ACTIVE ADJUSTMENT OF THE CERVICAL SPINE DURING PITCH PRODUCTION COMPENSATES FOR SHAPE: THE ARTIVARK STUDY

Scott Reid Moisik[1], Dawn Poh Zhi Yun[1], & Dan Dediu[2,3,4]

[1]*Linguistics and Multilingual Studies, Nanyang Technological University, Singapore*
[2]*Laboratoire Dynamique Du Langage UMR5596, Université Lumière Lyon 2*
[3]*Language and Genetics Department, Max Planck Institute for Psycholinguistics*
[4]*Donders Institute for Brain, Cognition and Behaviour, Radboud University*
scott.moisik@ntu.edu.sg, ZHIYUN001@e.ntu.edu.sg, dan.dediu@univ-lyon2.fr

ABSTRACT

The anterior lordosis of the cervical spine is thought to contribute to pitch (f_o) production by influencing cricoid rotation as a function of larynx height. This study examines the matter of inter-individual variation in cervical spine shape and whether this has an influence on how f_o is produced along increasing or decreasing scales, using the ArtiVarK dataset, which contains real-time MRI pitch production data. We find that the cervical spine actively participates in f_o production, but the amount of displacement depends on individual shape. In general, anterior spine motion (tending toward cervical lordosis) occurs for low f_o, while posterior movement (tending towards cervical kyphosis) occurs for high f_o.

Keywords: pitch production, cervical spine, larynx height, MRI, ArtiVarK.

1. INTRODUCTION

The role of the cervical spine in pitch (or fundamental frequency, f_o) production has been the subject of interest in several articulatory studies (e.g. [3, 4, 13]). In particular, the anterior curvature or lordosis of the cervical spine evidently [4] causes rotation of the cricoid cartilage as the larynx moves downwards, which contributes to pitch production by reducing tension on the vocal folds without concomitant intrinsic build-up of tension (as would occur under thyroarytenoid muscle contraction). The problem is that there is variation in spinal curvature [5], so it is unclear how individual anatomical variation influences this mechanism of pitch control.

In [4], the vertical movement of the larynx along the cervical spine was observed to facilitate the rotation of the cricoid cartilage, an important mechanism in changing vocal fold tension. Additionally, MRI data from this study shows that the lordosis of the cervical spine appeared to be less pronounced during the production of higher pitches—a slight kyphosis appeared to aid in backward rotation of the cricoid cartilage. Conversely, spinal lordosis appears to be enhanced in low f_o ranges, facilitating the rotation of the cricoid cartilage to shorten vocal folds. A similar interaction between pitch production and spinal curvature was also noted in [7]. These subtle changes in spinal morphology in different f_o ranges suggest an active manipulation of the cervical spine in pitch adjustment [4].

Although previous studies have established that the shape and movement of the cervical spine plays a key role in controlling f_o, variations in spinal shape remain largely unknown. Honda et al. [4] discuss some inconsistencies in a previous investigation [3] arising from the different degrees of lordosis amongst subjects. We explore further here the question of whether inter-individual variation in spinal morphology has an effect on cervical adjustments during pitch production tasks.

2. METHODS

2.1. The ArtiVarK study

This study makes use of the ArtiVarK dataset (ethics approval 45659.091.14, 1 June 2015, Donders Institute for Brain, Cognition and Behaviour [DIBCB], Nijmegen), a large multi-ethnic sample (n = 90; 35 female; broad ethnic groups are 'Chinese' (C), 'North Indian' (NI), 'South Indian' (SI), and 'Dutch') of anatomical and speech production data. It includes high-resolution intraoral (three-dimensional) optical scans, magnetic resonance imaging (MRI) featuring structural anatomical scans and static and dynamic ('real-time') articulatory scans, and audio data collected during a phonetic training phase (used to elicit a wide range of familiar sounds, give training for several 'new' sounds, and provide preparation for MRI scanning) and during the real-time MRI (rtMRI) scans. Note that only a subset (n = 80) participated in the MRI component.

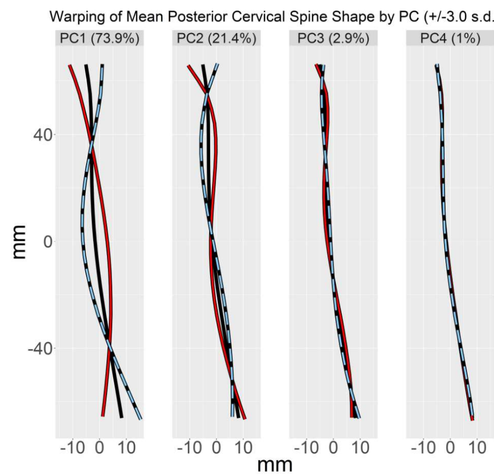
Here we only discuss details of the ArtiVarK methodology (see [1]) relevant here. Among various other tasks, participants carried out a pitch task in which they produced sequences of [afa] said at incrementally increasing or decreasing pitch along a 7- step scale. Participants were told they would hear several series of (decreasing or increasing) guide tones (sine wave tones ranging, in 30 Hz increments, between 100 Hz and 270 Hz) and instructed to follow each sequence to the best of their ability (no assessment was given on how well they performed). Note that given the very wide ethnic sample, the guide tones were uniform across all participants with no accommodation to the expected ranges of males and females; increments were made in Hz for simplicity. During the phonetic training phase, participants practiced two descending and two ascending sequences. During the MRI scanning phase, participants performed each pitch direction only once. Scanning was conducted on a 1.5T MRI system (Avanto, Siemens Healthcare) at the DIBCB (in the Donders Centre for Cognitive Neuroimaging or DCCN). The sequence was a Siemens' 2D single sagittal slice true fast imaging with steady state precession ("true FISP"; acquisition time = 15 s, frame rate = 6.67 fps, TE = 1.09 ms, TR = 148.48 ms, flip angle = 49°, slice thickness = 5 mm, field of view = 160 mm × 160 mm, voxel size = 2.0 mm × 2.0 mm × 2.0 mm). Because of the low frame rate, participants were told to produce the sequence slowly. Sound was recorded using a FORMI-III (Optoacoustics Ltd.) dual-channel optical microphone system with noise cancelling (and was later Wiener filtered to improve signal quality).

2.2. Analysis methods

The rtMRI data for the pitch task include (80 x 100 x 2 =) 16K frames, and, thus, it was deemed infeasible to manually segment the data for vocal tract structures of interest. For nearly semi-automated segmentation, we used the method outlined in [11], based on a MATLAB R2018b implementation and aided by starting contours of the vocal tract derived from manual tracing of the mean image taken across all rtMRI scans collected for a given participant (thus including scans not related to the pitch task described here). Owing to complexity of algorithm parameters, three segmentation passes were made with different parameter settings (discovered prior through experimentation) in an attempt to improve the odds of getting the largest number of good segmentations. Manual inspection of each segmentation resulted in discarding of 16 participants whose scans were too poor to be amenable to segmentation (or where the segmentation failed for other, unclear reasons).

With the vocal tract segmentation, we then automatically identified the location of the larynx in each image and obtained its mean location across all scans within a trial. From this location, we were then able to identify regions of interest within each video to form brightness distributions for larynx height (vertical) and for the cervical spine (horizontal). The mean of each such distribution was used as an estimate of vertical larynx displacement and cervical spine posteriority over time.

Figure 1: PCA of cervical spine shape illustrating warps: dashed/solid lines for +/–3.0 s.d. from the mean shape (middle black line) for PCs 1-4, which together explain most of the variation in the data.



In addition, for each frame, a fifth degree polynomial (chosen to provide a low-bias characterization of the raw trace data) was fit to the posterior contour of the cervical vertebrae (which is adjacent to the spinal cord), as it was judged to be less noisy than the part of the segmentation over the anterior contour of the cervical spine. The frame-wise means by trial of these contours were taken to characterize the shape of the spine via Procrustes superimposition with rescaling but no reflection (using the shapes package [2] in R [8]). This yielded a PCA (e.g., Fig. 1) of cervical spine shape, and the first three PCs (together explaining 98% of the variance) were used in subsequent analyses to determine the effect of shape variation on spine displacement and larynx height (our DVs of interest). Each participant is represented twice in this PCA (once for each pitch direction trial), and the paired-sample correlations for each PC from 1 to 5 were very high (> 0.65) and all significant, with PC1 (73.9% of the variance) showing a between-trial correlation of 0.96 ($p = 1.3e{-}39$). We took this to indicate that the PCA of cervical spine shape was consistent within participants.

Measures of f_o and intensity were obtained from the audio recordings. For f_o, 2nd-order polynomial curves were iteratively fit (with an ever-narrowing exclusion region starting at +/–25 Hz above and below the curve) to help improve the f_o signal. These signals were interpolated to a 100-step time-normalized scale.

2.3. Statistical modelling

We used Generalized Additive Mixed Models (GAMMs), following [10, 12] and aided by the `mgcv` package in R [14], to model (i) spine posteriority and (ii) vertical larynx displacement as DVs in two series of separate models with different effects structure (the full details of which cannot be given here). The general IV set includes the continuous covariates, PCs 1-2, f_o, and intensity (giving varying coefficient models for these predictors), and the factors, sex, group, and (pitch) direction. By-participant-and-by-direction random smooths were employed in every model. Each model was given a preliminary assessment for whether residual autocorrelation needed to be accounted for. Fitting was done with the scaled t-family to address non-normality of the residuals discovered with preliminary models. Model selection was done using the shrinkage smoother method [6], which is advantageous as it is carried out in a single step. Model diagnostics and residual analysis were performed for each model using the `gam.check()` function (in the `itsadug` package [9]) to ensure a good fit, normally distributed residuals, and that the basis dimensions of the smooths were adequate. Deviance explained was near 35% for all models, suggesting further parameters might help provide an even better fit. For the sake of generating plots across the factors sex and group, two additional, nested models were run with these variables excluded (otherwise plotting is forced to occur at specific levels of these predictor variables).

3. RESULTS

In both models, most of the IVs for the smooth terms are significant. We focus here though on the pattern over time and in relation to cervical spine shape.

3.1. Spine displacement

A GAMM run with spine posteriority as DV shows that smooths by direction are significantly different from zero, meaning that the cervical spine changes its position over the course of the task. Visualization of the smooths allows us to see that, across sexes and groups, the spine becomes more posterior as a function of increasing pitch, and more anterior with decreasing pitch (Fig. 2).

The smooths are ‘wobbly’ because of the cyclic nature of the task (with participants sometimes breathing in between each utterance of [afa]). More importantly, the amount of displacement over the course of the video is, on average not very large, about 1 mm either way. Direct inspection of the videos indeed reveals that some participants do not seem to adjust the spine at all, while others show fairly large displacements.

Figure 2: Smooths for spine posteriority over time (frame) for increasing (dotted) and decreasing (solid) pitch by sex (top row = female) and group.

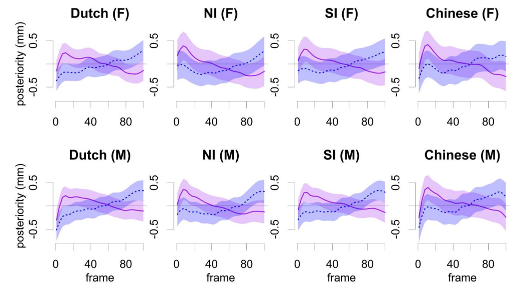
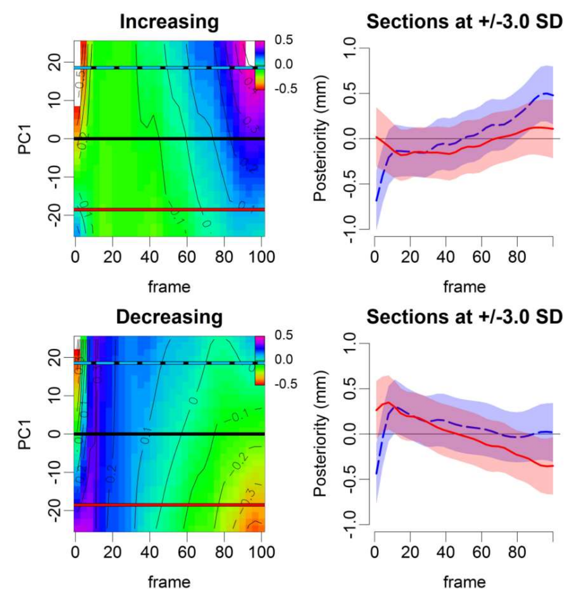


Figure 3: Contour and section plots of spine posteriority for PC1 in relation to frame number (time), increasing (top) and decreasing (bottom) pitch across sex and group (based on the model without these predictors). Mean (black line) and +/– 3.0 s.d. (dashed/solid lines) away from the mean (see Fig. 1 for PC1 interpretation).



When pitch is increasing, the spine adjusts towards kyphosis (concavity); for decreasing pitch, the lordosis is enhanced. Critically, however, as shown in Fig. 3, the degree of cervical spine displacement over time varies non-linearly as a function of shape of the cervical spine. Specifically, for PC1, the GAMM indicates this variation is significant (edf = 6.2, ref.df = 40, F = 0.5, p = 1.7e–5). (PC2 is also significant but we will not discuss this here.) If the spine is straight (or even kyphotic), as indicated by low values of PC1, then extra anterior spine movement is observed during pitch lowering. If the spine is convex (anterior lordosis), then extra posterior movement occurs during pitch raising.

3.2. Larynx height

A GAMM with larynx height as DV indicates that it varies nonlinearly over time. Fig 4 shows that indeed larynx height changes in correspondence with pitch (edf = 3.6, ref.df = 40, F = 0.5, p = 2.7e–10). Larynx height ranges over about 7 mm during the course of either pitch direction. Unlike the spine, for most groups (except non-Dutch females, all having small sample sizes), the larynx oscillates because of the periodic nature of the task (and a tendency for larynx lowering during inspiratory breaths between utterances of [afa]).

Figure 4: Smooths for larynx height over time (frame) for increasing (dotted) and decreasing (solid) pitch by sex (top row = female) and group.

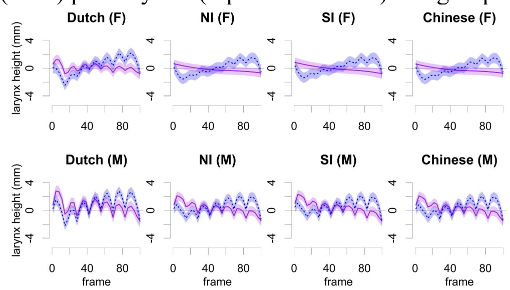


Figure 5: Time-evolution plots for a SI male comparing posterior cervical spine shape to larynx height (frame = shading). Left is anterior (negative x-axis values). Dashed line is mean larynx height for pitch decrease.

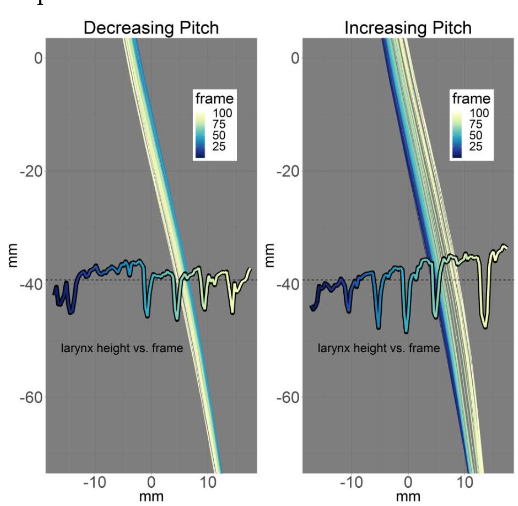


Fig. 5 shows cervical spine contour and larynx height as a function of time (note that, larynx height is also depicted across the x-axis, but this is just for visualization purposes). This participant is chosen somewhat arbitrarily (as the first participant) but happens to show the cervical spine displacement (anterior and posterior) and larynx height (lowering and raising) patterns for both pitch task directions (decreasing and increasing).

4. DISCUSSION & CONCLUSION

We hesitate to view the pitch tasks in our study as either particularly speech-like or singing-like, but rather somewhere in-between, and so we must reserve some caution before generalizing these results to either of those contexts (for instance, to the execution of intonation contours or tone contrasts in natural speech). The range of f_o excursion, however, was not particularly large, and can be taken to be within the normal operating range for most participants. Occasionally, participants were observed to perform extreme pitches (presumably by accident or out of straining at the edges of their pitch range). Some participants had difficulty adjusting pitch and would instead alter their intensity (or a combination depending on the specific step along the pitch scale). Despite the universal (and somewhat low) pitch scale, males and females seemed to perform the task equally well.

It is not entirely certain what mechanisms underlie the adjustments to cervical spine position. We might posit that when participants enhance anteriority of the spine, they do so for reasons similar to those put forth by Honda et al. [4], that cervical lordosis interacts with larynx lowering to produce a rotation about the cricothyroid joint favourable for lowered stress on the vocal folds and hence lowered f_o. Even if lordosis is not achieved (but rather the spine simply moves anteriorly), the cricoid cartilage may still be pushed towards the thyroid cartilage by the spine, achieving a similar effect. But why should posterior displacement help with increasing pitch? We can only suggest that, through connective linkage between the pharynx and the larynx, this action may increase stresses throughout the laryngeal system, possibly even pulling the cricoid cartilage differentially backwards (in relation to the thyroid cartilage) and thereby increasing vocal fold tension. We did not observe any obvious posterior motion of the cricoid in contrast to thyroid position, but the resolution of the MRI scans might make this infeasible. Also it should be kept in mind that the displacement is of a small magnitude (about 1 mm on average across the pitch range), and it is unclear whether this amount is enough to be of practical significance.

We consider this only a preliminary step towards a full analysis of the data, which are complex and include many factors (e.g., sex and group) and covariates (e.g., f_o, intensity, formants, other anatomical measures) of interest that could not be addressed here. The measure of spine displacement (spine posteriority) is rather simple, but it would also be possible to gauge actual lordosis/kyphosis more directly by looking at curvature.

5. REFERENCES

[1] Dediu, D. and Moisik, S.R. 2019. Pushes and pulls from below: Anatomical variation, articulation and sound change. *Glossa: a journal of general linguistics* 4, 1, 7 DOI:https://doi.org/10.5334/gjgl.646.

[2] Dryden, I.L. 2018. *shapes: Statistical Shape Analysis*. R Package.

[3] Hirai, H., Honda, K., Fujimoto, I. and Shimada, Y. 1994. Analysis of magnetic resonance images on the physiological mechanisms of fundamental frequency control. *Journal of the Acoustical Society of Japan* 50, 296–304.

[4] Honda, K., Hirai, H., Masaki, S. and Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in f0 control. *Language and Speech* 42, 4, 401–411.

[5] Lammert, A., Proctor, M. and Narayanan, S. 2013. Morphological Variation in the Adult Hard Palate and Posterior Pharyngeal Wall. *Journal of Speech, Language, and Hearing Research* 56, 2, 521–530 DOI:https://doi.org/10.1044/1092-4388(2012/12-0059).

[6] Marra, G. and Wood, S.N. 2011. Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* 55, 7, 2372–2387 DOI:https://doi.org/10.1016/j.csda.2011.02.004.

[7] Miller, N.A., Gregory, J.S., Semple, S.I.K., Aspden, R.M., Stollery, P.J. and Gilbert, F.J. 2012. The Effects of Humming and Pitch on Craniofacial and Craniocervical Morphology Measured Using MRI. *Journal of Voice* 26, 1, 90–101 DOI:https://doi.org/10.1016/j.jvoice.2010.10.017.

[8] R Core Team 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

[9] van Rij, J., Wieling, M., Baayen, R.H. and van Rijn, H. 2016. *itsadug: Interpreting Time Series and Autocorrelated Data using GAMMs*. R Package.

[10] Sóskuthy, M. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv:1703.05339 [stat]*.

[11] Wang, L., He, L., Mishra, A. and Li, C. 2009. Active contours driven by local Gaussian distribution fitting energy. *Signal Processing* 89, 12, 2435–2447 DOI:https://doi.org/10.1016/j.sigpro.2009.03.014.

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

[13] Wilson Arboleda, B.M. and Frederick, A.L. 2008. Considerations for Maintenance of Postural Alignment for Voice Production. *Journal of Voice* 22, 1, 90–99 DOI:https://doi.org/10.1016/j.jvoice.2006.08.001.

[14] Wood, S.N. 2017. *mgcv: mixed GAM computation vehicle with automatic smoothness*. R Package.