speakers. The results remained similar.

R1 also made a number of other important points: you need to make sure that the dialectal differences observed do not stem from (1) the phonological differences of the words that are chosen between the two dialects; and also not from

(2) the different measurement points between the speakers of the different dialects as the data were collected by different experimenters (for this, perhaps you need to provide further analysis or justify how your normalization procedure would resolve this issue).

Our data was collected by the same experimenter. The other people present were supporting the main experimenter (MW). This is now indicated in the text.

R1 also noted that you need to make the introduction/discussion "a bit more even-handed" in terms of strengths and weaknesses of EMA vs. other techniques.  
OK, we have done this.

R2 additionally pointed out that the observed asymmetric patterns between articulatory and acoustic (especially F1) data might be due to the fact that formant values were obtained from vowels (or possibly sonorants) while the articulatory data came from the whole articulatory trajectory including the consonantal components.

We have also looked at the articulatory data of only the vowels, and there still is an inconsistency between the formant measurements and the position of the tongue. We have reported this in the manuscript.

I would like to add here that we as experienced phoneticians know that automatically obtained formant frequencies are not really reliably (and even worse for F1). So I would like to ask you to obtain more (objectively) reliable (hand-corrected) formant values for the comparison, or the comparisons should be based with some subsets of data (for which only vowels are included, for example, maybe for particular words, and formants are corrected in some ways).

We have done this (for a subset of the data), and while the correlations are 0.87 for F1 and 0.83 for F1, the manual measurements show mostly higher correlations with the tongue position. Consquently, we have based the formant discussion on manual measurements. We have added this to the text.

Reviewers also made some other very useful comments that all merit your close attention.  
  
I have just a few minor points that are all essentially related to reviewers' comments:  
- On p. 2 "which means that coarticulation effects are generally ignore" -> this sounds like the coarticulation effect is the only thing this is missing. Can you make it more general, perhaps by saying something like "fine-grained subphonemic (phonetic) differences (e.g., coarticulation effects) are generally ignored"?  
Done.

- On p.4, if possible, could you not mention exactly how much was paid?  
Done.

- On p.5, please explain how one can be assured that the subjects (high school students) know how to pronounce standard Dutch;  
They are taught in standard Dutch; we added this to the text.

- On p.5, provide a bit more explanation on "spline" and "generalized cross-validation" in layman's terms preferably on footnote;  
Done.

- On pp.7-8, in the results section, you mentioned there were two approaches to be used to formally test the group distinction statistically, but in the subsequent text you sounded like you followed the second approach but being very vague about your choice (or at least not being clear about the use of the first approach);

We have clarified this.

- please be more specific on how the results on Table 2 are obtained;

We have specified this: it is the non-linear part of the model.

- spell out ACF in Figure 4;

Autocorrelation magnitude; we have specified this.

- the comparisons to linear discriminant analysis and formants came rather unexpectedly; please make a signpost earlier in the paper;  
  
We have done this at the end of the introduction.

So there appear to be substantial revisions to undertake (falling under the category of "major revisions"), and the biggest challenge for you is how to make the paper linguistically relevant which requires further detailed analyses of some subsets of the data as suggested by reviewers. But I feel that nothing is really insurmountable, and I really hope that you come up with ways of improving the paper along the line of reviewers' comments, so that your innovative methods can be readily accessible to the general readers of JPhon with greater impact.  
  
If you decide to revise the work, which I hope, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.  
  
(R1's comments are included in a separate file attached, and R2's comments are appended below. If you fail to retrieve the R1's file, please let us know by returning this email, so that our journal manager can send you the file directly.)  
  
In order to maintain the continuity of the editorial process, we have found it very beneficial to have revisions completed as soon as practical.  
  
I request that the revision be completed by 01 May 2016.  
  
Given the major nature of the revisions, however, I understand that a longer time frame may be necessary; if so, please contact the Journal Office concerning an alternative time frame.  
  
If you choose to revise your manuscript, please consult both the journal's Guide for Authors, and the information on electronic artwork preparation at the following website: <http://www.elsevier.com/artworkinstructions>.  
  
To submit a revision, go to <http://ees.elsevier.com/phonetics/> and log in as an Author.  You will see a menu item called "Submission Needing Revision".  You will find your submission record there.  
  
If for any reason you do not intend to submit a revision, please could you kindly let the Journal Office know immediately.  
  
NOTE: Upon submitting your revised manuscript, please upload the source files for your article. For additional details regarding acceptable file formats, please refer to the Guide for Authors at: <http://www.elsevier.com/journals/journal-of-phonetics/0095-4470/guide-for-authors>  
  
When submitting your revised paper, we ask that you include the following items:  
  
Manuscript and Figure Source Files (mandatory)  
  
We cannot accommodate PDF manuscript files for production purposes. We also ask that when submitting your revision you follow the journal formatting guidelines.  Figures and tables may be embedded within the source file for the submission as long as they are of sufficient resolution for Production. For any figure that cannot be embedded within the source file (such as \*.PSD Photoshop files), the original figure needs to be uploaded separately. Refer to the Guide for Authors for additional information.  
<http://www.elsevier.com/journals/journal-of-phonetics/0095-4470/guide-for-authors>  
  
Highlights (mandatory)  
  
Highlights consist of a short collection of bullet points that convey the core findings of the article and should be submitted in a separate file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point). See the following website for more information  
<http://www.elsevier.com/highlights>  
  
Please note that this journal offers a new, free service called AudioSlides: brief, webcast-style presentations that are shown next to published articles on ScienceDirect (see also <http://www.elsevier.com/audioslides>). If your paper is accepted for publication, you will automatically receive an invitation to create an AudioSlides presentation.  
  
Yours sincerely,  
Taehong Cho, PhD,  
Editor in Chief  
Journal of Phonetics  
  
Email: [phonetics@elsevier.com](mailto:phonetics@elsevier.com)  
Homepage: <http://www.journals.elsevier.com/journal-of-phonetics/>  
Online submissions: <http://ees.elsevier.com/phonetics/>  
  
Reviewers' comments:  
  
Reviewer #1: see attached file.

Reviewer #2: Investigating dialectal differences using articulography  
  
Review:  
The paper applies a new statistical modeling approach (generalized additive modeling) to a large articulatory data set as a method for studying dialectal variation.  The proposed statistical method proposed could well be an important advancement in the analysis of certain data types, but to warrant publication in the Journal of Phonetics, I believe the dialectal variation part should be interpreted/explored in more detail.  
  
The main result is that speakers from one Dutch dialect (Ter Apel) have a more retracted tongue position than speakers from a different dialect (Ubbergen), and this both when they speak dialect words and when they speak CVC sequences in standard Dutch. This result is however not interpreted (and perhaps it is indeed uninterpretable) with respect to what is known about the respective dialects. If the result is just showing that speakers in the two groups have different tongue postures (suggested by the similar results for dialect and standard speech), how could research such as this reveal articulatory differences between dialects? Is it assumed that pronunciation differences actually stem from tongue posture differences? Or is it actually the case that these uncovered posture differences reflect different pronunciation differences between dialects?

In our new analysis (showing similar results), we have taken into account the non-speech resting positions as well. We interpret these results as them really being caused by differences in articulation during pronunciation, which in turn reflect pronunciation differences.

I believe the authors should discuss their result more and in addition they may want to provide results of subsets of their data, if these could make it easier to provide an interpretation of what these tongue posture differences actually mean linguistically. To exemplify what I mean, I am going to use the word "taarten", which the authors provide with no interpretation as to what dialectal pronunciation differences the tongue postures may reflect. This word differs (among others) in vowel quality (back vs. front), which may account for the overall more posterior position in Ter Apel (with the back vowel) compared to the the Ubbergen position (with the front vowel). No such difference is observed for "boor" which has the same vowel quality in the two dialects. My speculation is that if the two dialects actually often differ in one favoring a front vowel and one a back vowel for the same word (and similarly for other sounds one favoring a more front variant of a consonant), the results could be interpreted in a way that is also linguistically relevant (i.e. capturing how the dialects may differ). This could in term make such an analysis more useful for dialectal research, and the paper of heightened interest for JPhon readers. So I suggest that the authors include a more thorough discussion of this issue, possibly with more detailed analyses of subsets of data if this makes a linguistically-relevant interpretation easier.

We have now added some single word comparisons for words which have an expected similar pronunciation (also shown in an additional table with the material). These also show the same pattern. Also if we fit a single model on the basis of all dialect words which are supposed to be pronounced similarly, we find the distinction between the two groups. Furthermore, if we focus the analysis on a single segment (/t/), we also find the same distinctions. So in sum, there appears to be a real difference in tongue movement between the two groups, which is also visible in subsets of the data.

I also think the authors should include a comparison within group across the two tasks: dialect and standard pronunciation (if unbalanced number of observation and standard CVC vs. a freer structure in dialect productions is a concern for the given statistical method, a subset of the dialect data could be selected that would resemble the standard data to some extent). It would be good to see that speakers do show a difference between the two tasks (i.e. that they at least attempted a standard pronunciation in the standard condition, even if the two respective standard conditions still bear the influence of dialect).

It was not really possible to select a subset of dialect data resembling the CVC sequences. However, the method does allow a comparison of the two tasks and on the basis of this analysis we see that the difference in posterior position is similar in both cases. We have reported this new model in the results.

For the comparison with formants, the authors should include in the articulatory analysis only the portions for which formants exist. Otherwise, it may be that the F1 pattern is determined by vowels, while tongue height position overall is determined by both vowel and consonantal positions (with consonants having a more robust height position), hence the unexpected result for F1.

We have done this and updated the manuscript to reflect this. There still remains a discrepancy between the formant measurements and the articulatory measurements.

Smaller issues:  
-       The very different ages in one group, with 16 teenagers and 6 adults over 40: could there be any differences especially with respect to dialect/standard usage? I think the authors should address this issue (I understand that the statistical model factors these differences out, but it may be one interesting potential grouping to further pursue).

Given that the older speakers were only included in the Ter Apel group, we have excluded them and only focus on the younger speakers.

-       Figure 2 is almost unreadable - fonts should be enlarged; it is not clear to me which is the jaw sensor and what the blue line is (maybe this is the jaw and if so, it means it has been attached somewhere under the jaw and not below the lower incisors as is customary with EMA, if so, this needs to be stated). )t would be useful to name all sensors and structures on the figure itself.

We have modified this graph and adapted the caption in order to improve the clarity of the figure. We removed the jaw sensor in order to provide a clearer view of the other sensors.

-       For the figures, it would be good to write on the figure or specify in the caption that 0 is anterior 1 posterior, 0 -low, 1 high (it is given in the text, but could be useful to be reminded of it when seeing the pictures themselves).

We have added this to the caption.

-       Correlations in comparisons with formant-based patterns. R-square are also useful to report, as they are easy to interpret as % of data explained.

We have added the r-square values of the correlations to facilitate interpretation

Reviewer #1

This is an excellent paper with complex analysis and which opens up new lines of research. It has very interesting methods, and potentially excellent results about vocal profile differences across dialects, extracted from EMA data. Moreover, the EMA data collection, from a huge sample, was done in the field, which is superb advance. It should certainly be published, if the questions I have can be resolved, particularly point C. I also have some minor quibbles about the text (See major point B) and the way that EMA is promoted but perhaps I am being overly picky with those comments – I hope not, because they meant in a collegiate way. Point A should be easy to resolve and really relates to whether the dialect word sample skews that part of the research. So long as the non-words and the dialect words behave alike, there is not too much of a problem.

Major questions.

1. The paper is very quantitative, and sets out very well the complex analytic requirements for comparing EMA across multiple speakers and dialects. There is rather a naïve perspective, however, on what benefits this approach might bring to people genuinely interested in dialect variation, and on the shortcomings of transcription. The greatest benefit will be in the level of detail which can be quantified, it seems to me – whereas the quantitative statistical comparison of words that are already audibly very different might not be very interesting, in the absence of a phonetic or dialectological research question.

OK, we have included the level of detail as a benefit in the paper.

It is good that the sample results in Figure 5 exemplify two forms which are clearly different and two which seem (from the transcription) to be more similar. However, choosing some dialect words which have radically different phonology makes it problematic to understand the averaged differences between the accents – are they skewed by the inclusion, for example, of /r/ allophones or different vowel forms? What is being reported here? Could the results be predicted from segmental phonological transcriptions? Is this a new dialectal result, or a quantification of an already-known one? Take the “boor” example. The two forms given are [bʊːr] for Ter Apel and [bʊːʀ] for Ubbergen – so we might predict EMA coils to be different in Ubbergen due to the use of a uvular rather than an alveolar. And so it is – but a better phonetic explanation would be required for the actual result.

We have changed the examples somewhat and also added two dialect items which have the same expected pronunciations in both dialects. Here the same effects can also be seen.

As we don’t have the transcriptions of each individual utterance, we tabulated the standard dialect transcriptions (shown in the added table) and categorized the vowels and consonants in front vs. back. This result shows that the Ter Apel group has a greater proportion of front consonants. The proportion of front vowels and front and back consonants hardly differs. Consequently, the results cannot be predicted from the phonological transcription (as the Ter Apel group has a pronunciation which is further back, rather than front).

The results for dialect words in Table 2 / Figure 7 & 8 show an overall result of this nature for the 70 dialect words. Without knowing the phonological make-up of these 70 dialect words, it is impossible to tell whether the overall result is due to segmental differences in the word-set. The wordlist was not included with the paper, though it is available online, though that might cause problems for blinding the review.

We have now included the word list with transcriptions in the manuscript.

So: it is essential to have a meaningfully representative set of words – this is achieved by the 27 item Dutch pseudo/real word set – and this was spoken in a standard accent, which is better. (The design might have been even better in both the dialect accent and the standard accent, with far fewer dialect words.) And, more importantly, the same result found for the dialect words is found in the non-word (CVC) set, which is very convincing, so long as it’s not based all around a huge difference in the /a/ vowel, say.

Given that our analysis takes into account the structural variability associated with each individual word, the results should not be contingent on a single vowel. Also, the difference is most clearly visible for the CVC sequences, which were balanced with respect to the three vowels.

1. While it is true that “Unfortunately, there are several drawbacks associated with the two articulatory observational methods described above”, there are also some benefits. There are also some drawbacks with EMA, see below. The introduction should be a bit more even-handed – there is no need to construct such a one-sided argument for EMA with this very obvious rhetoric. All the techniques have something to offer, and all have flaws. What is interesting is how their benefits and flaws DIFFER, making them useful for different research goals, as research teams find appropriate.

We have made this discussion more balanced; thanks for this criticism.

1. Dialect differences are found in normalised coil locations, in normalised time, during speech. But where is the baseline data which shows that there is no difference in coil location during non-speech resting positions? It’s not given at all. This is a fundamental confound problem, if it cannot be resolved, which might undo all the results, I’m afraid. The normalisation is as follows: “To enable an appropriate comparison between speakers, the positional information was normalized in such a way that 0 in the inferior-superior direction indicated the lowest (i.e. most inferior) point of the three tongue sensors and 1 the highest point (i.e. most superior). Similarly, 0 in the anterior-posterior direction indicated the most anterior position of the three tongue sensors, while 1 in this direction indicated the most posterior position” (page 6 line 21).

When we take the position of the non-speech resting position into account in the analysis, the results still show the same front-back pattern. This non-speech resting position was obtained in a separate recording at the start of the experiment and is taken into account as follows: For each speaker, we first subtracted the raw non-speech resting position (for the corresponding sensor and axis) from the raw position of each sensor and axis. Subsequently, this value was normalized in such a way that the total range was 1 (by dividing by the maximum minus the minimum value, as before). However, in contrast to the previous normalization, we did not subtract the minimum value before the division. Consequently, the range did not lie between 0 and 1, but could also contain negative values (e.g., from -0.4 to 0.6). Negative values indicate positions in front of (or lower than) the non-speech resting position, while positive values indicate positions behind (or higher than) the non-speech resting position of that sensor. As this normalization method appears to be more conservative (i.e. larger confidence bands) than the original method, we report his result in the paper.

And the first author MW plus author DA did coil placement in Ter Apel while MW did coil placement with a different author FT in Ubbergen. IT is therefore possible that the tongue or reference coil placement strategy was slightly different in the two places.

MW always did sensor coil placement, which is now indicated in the manuscript.

Also, the 15 school pupils varied in age from 13 to 19 in Ter Apel, plus 6 adults were included. 12 were male and 9 female. In Ubbergen, all participants are students, of whom 17 are male and 2 are female. There is no table of speaker age/sex to see if they are roughly equally distributed, but it seems there are age/sex confounds. In any case, adolescents’ heads vary hugely, both in absolute size and proportions – there is lots of mandibular growth in adolescent males. So while a table with external speaker characteristics – which is worryingly absent from the paper – should be included, more needs to be done.

Unfortunately, we do not have head measurements. However, we have excluded the adult speakers for a fairer comparison (and the results still remained the same). We also ran an analysis with only men and this did not affect the results. We have reported this in the paper.

Further, coil placement is described, but not quantified (despite comments in the paper about EMA being easy to quantify). Thus: “One sensor was positioned as far backward as possible without causing discomfort for the speaker. Another sensor was positioned about 0.5 cm behind the tongue tip.” Well – HOW FAR APART? It’s not acceptable that the mean distances between T1, T2, T3 are not reported and analysed statistically to see if there is a difference between the dialects – with perhaps some speaker age/sex angle built in as well.

Given that we included the non-speech resting position in the analysis, we do not believe this to be a problem. Nevertheless, we also obtained the difference between the T1 and T3 sensors for each young speaker and the average was 23.5 mm. for Ter Apel and 24.2 mm. for Ubbergen. These differences were not significant.

The data analysed is purely speech data. IF there is also a dialect effect for nonspeech data, then I’m afraid there would probably be a confound big enough to rule out publication until the issue could be dealt with. Interspeaker differences are not really all THAT likely to be responsible for these results, but at present there is no attempt to deal with the issue. This is an absolute crisis for the paper at present – the lack of reporting of speaker head size/coil placement s e.g. absolute locations of coils, is unacceptable, but is, hopefully, easy to fix.

We now have included the non-speech resting position of the coils in the analysis, and this showed the same pattern as before. Consequently, we do not believe these interspeaker differences to be responsible for these results.

Smaller issues and changes

Page 2

Abstract – change “introduces” – the paper does not introduce EMA! Also, omit “the measurement of the position of tongue and lips during speech,” since this is not a definition of EMA. While the paper is innovative - it most certainly does not introduce the idea that phoneticians might measure articulators in a dialectal study. Don’t over-claim the novelty or scientific value of the work, especially in an abstract.

This sentence has been rephrased.

Page 2 line 32 – “not perfect” – in fact it is a simplistic assumption that is only useful in the most informal way – both the use of a single formant such as F2 on the acoustic side, and the use of horizontal and vertical axes in a planar 2D analysis of the tongue and lips on the other. Your approach has real promise in improving on these terrible old analytic parameters that are really only used because (a) they are traditional and (b) because they are easy to measure. This comes up again at page 17 line 6. It’s a simple assumption with huge limitations if it actually meant to be anything more than that. It’s good enough for correlations for certain sorts of vowels, sure, particularly from the periphery of the vowel space in comparison to each other. Like the size of human heads and feet with a sample that includes adults and infants. See also below for p17. I am very pleased to see the more sceptical wording on page 18 line 48, but this is hardly a new criticism that follows just from your results.

We have rephrased the text somewhat at these places.

Page 2 line 43 – Aren’t there some more recent normalisation studies than Adank et al?

Yes, certainly, but Adank also focused on Dutch, so we prefer to keep the reference. We have added this to the text.

Page 2 line 59 – transcription does not address the underlying speech signal. It is a holistic descriptive representative encoding of the impression of the speech, combining articulatory and acoustic aspects, at an arbitrary level of detail. A transcription is not the same thing as an underlying lexical representation, which relates only to the most broad, abstract, segmental, phonemic level of transcription. You would NEVER rely on broad transcription for any meaningful phonetic analysis.

OK, we have adapted this part reflecting these suggestions.

Page 3 line 5 – so, as implied by my last criticism, a narrow transcription does NOT ignore coarticulation. You need to be more tolerant of other approaches and recognise what they are good for, and good at, not just present some negative characteristics of each.

Thank you, we have rephrased this.

Page 3 line 9 – is ease of articulation really KNOWN to be a driver of change? It is widely thought to be relevant, in a wide body of work including Sweet – and why use such an old reference for just this point in the paper, not throughout for a range of other long-held observations? It might be better to do what you do elsewhere and stick to rather more recent and/or modern approaches to lenition in change, such as Sebregts (2015) which is clearly relevant descriptively to your work as well as thematically. It’s also a good work for highlighting the importance of narrow transcription in Dutch dialectology, just for one segment and taking as a comparison point.

We have rephrased this somewhat and added the reference, also w.r.t. the narrow transcriptions..

Page 3 line 36 – Again the benefits of EMA are being highlighted, in a rather rhetorical strong way. Yes, tracking a single fleshpoint with ultrasound in the way EMA works is impossible – but ultrasound shows the tongue shape at a series of points through the vocal tract – something EMA can’t easily do, since the coils move in 2D. They are usefully different techniques. Also, you understate the drawbacks of EMA by saying the tongue cannot be tracked “completely”. No technique can do that – but EMA typically tracks just three locations on the tongue – that’s a long, long, long way from “completely”. Also, don’t note that sensors can’t be placed “too” far back – a colloquially disarming phrase. Nobody would want to place a coil “too” far back. In fact, these three coils are being places within a few cm of the tongue tip.

We have made this section more even-handed and rephrased these sentences.

Please state the average distance between the coils T1/T3, to make it clear how much and which part of the mid-sagittal tongue surface is being tracked (with range or some indication of variance). It’s usually about 4cm from about 0.5cm to 1cm behind the tip, which is around the front quarter of the tongue surface.

We added this to the section about articulatory data collection.

The reason EMA is “excellently suited for quantitatively analysis” is because established techniques exist already and are used by many researchers, not because EMA is somehow superior to other instrumental articulatory techniques – most of which are also excellently suited to quantification! – how about rewording to explain the current state of affairs… this is more rhetoric. See my point above – it is very easy to quantify the amount of surface modelled in EMA, but because it is a relative weakness, this simple bit of quantification is lacking in this (and most other) EMA papers.

We have rephrased this.

Another weakness of EMA is that there are only around 3-5 coils used – in a review like this, which is arguing for EMA’s usefulness to a potential new group of researchers, and where it is being compared to other techniques, it would be better to be upfront about this too.

We have added this.

Page 4 line 58 – there are indeed very few articulatory studies of Dutch looking at supralaryngeal articulation, but there is at least one previous big corpus, EUROM.1 (1990) which used laryngograph with 64 participants http://www.phon.ucl.ac.uk/resource/eurom1/DutchProject.pdf and a clinical EMA study, Nijland et al (2009) DOI: 10.1080/1476967031000091015

We have added the references. Though note that the EUROM corpus only included 9 speakers for the laryngograph recordings.

Also, FYI – though there are no published papers (so no need to cite if you don’t want to), there is work ongoing on Dutch in ultrasound by Ooijevaar as well as by Sebregts and Strycharczuk which I presume you are aware of.

http://www.meertens.knaw.nl/cms/en/staff/143931-etskeo

http://www.ultrafest2015.hku.hk/programme.html

Thank you for the links. We have included them.

Your paper, when it talks about “articulation” is really talking about LINGUAL instrumentation, specifically – not airflow, EGG, EPG or anything else.

We have specified this more clearly.

Page 4 line 2 – it might be nice to finish off this introduction with an ultrasound image to compare to the current Figure 1.

Given that the paper is about EMA and not ultrasound, we have decided not to include this.

More importantly, there have been huge developments in ultrasound since the Sebregts articulatory study’s data collection and analysis were undertaken. That paper was published in 2011, not 2010, according to this online preliminary version: http://www.qmu.ac.uk/casl/pubs/Scobbie%20Sebregts%20and%20Stuart-Smith%20WP18%202009.pdf

According to the publisher (https://global.oup.com/academic/product/interfaces-in-linguistics-9780199567249?cc=nl&lang=en&#), the book (in which the chapter appeared) was published on December 2, 2010. So we have left it as it is.

Page 5 line 12. Omit “a total of”

Done.

Page 6 line10. change “obtained data” to “data obtained”

Done.

Page 6 line 18. What is “start” of word? Specify if this is acoustic start or an articulatory start and note that these are not the same thing.

Acoustic start, due to the acoustic segmentation; changed.

Page 8 line 57 – typo for “package”

Corrected.

Page 10 line 4

The way 6 parameters from T1, T2, T3 relative to two axes is rather opaquely worded.

This is rephrased.

Page 11 Figure 6 – while “taarten” made a good example earlier on in the paper in combination with others, specifically in Figure 5, you cannot use it here: The two dialects have two different phonologisations for “taarten”. You might as well present a figure showing that “tand” and “koe” differ, or that Dutch “koe” and English “cow” differ. The whole point of your analytic superstructure is to reveal differences that are more subtle – we already know how the two forms in Figure 6 differ: [tʊːtn] vs. [toeʀtə]. Also, in Figure 6, when you use two more similar variants, do also add a rough segmental transcription time-aligned to the coil movement, or perhaps if not segmental, at least a transcription relevant to the coil in question.

We have changed the example to taat and explained in the caption where the differences were significant (at the /t/’s).

Page 17 – why would you expect an F2 correlation at all when the coil positions reflect consonants as well as vowels. I am again rather surprised that any results were found at all.

The correlations were only conducted using the vowel positions and F2 values for the vowels. This is now made more explicit in the text.

Renée van Bezooijen did a small VPA analysis of Dutch in comparison to other languages, but included articulatory parameters. Perhaps some dialect variation data is also available?

If you mean the paper: Van Bezooijen, R. (1988). The relative importance of pronunciation prosody and voice quality for the attribution of social status and personality characteristics. In: Roeland van Hout et al. (eds.) Language attitudes in the Dutch language area., then this only focused on speakers from around the area of Nijmegen, so dialectal variation will be limited. We have not changed the manuscript on the basis of this comment.