THE ACOUSTICS AND ARTICULATION OF MANDARIN SIBILANTS: IMPROVING OUR DATA BY MODELING THE PALATE WITH EMA

Chris Neufeld[1] and Andrei Anghelescu[2]
[1]Dept. of Speech and Language Pathology, University of Toronto, christopher.neufeld@utoronto.ca
[2]Dept. of Linguistics, University of Toronto, aanghelescu@gmail.com

1. INTRODUCTION

Electromagnetic articulography (EMA) is a tool for tracking the motion of the articulators during speech (1, 2). One drawback of fleshpoint measurements is the absence of clear anatomical boundaries. Although we may know the location of EMA coils in 3D space (here *x* is horizontal displacement, *y* lateral, *z* vertical), we have only a rough idea of the tongue coils' location relative to the palate. Here we outline a procedure for creating a 3D model of the hard palate, and offer a small exploratory dataset examining Mandarin sibilants (3), finding that our articulatory measurements are improved if they are redefined relative to the palate.

2. METHODS

Mandarin has a three-way place contrast for sibilants: alveolar, palatal and retroflex. For each place there are three possible manners of articulation: fricatives, unaspirated affricates and aspirated affricates. Phonotactically licit monosyllables were created by pairing target consonants with vowels [a] and [ɤʊ̯] with high level tone. Target words were put in the frame *'tā shuō __ ma?'* "Does he say__?" Stimuli were presented in a random order on a screen in Pinyin for 20 seconds each. One male participant repeated each phrase continuously while it was displayed, and the set was repeated twice, yielding roughly 25 tokens per condition.

Speech was recorded at 64 kHz with a head-mounted mic. Consonants were manually annotated in Praat, and the centre of gravity (COG) was extracted with a 25 ms Hanning window every 5 ms for the duration of each consonant. For each consonant, we found the peak COG and the relative time of peak COG (consonant = 0, offset = 1).

Data was collected using the AG-500 system, with a sampling rate of 200 Hz. All trials were normalized to the occlusal bite plane, filtered and smoothed according to standard procedures (4). We constructed a 3D model of the participant's palate by tracing its surface with an EMA coil attached to a stick. A 3D surface was fit to the palate data using the lowess algorithm with a span of 5%. The constriction location (*cl*) and constriction degree (*cd*) of the tongue tip (TT) and tongue body (TB) coils were calculated using this model. At each sample, the shortest line in the sagittal plane between the tongue coil and the palate model was found. The line's length is *cl*, and its point of intersection with the palate model is *cd* (5). For EMA data in both absolute and palate-relative space, we found the point of maximum closure within the span of the consonant (highest *z*-value for absolute space, lowest *cd* for palate-space). At this point, we extracted (*x*, *z*)/(*cl*, *cd*) and the relative time of maximum closure. All articulatory measurements discussed below are extracted at the point of maximum stricture.

3. RESULTS

Place of articulation (PoA) of Mandarin Sibilants is readily distinguished by peak COG. For all manners of articulation, alveolar sibilants had the highest COG, retroflex the lowest, and palatal in the middle. Manner of articulation is distinguished by the relative time of the COG peak, with fricatives' COG peak occurring roughly in the middle of the consonant, aspirated affricates' peak COG occurring quite early, and unaspirated affricates' coming late.

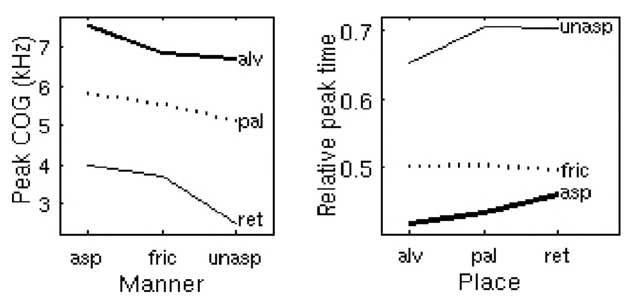

Figure 1. Mean peak COG & relative time by consonant

In figure 2, we show *x*/*cl* at the point of maximum constriction for TT and TB. While there is some overlap in all four boxplots, it can be seen that TB*cl* in palate space gives the most categorical picture of PoA for Mandarin Sibilants, and conforms to descriptive accounts of these speech segments (3), with alveolar being the most front, retroflex the most back, and palatal in the middle. In figure 3, we show a series of scatterplots, with peak COG on the abscissa, and spatial variables on the ordinates. A line of best fit is shown, and the corresponding Pearson's *r* is written in each panel. Examining these scatterplots we should hope to see either a clear linear relationship between acoustics and articulation, or three clusters which correspond to PoA. Here we can have our cake and eat it too: for TB in palate-relative space, *cl* exhibits a fairly linear relationship with peak COG. When peak COG is plotted as a function of TB*cd*, we see three fairly obvious clusters. In figure 4 we present this scatterplot sorted by PoA, and see a highly categorical grouping. For all other articulatory variables, we do not see as high a linear correlation, nor

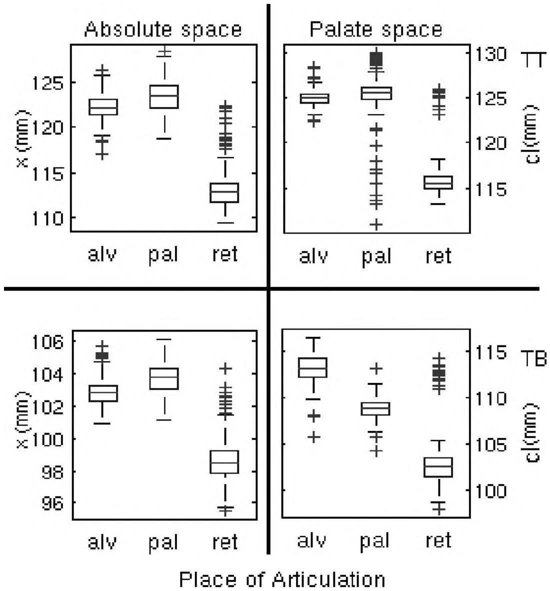

Figure 2: x/CL by PoA and coil in different co-ordinate spaces

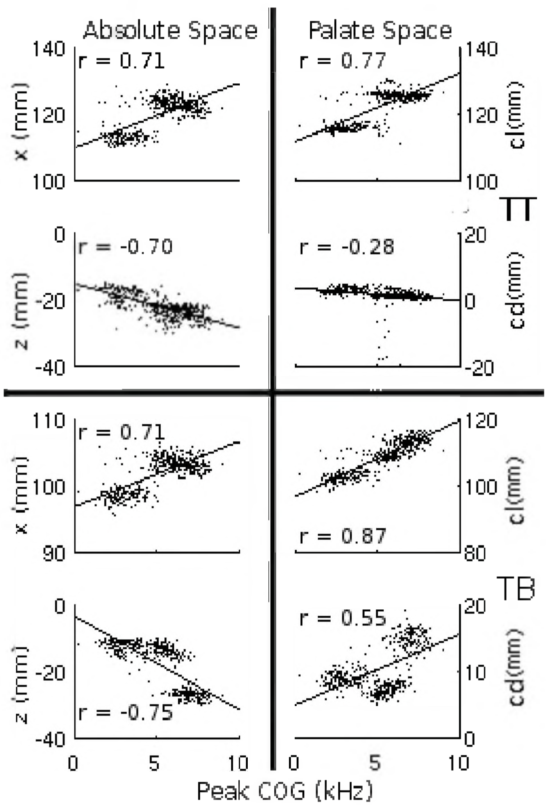

Figure 3: COG and articulatory variables

do we observe such a clear 3-way grouping. These findings suggest that what may *appear* to be a quantal effect (6) such as the relationship between COG and TB*x* in absolute space, may actually be a simple linear relationship if translated into palate-space. Furthermore, true quantal effects may also be *more* quantal in palate-space, as in the relationship between COG and TB*cd*. Or, to put it another way, since palate-relative space more accurately captures the shape of the oral cavity, acoustic-articulatory correspondences are more accurately revealed, and true non-linearities are amplified and false non-linearities may be 'linearized.'

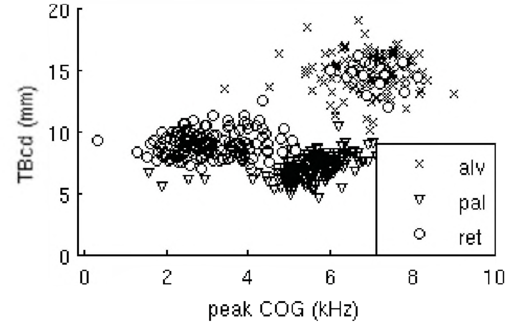

Figure 4: Correlation between COG and TBcd by PoA

4. DISCUSSION AND CONCLUSIONS

In addition to providing some descriptive data on Mandarin sibilants, we believe we have also demonstrated several benefits of redefining the location of EMA tongue coils relative to the palate. Describing EMA data in this fashion both gives a clearer picture of the physical state of the vocal tract, and has a closer affinity with several theoretical approaches to phonology or speech motor control (i.e. Articulatory Phonology (7), Task Dynamics (8)) which are couched in terms of variables or goals of the vocal tract, rather than purely spatial parameters. We conclude that this may be a useful method for deriving vocal tract variables from EMA data, and hope this continues to be found to be a productive research tool.

REFERENCES

[1] Yunusova Y, Green J.R. & Mefferd A. (2009). Accuracy Assessment for AG500, Electromagnetic Articulograph. *Journal of Speech, Language and Hearing Research*. 52, 547-555.

[2] Hoole, P. & Zierdt, A. (2010). Five-dimensional articulography. In P.H.H.M van Lieshout & B. Maassen (Eds.), *Speech Motor Control: New developments in basic and applied research* (pp. 331-349). Oxford University Press.

[3] Ladefoged, P. & Wu, Z. (1984) Places of articulation: an investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12: 267-278, 1984.

[4] Van Lieshout, P. H. H. M., & Moussa, W. (2000). The assessment of speech motor behaviors using electromagnetic articulography. T*he Phonetician*, 81(1), 9-22.

[5] Neufeld, C. & van Lieshout, P.H.H.M. (2011). Alternative co-ordinate spaces for EMA. Submitted to JASA.

[6] Stevens, K. N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In E.E. David & P.D. Denes (Eds.), *Human communication: a unified view*. (pp. 51-66)

[7] Browman, C.P. & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan & J. Anderson (Eds.), *Phonology Yearbook 3*. Cambridge: Cambridge University Press, pp. 219-252.

[8] Saltzman, E.L. & Munhall, K.G. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333-382.