A Data-Driven Phoneme-Specific Analysis of Articulatory Importance

Anusuya P K, Aravind Illa, Prasanta Kumar Ghosh

Electrical Engineering, Indian Institute of Science (IISc), Bangalore, India

anuknkp@gmail.com, aravindi@iisc.ac.in, prasantg@iisc.ac.in

Abstract

Speech can be seen as a result of temporally overlapping gestures of articulators comprising lips, jaw, tongue and velum. The spectrum of acoustic speech signals is influenced by these articulatory gestures. In the articulatory space, few articulators exhibit minimal variance at their target position for a specific phoneme, which are called critical articulators. The positions of these critical articulators govern the acoustic characteristics of that particular phoneme articulation. This results in a categorical decision on the articulators as critical or non-critical articulators (Jackson and Singampalli (2009)). Instead of assigning a binary decision on articulators being critical or non-critical, in this work we attempt to assign an articulatory importance value between 0 to 1, in a data-driven manner. We denote it as articulatory importance function (AIF). Experimental analyses are performed on 38 subjects' acoustic-articulatory data and AIF values are reported. Findings from this analysis could contribute to the understanding of inter speaker variability in speech production and variability across different languages.

Keywords: speech production, critical articulators, electromagnetic articulograph

1. Introduction

In speech production, the generation of a particular phoneme is contributed by specific articulators, critical for producing that phoneme. Articulators that influence the production of a particular phoneme without which the sound cannot be produced completely are called critical articulators. Analysing critical articulators is important because by enabling the identification of which articulators are involved in producing specific sounds, it can help improve the speech production models. In the articulatory phonology by Browman and L. Goldstein (1992), gestural scores record the gestures of each articulator in the production of a word. This study clearly describes that not all the articulators are very important to produce a particular phoneme, thus making specific articulators critical.

Recasens, Pallarès, and Fontdevila (1997) explained critical articulators using phonetic invariance in the articulatory space, while Bladon and Al-Bamerni (1976) used articulatory resistance for the phoneme /l/ to explain the same phenomenon. Attempt to model the dynamic movements of articulators analytically towards phone-specific goals led to gestural approaches (Ohala, Browman, and L. M. Goldstein 1986; Saltzman and Munhall 1989; MacNeilage 1970; Liberman 1970). Jackson and Singampalli (2009) suggested a statistical approach to measure the criticality of the articulators in which the Kullback-Leibler distance between the distributions of different articulators was used to identify articulators as critical, dependent or redundant. In the work by Ananthakrishnan and Engwall (2008), the importance function is calculated based on change of velocity and angle of the articulatory trajectory. They analysed that the articulator reaches the critical position when there is a drop in velocity or change in angle. To reach the next critical location, the velocity increases and probably with a change in the angle. Wang, Green, and Samal 2013 used SVM model to identify important articulators. Three levels are used by Mermelstein to rank how critical an articulatory gesture was to a given phone (Mermelstein (1973)). Instead of assigning a binary decision on articulators being critical or non-critical, in this work we attempt to assign an articulatory importance value between 0 to 1, in a data-driven manner. We denote it as articulatory importance function (AIF).

2. Data collection and preprocessing

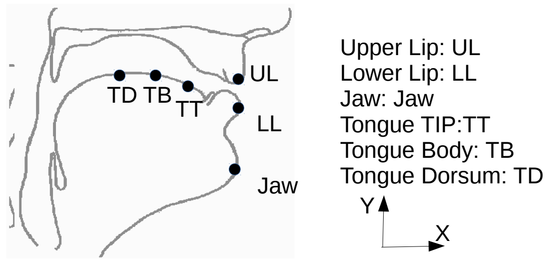

Figure 1: *Schematic diagram indicating the placement of EMA sensors*

For experiments, 460 MOCHA TIMIT sentences were chosen as speech stimuli to collect acoustic-articulatory data using electro-magnetic articulograph (EMA) AG501[1]. To capture articulatory movements, six sensors were attached to speech articulators namely, upper lip (UL), lower lip (LL), jaw (Jaw), tongue tip (TT), tongue body (TB) and tongue dorsum (TD) as shown in Figure 1. For head movement correction two sensors were placed behind the ears. We considered the articulatory movements in the midsagittal plane in X and Y directions which indicate horizontal and vertical directional movements of articulators, respectively. This results in 12 articulatory features denoted by, UL_x, UL_y, LL_x, LL_y, Jaw_x, Jaw_y, TT_x, TT_y, TB_x, TB_y, TD_x, TD_y. A total of 38 Indian subjects' data was considered in this study, out of which 24 were male and 14 were female. All the subjects were from an age group of 21-28 years and fluent speakers of English with no record of speech disorders in the past. We followed a recording setup and post-processing of articulatory data similar to that described in Illa and Ghosh (2018). The phonetic boundaries were obtained by force alignment using Kaldi speech recognition tool kit (Povey

[1] http://www.articulograph.de/

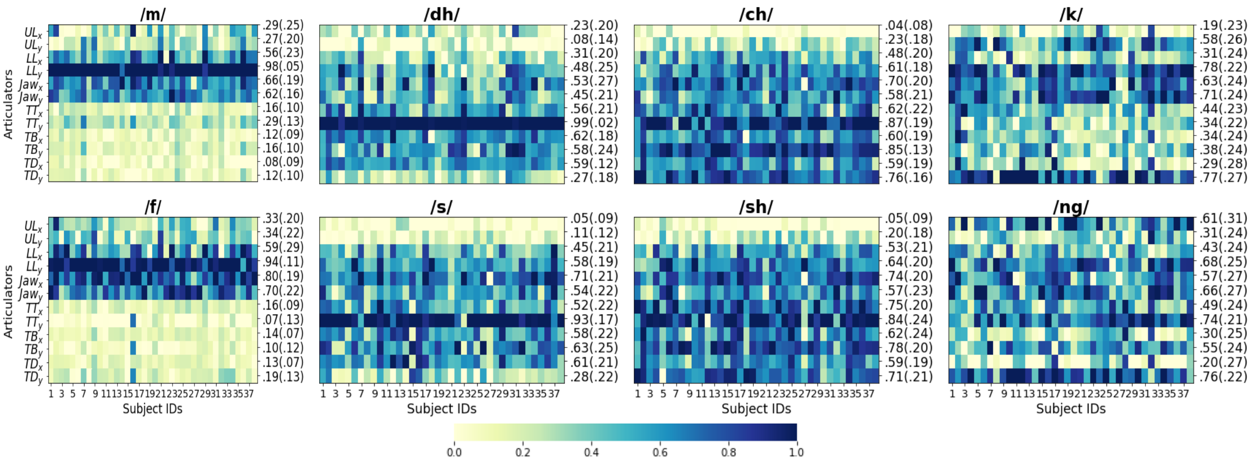

Figure 2: Normalized articulatory importance function values (color-coded from 0-1) for different consonants for each of 38 subjects (x-axis shows subject index, y-axis indicates different articulators)

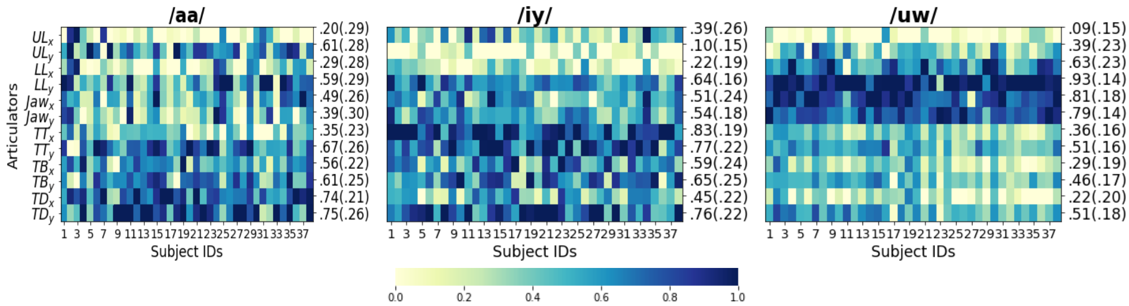

Figure 3: Normalized articulatory importance function values (color-coded from 0-1) for different vowels for each of 38 subjects (x-axis shows subject index, y-axis indicates different articulators)

et al. 2011). The phonetic transcription of the data set consists of 39 ARPABET symbols.

3. Articulatory Importance Function

The importance of k^{th} articulator (a_k) is defined using the phoneme specific positional variance ($\sigma_p^{a_k}$) calculated from the collection of samples from the midpoints between the corresponding phoneme boundaries, and global variance ($\sigma_g^{a_k}$) calculated from the collection of samples from midpoints between the boundaries of all phonemes. The Importance of an articulator is calculated using negative logarithm of the ratio of phoneme specific variance to the global variance as given by

$$i^{a_k} = -\log \frac{\sigma_p^{a_k}}{\sigma_g^{a_k}} \tag{1}$$

i^{a_k} is the importance function of an articulator that takes values greater than zero for $\sigma_p^{a_k} < \sigma_g^{a_k}$. The lesser the $\sigma_p^{a_k}$ than the $\sigma_g^{a_k}$, more is the importance of the corresponding articulator. To bound the range of values between 0 to 1, we further normalize i^{a_k} to I^{a_k} using the equation below to obtain AIF.

$$I^{a_k} = \frac{i^{a_k} - min_k(\{i^{a_k}\})}{max_k(\{i^{a_k}\}) - min_k(\{i^{a_k}\})} \tag{2}$$

4. Results

In the database of 460 sentences, 39 phonemes are observed and those include /aa/, /ae/, /ah/, /ao/, /aw/, /ay/, /b/, /ch/, /d/, /dh/, /eh/, /er/, /ey/, /f/, /g/, /hh/, /ih/, /iy/, /jh/, /k/, /l/, /m/, /n/, /ng/, /ow/, /oy/, /p/, /r/, /s/, /sh/, /t/, /th/, /uh/, /uw/, /v/, /w/, /y/, /z/ and /zh/. Articulatory importance function is examined for each of 39 phonemes using articulatory data of all 38 subjects separately.

In the first step, phoneme-specific and global variance are calculated from the collection of samples from the midpoint of a/all phoneme segment(s) followed by the importance function. We present the results for consonants followed by vowels. Figure 2 illustrates the articulatory importance value for each subject for a subset of consonants. These consonants and vowels are chosen such that a wide variety of critical articulators are covered. For example, high AIF value is observed for lip and Jaw in case of /m/ and /f/, similarly, tongue tip for /dh/ and /s/, tongue body for /ch/ and /sh/, and tongue dorsum for /k/ and /ng/. The average of all entries in a row (corresponds to one articulator) is mentioned in the right y-axis with standard deviation in the bracket. These average AIF values across all the subjects are observed to be consistent with the critical articulators reported in Kim et al. (2015). In general, the maximum values are obtained for the articulators which are considered to be critical. In Figure 3, vowel-specific AIF values are reported. For example, for a back vowel say /aa/ the maximum impor-

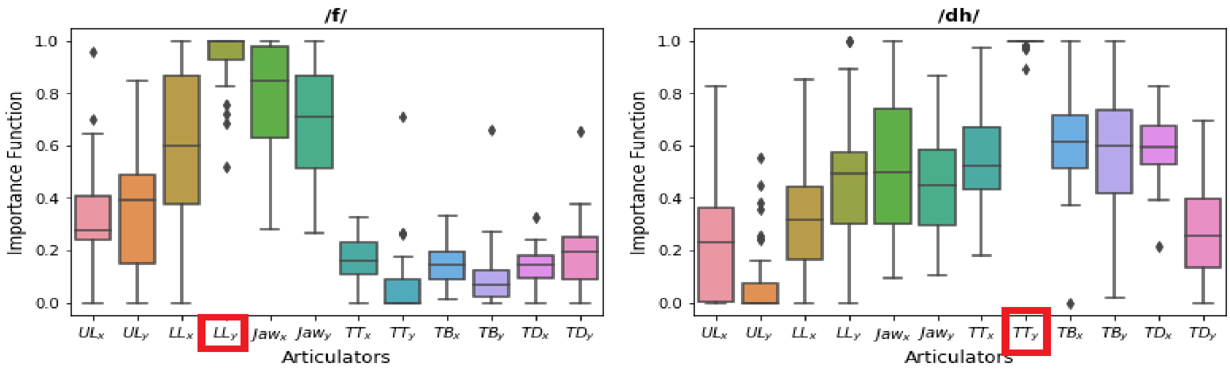

Figure 4: Articulatory Importance Function variation across subjects

tance is observed at tongue dorsum, similarly at TT_x for front vowel /iy/. For a back rounded vowel /uw/, high value of AIF is observed for LL_y and TD_y. The articulatory importance for vowels are found to be consistent with their place of articulation as highlighted in the IPA chart. Figure 4 illustrates the importance function values plotted against various articulators across all subjects. In the phoneme /m/ and /dh/, for LL_y and TT_y the importance value is almost 1 where all the values in the interquartile range are very close to 1 making the spread across subjects minimum. Thus critical articulators show very less variance across subjects.

Note that similar observations are made by calculating mean and median in place of midpoint (for computation of importance in equation 1). That is, in these observations, phoneme-specific and global variance are calculated from the collection of mean/median of one third of the samples from either side of the midpoint of a/all phoneme segment(s). The results are found to be consistent across all three different choices. Further, results without performing normalization is also explored. It is found that normalization makes articulatory importance more consistent across subjects.

5. Conclusion

In this work, we proposed a metric to assign an articulatory importance value between 0 to 1, instead of assigning a binary decision on articulators being critical or non-critical for phoneme production. Experiments are performed on 38 subjects' acoustic and articulatory data. This work gives a better understanding of the importance of various articulators in phoneme production. This analysis could benefit to the understanding of inter speaker variability in speech production mechanisms and to provide articulatory feedback in language learning tasks. We plan to investigate these directions as parts of future work.

6. Acknowledgements

Authors thank all the subjects for their participation in the data collection. We also thank the Pratiksha Trust and the Department of Science and Technology, Govt. of India for their support.

7. References

Ananthakrishnan, Gopal and Olov Engwall (2008). "Important regions in the articulator trajectory". In: *Proc. 8th ISSP*, pp. 305–308.

Bladon, R Anthony W and Ameen Al-Bamerni (1976). "Coarticulation resistance in English/l". In: *Journal of Phonetics* 4.2, pp. 137–150.

Browman, Catherine P and Louis Goldstein (1992). "Articulatory phonology: An overview". In: *Phonetica* 49.3-4, pp. 155–180.

Illa, Aravind and Prasanta Kumar Ghosh (2018). "Low Resource Acoustic-to-articulatory Inversion Using Bi-directional Long Short Term Memory". In: *Proc. Interspeech*, pp. 3122–3126.

Jackson, Philip JB and Veena D Singampalli (2009). "Statistical identification of articulation constraints in the production of speech". In: *Speech Communication* 51.8, pp. 695–710.

Kim, Jangwon, Asterios Toutios, Sungbok Lee, and Shrikanth S Narayanan (2015). "A kinematic study of critical and non-critical articulators in emotional speech production". In: *The Journal of the Acoustical Society of America* 137.3, pp. 1411–1429.

Liberman, Alvin M (1970). "The grammars of speech and language". In: *Cognitive Psychology* 1.4, pp. 301–323.

MacNeilage, Peter F (1970). "Motor control of serial ordering of speech." In: *Psychological review* 77.3, p. 182.

Mermelstein, Paul (1973). "Articulatory model for the study of speech production". In: *The Journal of the Acoustical Society of America* 53.4, pp. 1070–1082.

Ohala, John J, Catherine P Browman, and Louis M Goldstein (1986). "Towards an articulatory phonology". In: *Phonology* 3, pp. 219–252.

Povey, Daniel et al. (2011). "The Kaldi speech recognition toolkit". In: *IEEE workshop on automatic speech recognition and understanding*.

Recasens, Daniel, Maria Dolors Pallarès, and Jordi Fontdevila (1997). "A model of lingual coarticulation based on articulatory constraints". In: *The Journal of the Acoustical Society of America* 102.1, pp. 544–561.

Saltzman, Elliot L and Kevin G Munhall (1989). "A dynamical approach to gestural patterning in speech production". In: *Ecological psychology* 1.4, pp. 333–382.

Wang, Jun, Jordan R Green, and Ashok Samal (2013). "Individual articulator's contribution to phoneme production". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 7785–7789.