Acta Logopaedica (2025), Vol. 2 | 121–134
DOI: 10.58986/al.2025.41035

Acta Logopaedica

ORIGINAL RESEARCH

Automatic evaluation of PaTaKa test using machine learning and audio signal processing

Birger Moëll, Fredrik Sand Aronsson

KTH Royal Institute of Technology (KTH) and Karolinska Institutet (KI), Stockholm, SwedenForskningsintresse: Machine learning for speech and language in a clinical setting.

***Corresponding authors:** Moëll B. and Aronsson, F.S., KTH Royal Institute of Technology (KTH) and Karolinska Institutet (KI), Stockholm, SwedenForskningsintresse: Machine learning for speech and language in a clinical setting. Emails: bmoell@kth.se, fredrik.sand@ki.se

Publication date: 8 August 2025

Abstract

This study presents an automated deep learning approach to evaluate the oral diadochokinesis, a widely used clinical tool for assessing syllable repetition speed in motor speech disorders. Addressing the limitations of manual assessments—including subjectivity, time constraints, and inter-rater variability—we developed a system leveraging the Wav2Vec2 speech recognition model, combined with audio preprocessing (resampling, mono conversion, and normalisation) and temporal alignment techniques for syllable detection. In an initial assessment of the developed method, the system was evaluated on 16 recordings from two healthy speakers, analysed by three speech and language pathologists (SLPs) and compared to ground truth measurements. Results demonstrated superior accuracy of the machine learning system, with a mean squared error (MSE) of 0.07, compared to 1.18 for human raters. Statistical analysis (Wilcoxon signed-rank test: $p = 0.98$ for model vs. $p = 0.00043$ for SLPs) confirmed the model's alignment with ground truth. While the system occasionally missed syllables (1–2 per recording), its precision in calculating syllables per second (SPS) and temporal consistency highlights its potential as a supplementary clinical tool. Key innovations include a user-friendly offline interface for data security and visualisations (Mel spectrograms, timing evenness, and distinctness metrics) to support clinical interpretation. The present study is subject to certain limitations. The study's methodology is constrained by a small and homogeneous sample. Separately, the developed system's performance is limited by unresolved challenges in the detection of subtle articulation errors. Future work will expand validation to diverse populations, including speakers with dysarthria, and refine human-in-the-loop integration to mitigate missed syllables. This study underscores the feasibility of combining deep learning with signal processing to enhance objectivity in speech assessments, offering a scalable solution to standardise the oral diadochokinesis test while preserving clinical expertise.

Keywords: PaTaKa test; oral diadochokinesis; speech motor control; deep learning; Wav2Vec2; automated assessment; clinical decision support

Sammanfattning

Denna studie presenterar en automatiserad metod för att utvärdera Pataka-testet, en central bedömning av stavelserepetitionshastighet vid motoriska talstörningar. Det föreslagna systemet använder Wav2Vec2-modellen för funktionsextraktion och transkription, kompletterat med ljudförbearbetning (omprovtagning, konvertering till mono, normalisering) för att förbättra datakonsistensen. Sexton ljudinspelningar, innehållande upprepade stavelser från två deltagare, analyserades av det automatiserade verktyget och av tre logopeder. Referensmätningar fastställdes med hjälp av mel-spektrogram för att identifiera stavelsernas gränser, vilket gav tillförlitliga referenspunkter för repetitionshastigheter. Resultaten visade att maskininlärningsmodellen presterade avsevärt bättre än manuella bedömningar: medelkvadratfelet (MSE) var 0,07 för modellen jämfört med 1,18 för de mänskliga utvärderingarna. Dessa resultat belyser de kliniska fördelarna med automatiserade verktyg för att förbättra diagnostisk precision och minska variabilitet i talbedömningar. Framtida arbete bör adressera systemets tendens att missa enstaka stavelser och utöka datasetet för att inkludera en bredare demografisk representation. Sammanfattningsvis visar studien att integrationen av djupinlärning och ljudsignalbehandling erbjuder ett skalbart, objektivt alternativ till traditionella bedömningar av Pataka-testet.

Keywords: PaTaKa-test; oral diadokinesis; talmotorisk kontroll; djupinlärning; Wav2Vec2; automatiserad bedömning; kliniskt beslutsstöd

Introduction

Analysing speech is essential for diagnosing and monitoring motor speech disorders. Speech and language pathologists (SLPs) commonly assess motor speech control using diadochokinetic (DDK) tasks, such as the ‘PaTaKa test’, where individuals rapidly repeat the syllables ‘pa’, ‘ta’, and ‘ka’ (Lancheros *et al.*, 2022). This task evaluates speech coordination, rhythm, and articulation, aiding in the assessment of motor speech disorders in conditions such as Parkinson’s disease (PD), amyotrophic lateral sclerosis (ALS), and multiple sclerosis (Ong *et al.*, 2024; Pinto *et al.*, 2024; Rong and Heidrick, 2021; Rozenstoks *et al.*, 2024). Traditionally, SLPs analyse DDK performance manually, counting repetitions over time. However, this method is susceptible to variability due to differences in training, perception of syllable boundaries and fatigue (Solomon *et al.*, 2021; Tanchip *et al.*, 2021). Automated analysis tools offer a promising alternative by providing objective, reliable, and efficient assessments, minimising inter-rater discrepancies, and enabling widespread clinical applications.

Machine learning (ML) techniques, particularly deep learning models, such as Wav2Vec2, leverage large-scale speech data to accurately transcribe phonemes and syllables, surpassing traditional methods in reliability and scalability (Baevski *et al.*, 2020). Prior studies have demonstrated the effectiveness of automated systems in detecting syllable boundaries and analysing speech patterns in clinical populations, using acoustic measures, such as oral diadochokinesis rate and articulatory precision (Pinto *et al.*, 2024; Rozenstoks *et al.*, 2024).

Previous work used computer tools to analyse speech in clinical settings. For example, studies have shown that visual representations of sound, such as Mel

https://doi.org/10.58986/al.2025.41035

spectrograms, and other signal-processing techniques can help to automatically identify syllable boundaries (Eyben *et al.*, 2015; Rudzicz, 2010). Automated systems have also shown potential for tracking speech changes in people with progressive neurological conditions (Hecker *et al.*, 2022). Studies on PD and ALS indicate that altered DDK rates and irregular syllable repetitions are sensitive markers of motor speech deterioration (Ong *et al.*, 2024; Tanchip *et al.*, 2021), showing the clinical relevance of automated speech analysis in tracking disease progression. Moreover, automated assessment methods incorporating spectrogram analysis and machine learning have successfully identified subtle speech impairments that may not be evident in perceptual evaluations (Lancheros *et al.*, 2022). However, there are still challenges, such as ensuring that these tools work well for people with different speech characteristics and addressing specific errors, like sometimes misidentifying syllables. Integrating these tools using a human-in-the-loop system (Kumar *et al.*, 2024), where humans make decisions guided by outputs from machine learning systems is a standard way to ensure human oversight.

The development of such a method, if validated, has the potential to change clinical practice by providing an objective scalable tool for evaluating speech function in individuals with neurological conditions. This study presents an automated method for analysing the PaTaKa test using Wav2Vec2 and advanced audio processing techniques. The aim is to make the first assessment of the accuracy and consistency of this automated system, compared to manual assessments performed by SLPs on a small dataset of healthy speakers.

Methodology

Participants and data

Recordings

Eight recordings were made each by the authors for a total of 16 audio recordings with variations in pace and pronunciation. The recordings were made with a high-quality microphone. Speaker 1 (B1–B8) had an average syllables per second of 6.59 with a median of 6.89, while speaker 2 (F1–F8) had a lower average of 5.02 and a median of 5.69. The standard deviation (SD) was higher for speaker 2 (2.24 vs. 1.55 for speaker 1), indicating greater variability in their speech rate. Speaker 1's syllable lengths ranged from 3.60 to 8.50, covering 4.90 units, whereas speaker 2's values spanned from 1.60 to 7.80, with a wider range of 6.20 units. Additionally, speaker 2 exhibited a larger interquartile range (3.76 vs. 1.96 for speaker 1), suggesting a less consistent speech pattern. For more information about individual audio samples, see Table 1.

Evaluation by speech and language pathologists

Three experienced SLPs assessed the audio recordings, measuring the number of syllables per second. The recordings were provided as separate Waveform (WAV) files, and the raters assessed the files independently of each other with the ability to playback using computer speakers or headphones. To ensure consistency with clinical practice, the raters followed the standard syllable-counting procedure used

Table 1. Average human error compared to ML system error on calculations of average syllables per second.

Clip	Average human	ML system	Ground truth	Human squared error	ML squared error
Bl	6.37	5.43	5.54	0.69	0.01
B2	5.10	3.33	3.23	3.51	0.01
B3	3.60	2.31	2.33	1.60	0.00
B4	8.13	8.37	8.40	0.07	0.00
BS	7.60	8.99	9.13	2.34	0.02
B6	8.50	7.04	6.86	2.70	0.03
B7	6.00	5.78	5.77	0.05	0.00
B8	7.40	7.27	6.95	0.20	0.10
Fl	2.30	2.00	1.85	0.21	0.02
F2	1.60	1.11	1.22	0.15	0.01
F3	3.25	3.32	3.33	0.01	0.00
F4	4.80	4.47	4.62	0.03	0.02
FS	6.60	6.72	6.60	0.00	0.01
F6	7.80	7.50	7.71	0.01	0.04
F7	7.27	9.17	8.25	0.97	0.85
F8	6.57	3.89	4.04	6.37	0.02

in settings where patient recordings are unavailable. However, no specific instructions were given, allowing them to use their preferred techniques. They were permitted to listen to the recordings as many times as needed. Lack of instruction could potentially be a source of bias but was chosen to have a naturalistic sample closer to clinical practice.

Ground truth

Ground truth was calculated by counting syllables and measuring the syllable-spoken duration during the audio file. By using both Mel spectrograms and audio files, and re-listening the audio clips for many times and comparing evaluations from two evaluators (FS and BM), we achieved a reliable ground truth measurement. The Mel spectrograms were used as an aid during ground truth calculations mainly for detecting the beginning and end of spoken utterances.

Error

The error was calculated as MSE of the difference between the ground truth syllables per second and the prediction of the model/SLPs. For instance, a prediction of 7.5 with a ground truth of 7 would give an error of 0.5×0.5 = 0.25. This error metric was used to standardise error direction as well as penalise larger errors more severely.

https://doi.org/10.58986/al.2025.41035

Model and tools

The system utilises Wav2Vec 2.0 XLS-R, a self-supervised learning model designed for multilingual and cross-lingual speech representation learning (Xu *et al.*, 2022). The model is pre-trained on a large corpus of unlabelled speech data across many languages, enabling it to generalise effectively to various speech tasks, including syllable detection in the PaTaKa test. Unlike traditional supervised models, Wav2Vec 2.0 XLS-R learns representations directly from raw audio waveforms, using contrastive learning to distinguish between similar and dissimilar speech segments (Xu *et al.*, 2022).

For this study, Wav2Vec 2.0 XLS-R was used without fine-tuning, relying on its pre-trained capabilities for speech-to-text transcription. The system was combined with a feature extraction and processing pipeline to enhance transcription accuracy and syllable segmentation

Audio preprocessing

To make sure the audio recordings are consistent and suitable for analysis, we performed the following steps:

- Resampling: We changed all audio files to a standard sampling rate of 16 kHz. This is like adjusting the 'resolution' of an audio to match what Wav2Vec2 expects.
- Stereo to mono conversion: If a recording had sound from two channels (stereo), we combined them into one channel (mono). This simplifies the audio data.
- Normalisation: We adjusted the loudness of the audio to a consistent level across all recordings.
- These steps help to minimise the effects of differences in recording quality and background noise, making the system more reliable.

Syllable detection

The model identifies syllables ('p', 't', and 'k') by analysing character-level transcription offsets provided by the Wav2Vec2 model. This detailed temporal alignment allows for precise segmentation and identification of syllable boundaries, a critical component in calculating repetition rates accurately.

Rate calculation

Syllables per second (SPS) is calculated as follows:
SPS = Number of syllables/Duration of syllable sequence.
This metric provides a direct measure of motor speech control, which is sensitive to both neurological conditions and treatment effects (Jain *et al.*, 2021).

Mel spectogram with syllables highlighted

For each recording, a Mel spectrogram was constructed with syllables highlighted (see Figure 1). A Mel spectrogram is a visual representation of sound that shows how its frequency content changes over time, using a scale that mimics human hearing sensitivity. Unlike a regular spectrogram, it emphasises lower frequencies where

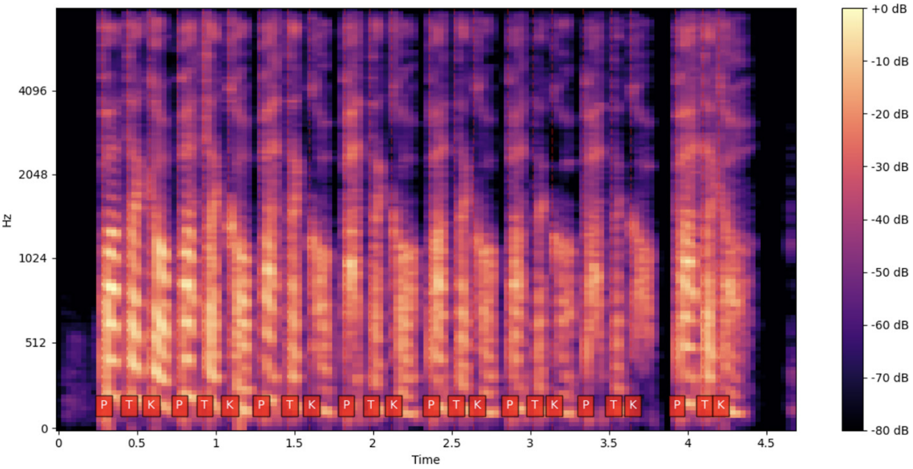


Figure 1. Mel spectrogram with highlighted syllables.

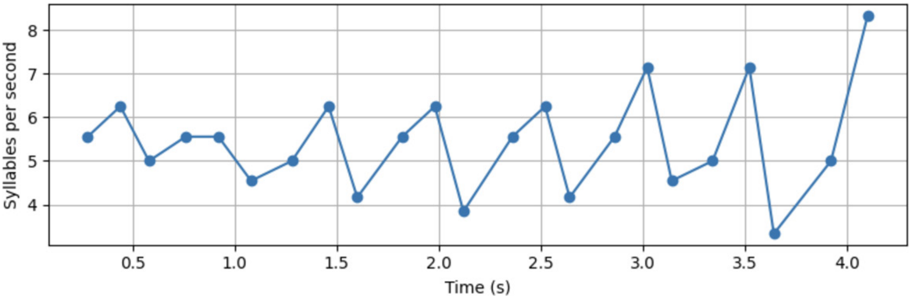


Figure 2. Syllables per second over time.

speech information is most concentrated, making it especially useful for analysing phonetic details. Speech-language pathologists use Mel spectrograms to examine speech patterns, detect articulation differences, and assess voice disorders.

Syllables per seconds

A plot is constructed that shows syllables per second over time (Figure 2). This plot helps to highlight changes in speaking patterns that can influence the rate of pronunciation.

Timing evenness analysis

Timing evenness analysis measures the consistency of timing between syllable repetitions by calculating the mean interval between repetitions for each syllable type. It shows the percentage deviation from the mean interval (Figure 3). The coefficient of variation (CV) is used to indicate timing regularity. A CV less than 0.1 signifies

https://doi.org/10.58986/al.2025.41035

highly regular timing; a CV between 0.1 and 0.3 denotes moderately regular timing; and a CV greater than 0.3 indicates irregular timing. The visualisation incorporates colour-coded zones to display acceptable range of variations.

Articulation distinctness analysis

In the articulation distinctness analysis, we used the probability assigned by the model for each syllable and compared it to the top probabilities of other syllables to measure the system's certainty (Figure 4). This generates a score of articulation based on the premise that a high likelihood of the model's prediction indicates proper pronunciation of that syllable. The analysis quantifies the clarity and precision of each

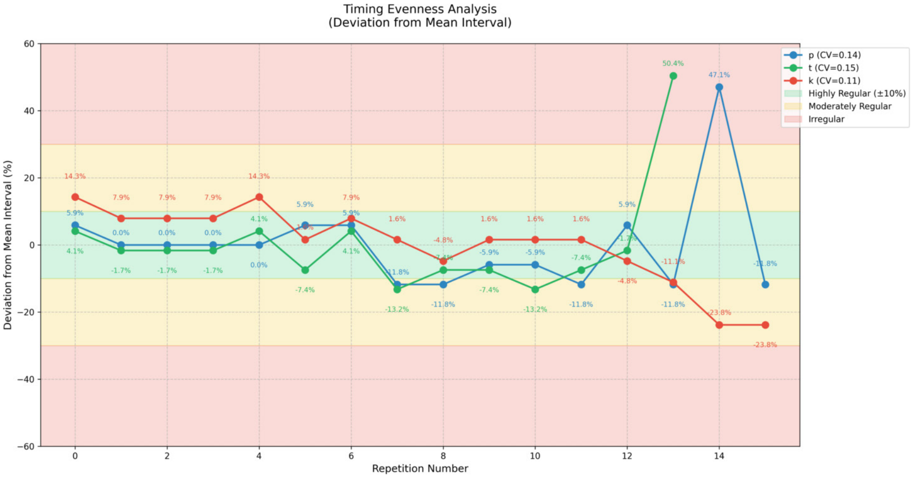


Figure 3. Timing evenness analysis.

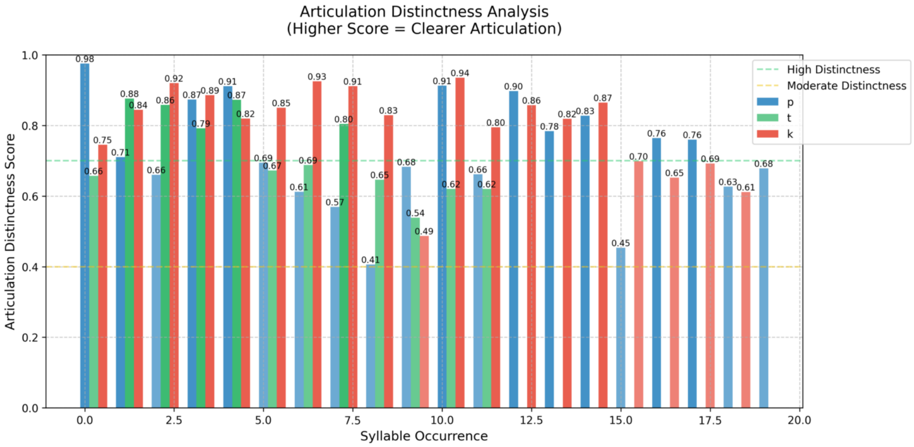


Figure 4. Articulation distinctness analysis.

syllable by measuring the model's confidence in detecting each syllable, with scores ranging from 0 (unclear) to 1 (very distinct). Reference levels are set as follows: greater than 0.7 indicates high distinctness; between 0.4 and 0.7 indicates moderate distinctness; and less than 0.4 indicates low distinctness. Colour-coded bars are used to show both syllable type and level of distinctness, helping to identify patterns of unclear articulation.

Energy analysis calculations

To analyse articulation strength, energy calculations are performed for each syllable. Measuring energy involves calculating the root mean square (RMS) energy for each syllable, which represents the intensity of articulation (Figure 5). Calculation of mean energy is done by determining the average energy across all occurrences of each syllable type. Assessing energy variation involves dividing the SD of energy by the mean energy, providing insight into the consistency of articulation strength. Determining percentage deviation involves calculating for each syllable, how much its energy deviates from the average. This approach helps to quantify articulation consistency and intensity across different syllables.

Programming

The tool is built in the Python programming language with the Streamlit library used for the interface.

Interface

The tool is available in a web interface where users can upload or record audio and results are shown and can be downloaded (Figure 6). The code is available open

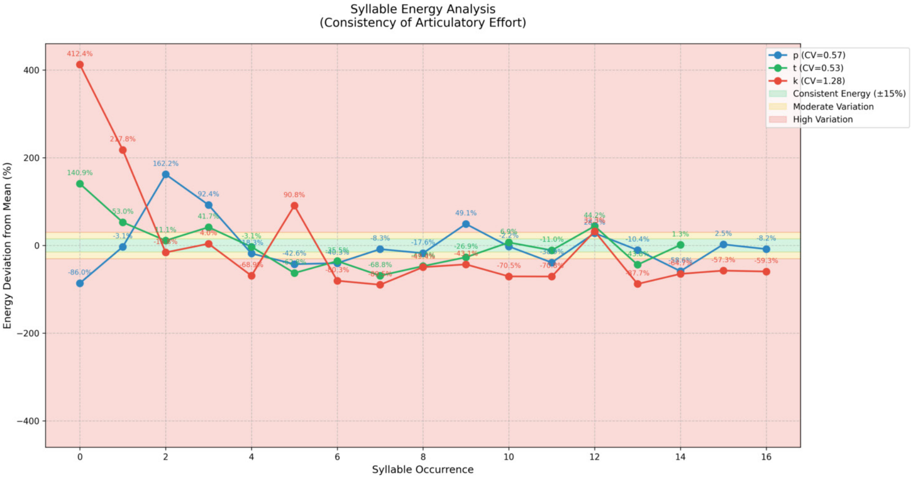
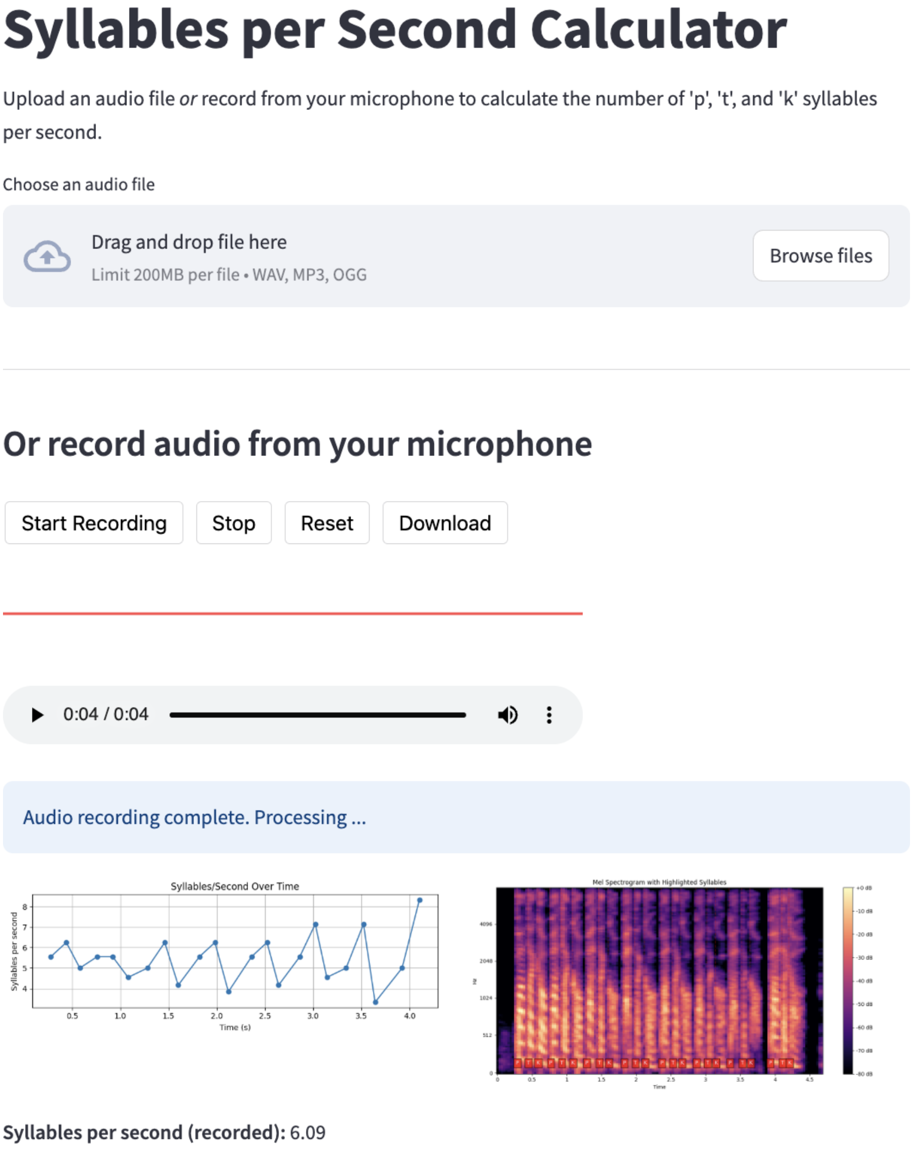


Figure 5. Syllable energy analysis.



Syllables per Second Calculator

Upload an audio file *or* record from your microphone to calculate the number of 'p', 't', and 'k' syllables per second.

Choose an audio file

Drag and drop file here

Limit 200MB per file • WAV, MP3, OGG

Or record audio from your microphone

Audio recording complete. Processing ...

Syllables per second (recorded): 6.09

Figure 6. Web interface.

source, and the system is designed to run completely offline. This is made possible by leveraging a graphical processing unit (GPU)—a specialised processor originally designed for graphics but now widely used for the intensive calculations required by machine learning models. By processing data on a local computer equipped with a GPU, the tool ensures that sensitive patient audio never leaves the clinical environment, thus complying with data protection regulations. While high-end GPUs were once limited to research settings, they are becoming increasingly common in standard desktop computers and even laptops. This trend suggests that running such artificial intelligence (AI) tools directly on device within the healthcare sector is a feasible and secure approach for the near future.

Since the tool can be run completely offline, it can be used to handle sensitive patient data. No patient data was used in the evaluation of this tool.

Statistical analysis

Statistical analyses were performed to compare the performance of the automated system with human raters relative to a ground truth.

The assumption of data normality was assessed using the Shapiro–Wilk test. Based on the outcome of this test, non-parametric analyses were chosen. To evaluate the difference between each assessment method (human and automated) and the ground truth, the Wilcoxon signed-rank test was employed. Subsequently, the Mann–Whitney U-test was used to directly compare the magnitude of deviations from the ground truth between the human rater group and the automated analysis group.

Results

Table 1 shows the ratings and errors for each recording, including ground truth, with Figure 7 highlighting the difference between average human, machine learning system, and ground truth. The average MSE for human raters was 1.18 versus 0.07 for machine learning model. This shows that machine learning model outperformed humans. The machine learning model's error was lower for all tasks, except one. For F5, the human squared error was 0, with the machine learning model squared error

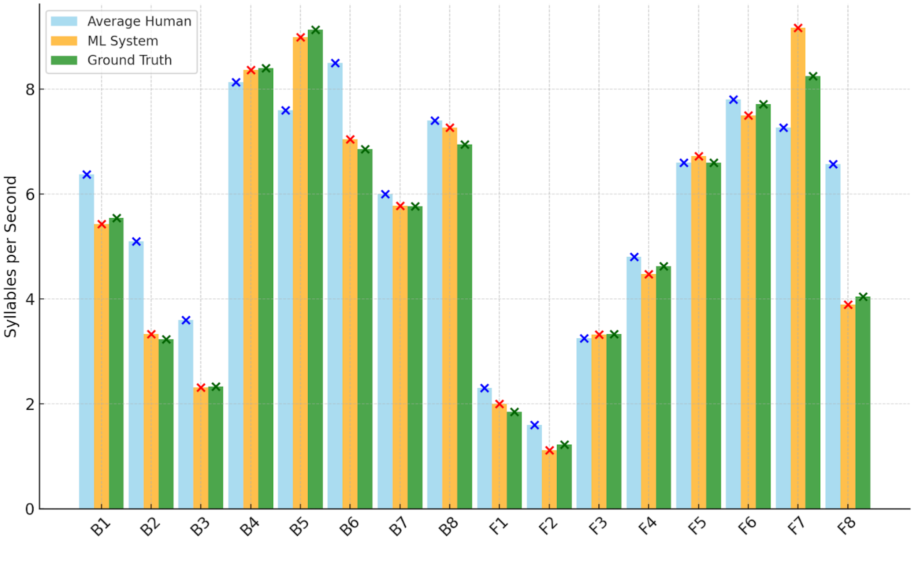


Figure 7. Comparison of syllable rate estimations.

https://doi.org/10.58986/al.2025.41035

at 0.01, a small difference. These results suggest that, under the chosen error definitions, the machine learning system's cumulative squared error is considerably lower than the human average error, indicating that the machine learning system exhibited lower overall deviation from the reference metric, compared to human evaluations for these specific recordings.

Comparison of automated and human assessment

The data from human raters was found to be not normally distributed (Shapiro–Wilk test, W = 0.72, $p < 0.001$).

The Wilcoxon signed-rank test, used to compare ratings with the ground truth, indicated a significant discrepancy for human raters (W = 253, $p = 0.00043$). In contrast, the automated analysis showed no significant difference from the ground truth (W = 253, $p = 0.98$).

A subsequent Mann–Whitney U-test confirmed a significant difference between the two groups, revealing that the deviations from the ground truth were markedly smaller for automated analysis than for human raters (U = 253, $p = 0.0000027$).

Errors made by the machine learning model

The machine learning model often missed a single syllable in the utterance. This can be seen in Figure 8, where the beginning P syllable is missing. In 12 recordings, one syllable was missed, one recording missed two syllables, and three recordings had all their syllables counted. As can be seen in Figure 8, the missing syllable in the beginning (P) is trivial for a human to catch and a human-in-the-loop system would be a great way to work with this tool. A human-in-the-loop approach would allow SLPs to look at annotated Mel spectrograms and listen to the recordings for the spectrogram. If they spot a missed or incorrectly identified syllable, the interface could

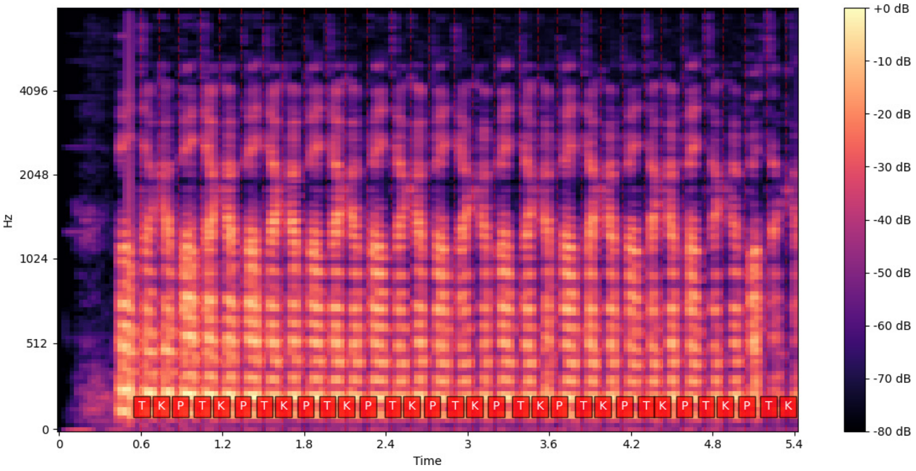


Figure 8. Mel spectogram with first P syllable missing.

https://doi.org/10.58986/al.2025.41035

allow them to simply click on the corresponding segment of the spectrogram to add, delete, or re-label a syllable. This correction would instantly update the overall syllables-per-second calculation, thereby combining the speed of automated analysis with the expert oversight needed for clinical accuracy. This would likely improve quality of assessment while reducing assessment time. Research has shown that similar human-in-the-loop approaches have worked well in other domains, with radiologists with AI support having high accuracy in breast cancer diagnostics evaluated in a study with over 50,000 participants (Hernström *et al*. 2025).

The PaTaKa test is a good place to start for AI assessment in speech pathology because it is a relatively simple test where the results of the test in isolation do not constitute clinical diagnostics. Over time, as both AI models improve and SLPs become comfortable with these tools, similar interfaces can be broadened to also do diagnostics of disorders based on speech assessment.

Notably, the accuracy of the tool is still high, because a single file contains many syllables (mean syllables: 29.31) and more importantly, the machine learning tool misses less syllables than humans, because the machine learning error is lower, compared to human evaluators.

Discussion

The findings from this study suggest that a machine learning approach leveraging Wav2Vec2 shows promise in assessing the PaTaKa test within a controlled, small-scale evaluation on healthy speakers, demonstrating higher consistency and lower error rates, compared to human raters. While these results are encouraging, the clinical implications—particularly regarding the detection of subtle speech changes in neurological conditions—require cautious interpretation.

We strongly recommend employing a human-in-the-loop approach with this tool to maximise the accuracy and reliability of assessments. Machine learning systems work best when seen as aids to clinical work and should be thought of as technical tools that augment the capabilities of clinical workers.

The integration of this tool into Swedish healthcare requires practical and technical considerations. While the offline interface ensures compliance with data security regulations (e.g. general data protection regulation [GDPR]), its adoption would require validation in Swedish clinical settings, where dialectal variations and multilingual speakers are common. A human-in-the-loop approach, where clinicians verify automated results, could mitigate missed syllables. For example, correcting one to two missed syllables per recording would likely take seconds, compared to manual counting (typically 1–2 minutes per recording), preserving time efficiency. However, this assumes, clinicians trust the system's output; training and iterative feedback mechanisms would be critical for adoption. Future work should also explore how visualisation tools (e.g. timing evenness metrics) could augment—rather than replace—clinical expertise.

Limitations

A key limitation of this study is the lack of specific standardised instructions for the SLPs performing manual assessments. While this approach was chosen to reflect the naturalistic variability of current clinical practice, it may have contributed to the higher inter-rater discrepancies observed and could be considered a source of bias.

The study's methodology, which relied on a limited dataset of 16 recordings from two healthy speakers, precludes broad generalisations to patient populations. The use of pre-trained Wav2Vec2 without fine-tuning on disordered speech may introduce biases in syllable detection accuracy. For instance, the model's occasional missed syllables (1–2 per recording) could reflect its training on typical speech patterns, potentially limiting its sensitivity to atypical articulation in dysarthria.

Conclusion

This study demonstrates the potential of a Wav2Vec2-based system to enhance the precision of PaTaKa test assessments in a controlled experimental setting with healthy speakers. However, the generalisability of these findings is constrained by the limited sample size, homogeneity of participants, and absence of pathological speech data. While the model's lower error rate (MSE = 0.07 vs. 1.18 for human raters) highlights its technical feasibility, clinical applicability remains speculative until validated across diverse populations, including speakers with dysarthria, apraxia, and neurodegenerative conditions. Key next steps include: (1) evaluating the system's robustness to speech disruptions (e.g. imprecise consonants and irregular rhythms), (2) optimising human–AI collaboration to balance efficiency and accuracy, and (3) assessing cross-linguistic performance in Swedish-speaking cohorts. These advancements could position automated tools as scalable adjuncts to—rather than replacements for—clinical expertise, particularly in resource-constrained settings.

References

Baevski, A., Zhou, Y., Mohamed, A. and Auli, M. 2020. Wav2Vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems* 33, 12449–12460. https://doi.org/10.48550/arXiv.2006.11477

Eyben, F., Wöllmer, M. and Schuller, B. 2015. openSMILE – The Munich versatile and fast open-source audio feature extractor. In MM'10 – *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 1459–1462. https://doi.org/10.1145/1873951.1874246

Hecker, P., Steckhan, N., Eyben, F., Schuller, B.W. and Arnrich, B. 2022. Voice analysis for neurological disorder recognition – A systematic review and perspective on emerging trends. *Frontiers in Digital Health* 4, 842301. https://doi.org/10.3389/fdgth.2022.842301

Hernström V, Josefsson V, Sartor H, Schmidt D, Larsson A-M, Hofvind S, ***et al.*** 2025. Screening performance and characteristics of breast cancer detected in the mammography screening with artificial intelligence trial (MASAI): A randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. *Lancet Digit Health* 7(3), e175–e183. https://doi.org/10.1016/S2589-7500(24)00267-X

Jain, A., Abedinpour, K., Polat, O., Çalışkan, M.M., Asaei, A., Pfister, F. M., … and Cernak, M. 2021. Voice analysis to differentiate the dopaminergic response in people with Parkinson's disease. *Frontiers in Human Neuroscience* 15, 667997. https://doi.org/10.3389/fnhum.2021.667997

Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S. K. and Sharma, R. 2024. Applications, challenges, and future directions of human-in-the-loop learning. *IEEE Access* 12, 75735–75760. https://doi.org/10.1109/ACCESS.2024.3401547

Lancheros, M., Pernon, M. and Laganaro, M. 2022. Is there a continuum between speech and other oromotor tasks? Evidence from motor speech disorders. *Aphasiology* 37(5), 715–734. https://doi.org/10.1080/02687038.2022.2038367

Ong, Y.Q., Lee, J., Chu, S.Y., Chai, S.C., Gan, K.B., Ibrahim, N.M. and Barlow, S.M. 2024. Oral-diadochokinesis between Parkinson's disease and neurotypical elderly among Malaysian-Malay speakers. *International Journal of Language & Communication Disorders* 59(5), 1701–1714. https://doi.org/10.1111/1460-6984.13025

Pinto, S., Cardoso, R., Atkinson-Clement, C., Guimarães, I., Sadat, J., Santos, H., Mercier, C., Carvalho, J., Cuartero, M.-C., Oliveira, P., Welby, P., Frota, S., Cavazzini, E., Vigário, M., Letanneux, A., Cruz, M., Brulefert, C., Desmoulins, M., Martins, I.P., Rothe-Neves, R., Viallet, F. and Ferreira, J.J. 2024. Do acoustic characteristics of dysarthria in people with Parkinson's disease differ across languages? *Journal of Speech, Language, and Hearing Research* 67, 2822–2841. https://doi.org/10.1044/2024_JSLHR-23-00525

Rong, P. and Heidrick, L. 2021. Spatiotemporal control of articulation during speech and speechlike tasks in amyotrophic lateral sclerosis. *American Journal of Speech-Language Pathology* 30(3S), 1382–1399. https://doi.org/10.1044/2020_AJSLP-20-00136

Rozenstoks, K., Novotny, M., Horakova, D. and Rusz, J. 2024. Automated assessment of oral diadochokinesis in multiple sclerosis using a neural network approach: Effect of different syllable repetition paradigms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28(1), 32–40. https://doi.org/10.1109/TNSRE.2019.2943064

Rudzicz, F. 2010. Articulatory knowledge in the recognition of dysarthric speech. *IEEE Transactions on Audio, Speech, and Language Processing* 19(4), 947–960. https://doi.org/10.1109/TASL.2010.2072499

Solomon, N.P., Brungart, D.S., Wince, J.R., Abramowitz, J.C., Eitel, M.M., Cohen, J., Lippa, S.M., Brickell, T.A., French, L. M. and Lange, R.T. 2021. Syllabic diadochokinesis in adults with and without traumatic brain injury: Severity, stability, and speech considerations. *American Journal of Speech-Language Pathology* 30(4), 1400–1409. https://doi.org/10.1044/2020_AJSLP-20-00158

Tanchip, C., Guarin, D.L., McKinlay, S., Barnett, C., Kalra, S., Genge, A., Korngut, L., Green, J.R., Berry, J., Zinman, L., Yadollahi, A., Abrahao, A. and Yunusova, Y. 2021. Validating automatic diadochokinesis analysis methods across dysarthria severity and syllable task in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research* 65(3), 940–953. https://doi.org/10.1044/2021_JSLHR-21-00503

Xu, Q., Baevski, A. and Auli, M. 2022. Simple and effective zero-shot cross-lingual phoneme recognition. In Interspeech 2022, 18–22 September 2022, Incheon, South Korea. ISCA. https://doi.org/10.21437/Interspeech.2022-60