



Parameterization of articulatory pattern in speakers with ALS

Panying Rong¹, Yana Yunusova², James D. Berry³, Lorne Zinman⁴, Jordan R. Green¹

¹ MGH Institute of Health Professions, Boston, USA

² Department of Speech-Language Pathology, University of Toronto, Toronto, Canada

³ MDA ALS Clinic, Massachusetts General Hospital, Boston, USA

⁴ ALS/MND Clinic, Sunnybrook Health Sciences Centre,

Department of Neurology, University of Toronto, Toronto, Canada

prong@mghihp.edu, yana.yunusova@utoronto.ca, jgreen2@mghihp.edu

Abstract

A combination of parallel factor analysis (PARAFAC) and principal component analysis (PCA) was used to parameterize the articulatory pattern of tongue, jaw and lip movements in 8 English vowels produced by 7 subjects with amyotrophic lateral sclerosis (ALS). A two-factor PARAFAC model derived an overall articulatory pattern represented by two basic modes dominated by tongue raising and advancement, respectively. The relation between the two articulatory modes and the acoustic formants (F1, F2) followed a simple one-to-one linear mapping. The PCA on the residuals of the PARAFAC model showed various individualized articulatory features superimposed on the overall pattern. These articulatory features contributed in a systematic way to the acoustic deviation across different subjects. The parameterization approach (1) provided a simple and generalizable way to explore the underlying articulatory mechanism of speech decline in ALS and (2) accounted for the articulatory features across affected individuals. With further development of the approach and a comparison with the articulatory pattern for healthy subjects, it is possible to derive a set of quantitative articulatory indicators of speech impairment in ALS.

Index Terms: parameterization, articulatory-acoustic mapping, ALS

1. Introduction

A one-to-many mapping between acoustics and articulation has been demonstrated by a variety of theoretical and experimental studies on speech production [4, 8, 10]. The one-to-many mappings imply that talkers are capable of producing the same acoustic output using different articulatory configurations [5, 8]. Of course, many of the theoretically possible articulatory configurations are not biologically plausible or preferred. Moreover, the articulatory configurations for phonemes vary across different languages and across different individuals due to a variety of anatomic and linguistic factors. To reduce the high-dimensional articulatory features and identify biologically plausible articulatory configurations, prior studies have parameterized articulatory configurations into a low-dimensional set of features that can be mapped directly to speech acoustic features based on empirical data.

Story [11] decomposed a set of empirical orthogonal modes from 10 English vowel area functions measured from the structural MRI images of one subject. These orthogonal modes comprised the basis functions of vocal tract shape, which was found to correlate with the F1-F2 formant pairs through a one-to-one

mapping under the constraint of a constant vocal tract length. Harshman et al. [6] used a PARAFAC procedure to analyze the tongue shape based on the midsagittal x-ray tracings of the vocal tract in 10 English vowels by five speakers. Two factors that corresponded to a forward tongue root movement accompanied by an upward tongue front movement and an upward-backward tongue movement, respectively, were identified to account for a majority of variance in the tongue movement. Compared to other factorial analyses such as PCA, PARAFAC introduces an additional level to account for the subject effect so that the solution from PARAFAC represents a more generalizable articulatory pattern that is not limited by subject-dependent factors. The cross-linguistic generality of PARAFAC parameterization was demonstrated in Hoole [7], which showed two factors accounted for over 90% variance in 15 German vowels imbedded in different consonantal contexts based on the positions of four electromagnetic transducers on the midsagittal surface of the tongue. It was also found that the additional variability of tongue shape introduced by the consonantal context can be represented by a further articulatory component derived from individual-based PCA.

In this study, we used a combined PARAFAC-PCA approach similar to Hoole [7] to parameterize articulatory patterns of persons with amyotrophic lateral sclerosis (ALS). ALS is a neurodegenerative disease due to the degeneration of upper and low motor neurons. One of the most devastating signs of ALS is loss of speaking ability, which significantly impacts the quality of life among individuals with ALS.

To explore the physiological disturbances underlying speech decline in ALS, kinematic studies have been conducted on the articulatory system, which identified a number of articulatory parameters that represent reduced and slowed movement of the primary articulators (e.g., tongue, lips, jaw) as the major deficits [3, 13]. However, due to the heterogeneity of ALS, the factors that contribute to speech decline usually vary across different individuals [9]. As a result, a few isolated articulatory parameters might not capture the whole picture of the articulatory pattern in individuals with ALS. A more comprehensive and systematic approach is needed that accounts for not only the overall articulatory pattern in ALS but also individual articulatory variations across different individuals. Therefore, we first applied a two-factor PARAFAC model to the midsagittal positions of four primary articulators for 8 vowels across 7 subjects with ALS. Then the error of the PARAFAC model was subjected to PCA for individual subjects. This two-step approach was used (1) to derive a set of articulatory modes that represented the characteristic articulatory movement pattern of per-

sons with ALS and (2) to determine how articulatory features varied across affected individuals.

2. Method

2.1. Data collection & processing

Seven participants with ALS (6 males and 1 female) that varied in severity of speech impairment were studied. The mean speech intelligibility was 98.71% ($SD = 2.06\%$) and the mean speaking rate was 167 Words per minute (WPM) ($SD = 70$ WPM). During the recording, we used the electromagnetic articulography (EMA) AG500 system (Carstens, Germany) to track the kinematic movements of four primary articulators (tongue tip, tongue body, jaw, lower lip) at a sampling rate of 200 Hz using electromagnetic sensors. For each subject, a total of 9 sensors were attached (1) 1 cm behind the apex of the tongue in the midsagittal plane (TT), (2) 2-3 cm behind TT (TB), (3) on the left and right sides of low chin (JL, JR), (4) on the central vermilion borders of upper and lower lips (UL, LL), and (5) on the center, left and, right forehead (HC, HL, HR) (as shown in Figure 1). The three head sensors were used to subtract speech-unrelated forehead movement from the total articulatory movement. The articulatory sensors TT, TB, UL, and LL captured the corresponding movements of tongue and lips that combined the independent tongue/lip movement relative to the jaw and the contribution of jaw movement to the total tongue/lip movement. With the sensors attached and a microphone placed about 3 cm away from the lip corner, the subjects were asked to read 8 English vowels (*/a/, /i/, /e/, /æ/, /ɜ/, /ɔ/, /o/, /u/*) imbedded in the syllable */bVb/* for ten times.

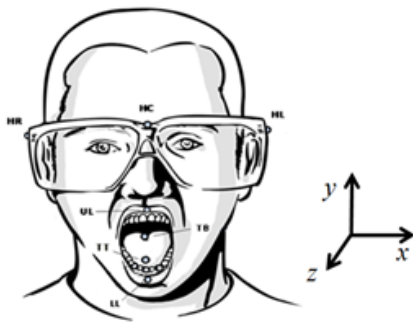


Figure 1: Graphical illustration of EMA sensor placement. HR: right forehead, HC: center of forehead, HL: left forehead, TB: tongue body, TT: tongue tip, UL: upper lip, LL: lower lip. JL and JR are not shown in this figure.

The vowels were parsed from the syllables based on the acoustic signal. The temporal midpoint of the parsed vowel was selected for both kinematic and acoustic signals. Then we averaged the kinematic data across the 10 repetitions of each vowel.

2.2. Factor analysis

First, a two-factor PARAFAC model was applied to a 3D dataset comprised of the kinematic data for 8 midsagittal coordinates of four primary articulators (TT, TB, JL, LL; each articulator has two coordinates corresponding to the vertical and horizontal movements) for 8 vowels across 7 subjects. We used the PARAFAC function in the N-way toolbox for MATLAB developed by Andersson and Bro [1]. Through this step, we derived

three sets of “modes” that corresponded to (1) the basis modes of articulatory movement (referred as “articulatory modes” in the following context), (2) the coefficients for the 8 vowels, and (3) the weights for the 7 subjects, respectively. In the second step, the error of the PARAFAC model was subjected to PCA for each subject in MATLAB (R2013b). We selected the first principal component (PC) to represent the individualized articulatory feature of the subject and accordingly, the corresponding coefficient of the PC represented the magnitude of the feature.

2.3. Formant tracking

The first two formant frequencies F1 and F2 of each vowel were automatically tracked using a custom LPC function in MATLAB. The automatically tracked formants were visually verified by the first author. Any values that were substantially deviant from the typical formant frequencies of adults were recalculated by plotting the FFT spectrum of the vowel and manually measuring the frequency of the closest harmonic to the first or second peak of the spectral envelop.

3. Results

3.1. Articulatory modes from the PARAFAC model

The PARAFAC model accounted for 93.7% of the variance in the data of 7 subjects. The average RMS error was 4.64 mm. Figure 2 shows the articulatory patterns represented by the two modes derived from the PARAFAC model. Specifically, the first articulatory mode (Figure 2, Panel b) was comprised of tongue raising and lip opening; and the second articulatory mode (Figure 2, Panel c) was comprised of tongue advancement and lip protrusion combined with some tongue back raising.

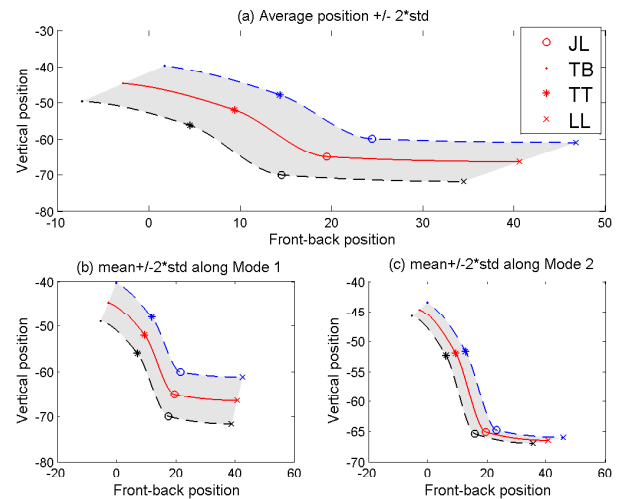


Figure 2: (a) Average positions (in mm) of four articulators and the corresponding positions when the articulators moved ± 2 standard deviations from the average position. (b) Average articulatory positions (in mm) and the articulatory positions when moved ± 2 standard deviations from the average position along 1st articulatory mode. (c) Average articulatory positions (in mm) and the articulatory positions when moved ± 2 standard deviations from the average position along 2nd articulatory mode. The solid (average) and dashed (± 2 standard deviations) curves are the shape-preserving cubic Hermite interpolations of the articulatory positions.

3.2. Acoustic-articulatory mapping

The relation between acoustic formants and articulatory configuration was illustrated in Figures 3 & 4. In Figure 3, the average acoustic vowel space across the 7 subjects is shown in parallel with the “articulatory vowel space”, which was comprised of the coefficients of the 8 vowels along the first and second articulatory modes. The correlations between the coefficients along the two articulatory modes and acoustic formants for all subjects are shown in Figure 4. By fitting a linear mixed-effects (LME) model, the correlation between the coefficient along the first articulatory mode (c1) and F1 was demonstrated to be statistically significant ($R^2 = 0.87, p < .001$); the correlation between the coefficient along the second articulatory mode (c2) and F2 was also significant ($R^2 = 0.88, p < .001$).

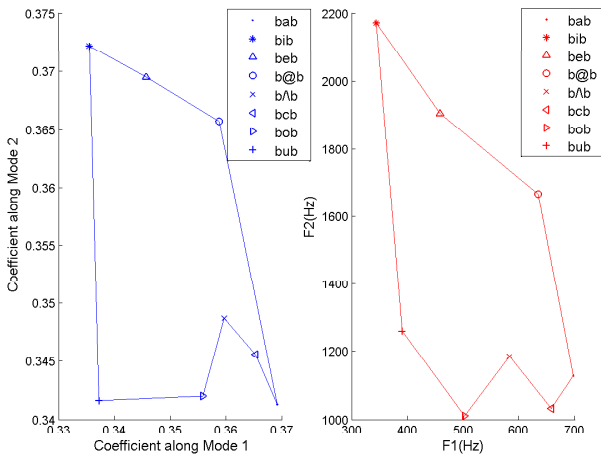


Figure 3: Articulatory (left) and acoustic (right) vowel space for 8 vowels. The articulatory vowel space is comprised of the coefficient along 1st articulatory mode on the x-axis and the coefficient along 2nd articulatory mode on the y-axis.

3.3. Individualized articulatory feature

The first principal component of PCA explained 83.58%, 91.02%, 68.85%, 90.56%, 68.66%, 57.48%, and 87.63% variance in the error of the PARAFAC model for each of the 7 subject, respectively. The articulatory feature represented by the first principal component (PC1) was found to differ across subjects in terms of both direction and magnitude. Figure 5 show the articulatory features of two subjects (Subject 3 & Subject 4) as typical examples for comparison. In addition to the tongue-jaw-lip movement shown in the overall articulatory pattern, (1) Subject 3 (left) involved additional upward movement of tongue tip relative to the jaw; and (2) Subject 4 (right) moved tongue tip backward and lower lip downward relative to the jaw.

3.4. Relation between articulatory and acoustic deviations

Acoustic deviation was calculated for both F1 and F2 by subtracting the average formant frequency across 7 subjects from the corresponding formant frequency for each individual subject. The relation between the acoustic and articulatory deviations was examined by fitting a linear mixed-effects (LME) model with the acoustic deviation as the dependent variable, the coefficient along PC1 as the fixed effect, and subject as a random effect for the intercept. The LME statistics showed a

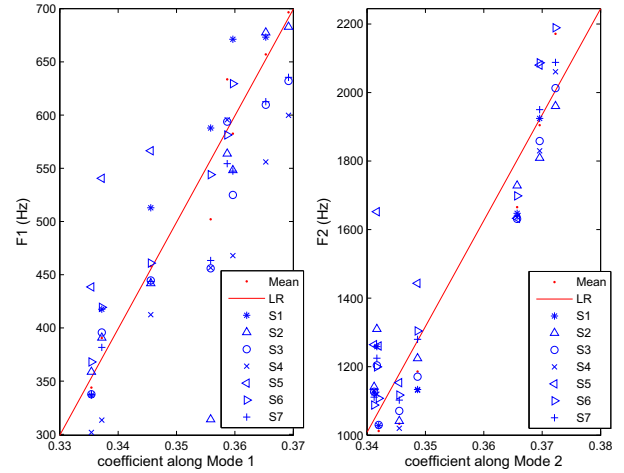


Figure 4: Correlation between F1 and the coefficient along 1st articulatory mode (left); correlation between F2 and the coefficient along 2nd articulatory mode (right) for the 7 subjects using different symbols. The dots are the average formant frequencies across the 7 subjects. The lines are the linear regression (LR) fits to both correlations.

significant correlation between F2 deviation and the coefficient along PC1 ($R^2 = 0.56, p = 0.015$); the correlation between F1 deviation and the coefficient along PC1 was marginally significant ($R^2 = 0.83, p = 0.066$). Based on the LME results, we computed the normalized acoustic deviation by subtracting the random effect of the intercept from the acoustic deviation for each subject. The relation between the normalized acoustic deviation and the coefficient along PC1 was shown in Figure 6 for all subjects, where the lines were the LME fits.

4. Discussion

4.1. Overall articulatory pattern in ALS

Based on the PARAFAC model, we derived an overall articulatory pattern in ALS represented by two articulatory modes that accounted for a majority of variance. Compared to the findings of Harshman et al. [6] and Hoole [7] that both showed combined raising and forward movements across different parts of the tongue in healthy speakers, our results separated the raising and forward movements of the tongue in two modes with the first mode dominated primarily by tongue raising and the second mode dominated mostly by tongue advancement. The difference of tongue movement pattern captured by the current versus prior studies may be attributed to different measurement techniques used in these studies. Both Harshman and Hoole’s studies measured the tongue shape with finer grids (9 grids based on x-ray image in [6]; 4 points based on EMMA in [7]) than did this study (2 points based on EMA), so they captured more detailed movements of different parts of the tongue. Placing more sensors on the tongue might provide additional information on the coordination between different parts of the tongue beyond the current finding, but it is also practically more challenging because ALS is characterized by muscle atrophy and fasciculations, making it challenging during sensor placement. Moreover, individuals with ALS might have different articulatory patterns compared to healthy speakers. Specifically, the movement directions of TT, TB, and JL in the two articu-

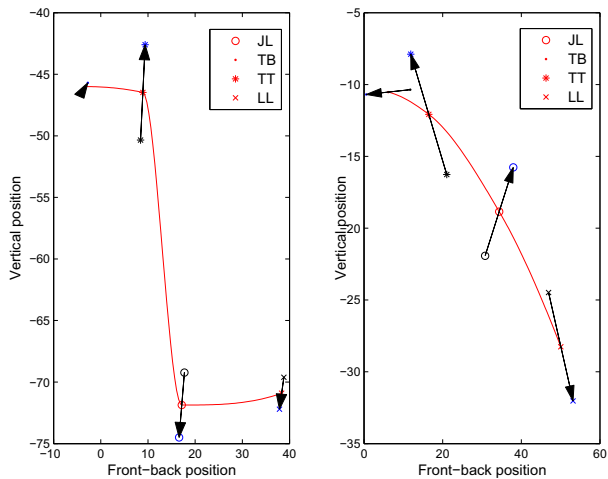


Figure 5: Two examples of individualized articulatory features (Subject 3 on the left, Subject 4 on the right). The curves are the average articulatory deviation (in mm) across 8 vowels superimposed on the neutral articulatory positions from the PARAFAC model. The arrows mark out the articulatory deviation comprised of ± 2 standard deviations from the average position along the direction specified by PC1.

latory modes showed synergistic tongue-jaw motion, suggesting that tongue movement was coupled with jaw movement. Increased tongue-jaw coupling in individuals with ALS could be attributed to two factors. First, compared to the deformable tongue, jaw is a rigid structure with less degree of freedom. Second, jaw function was demonstrated to be less affected by ALS than tongue function [2, 3, 13]. Therefore, it is likely that individuals with ALS rely more on jaw movement to achieve articulatory targets, resulting in synergistic movements between jaw and different parts of the tongue in Figure 2 (Note that the tongue and jaw movements in this study were not decoupled).

The articulatory vowel space comprised of c1 and c2 aligned well with the acoustic vowel space of F1 and F2 (Figure 3) for all vowels except /o/ and /ɔ/, which are naturally more variable than other vowels in English [12]. The alignment of vowel space suggested the possibility of a direct mapping between articulatory configuration and acoustic formants. According to the correlation between F1 and c1 and the correlation between F2 and c2 in Figure 4, we demonstrated that the relation between the articulatory pattern and acoustic features in ALS can be simplified as a one-to-one linear mapping between the two articulatory modes derived from the PARAFAC model and the first two acoustic formants. The direct articulatory-acoustic mapping provides a theoretical basis for real-time articulatory synthesis of speech in ALS, which relies on only two coefficients rather than the computationally intensive inverse mapping to determine the articulatory-acoustic relation. This finding might have potential practical applications such as feedback-based learning and behavioral treatment of ALS.

4.2. Individualized articulatory feature

PCA on the error of the PARAFAC model suggested articulatory features varied across different individuals, which were superimposed on the overall articulatory pattern to account for individual variability in ALS. For the two examples in Figure 5, Subject 3 (left) showed upward movement of the tongue

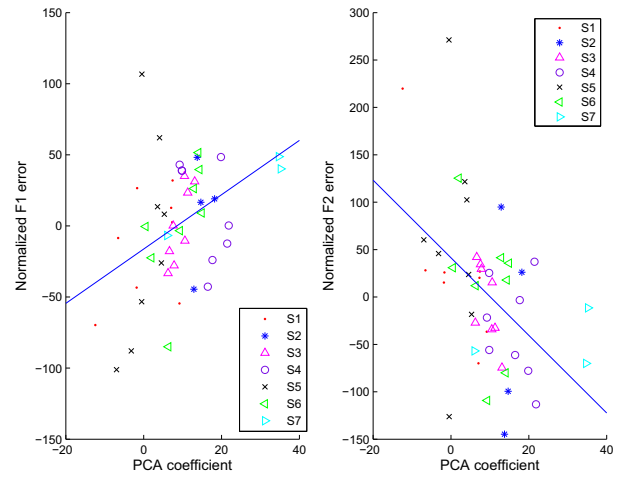


Figure 6: Correlation between the normalized F1 deviation and the coefficient along PC1 (left) and correlation between the normalized F2 deviation and the coefficient along PC1 (right) for all subjects using different symbols. The lines are the linear fits to the correlations.

independent of the jaw; and Subject 4 (right) showed backward tongue movement and downward lip movement independent of the jaw. Compared to the overall pattern in Figure 2, these subjects showed reduced inter-articulatory coupling, which could lead to at least three possibilities: (1) random errors of the PARAFAC model, (2) voluntary tongue movement for fine-tuning of speech production, and (3) discoordination over the vocal tract musculature.

Based on the correlations between the principal component coefficient and the acoustic formant deviations (Figure 6), the articulatory features observed in different individuals were not just random errors of the PARAFAC model but they contributed in a systematic way to the acoustic deviation in speech. Therefore, the pattern of individual articulatory feature derived by PCA might serve as a potential articulatory indicator of speech impairment in persons with ALS. The follow-up question addressing the contribution of individualized articulatory features to speech impairment requires additional data on healthy control subjects and individuals with a wider range of speech impairment severity due to ALS.

5. Conclusion

Through a PARAFAC-PCA combined approach, we parameterized the articulation of vowels in ALS with an overall articulatory pattern comprised of two basic modes and a set of individualized articulatory features superimposed on the overall pattern. The parametrization provided a simple and generalizable way to explore the underlying articulatory mechanism of speech decline. In future studies, we need to collect data from healthy control subjects and compare their articulatory pattern with the pattern found for individuals with ALS.

6. Acknowledgements

The authors would like to acknowledge the funding support from NIH grants R01 DC009890 & R01 DC0135470, and ALS Society of Canada Bernice Ramsey Discovery Grant.

7. References

- [1] Andersson, C.A. and Bro, R., "The N-way Toolbox for MATLAB", *Chemometrics and Intelligent Laboratory Systems*, 52(2000):1-4, 2000.
- [2] DePaul, R. and Abbs, J.H., "Manifestations of ALS in the cranial motor nerves: Dynametric, neuropathologic, and speech motor data", *Neurologic Clinics*, 5(2):231-250, 1987.
- [3] DePaul, R. and Brooks, B.R., "Multiple orofacial indices in amyotrophic lateral sclerosis", *Journal of Speech and Hearing Research*, 36:1158-1167, 1993.
- [4] Fant, G., "The acoustic theory of speech production", Mouton and Co., The Hague, 1960.
- [5] Gay, T., Lindblom, B. and Lubker, J., "Production of bite-block vowels: Acoustic equivalence by selective compensation", *J. Acoust. Soc. Am.*, 69(3):802-810, 1981.
- [6] Harshman, R., Ladefoged, P. and Goldstein, L., "Factor analysis of tongue shapes", *J. Acoust. Soc. Am.*, 62(3):693-707, 1977.
- [7] Hoole, P., "On the lingual organization of the German vowel system", *J. Acoust. Soc. Am.*, 106(2):1020-1032, 1999.
- [8] Hughes, O. and Abbs, J., "Labial-mandibular coordination in the production of speech: implications for the operation of motor equivalence", *Phonetica*, 33(3):199-221, 1976.
- [9] Kent, R.D., Sufit, R.L., Rosenbek, J.C., Kent, J.F., Weismer, G., Martin, R.E. and Brooks, B.R., "Speech deterioration in amyotrophic lateral sclerosis: A case study", *Journal of Speech and Hearing Research*, 34:1269-1275, 1991.
- [10] Maeda, S., "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model", in W. J. Hardcastle and A. Marchal [Ed], *Speech Production and Speech Modelling*, 131-149, Kluwer Academic, Dordrecht, 1990.
- [11] Story, B. and Titze, I., "Parameterization of vocal tract area functions by empirical orthogonal modes", *Journal of Phonetics*, 26:223-260, 1998.
- [12] Wells, J.C., "Accents of English", Cambridge, 1982.
- [13] Yunusova, Y., Green, J.R., Lindstrom, M.J., Ball, L.J., Pattee, G.L. and Zinman, L., "Kinematics of disease progression in bulbar ALS", *Journal of Communication Disorders*, 43(1):6-20, 2009.