Word or Phoneme? To Optimise Prosodic Features to Predict Lung Function with Helicopter Task

Biao Zeng[1], Hok Shing Lau[2], Mark Huntly[2], Tim Bashford[2], Nathan Morgan[2]

Chelsea Williams[1], Lauren Game[1]

[1] Psychology Department, University of South Wales, Pontypridd, UK
[2] Wales Institute of Digital Information, University of Wales Trinty Saint David, Swansea, UK

Abstract. The study aimed to decide whether word-based or phoneme-based acoustic features could significantly correlate to prosodic and lung function measures. "Speech breathing" usually means producing the airflow required for phonation by utilising expired air and lung mechanics. Voice analysis as a health indicator has been extensively documented. The "helicopter task" is an articulatory test that measures the endurance of the respiratory system by having participants quickly repeat the word "helicopter" for three 20-second runs, separated by two 20-second silent, relaxed breathing intervals. Ten native English speakers' speech data that correlated with lung function measurements were used in the study. The study used ten native English speakers' speech data correlating to lung function measures. Specifically, both word-based ("helicopter") and phoneme-based (e.g., fricative consonant /h/, plosive consonant /k/, /p/, /t/) prosodic analysis was run to correlate to speech rate, word duration and lung function measures. Furthermore, the run effect on prosodic features at word and phoneme levels was investigated. The study found that, among nine phonemes in the word helicopter, /h/, /p/, /ɔ/ and /ə/ were significantly correlated with speech rate and word duration. In addition, it was found that the plosive phoneme /p/ duration became more variable in the third articulation run than in the first and second runs. It was explained that consonant /p/articulation change might reflect the taxed and exhausted respiratory system when the task was carried out. The consonant /p/ might be the best phoneme candidate to replace the whole helicopter to predict lung function measures.

Keywords: Speech Biomarkers, Speech Breathing, Helicopter Task, Lung Function.

1 Introduction

Speech breathing typically describes producing the airflow required for phonation using expired air and lung mechanics. [1]. A long, prolonged expiration during speech breathing follows a rapid inspiration. In a speaking exchange, prompt inspiration can shorten pauses and help a speaker keep their turn. To make speech sounds, air must first enter the lungs and then be forced out through the glottis. The different air volumes inhaled are determined by the type of contents to be produced.

Typically, the speech starts when inspiration stops and the lungs are full. Firstly, a large lung volume (LV) is linked to glottal leakage, higher subglottal pressures, higher sound pressure levels, longer voice onset times, and fundamental sound frequency. On the other hand, compared to high LVs, speech produced at low LVs has been linked to a more adducted vocal state. For example, Iwarsson, Thomasson, and Sundberg [2] examined how lung volume affected the glottal voice source and discovered that the closed quotient rose as lung volume decreased. Glottal leakage, peak-to-peak flow amplitude, and subglottal pressure, on the other hand, tended to decrease. Furthermore, Murray et al. coerced speakers to read passages in two different speaking voices: breathy vocals during the experimental phase and typical vocals during the baseline and return phases [3]. Compared to the baseline phase, they discovered that during the experimental phase, the participants spoke with longer LV excursions, marked by increased LV initiation and decreased LV termination.

Many studies investigate the potential of speech breathing to diagnose and predict lung function based on the airstream mechanism involved in speech breathing. Speaking assessments provide potential means of diagnosis and monitoring for certain populations, such as those suffering from Parkinson's disease and asthma. Healthcare professionals could estimate FEV1 (predicted forced expiratory volume in one second) using speech samples from asthma patients, according to Tayler et al. [4]. This discovery offers proof that acute asthma affects speech. Speech breathing has been investigated as a single biomarker in many fields recently, including COPD [5][6], asthma ([7][8]), and mental health ([9]).

Most speech breathing measures have been developed and invented for COPD. According to the NHS (2020), the primary signs of COPD include worsening dyspnea, a persistent cough that produces mucus, frequent chest infections, and persistent wheezing. COPD is one of the leading causes of death worldwide, and there is currently no known cure. Tobacco smoking is the primary cause of COPD, and projections indicate that COPD will rise to the third most common cause of death globally by 2030.

Researchers who are interested in speech breathing have focused on two areas. Many techniques and assignments were initially created to gather voice and speech samples. Rainbow Passage [10] is one commonly used technique for gathering data. It is a brief passage that speech therapists use in clinics to evaluate patients' vocal abilities. The first four lines of the passage capture every phoneme for American English as well as a large number of r-coloured vowels. It is full of alliteration, unusual consonant and vowel combinations and short and extended passages to test breathing and speech patterns [3][11].

Furthermore, scientists identify and refine speech characteristics associated with particular symptoms and illnesses. Advances in computerised deep learning techniques have provided novel approaches to speech modelling and analysis for healthcare applications [12]. For example, Nallanthighal et al. suggested analysing breathing using deep learning [13]. Speech production is primarily driven by three essential mechanisms: inhalation, expiration, and breathing. These authors investigated methods for detecting the breathing cycle and breathing metrics from speech using deep learning architectures. They talked about the difficulties in proving that using this technology would be beneficial. Respiratory parameters can be determined by estimating the breathing pattern from speech, enabling a speech-only evaluation of the speaker's respiratory health.

1.1 Helicopter Task

Zeng et al. [14] proposed a fast single-word ("helicopter") articulation task to test lung health. The diadochokinesis (DDK) task served as the basis for the design. DDK is one of the earliest and most popular tasks for assessing different speech communication issues. It frequently entails rapid word repetition or non-verbal oral motions like lip-opening and lip-closing. It also goes by the names verbal, oral, or phonoarticulatory DDK and has many variations. It has cross-disciplinary applications in biomedical engineering, biological sciences, communication sciences and disorders, computer methods in biomedicine, craniofacial surgery, dentistry, neurology and neurosciences, and oral surgery [15].

Producing speech is a complex sequential process involving memory, motor skills, and cognition. Compared to speech production, articulation is far more focused on the later motor process. Articulation can be considered an explicit motor event in the sequence, in addition to encoding in the motor cortex. During processing, speech breathing provides oxygen to the brain, memory, and motor components. It also improves articulation and may be a sign of lung health. Articulation events provide an explicit means of measuring lung function. It is necessary to create a single, pure articulation task to index lung functions as much as possible and reduce cognitive load and memory, both of which use oxygen during speech production.

As a result, we created the single-word articulation task and suggested several selection criteria for the target word, "helicopter." First, a list of words frequently used in speech therapy assessments was used to select the word "helicopter." The length of four syllables may create manageable articulation difficulties. Secondly, the task ought to exert maximal strain on the respiratory system. Helicopter is a target word with three plosive phonemes: The voiceless glottal fricatives, /k/, /p/, and /t/, and /h/is the voiceless glottal fricative. Combining them requires the greatest airflow to maintain articulation in contrast to the other consonants, mainly stops, nasals, or approximants.

Third, the target word should be high-frequency and retrieved from the mental lexicon with little effort. Such psycholinguistic characteristics, especially high frequency, will make articulating fluent and reduce the oxygen demand, which results in cognition and memory loads. For instance, "hippopotamus" and "helicopter" contain/h/ and /p/. Between the two words, "hippopotamus" is a low-frequency word and may prevent speakers from articulating fluently as it brings more cognition and memory loads. Therefore, "helicopter" was chosen as the word for our task.

1.2 Three-tier Features at Word and Phoneme Levels

The helicopter task is one articulation measure. As a word, "helicopter" can be analysed at syllable and phoneme levels. In linguistics, a syllable is a sequence of speech sounds (formed from vowels and consonants) organised into a single unit; a phoneme is the smallest phonetic unit in a language which can distinguish two words. For example, in the English word helicopter, the four syllables /he/, /lI /, /kɔp/ and /tə/ are composed of nine phonemes (i.e., /h/, /e/, /l/, /I/, /k/, /ɔ/, /p/ /t/, and /ə/). In the phonological encoding of the word, speakers parallelly activate phonemes and metrical frame (i.e., stress information) corresponding to the target word, followed by a serial association of phonemes to the metrical frame.

One type of breathing is speech breathing. It is a complex phenomenon that combines articulation, speech production, and breathing. Different kinds of information, such as acoustic, prosodic, and breathing, could be extracted from speech sounds by their resources. As a result, we would create a three-tier feature measure using a systematic analysis method. The prosodic, breathing, and acoustics are the three levels. The study of acoustics examines the physical characteristics of sounds and measures speech-related information such as vowel formants, intensity (interpreted as loudness), or fundamental frequency (interpreted as pitch). On the other hand, prosodic features deal with the organisation of speech sounds, such as length of run, pause ratio, etc. Particularly when it comes to speech breathing, the breathing features are underdeveloped. When at rest, people typically breathe ten to fifteen times per minute. We refer to this as the respiratory rate. We use respiratory rate as the primary breathing measure in the current investigation.

Numerous aspects of speech have been validated and examined in relation to health status. Two types of speech features—acoustic and prosodic information—were proposed by Farrús, Codina-Fibà, and Escudero and used for bipolar disorder detection and classification [9]. They contended that prosodic elements such as rhyme, stress, and intonation could represent an individual's emotional facets. This study concentrated on the prosodic information using acoustic and prosodic features.

A critical aspect of rhythm in this study was the exact measurement of the pause ratio. The respiratory underpinnings of spoken language were reviewed by Fuchs and Rochet-Capellan [16], who also emphasised the interactions between breathing, respiration, syntax, and planning. In a breathing cycle, we should distinguish between linguistic and respiratory pauses. A normal respiratory pause happens in the middle of a breath cycle. An inhalation and an exhalation occur in a typical breathing cycle when the body is at rest. An automatic pause, or period of no breathing, occurs for one to two seconds after the out-breath. On the other hand, a linguistic pause is either filled with um or uh or silent. Such a linguistic pause may function as a breath-in for speech breathing, igniting speech and bolstering the subsequent articulation. According to Zeng et al. [14], the pause ratio rose from the first to the third run, indicating a rise in breathing demand. These results demonstrated that the repeated articulation task quantitatively challenges speech articulation.

The present study adopted the "helicopter task" and extracted different characteristics. It aimed to identify a single phoneme rather than a whole word and investigate whether such phoneme features significantly predict lung function.

In the present study, two main research hypotheses were addressed:

1. Run effect: The later runs are assumed to have more breath-taking airflow due to the airflow consumed in the task. The pause ratio will significantly correlate to lung function measures in the later run(s).

2. Significant phoneme: As the helicopter word consists of nine phonemes, we predict some phonemes' prosodic features (duration) could be correlated to the whole

word's prosodic features. If we could identify these phonemes, we could develop phoneme-based predictors of lung function measures.

2 Method

2.1 Participants

Ten healthy, native English-speaking participants (M5: F5, mean age 51, range 42 – 65 years, height 171± 6.82 CM, weight 79 ± 12.38 Kg, n=10) were recruited through random sampling from the University of South Wales. All participants filled out a Clinical Report Form One, which consists of seven questions, including age, sex, height, weight, respiratory condition, smoking history and breathing status. No other general health status, medication or physical activities were investigated. The Faculty of Life Science and Education Ethics Panel approved the study, University of South Wales (No 210901HR), under the Declaration of Helsinki. Written and verbal informed consent were obtained from each participant.

2.2 Procedure

The subjects were instructed to remain motionless while their computers were placed about 40 centimetres away in a quiet room. The students were given a 20-second production time and instructed to produce the word "helicopter" as quickly and as many as possible. Every participant received three 20-second production intervals, separated by a forced 20-second break from the first to the second and the second to the third. The experimenters signalled the beginning of the production intervals and their breaks. The guidelines were as follows:

"For this exercise, you will be asked to repeat the word "helicopter," as before, but this time for only 20 seconds at a time. For this exercise, I want to do this three times but with a short break in between. When I say "go" please start repeating "helicopter" until I say "stop." After a short interval I will say "go," so like before keep going until I say "stop." After another short rest, I will ask you to do this one more time. I will time and record the whole exercise, so please sit quietly during the two resting intervals."

"Do you have any questions you would like to ask?"

"Are you ready to begin?"

The experimenter and the participants were recorded online using Microsoft Teams. The camera was turned off during the test, and only speech was recorded. Speech from participants alone was examined.

Using the audio editing program Audacity, the audio recordings were transformed into .wav files. The three 20-second runs were extracted and saved for acoustic analysis as separate files. The acoustic analysis only included instances that were fully spoken; disfluencies, loud background noise, and instances that overlapped with the experimenters' instructions were not included. The number and duration of each audible breath (inspiration) in each run were determined by identifying and matching them to the digital recording. All audio files had their word and pause boundaries manually segmented in VoxLab [17].

3 Results

Prosodic features., e.g., speech rate (SR, number of words per second for the entire run, word duration (WD, mean "helicopter" word duration in the entire run) and pause ratio (mean pause duration of the entire run) are reported below (Table 1-3). Predicted lung function was calculated using the Global Lung Function Initiative index (European Respiratory Society). Weight and height data predicted two measures: forced expired volume (FEV1) and forced vital capacity (FVC). Run (the repetition order in the task) is one independent variable in the study. Correlation analysis was run in three separate runs among phoneme durations, three prosodic features (speech rate, word duration and pause ratio), and three lung function measures (FEV1, FVC, and ratio of FEV1/FVC). All nine phonemes were input into the analysis. Tables 1, 2, and 3 present all significant correlations.

Initially, across all three runs, a significant negative correlation was found between the FEV1/FVC ratio and age, which was stable and consistent ($r= -.69$, $p<.05$). Another significant positive correlation was found between the 5th phoneme (/k/) duration and age ($r_1=.76$, $p<.01$; $r_2=.77$, $p<.01$; $r_3=.81$, $p<.05$), which indicates that the 5th phoneme (/k/) might be a sensitive phonetic mark for age.

Furthermore, the correlation analysis was run among individual phonemes, prosodic features, and lung functions. In the first run, for speech rate, two significant negative correlations were found in the 6th phoneme (/ɔ/, $r = -.64$, $p<.05$) and 9th (/ə/, $r=-.95$, $p<.01$). For the mean of syllable duration, three significant positive correlations were found in the 1st phoneme (/h/, r=71, $p<.05$), 7th (/p/, $r = .66$, $p<.05$) and 9th phoneme (/ə/, r=. 86, $p<.01$).

Table 1. Correlation matrix among phonemes, word, age and FEV1/FVC ratio in the first run.

	Age	SR	WD	FEV1/FVC
1st			.71*	
5th	.76*			
6th		-.64*		
7th			.66*	
8th				
9th		-.95**	.86**	
FEV1/FVC	-.69*			

*Phoneme-based measure: 1st phoneme = /h/, 2nd phoneme=/e/, 3rd phoneme=/l/, 4th phoneme= /I/,5th phoneme=/k/,6th phoneme=/ɔ/, 7th phoneme=/p/, 8th phoneme=/t/, and 9th phoneme= /ə/; word-based measure; SR=speech rate, MD = mean duration.

In the second run, for the speech rate, four significant negative correlations were found in the 1st phoneme (/h/, r= -.75, p<.05), 6th phoneme (/ɔ/, r=-.78, p<.05), 7th phoneme (/p/, r= -.88, p<.01) and 9th phoneme (/ə/, r= -.94, p<.01). For the mean of syllable duration, four significant negative correlations were found in the 1st phoneme (/h/, r= .80, p<.05), 6th phoneme (/ɔ/, r= .80, p<.05), 7th phoneme (/p/, r= .80, p<.01) and 9th phoneme (/ə/, r= .93, p<.05).

In addition, in the second run, we also found that the fourth phoneme (/i/) was significantly correlated with the FEV1/FEV ratio (r= -.65, p=.043); age was significantly correlated with the pause ratio (r= .68, *p*=.043) only in the second run.

Table 2. Correlation matrix among phoneme, word, age and FEV1/FVC ratio in the second run.

	Age	SR	WD	FEV1/FVC
1st		-.75*	.80*	
5th	.77**			-.698*
6th		-.78*	.80*	
7th		-.88**	.80**	
8th				
9th		-.94*	.93*	
FEV1/FVC	-.69*			

In the third run, for the speech rate, four significant negative correlations were found in the 1st phoneme (/h/, r= -.80, p<.01), 6th phoneme (/ɔ/, r=-.82, p<.05), 7th phoneme (/p/, r= -.83, p<.05) and 9th phoneme (/ə/, r= -.96, p<.01). For the mean of syllable duration, three significant negative correlations were found in the 1st phoneme (/h/, r= .84, p<.01), 7th phoneme (/p/, r= .75, p<.05) and 9th phoneme (/ə/, r= .93, p<.01).

In addition, both FEV1 and FVC were negatively correlated with the pause ratio (r=-.65, p=.041; r=-.65, p=.042). This indicated that a longer pause is correlated to weaker lung function. Such a pattern only emerges in a late third run.

Table 3. Correlation matrix among phoneme, word and FEV1/FVC ratio in the third run.

	Age	SR	WD	FEV1/FVC
1st		-.80**	.84**	
5th	.81*			
6th		-.82*		
7th		-.83*	.75*	
8th				
9th		-.96**	.95**	
FEV1/FVC	-.69*			

Based on the findings above, we summarise all single phoneme correlations to age, lung function, speech rate, and mean duration in each run. We found that two consonants (/h/ and /p/) and two vowels (/ɔ/ and /ə/) were correlated to prosodic features in all three runs. The results showed that, except for /h/, all phoneme contributors were in the late section of the word "helicopter." Compared to the other plosive phonemes /k/ and /t/, /p/ might be the most predictable consonant to the whole word.

Table 4. Summary of individual phoneme in word-based measure, lung functions and age.

	Age	Lung function	Run 1	Run 2	Run3
1st /h/			WD	WD, SR	WD, SR
2nd /e/					
3rd /l/					
4th /i/		√			
5th /k/	√				
6th /ɔ/			SR	WD, SR	WD, SR
7th /p/			WD	WD, SR	WD, SR
8th /t/					
9th /ə/			WD, SR	WD, SR	WD, SR

Furthermore, we investigate how the durations of /p/ change over the runs. We found that the standard deviation of the duration of phoneme /p/ and /t/ at the third run was significantly different to first run (p=0.007 for /p/ and 0.043 for /t/ respectively) and second run (p=0.025 for /p/ and 0.049 for /t/ respectively) while no such significance can be found between the first and second run (p=0.660 for /p/ and 0.852 for /t/ respectively). In other words, the duration of phonemes/p/ and /t/ become more varied as the run goes on, i.e., it is difficult for participants to keep the constant speech rate for specific phonemes as they become more exhausted throughout the runs.

4 Conclusion

In sum, we adopted the helicopter task to analyse one prosodic feature duration at word and phoneme levels—some paramilitary findings are reviewed. First, the run effect was observed in this study. In the third run, both FEV1 and FVC were significantly correlated with the pause ratio (r=-.65, p=.041; r=-.652, p=.042), which indicates that the third run was the most sensitive session for predicting lung function. This replicated our previous findings [14]. In addition, we also found that /p/ and /t/ durations in the third became more variant than in the first and second runs. Second. Between phoneme- and word-based prosodic features, there is a significant correlation between the second-half phonemes, e.g. consonants (/p/) and vowels (/ɔ/ and /ə/), and the whole word features, e.g. speech rate or word duration. Specifically, /p/ might be the best single phoneme predictor for prosodic features. More studies will be conducted to investigate other acoustic, prosodic, and breathing features on a single phoneme.

However, neither phoneme- nor word-based prosodic features were significantly correlated to lung function measures. (Except in the second run, the fourth phoneme (/i/) was significantly correlated with the FEV1/FEV ratio (r= -.648, *p*=.043). The mechanism between vowel and speech breathing is unclear so far.

In future studies, vowel and consonant articulation mechanisms could be investigated, especially the airflow allocation between speech components or articulatory manners. Furthermore, the speech breathing pattern from special groups, e.g. COPD or asthma, could be investigated to create the causation between speech breathing and lung functions. The current study offers a case to probe AI's contribution to a specific area: speech breathing. AI can significantly enhance processing, analysing, and deriving insights from massive speech-breathing data. By automating data processing, improving accuracy in detection, and providing real-time analysis and feedback, AI opens new avenues for research and practical applications in fields like speech therapy, linguistics, and health monitoring.

Acknowledgements.: This study was supported by the Wales Innovation Network.

Disclosure of Interests.: The authors have no competing interests to declare relevant to this article's content.

References

1. Winkworth, A. L., Davis, P. J., Adams, R. D., Ellis, E.: Breathing patterns during spontaneous speech. Journal of Speech, Language, and Hearing Research, **38**(1), 124-144 (1995)
2. Iwarsson, J.M., Thomasson, M. & Sundberg, J.: Effects of lung volume on the glottal voice source. Journal of Voice, **12**(4), 424-433(1998).
3. Murray, E.S.H., Michener, C.M., Enflo, L., Cler, G.J., Stepp, C.E.: The impact of glottal configuration on speech breathing. Journal of Voice, **32**(4), 420-427(2018).
4. Tayler, N., Grainge, C., Gove, K., Howarth, P., Holloway, J.: Clinical assessment of speech correlates well with lung function durnig induced bronchoconstriction. NPJ Primary Care Respiratory Medicine, **25**(1), 1-3(2015)
5. Tehrany, R.: Speech breathing patterns in health and chronic respiratory disease (Doctoral dissertation, University of Southampton) (2015)
6. Nallanthighal, V. S., Härmä, A., Strik, H.: Detection of COPD exacerbation from speech: comparison of acoustic features and deep learning based speech breathing models. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9097-9101). IEEE (2022)
7. Loudon, R. G., Lee, L., & Holcomb, B. J.: Volumes and breathing patterns during speech in healthy and asthmatic subjects. Journal of Speech, Language, and Hearing Research, **31**(2), 219-227 (1988)
8. Yadav, S., Gope, D., Maheswari, K. U., Ghosh, P. K.: Role of breath phase and breath boundaries for the classification between asthmatic and healthy subjects. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 870-873). IEEE (2021)

9. Farrús, M., Codina-Filbà, J. & Escudero, J.: Acoustic and prosodic information for home monitoring of bipolar disorder. Health Informatics Journal, **27**(1), 1460458220972755(2021)
10. Fairbanks, G.: Voice and articulation drill book, 2nd edn. New York: Harper & Row. pp124-139(1960).
11. Winkworth, A. L., Davis, P. J., Ellis, E., Adams, R. D.: Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. Journal of Speech, Language, and Hearing Research, **37**(3), 535-556(1994)
12. Cummins, N., Baird, A., Schuller, B. W.: Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. Methods, **151**, 41-54 (2018)
13. Nallanthighal, V. S., Mostaani, Z., Härmä, A., Strik, H., Magimai-Doss, M.: Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings. Neural Networks, **141**, 211-224(2021)
14. Zeng, B., Williams, E. M., Owen, C., Zhang, C., Davies, S. K., Evans, K., Preudhomme, S. R.: Exploring the acoustic and prosodic features of a lung-function-sensitive repeated-word speech articulation test. Frontiers in Psychology, **14**, 1167902 (2023)
15. Kent, R. D., Kim, Y., Chen, L. M.: Oral and laryngeal diadochokinesis across the life span: A scoping review of methods, reference data, and clinical applications. Journal of Speech, Language, and Hearing Research, **65**(2), 574-623 (2022)
16. Fuchs, S., Rochet-Capellan, A.: The respiratory foundations of spoken language. Annual Review of Linguistics, 7, 13-30 (2021)
17. Bashford, T., Lau, M. H. S., Huntly, M. M., Morgan, M. N., Iyenoma, M. A., Powell, T., Zeng, B.: AI Classification of Respiratory Illness through Vocal Biomarkers and A Bespoke Articulatory Speech Protocol. International Journal of Simulation--Systems, Science & Technology, 25(1) (2024).