

PAoS Markers: Trajectory Analysis of Selective Phonological Posteriors for Assessment of Progressive Apraxia of Speech

Afsaneh Asaei¹, Milos Cernak¹, Marina Laganaro²

¹Idiap Research Institute, Martigny, Switzerland

²University of Geneva, Switzerland

{aasaei,mcernak}@idiap.ch, marina.laganaro@unige.ch

Abstract

Progressive apraxia of Speech (PAoS) is a progressive motor speech disorder associated with neurodegenerative disease causing impairment of phonetic encoding and motor speech planning. Clinical observation and acoustic studies show that duration analysis provides reliable cues for diagnosis of the disease progression and severity of articulatory disruption. The goal of this paper is to develop computational methods for objective evaluation of duration and trajectory of speech articulation. We use phonological posteriors as speech features. Phonological posteriors consist of probabilities of phonological classes estimated for every short segment of the speech signal.

PAoS encompasses lengthening of duration which is more pronounced in vowels [1, 2]; we thus hypothesize that a small subset of phonological classes provide stronger evidence for duration and trajectory analysis. These classes are determined through analysis of linear prediction coefficients (LPC). To enable trajectory analysis without phonetic alignment, we exploit phonological structures defined through quantization of phonological posteriors. Duration and trajectory analysis are conducted on blocks of multiple consecutive segments possessing similar phonological structures. Moreover, unique phonological structures are identified for every severity condition.

Index Terms: Progressive apraxia of speech (PAoS), Phonological posterior features, Phonological structures, Linear prediction coefficient (LPC).

1. Introduction

Dysarthria and Progressive Apraxia of Speech (PAoS) are two common speech motor disorders observed in neurodegenerative diseases. While automatic processing (for assessment and assistive applications) of dysarthric speech is getting considerable attention in the speech community [3, 4, 5, 6], acoustic and automatic processing studies of PAoS are rather rare. It might be due to increased complexity of speech degradations of patients with PAoS, where production errors are more inconsistent and unpredictable [7].

PAoS is a speech motor disorder associated to several neuropathological conditions, which causes progressive degradation of the main speech characteristic and of speech intelligibility. The main symptoms of PAoS are phonetic distortions and phonemic errors, groping and effortful speech initiation with successive approximations, changes in inter- and intra-syllabic transitions, increased syllabic duration and decreased speech rate [6, 7].

PAoS has been associated with impaired phonetic encoding (planning of speech gestures) rather than to impaired motor execution [7, 2, 8]. Here we hypothesize that analysis of phono-

logical features extracted from the degraded speech signal could contain clues for assessment of the progressive disruption and severity levels of PAoS.

One particular contribution of this paper is selection of phonological classes using linear prediction analysis of phonological posteriors. In addition, we exploit phonological structures [9] to enable automatic analysis of duration and trajectory without any need for automatic alignment. Prior work on phonological structures demonstrate their relation to articulatory postures [9], thus considering the structure of multiple consecutive segments enables quantification of the dynamic and trajectory of articulatory movements and co-articulation. The studies presented in this paper exploit this structural property of phonological posteriors to obtain speech-based markers of PAoS severity.

The results obtained in Section 5 demonstrate a significant increase in duration and less consistency in articulatory movements as the neuro-degeneration thrives. Furthermore, we identify unique structures per severity condition which indicates that certain articulatory postures disappear in AoS progression and are replaced by new postures and trajectory of movements. This observation can lead to development of a novel automatic assessment method relying on the nearest neighbor rule of classification. Preliminary studies show the potential of this method. However, recordings of many patients is required to validate/endorse its usefulness for clinical applications.

To the best of authors' knowledge, prior work on application of phonological features for objective intelligibility prediction of pathological speech considered statistical measures of independent processing of segments of speech [10]. In contrast, we propose ranking and selection of phonological classes, and we study the relation between adjacent segments, or trajectory of articulation. The proposed approach provides simple objective tools that correlate with higher level speech production behaviors such as speaking rate and co-articulation without any requirement for speech alignment.

The rest of the paper is organized as follows. Section 2 describes the data available from 3 assessment sessions of a patient diagnosed with isolated PAoS. This data is used for estimation of phonological posteriors through the procedure explained in Section 3. We used linear prediction analysis to rank the phonological classes for trajectory analysis in Section 4. The trajectory analysis methods are explained in Section 5 where selection of phonological classes is found an effective approach to obtain more distinct markers of severity. Moreover, distinct structural patterns are observed for every severity condition. This observation leads to devising a classifier for detection of the level of severity which is elaborated in Section 6. The conclusions are drawn in Section 7.

2. PAoS Data

The data used for evaluation of the methods proposed in this paper consist in 3 recordings over a 28-months period of a 67 year-old french speaking woman diagnosed with isolated PAoS. The patient has been recorded for about 2 minutes while reading the same text (“La bise et le soleil” [11]). The total duration is thus about 7 minutes. Across the 3 sessions the severity of speech disruption progresses from mild, to medium and to severe impairment according to clinical assessment by speech and language therapists and to normative acoustic data. Diadochokinetic rate assessed with standard diadochokinetic tasks [12] and articulation rate are reported in Table 1.

Table 1: *Clinical PAoS pattern: speech rate and diadochokinetic rate (syll/sec) across the assessment sessions. The numbers in parenthesis shows the relative reduction in rates with respect to the mild condition. The patient’s production impairment in medium condition is assessed after 16 month from the mild condition; the severe impairment is evaluated after 12 month from the medium condition.*

Condition	Mild	Medium	Severe
Articulation rate	2.73	2.39 (13%)	2.06 (25%)
Diadochokinesis rate	2.85	2.22 (22%)	1.58 (45%)

The clinical and acoustic durational measurements show that this patient after 16 months from its initial mild AoS, exhibits increased impairment where the articulation rate is decreased by 13% and diadochokinetic rate is decreased by 22%. In the follow up assessment session after 28 months from the diagnosis of mild AoS, the patient reaches more severe impairment manifested in 25% reduction in articulation rate and 45% reduction in diadochokinetic rate.

In the rest of the paper, our goal is to quantify speech markers that correlate with the clinical markers. Motivated from the intuitive effects of PAoS on articulatory disruptions, and how the clinical assessments quantify this impairment, we focus our work on ranking and selection of speech representations, and their evolution through time in trajectory analysis.

3. Phonological Structures

We use deep neural network (DNNs) to estimate the phonological posterior features. As we have already seen in Section 1, PAoS affects phonetic planning. Hence, phonological posteriors are suitable representation of speech to enable assessment of these patients. Moreover, phonological posteriors exhibit highly constrained structures that are consistent for adjacent segments and change according to the speaking rate. In the next Section 3.1, we explain the framework for estimation of phonological posteriors.

3.1. Phonological Posteriors

Figure 1 illustrates the process of the phonological analysis [13, 14]. This process starts by converting a segment of speech samples into a sequence of acoustic features $X = \{x_1, \dots, x_n, \dots, x_N\}$ where N denotes the number of segments in the utterance. Conventional cepstral coefficients can be used as acoustic features. Then, a bank of phonological class analyzers realized via neural network classifiers converts the acoustic feature observation sequence X into a sequence of phonological posterior probabilities $Z =$

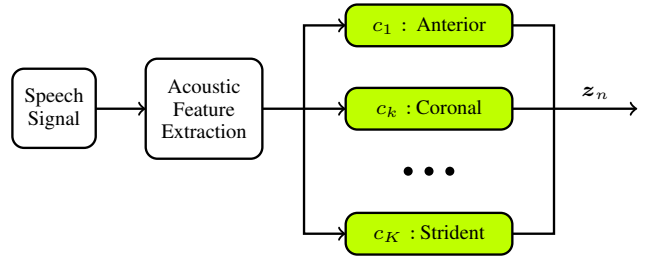


Figure 1: *The process of phonological analysis. Each segment of speech signal is represented by phonological posterior probabilities z_n that consist of K class-conditional posterior probabilities. For each phonological class, a DNN is trained to estimate its posterior probability given the input acoustic features.*

$\{z_1, \dots, z_n, \dots, z_N\}$; a posterior probability

$$z_n = [p(c_1|x_n), \dots, p(c_k|x_n), \dots, p(c_K|x_n)]^T \quad (1)$$

consists of K phonological class-conditional posterior probabilities where c_k denotes the phonological class and T stands for the transpose operator. The phonological posteriors Z yield a parametric speech representation.

Phonological analysis was performed with the PhonVoc: phonetic and phonological toolkit [16]. Probabilities of $K = 24$ phonological classes corresponding to the French version of the Sound Pattern of English [14] were extracted from 25 ms speech segments, using 10 ms steps [16].

3.2. Structured Sparsity

Phonological posteriors are indicators of the physiological posture of human articulation machinery. Due to the physical constraints, only few combinations can be realized in our vocalization. This physical limitation leads to a small number of unique patterns exhibited over the entire speech corpora [17]. We refer to this structure as *first-order structure* which is exhibited at *segmental* level. These structures can be quantified using binary (1-bit) quantization or finer quantization levels. We will compare both binary (Q1) and 2-bit (Q2) quantization levels to perform trajectory analysis in Section 5. We will see that binary structures are the best level of structural definition.

In addition to the first-order structures, the dynamic of the phonological posteriors can be quantified considering the higher-order structure underlying a sequence (trajectory) of phonological posteriors. This structure is exhibited at *supra-segmental* level which is associated to the syllabic information or more abstract linguistic attributes. We refer to this structure as *high-order structure*.

Previously we have shown that the trajectories of the articulatory-bound phonological posteriors correspond to the distal representation of the gestures in the gestural model of speech production (and perception) [9]. In this paper, we exploit these structures as markers for objective evaluation and assessment of the level of severity of speech motor disorder in patients diagnosed with progressive apraxia of speech. The details of our analysis are explained in Sections 5.

Unlike previous work on application of phonological posteriors for assessment of pathological speech, we hypothesize that not all phonological classes are equally important. In other words, a small subset of phonological classes may provide stronger cues for diagnosis of the level of PAoS. This hypothesis is supported by clinical investigations confined to the

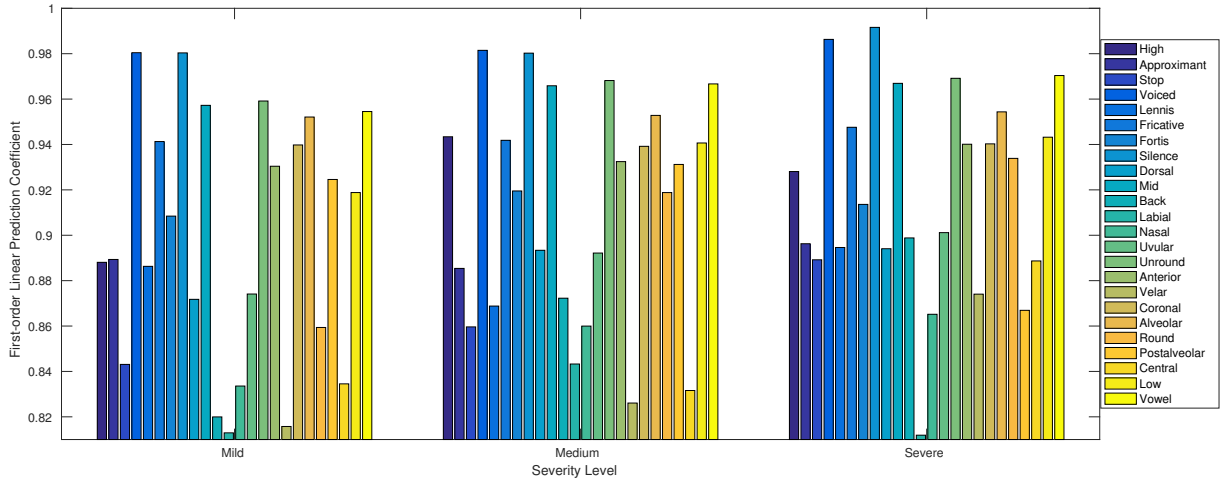


Figure 2: *Ranking and selection of important phonological classes: The first coefficient of the linear prediction analysis [15] is computed for the blocks of posteriors consisted on adjacent segments with similar binary structures. The average of the first LPC for each phonological classes is calculated per recording session. Finally, the mean of class-specific LPCs are computed for all the session and sorted. The classes with the highest mean values are considered as those contributing the most to the assessment of speech rate and trajectory in PAoS. The top-4 classes are “Voiced”, “Vowel”, “Unround” and “Mid”. Considering the phonological-phonetic mapping used for the neural network training [14], these classes capture phonological variation of vowel-like sounds (c.f. Table 2).*

vowels and in particular lengthening of the vowels as the PAoS thrives [1, 2]. In the next section, we elaborate on a method for ranking and selection of phonological classes using linear prediction analysis [15].

4. Selection of Phonological Classes

We hypothesize that phonological classes are not equally important for assessment of PAoS. The focus of the present work is on trajectory analysis of phonological posteriors; hence, we rely on linear prediction coefficients (LPC) to measure the dependency and predictability of consecutive phonological posteriors.

4.1. Linear Prediction Analysis

The goal of linear prediction analysis is to minimize prediction error of the current segment using the values of the posteriors from the past consecutive segments. The predicted posterior at segment n is thus obtained as

$$\hat{z}_n = \sum_{p=1}^P \alpha_p \odot z_{n-p} \quad (2)$$

where $\alpha_p = [\alpha_p^1 \dots \alpha_p^k]^\top$ is a K dimensional vector, and \odot stands for element-wise product. The LPCs are estimated to minimize the reconstruction error in mean square sense, i.e. $\|\hat{z}_n - z_n\|_2$.

The procedure for LPC analysis of phonological posteriors and selection of the most important classes is as follows:

1. *Blocking*: The posteriors are analyzed in blocks of consecutive segments which possess similar structures after quantization.
2. *Class-specific LPC*: The high-order LPC analysis is performed where the LPC order is chosen as the block length, i.e. number of segments.

3. *Ranking*: Means of the first LPC coefficient α_p^1 for all blocks and recording conditions are computed for every phonological class. The means are then sorted and the classes which exhibit largest means are considered as the most informative classes for trajectory analysis.

Figure 2 illustrates the average value of the first LPCs for each phonological posterior per recording sessions corresponding to mild, medium and severe condition. The top-4 most important classes are identified as “Voiced”, “Vowel”, “Unround” and “Mid”. We can see the consistent high value of LPCs estimated for these classes throughout progression of AoS.

4.2. Phonological-Phonetic Importance of Vowels

To have a better understanding of what the selection of phonological classes may imply, we refer to the phonological-phonetic mapping used for neural network training in posterior estimation [14]. Table 2 shows the mapping between selected classes and phonemes associated with these classes.

Table 2: *Association of selected French phonological classes and phonemes. The French phoneme set is taken from BDLex [18].*

Class c_k	Phonemes
Voiced	a ā ã ï y u e ē ø o ô ɔ ə ε œ ã̃ j l m ɥ w b ɲ ɲ̃ n v g ʁ d z ʒ
Vowel	ï y u e ē ø o ô ə ε œ ɔ a ā ã̃ ɔ̃
Unround	a ā i e ē ε ɔ
Mid	ø ē e œ ε

We can see that the selected top-4 classes capture phonological variability of all vowel-like speech sounds (all vowels and voiced consonants). This might indicate that vowel analysis is more important in PAoS than the consonant analysis. This observation is inline with the clinical assessment of PAoS [8].

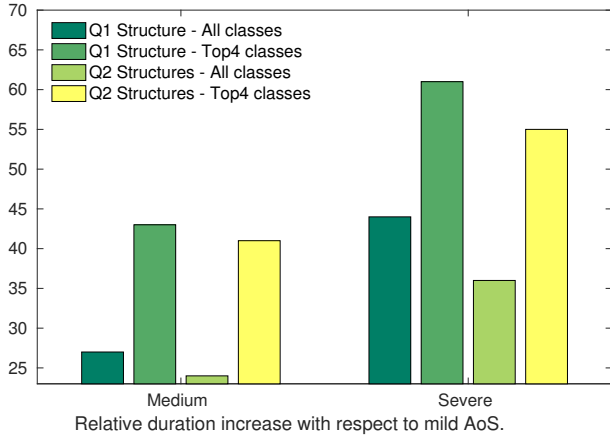


Figure 3: Relative increase in duration illustrated for (i) $Q1$ and $Q2$ quantizations to obtain phonological structures, and (ii) when phonological structures are obtained from all (24) classes vs. the case that only top-4 classes are used for duration modeling. We see clear benefit when $Q1$ binary structures are used for duration analysis, and when we use the selective top-4 classes. In the most distinctive case, using $Q1$ structure of top-4 classes, we observe 43% increase in duration in medium AoS and 61% increase in duration in severe AoS with respect to mild AoS (c.f. Table 3).

5. Trajectory Analysis

Analysis of trajectory of phonological posteriors is performed using three metrics defined through (5.1) Duration of phonological structures, (5.2) Predictability of phonological classes from the previous segments, and (5.3) Dynamic of posteriors quantified through high-order structures underlying consecutive segments.

Clinical assessment for diagnosis of PAoS asserts that speaking rate is reduced and the control over muscle movements is less consistent as the disease thrives [2, 8]. Accordingly, we expect to see an increase in duration and less predictability from mild to severe condition. Complying to the clinical emphasis put on vowel analysis, focusing our analysis on top-4 classes (c.f. Table 2) is expected to be advantageous in distinction of the severity conditions.

5.1. Structural Duration

Typically, duration analysis requires automatic alignment of speech with the actual transcription using automatic speech recognition (ASR). However, this can be a cumbersome method that requires ASR resources and expertise. Moreover, automatic alignment is affected by the ASR errors due to *progressive* mismatch between the training and testing conditions.

To alleviate this limitation, we propose to use the phonological structures for duration analysis. The structures are obtained through quantization of posteriors, and they are often similar for adjacent segments. As the phonological structures can be related to the articulatory postures of speech production [9], slower speaking rate indicates a slower dynamic in the structural changes.

Applying the same blocking procedure as explained in Section 4, we quantify the *structural duration* as the average number of the segments in one block.

Different level of quantization can be applied to obtain the

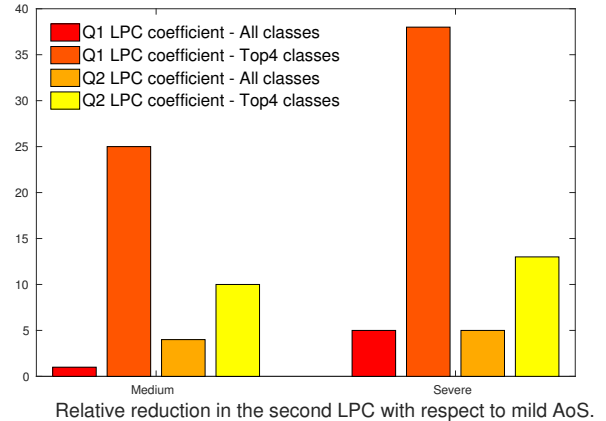


Figure 4: Relative reduction in the second coefficient of high-order LPC analysis for (i) $Q1$ and $Q2$ quantizations to obtain phonological structures, and (ii) when phonological structures are obtained from all 24 classes vs. the case that only top-4 classes are used for duration modeling. In the most distinctive case, using $Q1$ structure of top-4 classes, we observe 25% reduction in the second LPC in medium AoS and 38% reduction in the second LPC in severe AoS with respect to mild AoS (c.f. Table 4).

structures. We compare 1-bit ($Q1$) and 2-bit ($Q2$) quantization in our studies. Furthermore, to see the effectiveness of working on a small subset of phonological classes, we compare the results when the structures are obtained from all classes or only from the top-4 most important classes.

The results are illustrated in Figure 3. More details are listed in Table 3. We can see a clear benefit of binary structures

Table 3: Structural duration measured in terms of the average number of segments in all blocks of similar structures. The numbers in parenthesis show the relative increase in duration with respect to the mild condition.

Condition	Mild	Medium	Severe
Q1 duration (all)	2.3	3.1 (27%)	3.6 (44%)
Q1 duration (top-4)	4.7	6.7 (43%)	7.6 (61%)
Q2 duration (all)	1.5	1.97 (24%)	2.2 (36%)
Q2 duration (top-4)	2.3	3.3 (41%)	3.6 (55%)

over the 2-bit quantization. Moreover, obtaining the structural duration using a subset of most indicative phonological classes leads to higher distinction across different PAoS conditions.

5.2. Long Term Dependency

Similar to the method explained in Section 4, the linear prediction analysis is conducted on blocks of the same phonological structures. We perform high-order LPC analysis where the order is determined from the length of the block, i.e. number of segments. We measure the mean of the second LPC for every recording session. Less control over the muscle movement leads to less consistency of the articulation trajectories. Figure 4 illustrates the relative reduction of the second LPC with respect to the mild condition. The details of the results are listed in Table 4.

It is evident that high-order dependencies are reduced. This

Table 4: *High-order linear prediction analysis: The values for the second coefficient averaged for all segments are listed at different recording sessions. The numbers in parenthesis show the relative decrease in high-order dependency with respect to the mild condition.*

Condition	Mild	Medium	Severe
Q1 LPC (all)	0.53	0.51 (1%)	0.48 (5%)
Q1 LPC (top-4)	0.22	0.17 (25%)	0.14 (38%)
Q2 LPC (all)	0.78	0.74 (4%)	0.75 (5%)
Q2 LPC (top-4)	0.61	0.54 (10%)	0.52 (13%)

effect is much more pronounced if we only consider top-4 most important phonological classes (c.f. Table 2). Similar studies on other LPCs larger than the second coefficient shows that those values are very small and their changes are less distinctive for PAoS objective evaluation.

5.3. High-order Structures

Relying on the relation between phonological structures and articulatory postures, the dynamic/trajectory of articulation or co-articulation can be quantified considering high-order structures [9]. To that end, we append consecutive phonological posteriors to define the trajectories through quantization of augmented posteriors. The number of consecutive posteriors determine the level (order) of trajectory structures. More specifically, C adjacent posterior vectors are appended to define a new posterior which encode C -order dynamic of features as

$$\mathbf{z}_n^C = \left[\mathbf{z}_n^\top \dots \mathbf{z}_{n+C}^\top \right]^\top. \quad (3)$$

As the phonetic planning is disrupted in PAoS [8], we expect to see unique structures for distinct levels of severity. More intuitively, certain articulatory postures can only occur at an specific level of neurodegeneration.

The percentage of unique structures (number of unique structures / number of all segments) is listed in Table 5. Furthermore, we quantify the percentage of structures that only occur in *one* severity condition.

Table 5: *Ratio of unique structures (%) per condition. The numbers in parenthesis show the ratio of the structures that only occur in one particular condition, thus indicative of particular articulatory posture that may only occur at a specific level of impairment.*

Condition	Mild	Medium	Severe
Q1 structures 1-order	7.1 (32)	5.9 (29)	4 (36)
Q1 structures 2-order	28 (60)	21 (59)	17 (63)
Q1 Structures 3-order	44 (76)	34 (74)	28 (74)
Q2 structures 1-order	62 (90)	47 (88)	42 (90)

We can see that the number of distinct structures grow rapidly as the order is increased. This demonstrates that the trajectories of phonological posteriors exhibit more distinct properties as we consider larger context (C) or finer structures (Q2).

Nevertheless, even at a segment level (no augmentation, $C = 1$), we can see a significant number of structures that only occur in one specific severity condition: nearly 30% of structures are unique. This observation on structural differences motivates us to perform automatic assessment using nearest neighbor

rule of classification. The procedure is explained in the following Section 6.

6. Preliminary Automatic Assessment

To visualize the structural differences between phonological posteriors across recording sessions, we used the t -distributed stochastic neighbor embedding (tSNE) method [19] for visualization of high-dimensional (posterior) features. Figure 5.2 illustrates the results. Phonological posteriors without augmentation ($C = 1$ in (3)) are used for visualization. We contrast the visualization of posteriors where all 24 classes are used vs. only top-4 selective classes are considered. We can see that the distinction in posterior distribution is well preserved.

Exploiting the structural differences enables us to perform automatic assessment via classification. We consider the nearest neighbor classification rule for this purpose. To that end, we divide the data in two training and testing splits. Each testing segment is independently labeled based on the label of its nearest neighbor phonological posterior in the training set.

We use segmental posteriors without augmentation, i.e. $C = 1$. The *cosine* similarity metric is used to find the nearest neighboring vector which is defined as one minus cosine of the angle between two vectors; mathematically, it is expressed as

$$S_{\text{cosine}}(\mathbf{z}_1, \mathbf{z}_2) = 1 - \frac{\sum_k z_1^k z_2^k}{\sqrt{\sum_k (z_1^k)^2 \sum_k (z_2^k)^2}}. \quad (4)$$

where z_1^k denotes the k^{th} element of posterior vector \mathbf{z}_1 . This metric has been found a suitable choice when comparing the similarity of two posteriors vectors [20].

The segment-level labels are then pulled to make a decision about the severity of articulatory disruption based on majority voting. We observe that exploiting about 5 seconds for training and testing data, enables us to perfectly classify the session of recording. The training size in all severity conditions is equal.

This suggests that inter-patient PAoS severity might be automatically assessed using nearest neighbor classifier and the phonological posteriors as speech features. Of course, this can not be validated or endorsed clinically unless a sufficiently large number of patients are recorded and used for exhaustive evaluation of this method.

7. Conclusions

Trajectory analysis of phonological posteriors enables objective assessment of progressive apraxia of speech. We demonstrated that a selected set of phonological classes can be considered as strong indicators of PAoS. In this paper, we performed linear prediction analysis to select the most important classes for trajectory analysis. Interestingly, these classes highly correlate with the clinical observation on importance of vowels in PAoS diagnosis.

To enable trajectory analysis without any need for automatic alignment of speech, we build on our previous work on phonological structures obtained through quantization as a method for quantifying the articulatory postures [9]. Our investigations on structural duration shows a significant increase in duration if we consider mainly the top-4 important classes. This observation is inline with the clinical evidence of speech rate reduction more pronounced in production and lengthening of vowels. This has been also verified in scientific studies using acoustic analyses.

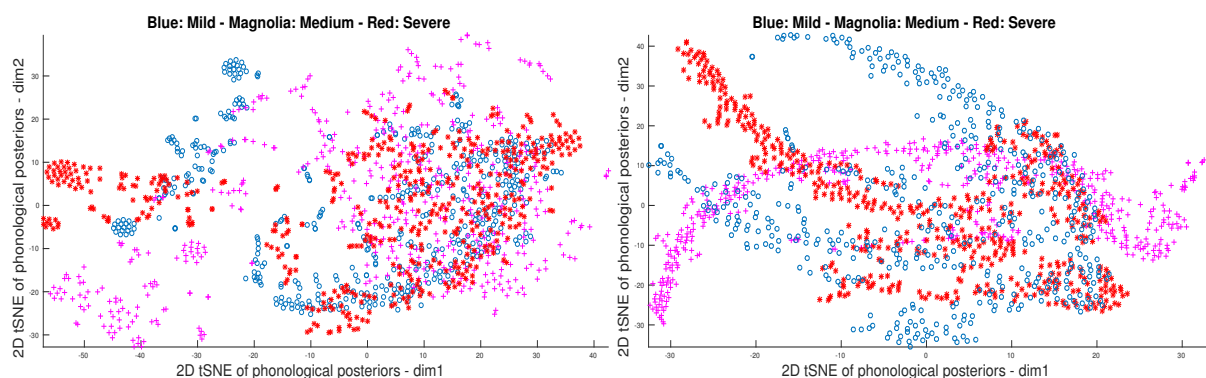


Figure 5: *tSNE* visualization of (left) all phonological posteriors and (right) selected top-4 classes (c.f. Table 2).

Furthermore, high-order LPC analysis demonstrate a significant decrease in consistency of phonological trajectories. The dynamic of phonological posteriors can be quantified by appending multiple adjacent phonological posteriors to form a super-vector of multiple phonetic classes. This method enables us to quantify the transitions or co-articulation among articulatory postures. The studies presented in this paper confirmed that unique phonological structures are exhibited for every severity condition. Exploiting this property can potentially enable us to perform automatic assessment of the level of severity in PAoS. Preliminary studies motivate us to explore this direction further.

8. Acknowledgments

The research leading to these results has received funding from by Swiss NSF project on “Parsimonious Hierarchical Automatic Speech Recognition (PHASER)” grant agreement number 200021-153507. We also thank the support of Swiss NSF projects on “Adaptive Multilingual Speech Processing (A-MUSE)” grant agreement number 200020-144281.

9. References

- [1] K. J. Ballard, S. Savage, C. E. Leyton, A. P. Vogel, M. Hornberger, and J. R. Hodges, “Logopenic and nonfluent variants of primary progressive aphasia are differentiated by acoustic measures of speech production.” *PLoS one*, vol. 9, no. 2, 2014. [Online]. Available: <http://view.ncbi.nlm.nih.gov/pubmed/24587083>
- [2] M. Laganaro, M. Croisier, O. Bagou, and F. Assal, “Progressive apraxia of speech as a window into the study of speech planning processes.” *cortex*, vol. 48, no. 8, pp. 963–971, 2012.
- [3] A. Rosenberg, “Introducing Objective Acoustic Metrics for the Frenchay Dysarthria Assessment Procedure,” Ph.D. dissertation, University of Sheffield, Sheffield, UK, 2007.
- [4] D. Martínez, P. Green, and H. Christensen, “Dysarthria Intelligibility Assessment in a Factor Analysis Total Variability Space,” in *Proc. of Interspeech*, 2013, pp. 2133–2137.
- [5] K. H. Wong, Y. T. Yeung, P. C. M. Wong, G. Levow, and H. Meng, “Analysis of Dysarthric Speech using Distinctive Feature Recognition,” in *Proc. of 6th Workshop on Speech and Language Processing for Assistive Technologies*, 2015.
- [6] L. Baghai-Ravary and S. W. Beet, *Automatic Speech Signal Analysis For Clinical Diagnosis And Assessment of Speech Disorders*. Springer, 2013.
- [7] M. R. McNeil, S. R. Pratt, and T. R. D. Fossett, “Speech motor control in normal and disordered speech, B. Maassen, R. D. Kent, H. Peters, P. H. H. M. Van Lieshout and W. Hulstijn (Eds.),” *Oxford University Press*, pp. 389–414, 2004.
- [8] M. Laganaro, “Patterns of impairments in aos and mechanisms of interaction between phonological and phonetic encoding,” *Journal of Speech, Language, and Hearing Research*, vol. 55, no. 5, pp. S1535–S1543, 2012.
- [9] M. Cernak, A. Asaei, and H. Bourlard, “On Structured Sparsity of Phonological Posteriors for Linguistic Parsing,” *Speech Communication (to appear)*, 2016.
- [10] C. Middag, Y. Saeys, and J.-P. Martens, “Towards an asr-free objective analysis of pathological speech,” in *INTERSPEECH*. International Speech Communication Association (ISCA), 2010, pp. 294–297.
- [11] Fougeron and S.-A. Jun, “Rate effects on French intonation: prosodic organization and phonetic realization,” *Journal of Phonetics*, vol. 26, no. 1, pp. 45–69, Jan. 1998. [Online]. Available: <http://dx.doi.org/10.1006/jpho.1997.0062>
- [12] G. Python, P. Pellet Cheneval, and M. Laganaro, “Dpistage norm des troubles de parole : apport des diadococinsies,” *Aphasie et domaines associs*, vol. 1, 2015.
- [13] D. Yu, S. Siniscalchi, L. Deng, and C.-H. Lee, “Boosting attribute and phone estimation accuracies with deep neural networks for detection-based speech recognition,” in *Proc. of ICASSP*. IEEE SPS, March 2012. [Online]. Available: <http://research.microsoft.com/apps/pubs/default.aspx?id=157585>
- [14] M. Cernak, B. Potard, and P. N. Garner, “Phonological vocoding using artificial neural networks,” in *Proc. of ICASSP*. IEEE, Apr. 2015, pp. 4844–4848.
- [15] J. Makhoul, “Linear prediction: A tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [16] M. Cernak and P. N. Garner, “PhonVoc: A Phonetic and Phonological Vocoding Toolkit,” in *Proc. of Interspeech*, 2016.
- [17] A. Asaei, M. Cernak, and H. Bourlard, “On Compressibility of Neural Network Phonological Features for Low Bit Rate Speech Coding,” in *Proc. of Interspeech*, Sep. 2015, pp. 418–422.
- [18] G. Perennou, “B.D.L.E.X. : A data and cognition base of spoken French,” in *Proc. of ICASSP*, vol. 11, 1986, pp. 325–328.
- [19] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [20] A. Asaei, H. Bourlard, and B. Picart, “Investigation of kNN Classifier on Posterior Features Towards Application in Automatic Speech Recognition,” Tech. Rep. Idiap-RR-11-2010, 2010, [online] <http://publications.idiap.ch/index.php/publications/show/1671>.