This may be the author's version of a work that was submitted/accepted for publication in the following source:

Duan, Shufei, Zhang, Xueying, Yan, Mimi, & Zhang, Jinglan (2020)
Statistical Distribution Exploration of Tongue Movement for Pathological Articulation on Word/Sentence Level.
IEEE Access, 8, Article number: 9091130 91057-91069.

This file was downloaded from: https://eprints.qut.edu.au/204811/

© 2020 The Author(s)

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution 4.0

Notice: Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.

https://doi.org/10.1109/ACCESS.2020.2993856
Statistical Distribution Exploration of Tongue Movement for Pathological Articulation on Word/Sentence Level

SHUFEI DUAN\(^1\), XUEYING ZHANG\(^1\), MIMI YAN\(^1\), AND JINGLAN ZHANG\(^2\)

\(^1\)College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China
\(^2\)College of Science and Engineering, Queensland University of Technology, Brisbane, QLD 4001, Australia

Corresponding author: Xueying Zhang (tyzhangxy@163.com)

This work was supported in part by the National Natural Science Foundation of China, under Grant 61371193, in part by the Natural Science Foundation of Shanxi Province, China, under Grant 201901D111095, and in part by the Taiyuan University of Technology Foundation, China, under Grant tyut-rc201405b.

Abstract
Pathological articulation exploration, especially the study of the kinematic characteristics of motor organ, is helpful to further reveal the essence of motor dysarthria. Due to the scarcity of the available pathological pronunciation database, there has little research working on the statistic distribution analysis for patients and normal controlled people. This paper applied the distribution method on TORGO database to discover the cognitive and motor rules of dysarthria patients. Single phoneme analysis is effective for locating the specific tongue muscle but ignoring cognitive ability assessment, particularly for the content understanding and the fluency degree of expression by patients. The paper focused on the word/sentence level rather than single phoneme analysis. The reaction time was designed to reveal the relationship between the brain cognition and motor neuron activation. The statistic distribution tells that the cerebral palsy or amyotrophic lateral sclerosis does affect people’s reflection and make the patients hard to control the tongue muscles effectively, resulting in unstable reaction time. The articulation velocity of patients appears 5mm/s faster than normal people, at 85mm/s, perhaps due to the factors of the word/sentence data and the big proportion of extra large displacement. It illustrates that the tongue moves relatively coherently and fluently once patients active the muscles, but hard to slow down as the muscle control ability decreases. The spatial occupancy was represented by the maximum articulation movement range (MAMR). We adapted the logarithmic normal distribution to find out the significant threshold for the diagnosis of dysarthria with the MAMR exceeding 7mm along left and right direction and the number of abnormal ranges surpassing 10% of total number. Primary test of MAMR, as an articulatory feature, for speech classification was carried out and it achieves 81% accuracy. These explorations convinced us to apply them to the pathological speech recognition task for improvement in future.

Index Terms
Dysarthria, statistical distribution, tongue movement, EMA, TORGO database.

I. INTRODUCTION
Speech function is considered to be a kind of motion with high degree of accuracy and speed [1], while response characteristics such as speed and acceleration of movement, are crucial in understanding the function of speech mechanisms, and changes in these parameters can better prompt the information of speech neurokinematics [2]. Motor dysarthria is speech dysfunctions caused by muscle weakness, motor incoordination, etc. The essence of motor dysfunction is dysfunction of motor organ [3]. Therefore, the study on kinematic characteristics of motor organ is helpful to further reveal the essence of motor dysarthria. As one critical articulation organ, the tongue could move in all directions within the mouth at a certain speed and varies its shape and size during speaking [4], reflecting the symptoms intuitively. Besides, there have been advanced instruments such as 3D Electromagnetic Articulgraph (EMA), NDI Wave to record the tongue movement while the subject talks. The device enables people to watch and learn the exact differences between normal controlled people and patients with dysarthria, which greatly facilitate the research of the pathological articulation,
and further provide guidance for intelligent assist devices’ development [5].

In the pathological articulation study, researchers often set up a targeted mission for subjects such as swallowing water [6], speaking given text [7]–[9], acting specific gestures [10] and collect the related kinematic parameters through sensors. The most kinematic features extracted and compared are velocity, acceleration, and movement size or space displacement. Among existing reported studies on the tongue movement in dysarthria, kinematic features vary a lot and the variation trend is inconsistent for different tasks. The speed and movement size seemed to be reduced and the duration of tongue movement was prolonged in patients with ALS when they completed a task of swallowing water. Language coordination was reduced as well [6]. In some phonological studies, the tongue movement tends to be reduced [3], [8], [9]. Whereas, increased tongue movement amplitude and speed were found in studies of opening/closing gestures [10], [11].

Considering the fact that there is little correlation between the nonverbal oral motor performance and speech behavior performance due to the different central nervous system control mechanism, it is necessary to evaluate the articulation movement in the process of speech production [12]. Current phonological studies are usually research on the single phoneme like vowels [4], [13], [14], consonants [15], [16]. Impaired articulation is, however, more likely to occur in connected speech in dysarthria patients than at the syllable level [17], [18]. Words and sentences are the basic and effective language for communication eventually. Short words pronunciation can well examine the speaker’s ability to organize vocabulary, grammar and semantic processing, as well as the performance of coherent and fluent pronunciation [19]. Therefore, the examination of articulation at the sentence level is required.

Similarly, existing studies about the sentence level pathological speech analysis also appear the phenomenon of inconsistency. Chen et al. reported that the maximum speed and acceleration of tongue movement were decreased but the duration was prolonged for dysarthric speakers following stroke when the subjects repeated five times of the tongue-tip and tongue-back sentences [20]. A study of the tongue tip and dorsum during a passage-reading task showed an overall reduction in tongue movement size and speed [7]. Whereas, in order to examine the association between the speech intelligibility and the articulatory movements of Parkinson’s disease patients on sentence level, Kearney et al. found that compared to the normal pronunciation, patients appears larger movement sizes for loud, clear and slow speech, and faster speeds for loud, clear speech [21].

Besides the inconsistency of research results, to the best of our knowledge, though there are considerable of kinematic exploration work on the pathological articulation analysis, they mostly are based on a single phenomenon analysis and lack of statistical distribution studies. What’s more, they limited to discuss appeared phenomena and explained the inducement mechanism [4], [22], [23]. Few work actually proposes the numerical indices for the dysarthria diagnosis. Currently, there are two main categories of dysarthria assessment: subjective and objective approaches [24]. The most common used assessments in recent rehabilitation practice and speech rehabilitation institutions are still those based on subjective auditory perception and/or subjective scales, with poor objectivity and stability [25]. Automatic pathological speech recognition has risen up to facilitate the diagnosis task. The INTERSPEECH, an international conference in the field of speech signal processing, has held two challenges on pathological speech research in 2012 [26] and 2015 [27], respectively. However, the features reported are mostly sourced from spectrum, prosodic features [28]–[31] and glottal features [32], lack of articulatory features.

Upon the above analysis, this paper works on the pathological articulation study on the word/sentence level by adapting statistical distribution method, aiming to explore the variation trend of tongue movement through kinematic features for pathological articulation and derive the numerical index for the dysarthria diagnosis. In the end, we expect to apply proposed articulatory features for better representation of pathological speech.

II. TORGO DATASET

A. DATASET OVERVIEW

In the current pathological phonetics research, the MEEI pathological voice database (by the Massachusetts Eye and Ear Hospital) [33] and the NKI-CCRT database [34] (by the University of Amsterdam) are widely used [35], [36], but they do not contain organ movement data that correspond to the audio data. The TORGO database [5], developed by the Holland-Bloorview Kids Rehab Hospital and the departments of Computer Science and Speech Language Pathology at the University of Toronto, is also widely studied [37], [38], which records the audio and articulation movement data of patients with cerebral palsy or amyotrophic lateral sclerosis (ALS) through the application of the EMA AG500 [39]. More important, this database is publicly available. However, in P.R. China, due to multiple reasons such as the medical qualification examination of equipments and ethical inspection for patients, the pathological articulation is rarely studied, and the public pathological database is unavailable. Thus, this study selects TORGO database as the research basis.

The TORGO database includes the audio data and synchronized pronunciation movement data for eight dysarthria patients (five males and three females) and seven normal people (four males and three females). The recorded corpus includes non-words, short words, restricted sentences, and unrestricted sentences [40]. The Euclidean distance of each sensor from the transmitter can be displayed by electromagnetic position data, which can then calculate XYZ coordinates, as well as measure, store and display the sensor’s position. The X axis represents the forward and backward
direction, Y axis represents the left and right direction, and Z axis represents the up and down direction. The sampling rate of recording action data is 200Hz. FIGURE 1(a) gives the 3D EMA AG500 device with a subject tested. FIGURE 1(b) shows the sensor configuration on the subject. Sensors were attached to the Tongue Tip (TT), Tongue Middle (TM), Tongue Back (TB), Left Mouth (LM), Right Mouth (RM), Upper Lip (UL), Lower Lip (LL), and jaw.

B. DATA SELECTION
To explore the similarities and differences of tongue movement between dysarthria patients and normal controlled group, it is essential to examine the two groups’ data at a peer platform. Since the raw dataset contains invalid data due to various reasons such as unstable sensor, data selection has to be conducted according to specific criteria, see TABLE 1.

TABLE 1. Dataset selection criteria.

| Selection Order | Selection Criterion                        | Dysarthria patients | Normal people |
|-----------------|-------------------------------------------|---------------------|---------------|
| 1               | Articulation data                          | ✓                   | ✓             |
| 2               | Corpus is short word or restrictive sentence. | ✓                   | ✓             |
| 3               | Pronounce same text                        | ✓                   | ✓             |
| 4               | Dropped sensor                             | ×                   | ×             |
| 5               | Abnormal data (movement range<50mm)       | ×                   | ×             |
| 6               | RMSE < 10.0mm                              | ✓                   | ✓             |

The selection processes in orders:
1) Articulation data. Confirm that the data contains both audio and articulate position file, especially the latter one.
2) Corpus is short word or restrictive sentence. Though the raw data is composed of non-words, short words, restricted sentences, and unrestricted sentences, the paper focuses on the short words and restricted sentences as the other two kinds of data is hard to control and compare for two groups in the aspect of the semantic and content.
3) Pronounce same text. Aiming to compare the kinetic characteristics between two group people requires data of the same text file.
4) Dropped sensor. Get rid of the data if there is dropped sensor.
5) Abnormal movement range. Based on the dataset and the experienced value of normal tongue movement range, get rid of the abnormal data with the range is greater than 50mm.
6) RMSE<10mm. The RMSE value is used to examine the stability of the dataset for each file, see equation (1).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2}{n}}
\]

where \(n\) represents position node number in one file. \((x_i, y_i, z_i)\) mean three-dimensional position data at the \(i\)th node. \((\bar{x}, \bar{y}, \bar{z})\) represent the mean value of all nodes in the x, y and z directions. Generally, the lower the RMSE, the more stable the data is. Here, 10mm is chosen as the threshold. Any file with RMSE>10mm is regarded [42].

C. DATASET PREPARATION
After the selection, the remained data are all named uniformly by the form of ‘content_number’ like bing_01 and zheng_01.

In the end, the dataset is prepared which in total contains five subjects with dysarthria (two females and three males) and two normal subjects (one female and one male). 200 restrictive sentences and 686 short words are suitable for comparison. More specifically, when it regards to the different tongue parts, the detailed data composition is listed in TABLE 2.

TABLE 2. Composition of identical text articulation dataset.

|               | tongue tip | tongue middle | tongue root |
|---------------|------------|---------------|-------------|
| dysarthria patients | 800        | 797           | 718         |
| normal people   | 640        | 639           | 632         |

III. THE OVERVIEW OF COMPARISON ANALYSIS BETWEEN DYSARTHRIA PATIENTS AND NORMAL CONTROLLED PEOPLE
The whole process of the comparison analysis between dysarthria patients and normal controlled people is described in FIGURE 2. According to the constitution of the TORGO dataset, the task is split into two parts. For audio data, the reaction time is calculated to explore the cognitive reflection of two groups. The articulatory data tells the movement of the tongue, which is measured by velocity and space distribution. The 3D point cloud analysis is firstly adopted to observe the tongue position differences between two groups when they spoke the same text. And then statistical calculation of the space distribution is carried out by extracting
the maximum articulation movement range (MAMR) as the
feature.
This paper is about to discover the distinguishment
between patients and normals through these features, and
probably to propose the diagnose criteria for dysarthria. In the
final stage, the velocity, reaction time, and the extracted
MAMR will be novel features feeding into the pathological
speech recognition task to improve the accuracy, which will
be our future work.

FIGURE 2. The overview of comparison analysis between dysarthria patients and normal controlled people.

IV. DISTRIBUTION OF REACTION TIME FOR
PRONUNCIATION STARTUP
Reaction time is duration between application of stimulus
to onset of response [43]. The reaction time of pronunciation
startup defined here refers to the reflecting time duration of
subjects from receiving the “START” instruction to actually
pronouncing the required text, see FIGURE 3. The purpose of
this study is to discover the otherness between patients and
normal people in the aspects of brain cognition and muscle
control through calculating the statistical distribution of the
reaction time $R_t$.

For uniform calculation, suppose the begin time $t_{\text{start}}$ of the
recording is the “START” instruction order given time. The
end time $t_{\text{end}}$ is bordered by the designed endpoint detection
algorithm.

A. ENDPOINT DETECTION
As the aim of endpoint detection is to guard the start voice,
the algorithm designed here simply takes the short-time
energy as the feature. Suppose $y_i(n)$ is the $i$th frame signal
obtained after the original sound wave $x(n)$ is windowed by
the hamming window function of $w(n)$. $y_i(n)$ should satisfy
the equation (2):

$$y_i(n) = w(n) * x((i-1)^{\text{inc}} + n), 1 \leq n \leq L, 1 \leq i \leq f_n$$  \hspace{1cm} (2)

where $n$ is the sample point, $L$ is the frame length, $\text{inc}$ is the
frame shift, and $f_n$ is the total number of frames after framing.
The short-term energy of the sound signal at the $i$th frame
is calculated by equation (3)

$$E(i) = \sum_{n=0}^{L-1} y_i^2(n), 1 \leq i \leq f_n$$  \hspace{1cm} (3)

A higher threshold $T_2$ is firstly selected on the short-term
energy envelope curve for a rough judgment. The energy
higher than $T_2$ must be sound, and the starting time point
should be outside the point of intersection of the threshold and
the energy envelope. Then a lower threshold $T_1$ is chosen, and
search towards both ends. Find the two points where the short-
term energy envelope intersects the threshold $T_1$. The two
points are the starting and ending point of the sound segment
determined by the first order decision.

FIGURE 3. Reaction time of pronunciation startup.
At this step, only the start point of the sound is used as \( t_{\text{end}} \) to determine the reaction time \( R_t \) of pronunciation startup.

### B. DISTRIBUTION HISTOGRAM

Once the reaction time of pronunciation startup is obtained, the statistical distribution histogram can be drawn in FIGURE 4. The maximum and minimum values are used to form the x axis, and 40 bins are chosen to equally split the x axis. The column bar represents the frequency of data fallen in the corresponding bin. The black dotted bar represents the normal people, while the purple bar illustrates the dysarthria patients.

![FIGURE 4. The distribution map of the reaction time of pronunciation startup.](image)

Through FIGURE 4, we can see that the pronunciation startup time of normal people is more concentrated than that of patients with dysarthria, about 0.8sec to 1.5sec, obeying normal distribution, fitted by normal distribution curve (green line in the FIGURE 4), the mean value is 1.19sec, and the variance is 0.11. Whereas, for the distribution of patients with dysarthria, although a certain number of data tends to concentrate at 1s to 1.5sec, there are plenty of data scattered along the both sides of the main peak, from 0sec to 6sec, exhibits diversity.

This tells that from hearing the “START” instruction to actually pronouncing, the brain of normal people needs about 1sec to convert the order and mobilize the relative muscle group to behave in a well controlled way. In the contrast, the reaction time of patients vary a lot. Sometimes they start to speak before the “START” instruction, and in some occasions, 4sec or 6sec are needed to be ready for talk. Reaction time is a useful physiological parameter which is affected by many physiological and pathological parameters [44]. Considering the patients of TORGO database suffer from the cerebral palsy or amyotrophic lateral sclerosis, two reasons may explain this: 1) the cognitive impairment of brain; 2) muscle atrophy. The analysis, in turn, reflects that the exploration of the reaction time of pronunciation startup is critical for the judgment of dysarthria in a certain degree.

### V. DISTRIBUTION OF ARTICULATION VELOCITY

Currently, most studies on the speed of articulation movement are about the instantaneous speed of single vowel or consonant. In order to further study the velocity of articulation organs when the corpus of pronunciation is phrases and restricted sentences, a method for calculating the average velocity of articulation is proposed. The articulation movement speed defined in this paper refers to the average movement speed of a articulation organ from the beginning to the end of the articulation of the subject. The time occupied by the preparation of the articulation organ before pronunciation and the time occupied by the restoration of the original state of the articulation organ after pronunciation are discarded.

#### A. ARTICULATION MOVEMENT SPEED CALCULATION

1. **Step1:** Conduct the endpoint detection algorithm designed in section III on the .wav file to get the start time \( t_{\text{start}} \) and end time \( t_{\text{end}} \) of pronunciation.

2. **Step2:** According to the time information achieved in Step1, and the recording frequency of 200Hz (EMA AG500), find out the corresponding pronunciation position nodes \( n_{\text{start}} \) and \( n_{\text{end}} \) in the .pos file, as equation (4) and (5).

\[
\begin{align*}
    n_{\text{start}} &= \left\lceil \frac{t_{\text{start}}}{0.005} \right\rceil \\
    n_{\text{end}} &= \left\lfloor \frac{t_{\text{end}}}{0.005} \right\rfloor
\end{align*}
\]

where, \( n_{\text{start}} \) needs to be rounded up and \( n_{\text{end}} \) needs to be rounded down, in order to make sure there has sufficient position nodes.

3. **Step 3:** Calculates the total displacement \( S \) of pronunciation movement from the start position node \( n_{\text{start}} \) to the end node \( n_{\text{end}} \) by Euclidean distance equation (6).

\[
s = \sum_{i=n_{\text{start}}}^{i=n_{\text{end}}-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}
\]

where the three-dimensional position information of the \( i_{\text{th}} \) node is \((x_i, y_i, z_i)\).

4. **Step4:** According to the pronunciation movement distance \( S \) and pronunciation time \( t_{\text{start}}, t_{\text{end}} \) of a corpus, calculate the average pronunciation movement speed.

\[
\bar{v} = \frac{S}{t_{\text{end}} - t_{\text{start}}}
\]
The tongue movement is split into three parts in the aim of comparison: tongue root, tongue middle and tongue tip. Thus, the movement speed calculation is also carried out for these parts separately. The distribution histogram is then derived and the curve fitting is applied as well. To better illustrate the difference between two groups, only the fit curves are shown in FIGURE 5.

We find that all the three tongue parts of patients act almost in consistent speed. But for normal people, the tongue middle moves slightly slower than the other two parts. More surprisingly, the whole tongue movement speed of patients is faster than normal people, at around 85mm/s, while that of normal is about 80mm/s. This discovery is different from single phoneme research, which reports that the speed of tongue movement gets slower than normal people [45]. But, it strongly proves Goozee’s opinion [46] that the reason the tongue moves slower while dysarthria patients speak is not because of the speed of movement itself slows down but the speed control ability decreases, which presents the difficulty to slow down, in the end resulting in inaccurate pronunciation. Besides, single phoneme like vowel or consonant articulation more focuses on the sole and specific motor neurons’ reflection, while words and sentences allows patients expressing the content in coherent and fluent way and to some extent relieves muscle stiffness.

The articulation position data of tongue root, tongue middle and tongue tip for dysarthria patients and normal controlled people are plotted in the same coordinate system in the condition of pronouncing the same text “Will Robin wear a yellow lily?”, as shown in FIGURE 6. Since it looks like a cloud in the sky, we named it as 3D point cloud. The blue color represents normal people while the black one means dysarthria patient. Through the observation, it could find that compare to the normal people, patient’s tongue moves lower, more left, and more backwards for the same text.

To dig more deeply, FIGURE 7 discovers the motion tendency of two groups only for the tongue root. Besides from what we have found in FIGURE 6, FIGURE 7 tells that normal controlled people can lift the tongue root along the Z axis which is the up-down direction, while the patient hardly lifts and the tongue root remains nearly flat. It reflects that due to the organic lesions of the neuromuscular, muscular weakness, paralysis, or dystonia of the vocal organs, the patient could not mobilize the tongue muscle effectively, so that results in uncoordinated movement.

VI. SPATIAL DISTRIBUTION

B. DISTRIBUTION HISTOGRAM

The articulation position data of tongue root, tongue middle and tongue tip for dysarthria patients and normal controlled people are plotted in the same coordinate system in the condition of pronouncing the same text “Will Robin wear a yellow lily?”, as shown in FIGURE 6. Since it looks like a cloud in the sky, we named it as 3D point cloud. The blue color represents normal people while the black one means dysarthria patient. Through the observation, it could find that compare to the normal people, patient’s tongue moves lower, more left, and more backwards for the same text.

To dig more deeply, FIGURE 7 discovers the motion tendency of two groups only for the tongue root. Besides from what we have found in FIGURE 6, FIGURE 7 tells that normal controlled people can lift the tongue root along the Z axis which is the up-down direction, while the patient hardly lifts and the tongue root remains nearly flat. It reflects that due to the organic lesions of the neuromuscular, muscular weakness, paralysis, or dystonia of the vocal organs, the patient could not mobilize the tongue muscle effectively, so that results in uncoordinated movement.

B. MAXIMUM ARTICULATION MOVEMENT SPACE DISTRIBUTION MAP

The 3D point cloud gives the intuitive observation of the differences between two groups, however, the result is improper to be generalized since it is just single phenomenon analysis and lack of statistical distribution studies. Thus, the distribution characteristics of articulation movement of tongue tip, tongue middle and tongue root in front and back, left and right, up and down directions are statistically explored in this section. The feature selected here is the maximum articulation movement range (MAMR), which is defined as the space between the maximum and minimum position values.
The distribution map displays that along the X axis, the MAMR increases from the minimum to maximum values. While most ranges tend to concentrate on less than 25 mm for all three directions, great convergence differences between two groups of people occur at the ranges greater than 25 mm. Clearly, the right side of the distribution map for patients converges slowly compared to normal people. Besides, among three direction movement, the articulation range appears the least distance along left and right direction with less than 10 mm, while the other two directions are around 0~25 mm. In order to evaluate the convergence difference between dysarthria patients and normal subjects on the right side of the map, 25 mm is chosen as the threshold for abnormal speech movement (the straight line drawn in FIGURE 9). The ratio of these abnormal data accounting in the total amount of data for two datasets is derived in FIGURE 10, where F/R, L/R, U/D denotes FRONT/REAR, LEFT/RIGHT, and UP/DOWN directions, respectively.

The bar chart tells that the ratio of abnormal MAMR for patients is much higher than normal people, with the highest number 9.79%, and the lowest ratio is 2.61%. Whereas for normal people it only has 1.88%, which is almost 8% less than that of patients. This study illustrates that the statistical analysis of MAMR distribution could facilitate the diagnosis of dysarthria. However, in this part, the abnormal articulation threshold 25 mm is selected by observation. To precisely describe the distribution features of abnormal movement, the statistic calculation needs to be conducted.

C. LOGARITHMIC NORMAL FITTING OF MAXIMUM RANGE DISTRIBUTION OF PRONUNCIATION MOVEMENT

By analyzing the distribution histogram, it is found that the range distribution of articulation movement obeys logarithmic normal distribution. The density function of the logarithmic normal distribution is equation (12).

\[ f(x) = \frac{1}{x\sigma \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \]  

where \( \mu \) and \( \sigma \) are the mean and standard deviation of the logarithm of the variable, respectively. To draw the probability density curve, the \( \mu \) and \( \sigma \) need to be carried out by equation (13) and (14), separately. And then apply these two indices to the equation (12).

\[ \mu = \frac{\sum_{j=1}^{n} (\ln I_j)}{n} \]  

\[ \sigma = \sqrt{\frac{\sum_{j=1}^{n} (\ln I_j - \mu)^2}{n}} \]  

To better illustrate and avoid duplicating, the distribution of MAMR for the tongue tip is taken as an example. FIGURE 11 gives the probability density curve for both dysarthria patients and normal subjects, where red solid lines
are curves for patients and green dotted lines are for normal persons. From the graphs, we could see that the probability density curve for the logarithmic normal distribution could fit the movement range distribution well. The statistics results for three articulation organs in three directions are listed in Table 3.

From the mean value $\mu$, it infers that the patients tend to pronounce harder than normal people since their MAMR is larger than that of normal controlled persons. With higher $\sigma$, patients exhibit more fluctuation than normal people. These two indices reflect that patients need to try harder than normal people to control tongue muscles for pronunciation due to the paralysis of muscles.

In addition, the tongue movement range is quite narrow at the left and right direction, comparing to the other two directions. The mean values are around 5.9mm and 4.46mm for patients and normal people, respectively. While the range of the other two directions is around 12mm and 15mm for two group people. Besides, the standard deviation also displays less fluctuation in left and right direction, especially for normal controlled people. This feature is critical that could be used later on for classification task to distinguish patients and normal people.

D. ABNORMAL ARTICULATION MOVEMENT RANGE ANALYSIS

Based on the previous exploration, we discovered that the maximum articulation movement range (MAMR) distribution converges slowly along the right side for patients and the tongue tends to move more stable along the left and right direction. This part will focus on the abnormal MAMR analysis in left and right direction by precisely locating the range threshold for abnormals.

The lognormal distribution has one critical property that if the random variable $\xi$ obeys the lognormal distribution, $\ln \xi$ would follow the normal distribution. Upon the theory, calculate the natural logarithm $\ln I_j$ for the MAMR $I_j$ firstly, draw the histogram of frequency distribution afterwards, and then apply the normal distribution to curve-fit the histogram, see Figure 12. Again, to avoid the redundancy, Figure 12 only gives out the distribution map of tongue tip in left and right direction. Clearly, patients’ distribution map has a long tail along right side while normals obeys normal distribution very well.

As $\ln I_j$ follows the normal distribution regarding to the property of lognormal distribution, the long tail along the right side of patients’ distribution is actually the abnormal articulation ranges. To locate the threshold is vital for distinguishing patients and normal people. Figure 13 gives out the determination algorithm design for the abnormal
S. Duan et al.: Statistical Distribution Exploration of Tongue Movement for Pathological Articulation on Word/Sentence Level

FIGURE 11. The probability density curve of movement range for both dysarthria patients and normal subjects.

TABLE 4. The proportion of abnormal articulation data in left/right direction for three organs.

| Organ (Direction) | Data Type | Dysarthria Patients | Normal People |
|-------------------|-----------|---------------------|---------------|
|                   |           | $\mu$  | $\sigma$ | $n_{+2.58\sigma}$ | $r$ | $\mu$  | $\sigma$ | $n_{+2.58\sigma}$ | $r$ |
| Tongue Tip (L/R)  | $\ln I_j$ | 1.23 | 0.40 | 7.30 | 155 | 19.38% | 1.48 | 0.40 | 5.4 | 4 | 0.63% |
|                   | $I_j$    | 3.43 | 1.50 |                 |     |          | 4.38 | 1.49 |           | |
| Tongue Middle (L/R) | $\ln I_j$ | 0.96 | 0.41 | 6.49 | 170 | 21.33% | 1.25 | 0.48 | 7.66 | 11 | 1.72% |
|                   | $I_j$    | 2.60 | 1.51 |                 |     |          | 3.48 | 1.62 |           | |
| Tongue Root (L/R) | $\ln I_j$ | 1.12 | 0.40 | 7.00 | 73 | 10.17% | 1.56 | 0.42 | 8.64 | 2 | 0.32% |
|                   | $I_j$    | 3.08 | 1.52 |                 |     |          | 4.77 | 1.50 |           | |

FIGURE 12. The frequency histograms of tongue tip movement ranges after taking the natural logarithm in left and right direction for both dysarthria patients and normal subjects.

MAMR threshold. The theoretical explanation and specific processing flow are as follows.

Since the density function of normal distribution is symmetrical about the mean value and reaches the maximum frequency at the mean point, we take the left half as reference to work out the tail threshold on the right side. The specific steps are following:

Step1: Calculate the mean value $\mu$.

$$\mu = \min(\ln I_j) + (p + 1) \times b$$  \hspace{1cm} (15)

where $p$ is the bin number with the largest frequency in the histogram and $b$ is the bin width.

Step2: Calculate one standard deviation $\sigma$. Due to the long tail on the right side of the distribution map, only the ranges on the left part are involved to derive the $\sigma$ as listed in equation (16). $N$ represents the number of ranges smaller than the mean value.

$$\sigma = \sqrt{\frac{\sum(\ln I_j - \mu)^2}{N}}$$  \hspace{1cm} (16)

Step3: Work out the $+2.58\sigma$ as the threshold for the abnormal articulation range since the confidence level achieves 99% under the confidence interval $(-2.58\sigma, +2.58\sigma)$ for normal distribution.

Once the threshold is determined, the ratio of abnormal ranges over total number of ranges can be solved. TABLE 4 lists the proportion of abnormal articulation data along the left and right direction for three organs. In the table, $I_j$ represents the original data which is before the logarithm. $\ln I_j$ means data after taking the logarithm. The abnormal articulation range threshold is only given for original data as this number is more usable for analysis. $n_{+2.58\sigma}$ denotes the abnormal range numbers and $r$ denotes the ratio.

From the table, it can be seen that the proportion of abnormal articulation movement range in the left and right direction of tongue in patients is more than 10%, even around 20%. This number is more than 10 times of normal persons, which is only about 1%. Besides, the abnormal articulation range threshold for two group people keeps consistent at.
FIGURE 13. The abnormal MAMR threshold determination algorithm.

approximately 7mm. This discovery is significant for the judgment of dysarthria through statistic distribution method. With this analysis, if the MAMR along the left and right direction exceeds 7mm and the number of abnormal ranges surpasses 10% of total number, the subject may be confirmed as dysarthria.

VII. DISCUSSION AND COMPARISON

Existing research about the dysarthria mainly works on the kinematic analysis for single phoneme like vowels and consonants. The kinematic features include velocity, acceleration, and displacement etc. Single phoneme analysis is effective and meaningful for locating the specific tongue muscle but ignoring cognitive ability assessment, particularly for the content understanding and the fluency degree of expression by patients. This paper aims to reveal the cognitive and kinematic features on the word and sentence level for dysarthria patients suffered from cerebral palsy or amyotrophic lateral sclerosis (ALS) based on the public TORGO database.

Perhaps due to the scarcity of the available pathological pronunciation database, there has little research working on the statistic distribution analysis of comparative study for patients and normal controlled people. This paper applied the distribution method to discover the cognitive and motor rules of dysarthria patients, which is more convincing compared to single phenomenon analysis.

A. THE REACTION TIME BEFORE PRONUNCIATION STARTUP

Section IV defines and calculates the reaction time before pronunciation start for both groups and drawn the distribution map in FIGURE 4. The results tell us that the brain of normal people needs about 1sec to convert the “START” order and mobilize the relative muscle group to behave in a well controlled way. The distribution obeys normal distribution. This discovery is consistent with Truskinger’s research about the spectrogram flashing reaction time [47] and Kosinski’s literature review about the reaction time [48]. Truskinger’s research was to measure subjects’ choice reaction time through the visual inspection of audio visualized as spectrograms for acoustic analysis. They tested 1sec, 2sec and 5sec for the spectrogram flashing. It has been verified that 2sec is the best option for accepted classification accuracy [47]. In Kosinski’s review, many researchers have confirmed that reaction to sound is faster than reaction to light, with mean auditory reaction time being 140-160 msec and visual reaction times being 180-200 msec [48]. Perhaps this is because an auditory stimulus only takes 8-10 msec to reach the brain [49], but a visual stimulus takes 20-40 msec [50]. When it regards to the TORGO dataset recording, subjects starts to speak when they heard the instruction “START UP”, which took the sound as the stimulus. In the end, the normal controlled group presented 1sec as the average reaction time, faster than the 2sec visual reaction time in [47]. From this point of view, the result is reasonable.

Whereas patients’ reaction time varies a lot, from 0sec to 6sec. The statistics indicate that the patients may exhibit the cognitive impairment of brain and muscle atrophy. As might be expected, brain injury slows reaction time [51]. Collins et al. found that high school athletes with concussions and headache a week after injury had worse performance on reaction time and memory tests than athletes with concussions but no headache a week after injury [52]. Eckner et al. cited several papers that studied the slowing of reaction time after concussion [53]. Soldiers and contractors in Iraq who suffered mild traumatic brain injury showed a marked impairment of reaction time when measured within 72 hours of the injury [54]. Considering the patients here suffer from the cerebral palsy and ALS, this result is reasonable and convincible.

To sum up, the defined reaction time could be viewed as one critical feature for distinguishing dysarthria patients and normal people.

B. ARTICULATION VELOCITY AND SPACE DISTRIBUTION

Section V concerns about the articulation velocity. Different from other speed definition of the single phoneme analysis, the velocity here is depended on the displacement for the whole sentence or word. The time is bordered by endpoint detection algorithm. The distribution map is also
depicted in FIGURE 5. From the graph, we discovered that for words/sentences, patients actually pronounce faster than normal people by about 5mm/s, at 85mm/s. This discovery is consistent with Kearney’s research [21] and can be well explained by Goozée’s perspective [46], [55].

Kearney et al. examined the association between the speech intelligibility and the articulatory movements of Parkinson’s disease (PD) patients on sentence level [21] as impaired articulation is more likely to occur in connected speech in PD than at the syllable level [17], [18]. They found that compared to the normal pronunciation, PD appears larger movement sizes for loud, clear and slow speech, and faster speeds for loud, clear speech. That’s why increasing loudness or clarity is a frequently used approach in the treatment of dysarthria to improve the speech intelligibility.

In Goozée’s research [46], the dysarthria patient was required to speak three single syllable real words consisting of the lingual consonants /l, s, k/. The study found that the patient was able to accelerate his tongue as quickly as the control subject in the production of consonants and reached a maximum velocity that was consistent with the control subject. However, the dysarthria patient exhibited difficulty in decelerating his tongue movements appropriately on the approach up to the palate during consonant production. The difficulty noted in deceleration resulted in inaccurate tongue movements and may have been instrumental in reducing the length of time that the tongue remained at the palate in comparison to the control subject. Thus, Goozée considers the tongue moves slower not because of the speed of movement itself slows down but the speed control ability decreases, which presents the difficulty to slow down. The discovery here indeed proofs that the speed of patients’ speaking does not slow down but faster when patients speak words or sentences, which is consistent with Goozée’s opinion. Besides, words, particularly, sentences allows patients expressing the content in coherent and fluent way and to some extent relieves muscle stiffness.

The displacement is another critical factor to lead the increasing velocity. Many research have noticed that the movement size of articulatory organs increases during the articulation for the tongue movement [10], [11], [21].

Section VI verified this phenomenon through the spatial distribution. The 3D point cloud analysis is upon the single phenomenon observation to give out the intuitive understanding of the differences between patients and normal people. The feature proposed to extract is the Maximum Articulation Movement Range (MAMR). It reveals the spatial occupancy of tongue movement. The right part of the statistical distribution map (FIGURE 11) has a long tail indicating the abnormal MAMR, which can be used to separate patients. As the distribution map obeys logarithmic normal distribution, we took the vital property of it and converted the map into the normal distribution. Under this condition, we would calculate the abnormal MAMR threshold to cut off the tail and derived the ratio that abnormals taking in the whole dataset. The analysis tells that the proportion of abnormal MAMR in the left and right direction of tongue in patients is more than 10 times of normal persons and the MAMR threshold is 7mm. Back to FIGURE 11, the long tail signifies that the maximum articulation movement gradually increases along X axis. In addition, because of the high proportion of these large movement ranges, the total displacement of the tongue actually greater than normal people, which results in the greater velocity as discussed above.

VIII. CONCLUSION AND FUTURE WORK

Kinematic studies provide a direct insight into the articulatory changes in dysarthria pronunciation. The work presented in this paper primarily aims to identify the distinguishing articulatory indicators for dysarthria patients and normal people by statistical distribution methods. As IT researchers, we hope to provide more scientific and numerical indices or tools to facilitate doctors in dysarthria diagnose and patients’ recovery treatment. Through the rigorous mathematical derivation, we finally put forward one numerical index for dysarthria, that is, if the MAMR along the left and right direction of the tongue exceeds 7mm and the number of abnormal ranges surpasses 10% of total number, the subject may be confirmed as dysarthria.

In addition, mouth/tongue massage and reading/speaking training are still the main methods used in the rehabilitation practice of patients with dysarthria at present. The word/sentence level research would help doctors design proper texts to better evaluate the fluency and intelligibility for patients compared to single syllables [21]. The tongue movement research enlightens doctors in designing appropriate gestures according to the patients’ situation to mobilize different tongue muscles. More importantly, the extraction of the articulatory features would greatly feed new blood into automatic pathological speech recognition. Thus, the exploration of the articulatory indicators is significant in promoting the development of medical diagnosis and treatment under AI era.

Traditional pathological speech analysis are based on spectrum and prosodic features, with little consideration of articulatory property. The paper extracted reaction time, the articulation velocity and the MAMR as potential features for pathological speech analysis. In future, we expect to apply these features to improve the recognition accuracy, as pointed at the right part in FIGURE 2.

As a trail job, also in order to verify the effectiveness of the maximum articulation movement range (MAMR) as one feature for classification task, the comparative study is conducted by simply applying the support vector machine as the classifier. The MAMR is extracted for tongue tip, tongue middle, and tongue root in three directions of front/rear, left/right, and up/down. The dataset is adopted the one we prepared in section II. The classification results are listed in TABLE 5.

As seen from the table, the MAMR can indeed be used to categorize the patients and normal people with the accuracy about 80%. Among three directions, the feature of
the left/right direction outperforms the other two directions slightly, which proves that the tongue moves more stable along left/right direction as discussed in part C of section VI. When all directions are combined, the recognition accuracy improves to 82.35%, higher than that of single direction classification.

Upon the existing research presented in this paper, the future work will be spread in two channels. First, we would like to build cooperation with one local hospital, and construct the dysarthria dataset like the TORGO dataset according to the characteristics of local people. Second, we would test all three extracted features in pathological speech recognition, and will compare them with spectrum and prosodic features, and so on. The feature fusion task is probably also required to improve the recognition accuracy.

**ACKNOWLEDGMENT**

The authors would like to specially thank the Holland-Bloorview Kids Rehab Hospital and the University of Toronto for their kindness in making the TORGO database public so that we can conduct this research.

**REFERENCES**

[1] B. E. Murdoch, *Dysarthria: A Physiological Approach to Assessment and Treatment*. Stanely Thomes, U.K.: Cheltenham, 1998.

[2] D. P. Kuehn and K. L. Moll, “A cineradiographic study of VC and CV articulatory velocities,” *J. Phonetics*, vol. 4, no. 4, pp. 303–320, Oct. 1976.

[3] X. Feng, W. Yue, Z. Chen, H. Wu, and X. Qin, “Articulatory kinematic characteristics during speech production for individuals with dysarthria: A study based on electromagnetic articulography,” *Chin. J. Rehabil. Theory Pract.*, vol. 25, pp. 7–16, Jan. 2019.

[4] P. Gómez, J. Mekyska, A. Gómez, D. Palacios, V. Rodellar, and A. Álvarez, “Characterization of Parkinson’s disease dysarthria in terms of speech articulation kinematics,” *Biomed. Signal Process. Control*, vol. 52, pp. 312–320, Jul. 2019.

[5] F. Rudzicz, A. K. Namasiyayam, and T. Wolff, “The TORGO database of acoustic and articulatory speech from speakers with dysarthria,” *Lang. Resour. Eval.*, vol. 46, no. 4, pp. 523–541, Dec. 2012.

[6] J. H. Perry, R. Martino, Y. Yunusova, E. K. Plowman, and J. R. Green, “Lingual and jaw kinematic abnormalities precede speech and swallowing impairments in ALS,” *Dysphagia*, vol. 33, no. 6, pp. 840–847, Dec. 2018.

[7] G. Weismer, Y. Yunusova, and K. Bunton, “Measures to evaluate the swallowing processes,” presented at the Interspeech, Portland, OR, USA, 2012.

[8] P. Xue, X. Zhang, J. Bai, and Z. Wang, “Acoustic and kinematic analyses of Mandarin vowels in speakers with hearing impairment,” *Clin. Linguistics Phonetics*, vol. 32, no. 7, pp. 622–639, Jul. 2018.

[9] Y. Yunusova, I. S. Rosenthal, K. Rudy, M. Baljko, and J. Daskalogniakakis, “Positional targets for lingual consonants defined using electromagnetic articulography,” *J. Acoust. Soc. Amer.*, vol. 132, no. 2, pp. 1027–1038, Aug. 2012.

[10] M. N. Wong, B. E. Murdoch, and B.-M. Whelan, “Kinematic analysis of lingual function in dysarthric speakers with Parkinson’s disease: A electromagnetic articulograph study,” *J. Austral. College Speech Therapists*, vol. 12, no. 5, pp. 414–425, 2010.

[11] M. N. Wong, B. E. Murdoch, and B.-M. Whelan, “Lingual kinematics in dysarthric and nondysarthric speakers with Parkinson’s disease,” *Parkinson’s Disease*, vol. 2011, pp. 1–8, Oct. 2011.

[12] W. Ziegler, “To speak or not to speak: Distinctions between speech and nonspeech motor control,” *Aphasiology*, vol. 17, no. 2, pp. 99–105, Jan. 2003.

[13] A. Mefferd, “Articulatory-to-Acoustic relations in talkers with dysarthria: A first analysis,” *J. Speech, Lang., Hearing Res.*, vol. 58, no. 3, pp. 576–589, Jun. 2015.

[14] R. J. J. H. V. Son, C. Middag, and K. Demuyckx, “Vowel space as a tool to evaluate articulation problems,” in *Proc. Interspeech*, 2018, pp. 357–361.

[15] H. Ackermann, J. Konczak, and I. Hertrich, “The temporal control of repetitive articulatory movements in Parkinson’s disease,” *Brain Lang.*, vol. 96, no. 2, pp. 312–319, Feb. 1997.

[16] Y. Yunusova, J. Westbury, and G. Weismer, “Articulatory movements during vowels produced by speakers with dysarthria and normal controls,” *J. Acoust. Soc. Amer.*, vol. 117, no. 4, p. 2605, Apr. 2005.

[17] D. Kempler and D. V. Lancker, “Effect of speech task on intelligibility in dysarthria: A case study of Parkinson’s disease,” *Brain Lang.*, vol. 80, no. 3, pp. 449–464, Mar. 2002.

[18] J. Jeng, J. S. Laures, and G. Weismer, “Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders,” *J. Acoust. Soc. Amer.*, vol. 102, no. 5, p. 3094, Nov. 1997.

[19] T. Wang, C. Lian, J. Pan, Q. Yan, F. Zhu, M. L. Ng, L. Wang, and N. Yan, “Towards the speech features of mild cognitive impairment: Universal evidence from structured and unstructured connected speech of Chinese,” in *Proc. Interspeech*, Graz, Austria, Sep. 2019, pp. 1–12.

[20] Y.-T. Chen, B. E. Murdoch, and J. V. Goozee, “Electromagnetic articulographic assessment of tongue function in dysarthric and nondysarthric speakers following Stroke,” *J. Med. Speech-Lang. Pathol.*, vol. 16, pp. 21–33, Mar. 2008.

[21] E. Kearney, R. Giles, B. Haworth, P. Faloutos, M. Baljko, and Y. Yunusova, “Sentence-level movements in Parkinson’s Disease: Loud, clear, and slow speech,” *J. Speech Lang. Hearing Res.*, vol. 60, no. 12, pp. 3426–3440, Dec. 2017.

[22] P. Gómez-Vilda, J. Mekyska, J. M. Ferrández, D. Palacios-Alonso, A. Gómez-Rodellar, V. Rodellar-Biarge, G. Galaz, S. Zmekal, I. Eliaisova, M. Kostalova, and I. Rektorova, “Parkinson disease detection from speech articulation neuromechanics,” *Frontiers Neuroinform.*, vol. 11, p. 56, Aug. 2017.

[23] T. Artís-Vergara, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, “Parkinson’s disease and aging: Analysis of their effect in phonation and articulation of speech,” *Cognit. Comput.*, vol. 9, no. 6, pp. 731–748, Dec. 2017.

[24] J. Carmichael, “Dysarthria diagnosis via respiration and phonation,” in *Proc. Int. Conf. Workshop Comput. Commun. (IEMCON)*, Oct. 2015, pp. 1–5.

[25] A. Mayle, Z. Mou, R. Bunesco, S. Mirshekarian, L. Xu, and C. Liu, “Distinguishing dysarthria with long short-term memory networks,” in *Proc. Interspeech*, Graz, Austria, Sep. 2019, pp. 4514–4518.

[26] B. Schuller, S. Steidl, A. Batliner, E. Noth, A. Vinciarelli, and F. Burkhardt, “The INTERSPEECH 2012 speaker trait challenge,” in *Proc. Interspeech*, 2012, pp. 1–5.

[27] J. R. Orozco, “The interspeech 2015 computational paralinguistics challenge: Nativeness, Parkinson’s & eating condition,” in *Proc. Conf. Int. Speech Commun. Assoc.*, 2015, pp. 1–5.

[28] D. Lu and F. Sha, “Predicting likability of speakers with Gaussian processes,” presented at the Interspeech, Portland, OR, USA, 2012.

[29] D. Y. Huang, Y. Zhu, D. Wu, and R. Yu, “Detecting intelligibility by linear dimensionality reduction and normalized voice quality hierarchical features,” in *Proc. Interspeech*, 2012, pp. 1–4.

[30] J. R. Williamson, T. F. Quatieri, B. S. Helfer, J. Perricone, S. S. Ghosh, and G. Ciccarelli, “Segment-dependent dynamics in predicting Parkinson’s disease,” in *Proc. Interspeech*, 2015, pp. 66–351.

[31] S. Hahn and J. Wang, “Parkinson’s condition estimation using speech acoustic and inversely mapped articulatory data,” in *Proc. Annu. Conf. Int. Speech Commun. Assoc.*, 2015, pp. 277–280.

[32] N. P. Narendra and P. Alku, “Dysarthric speech classification from coded telephone speech using glottal features,” *Speech Commun.*, vol. 110, pp. 47–55, Jul. 2019.

[33] Kay Elemetrics Corp. *Disordered Voice Database (CD-ROM)*, Massachusetts Eye and Ear Infirmary, Boston, MA, USA, 1994.
