

Original Paper

Automatic Speech Recognition and Acoustic Analysis for Dysarthria Assessment in Telerehabilitation: User-Centered Design and Usability Study

Pierre Vinet^{1, 2}, MSc; Pierre Dillenbourg³, Prof Dr; Amelie Slot⁴, MSc; Sharmila Selvanayakam¹, MSc; Sandra Giovanoli⁵, PhD; Elisa Du⁵, BSc; Julia Cardoso^{1, 4}, MSc; Meret Branscheidt^{4, 6}, Prof Dr; Chris Easthope Awai⁵, PhD; Christoph Michael Bauer¹, Prof Dr

¹Therapy Science Lab, Health, Lake Lucerne Institute, Vitznau, Switzerland

²School of Engineering, Section of Microtechniques, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

³School of Computer and Communication Sciences, Computer-Human Interaction Lab for Learning & Instruction, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

⁴Center for Neurorehabilitation, Cereneo, Vitznau, Switzerland

⁵Data Analytics & Rehabilitation Technology (DART), Health, Lake Lucerne Institute, Vitznau, Switzerland

⁶Lehre HEST, D-HEST, ETH Zurich, Zürich, Switzerland

Corresponding Author:

Christoph Michael Bauer, Prof Dr
Therapy Science Lab, Health
Lake Lucerne Institute
Rubistrasse 9
Vitznau 6354
Switzerland
Phone: 41 79 5272449
Email: christoph.bauer@llui.org

Abstract

Background: Dysarthria is a frequent motor speech disorder following a stroke, affecting up to 42% of survivors and resulting in reduced speech intelligibility and diminished quality of life. Clinical assessments, such as the Frenchay Dysarthria Assessment, Second Edition (FDA-2), rely heavily on the subjective judgment of speech-language pathologists (SLPs), which limits comparability and scalability. Telepractice solutions have the potential to extend access to care, but validated digital tools that combine automatic analysis with clinically usable interfaces remain scarce.

Objective: This study aimed to develop and evaluate a web-based application that integrates automatic speech recognition (ASR) and acoustic analysis into a user-centered dashboard for SLPs. Specifically, we investigated: (1) whether ASR can provide intelligibility scores comparable to those of human listeners; (2) the usability of the system in 2 iterative cycles with SLPs; and (3) the feasibility of presenting clinically relevant acoustic features to support telerehabilitation.

Methods: A user-centered design process was followed, involving contextual inquiry, requirements gathering, prototype development, and iterative testing with SLPs. The analytical core of the prototype included an ASR module (Whisper Large-v3) to compute intelligibility scores, combining word error rate-based accuracy with sentence-level and word-level alignment. Phoneme-level error highlighting was implemented to identify frequent substitution or deletion patterns. In parallel, an acoustic module extracted clinically relevant measures, including fundamental frequency (mean and range), intensity (mean and variability), and vowel formants (F1–F2 space), supplemented by sustained phonation duration. A pilot validation compared ASR-based intelligibility scores with transcriptions from 8 lay listeners for 3 patients with dysarthria performing the Frenchay Dysarthria Assessment–2 word and sentence tasks. Usability was evaluated in 2 cycles with 8 and 4 SLPs, respectively, using the System Usability Scale and structured questionnaires.

Results: In the pilot validation, ASR performance was comparable to, and in some cases better than, untrained human listeners for individuals with mild and moderate dysarthria, though performance declined with severe cases. Both usability cycles yielded excellent System Usability Scale scores (cycle 1: mean 88.4, SD 4.6; cycle 2: mean 91.7, SD 4.1). Core workflow elements, including navigation, session upload, and intelligibility score presentation, were consistently rated highly.

Feedback evolved from bug reports and requests for clearer terminology in cycle 1 to suggestions for advanced analytic features in cycle 2, such as additional voice-quality indices and integrated note-taking.

Conclusions: The prototype demonstrates that automatic intelligibility scoring and acoustic analysis can be integrated into a clinically usable, web-based dashboard. While current limitations include reliance on English-only phoneme analysis, limited advanced acoustic features, and lack of regulatory compliance, the application achieved excellent usability and shows promise for scalable telerehabilitation. Future work should expand multilingual support, incorporate additional acoustic measures, and validate the tool in larger clinical cohorts.

JMIR Form Res 2026;10:e85230; doi: [10.2196/85230](https://doi.org/10.2196/85230)

Keywords: speech and language therapy; user-centered design; telerehabilitation; dysarthria; web application; automatic speech recognition

Introduction

Dysarthria is a neuromotor speech disorder resulting from neurological damage and is present across many neurological diseases, including stroke, cerebral palsy, and amyotrophic lateral sclerosis [1,2]. It affects the speed, strength, accuracy, range, tone, or duration of the movements required for speech control and commonly reduces speech intelligibility [1], substantially impacting participation, psychosocial well-being, and quality of life [3,4]. Clinical assessment tools, such as the Frenchay Dysarthria Assessment, Second Edition (FDA-2), one of the most widely used assessments across different clinical systems [5], are widely used but rely heavily on subjective judgment, leading to interrater and intrarater variability in some items [6-8]. This creates variability and limits comparability across raters and time.

Speech-language pathologists (SLPs) play a vital role in diagnosing and treating dysarthria [9], yet their reach is constrained by time, geography, and patient accessibility. Telerehabilitation, which involves delivering rehabilitation services remotely, has emerged as a scalable, clinically endorsed solution to bridge this gap, with professional guidance and reimbursement pathways increasingly being established [9-12]. By enabling remote assessment and treatment, telerehabilitation can extend therapeutic support to underserved regions and improve continuity of care for individuals who experienced stroke.

Integrating acoustic analysis based on automatic speech recognition (ASR) into telerehabilitation platforms opens the possibility of continuous, noninvasive speech monitoring in naturalistic settings and timely feedback to patients. Advancements in signal processing and large self-supervised “audio or language” models now enable robust extraction of acoustic features (eg, fundamental frequency, intensity range, and speaking rate) and interpretation of speech. Such measures serve as quantifiable indicators of speech quality and progression, specifically the severity of dysarthria, providing data-driven support to SLPs and researchers during evaluation [13-15]. Modern representation-learning approaches [16] further strengthen automated analyses and downstream assessment pipelines [17-19]. However, the integration of these advances into clinical practice has not kept pace, creating a mismatch between technological potential and clinical uptake that motivates this study. Despite these advances, adoption in routine care remains limited.

Many available tools require installing desktop software and substantial technical training, have limited usability, fit poorly within clinical workflows, and lack standardized, shareable digital outcome measures. Furthermore, patient privacy needs and data protection concerns must be addressed. As a result, human perceptual judgment continues to serve as the de facto reference standard [17,20,21].

This study addresses this gap by developing a web-based tool that integrates ASR and acoustic analysis. While initially designed with stroke-related dysarthria in mind [22], the underlying analytical framework is intended to be generalizable across dysarthria etiologies. Specifically, we examined: (1) whether ASR can achieve comparable performance to human listeners in intelligibility assessment; (2) the usability of the tool for SLPs in a formative evaluation; and (3) implications for telerehabilitation and long-term monitoring.

Methods

Study Design

A user-centered design approach guided the project from early requirements gathering to final prototype validation [23, 24]. Initial requirements were collected before the frontend design to match the SLPs’ requirements for a web-based digitized speech and language assessment [22]. This study consisted of a three-step approach: (1) needs and context analysis including specifying the context of use and gathering user requirements; (2) preliminary testing; and (3) end-user testing and refinement. The 2 empirical components—prototype development and preliminary testing, and end-user testing and refinement—are reported in the “Results section.” Two iterative usability testing cycles were conducted to identify and address initial usability issues and to evaluate the effectiveness of implemented improvements in a refined prototype.

Ethical Considerations

Patient data were obtained from patients with stroke and the publicly available TORGO database, a database that includes individuals with cerebral palsy and amyotrophic lateral sclerosis, which permits academic use [25]. All procedures adhered to ethical guidelines and regulations. Approvals were obtained from the Ethics Committee of Northwestern and Central Switzerland (Req-2024-00103 for usability testing, R-2025-00538 for voice recordings). Written

informed consent was obtained from all participants. The participants did not receive compensation.

User-Centered Design

Needs and Context Analysis

Needs Assessment and Context Analysis

The iSpeak system was initially developed for SLPs conducting assessments of dysarthria in individuals who experienced stroke [22]. To capture dysarthria across neurological conditions more broadly, this study used the TORGO database [25]. Nine SLPs were recruited, and participant demographics are summarized in Table 1. A purposive sampling method was used, based on the assumption that each participant would offer unique insights [26]. Because the participants' roles were not interchangeable, the sample size was guided by data saturation rather than statistical power analysis [26]. The SLPs contributed through preliminary meetings, focus groups, prototype testing, and feedback.

The following 2 methods informed contextual understanding:

- Work shadowing sessions (n=6, 30-45 min each), including observation of live telerehabilitation sessions and mock sessions where 1 researcher acted as the patient, revealed workflow challenges and opportunities for automation. Field notes from the work shadowing sessions were summarized descriptively and reviewed to identify recurring workflow steps, contextual constraints, and opportunities for automation using Microsoft Excel.
- Two focus group interviews with a total of 6 SLPs, who practiced telerehabilitation with a mean experience of 4.6 years, were conducted to explore assessment practices, workflow bottlenecks, and requirements for automated tools. Focus group responses were summarized descriptively by the research team and reviewed to identify recurring needs, workflow barriers, and potential requirements for the prototype; no dedicated qualitative analysis software was used for this step. During the focus groups, the SLPs emphasized the need for asynchronous use. Typical practice involved conducting a live therapy session via videoconferencing, followed by posthoc review and analysis using the application.

Table 1. Overview of participants in study phases.

Phase	Participants	Region	sex (female), n	Work experience (y), mean (SD; range)
Needs and context analysis				
Work shadowing	3 patients and 3 mock sessions	Europe (4)–Arabic Peninsula (1)	3	— ^a
Focus group interviews	6 SLPs ^b	Europe (4)–Arabic Peninsula (1)	5	4.6 (2.1; 3-6)
User requirements	6 SLPs	Europe (4)–Arabic Peninsula (1)	5	4.6 (2.1; 3-6)
Preliminary testing				
Pilot validation	8 lay users	Europe (8)	5	—
End-user testing and refinement				
Usability study cycle 1	8 SLPs	Europe (7)–United States (1)	8	7.56 (3.2; 3-26)
Usability study cycle 2	5 SLPs	Europe (4)–Arabic Peninsula (1)	5	4.3 (2.0; 3-6)

^aNot applicable

^bSLP: speech and language pathologist.

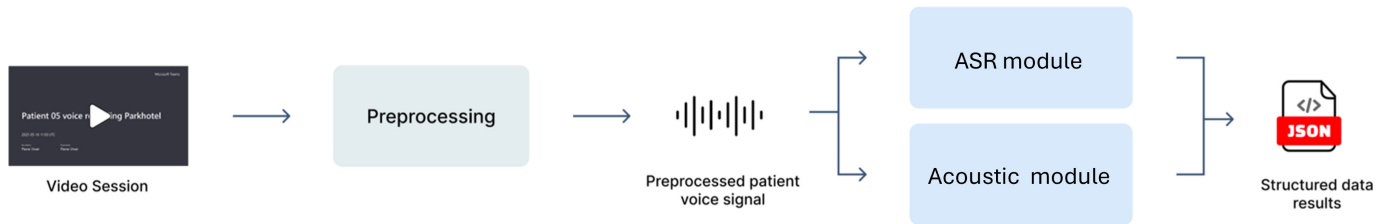
User Requirements

Requirements were identified from observations and focus groups, summarized by the research team, and classified into functional and nonfunctional categories (Tables S6-S8 in Multimedia Appendix 1). The requirements were then prioritized according to feasibility and perceived importance for SLPs using Microsoft Excel; functional requirements described expected features (eg, review of recordings, intelligibility scoring, and progress tracking), while nonfunctional requirements addressed qualities such as speed, reliability, and security. Prioritization considered both feasibility and user importance. These requirements informed prototype design and evaluation (Tables S6-S8 in Multimedia Appendix 1).

Design and Prototype Development

The prototype design was created in Figma using Shaden component libraries to ensure consistency and rapid iteration. Designs were validated with SLPs before implementation. The final prototype comprised a backend (Python, Python Software Foundation; FastAPI, Sebastián Ramírez [known online as @tiangolo]; PostgreSQL, PostgreSQL Global Development Group) hosted on a Swiss VPS, and a front-end (Next.js, Vercel Inc; Tailwind CSS, Tailwind Labs Inc) deployed via Vercel, Vercel Inc. These implementation details are provided in Multimedia Appendix 1. Figure 1 illustrates the prototype's backend modules for the analysis end point. The prototype's style guide is illustrated in Figure S6 in Multimedia Appendix 1.

Figure 1. Overview of the backend modules for the analysis end point. ASR: automatic speech recognition.



Evaluation of the Prototype

The prototype was evaluated in 2 stages: prototype development and preliminary testing with lay users to detect usability issues and bugs, as well as the technical validity of ASR-based intelligibility scoring; and end-user testing and refinement with SLPs across 2 iterative cycles, during which SLPs evaluated usability, workflow integration, and clinical relevance.

Preliminary Testing

Local community members familiar with laptops explored the prototype freely, uploading recordings and navigating the dashboard. A mock session ensured comparability. Minor issues (eg, broken links and unclear navigation) were corrected before expert testing.

Participants

Eight lay listeners were recruited from the faculty staff. None had training in speech-language pathology, but all possessed advanced English proficiency. They were selected as proxies for naïve intelligibility judgments, which are commonly used in dysarthria research to approximate real-world listener understanding rather than relying on expert clinical ratings [27]. Three patients with stroke with varying dysarthria severity were recruited from a neurorehabilitation hospital and provided speech material. Each had completed the FDA-2 reading tasks (10 words and 10 sentences) [5].

Materials and Procedure

Recordings were processed using OpenAI Whisper (Large-v3), configured for English with a temperature of 0.0 [28]. Participants listened via a laptop and transcribed the speech without access to reference texts. Both lay transcriptions and ASR output were normalized using the same procedures and compared against reference texts. Performance was quantified by word error rate (WER), with lower WER indicating higher intelligibility. The results were visualized using boxplots.

End-User Testing and Refinement

Usability Testing and Refinement

Overview

Two structured cycles followed a standardized protocol. Each participant received access credentials, a user manual with screenshots, and sample patient data (from TORGO). Usability was assessed through a questionnaire that included open-ended questions, Likert scales, and the System Usability Scale (SUS) [29]. Anonymous web analytics complemented self-reports by capturing interaction patterns

and navigation behavior. Likert-scale questionnaire items and SUS scores were analyzed descriptively by calculating item-level and overall mean scores. Open-ended questionnaire responses were summarized and categorized according to their content using Microsoft Excel. Anonymous web analytics were inspected descriptively to identify navigation patterns, task completion issues, and dead clicks. After each cycle, open-ended feedback and observed usability issues were categorized into bugs, improvements, and feature requests, which were used to guide the next refinement step. Critical issues were fixed before the next cycle. The analytical pipeline was designed to be condition-agnostic, operating on speech signal characteristics rather than disease-specific features, thereby supporting potential generalization across different dysarthria etiologies.

Results of the 2 cycles, including SUS scores and qualitative feedback, are presented in the “Results” section.

Participants

Eight SLPs who practiced telerehabilitation were recruited through an established professional network to participate in 2 usability cycles, with a mean of 7.57 years of clinical practice. The SLPs used the application at their workplaces or homes. The SLPs received audio recordings and reference transcripts from the TORGO database [25] to ensure compliance with health care patient data regulations while providing realistic clinical testing scenarios.

The Prototype

Preprocessing

Recordings were preprocessed by extracting audio, applying a band-pass filter (80 Hz-8 kHz), normalizing the signal amplitude, and resampling to 16 kHz mono. These steps ensured consistency across sessions and compliance with model requirements [28].

Transcription and Normalization

Whisper Large-v3 produced transcriptions, which, together with reference texts, were normalized (lowercased, punctuation removed, numbers expanded, and contractions expanded) to ensure scoring reflected content rather than formatting [30, 31].

Error Metrics

WER and character error rate were calculated using the standard edit-distance (Levenshtein) algorithm. WER is defined as the sum of substitutions, deletions, and insertions divided by the total number of reference words. Character

error rate follows the same logic at the character level. Detailed equations are provided in [Multimedia Appendix 1](#) [31].

Alignment

To compute sentence-level scores, a global alignment between concatenated reference and transcription strings was performed. Insertions, deletions, and substitutions were marked, and then boundaries were adjusted back to individual sentences. This allowed for per-sentence WER and alignment visualizations highlighting correctly and incorrectly recognized words.

Intelligibility Score

Intelligibility refers to how much of a speaker's intended message is understood by a listener and is commonly operationalized in dysarthria research through listener transcription accuracy at the word or sentence level [13,27,32,33]. In FDA-2 scoring, each item is judged as correct or incorrect. We implemented two scoring modes:

- Binary scoring (FDA-2 standard) [5]: words and sentences are scored as correct or incorrect.
- Word-level scoring: sentence scores are calculated from the average word-level accuracy.

The final intelligibility score (IS) was defined as follows:

$$IS = 1 - WER_{avg}(1)$$

Where WER_{avg} is the average WER across tasks. Global scores combine word and sentence results as follows:

$$IS_{global} = \frac{N_w * IS_w + N_s * IS_s}{N_w + N_s}(2)$$

where IS_w and IS_s are word and sentence intelligibility scores, and N_w and N_s are their respective counts.

Phoneme-Level Analysis

Phonemes are the smallest contrastive speech sound units that can distinguish meaning within a language; phoneme-level analysis, therefore, focuses on the sound structure of words rather than on the listener's global understanding of the message [34]. Reference words were converted to phoneme sequences using CMUdict and mapped to the International

Phonetic Alphabet. Errors were classified as substitutions or deletions, aggregated, and ranked by frequency and word position. This analysis provided SLPs with clinically relevant insights for tailoring therapy.

Acoustic Feature Extraction

Acoustic features were extracted to provide objective, interpretable measures of speech. Based on clinical input, 3 families were prioritized as follows [35-37]:

- Fundamental frequency (F0): mean, range, and variability.
- Intensity: mean and SD.
- Formants (F1, F2): vowel-space measures that indicate articulatory precision.

Maximum phonation time was derived from sustained /a/ tasks. These features capture prosodic control, respiratory support, and articulation. Sex-related differences in vocal tract anatomy, which influence formant frequencies (F1 and F2), were not controlled for in this study. Extraction was performed using Parselmouth, a Python interface to Praat, integrated into the backend pipeline.

Statistical Methods Used for Analyzing Acoustic Features

The extracted acoustic features were computed using Parselmouth and summarized descriptively to support clinical interpretation. No inferential statistical analyses were performed on these features, as the study focused on the feasibility and usability of the analytical pipeline rather than a hypothesis-driven evaluation of acoustic measures.

Dashboard and Visualization

The web application dashboard was designed to present results clearly to SLPs. Key components included (1) circular progress indicators showing intelligibility percentages; (2) bar charts of phoneme error frequencies, color-coded by word position; and (3) side-by-side alignment displays highlighting correctly recognized (green) versus incorrect (red) words.

Figures 2–4 illustrate the interface components and analytics dashboards.

Figure 2. Analysis and segment selection (of 1 TORGO database patient). FDA2: Frenchay Dysarthria Assessment, Second Edition.

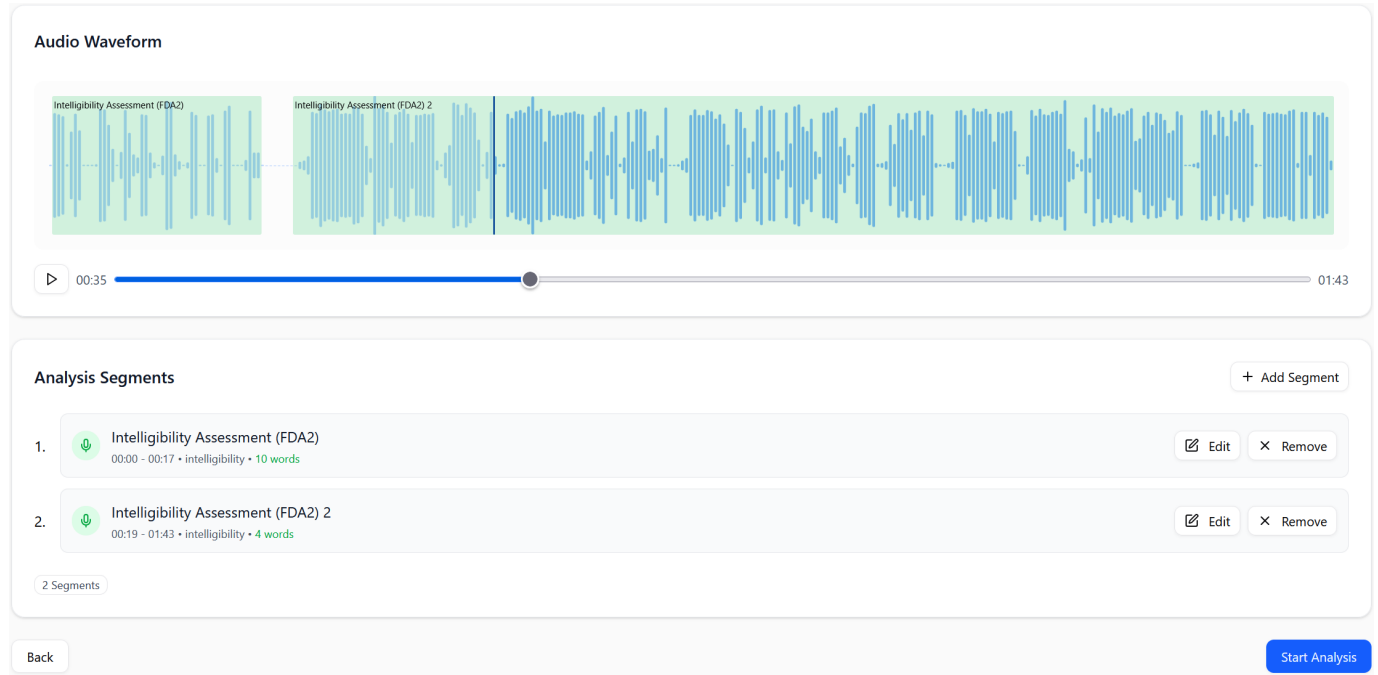


Figure 3. Intelligibility score and phonemes analysis (left) detailed phonemes analysis page (right) of a patient from the TORGO database.



Figure 4. Acoustic analysis: acoustic base frequency (top), acoustic intensity (middle), acoustic formants (bottom) of a patient from the TORGO database. F0: acoustic base frequency, F1-F5: acoustic formants.



Results

Table 1 provides an overview of the study phases and participant demographics. All SLPs who took part in the study practice telerehabilitation.

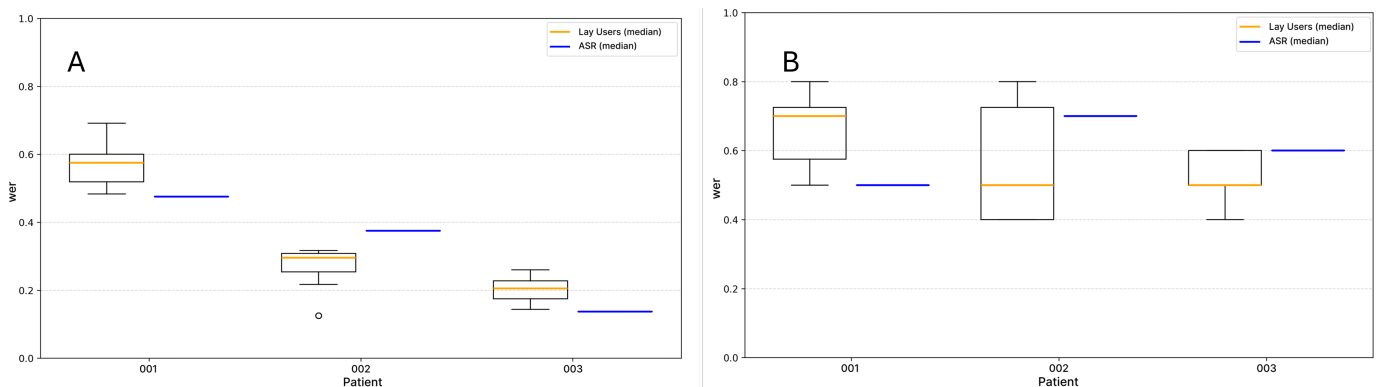
Preliminary Testing

Eight lay users (Table 1) transcribed audio recordings from 3 patients with dysarthria to validate the accuracy of the automatic intelligibility assessment. The recordings included the FDA-2 words and sentences tasks. Performance was quantified using WER, with lower values indicating greater accuracy.

The ASR system (Whisper Large-v3) performed comparably to, and in some cases better than, untrained listeners. For mild dysarthria, ASR achieved lower median WERs than most participants. For moderate and severe dysarthria, ASR errors increased but remained within the variability range of human listeners. Boxplots of WER distributions are shown in Figure 5.

These findings confirm that ASR can approximate human perceptual judgments of intelligibility and support its integration into the prototype.

Figure 5. Word error rate (WER) comparison between lay user transcribers and ASR transcriptions for the word (A) and sentences (B) tasks; ASR: automatic speech recognition system.



End-User Testing and Refinement-Cycle 1

Eight SLPs participated in the first usability cycle (mean clinical experience 7.56, SD 3.2 y; Table 1). Most accessed the application at the workplace, while 2 participated from home, using a mix of Chrome, Firefox, and Edge.

The mean SUS score was 88.4 (SD 4.6), which is considered excellent [38,39]. Average ratings across survey items were 4.26/5 (Table 2). Core workflow elements—including navigation, uploading sessions, segment selection, and processing time—scored between 4.50 and 4.63. Visual design received the highest rating (mean 4.88, SD 0.35). Understanding the intelligibility outputs scored slightly lower: the side-by-side transcription view was rated 4.50, and the

average intelligibility score clarity was 4.13 (SD 0.13). The lowest ratings were for the perceived accuracy of the automatic intelligibility score (mean 3.13, SD 1.25) and phoneme-level error highlighting (mean 3.88, SD 1.25). Qualitative feedback (Table 3; Table S9 in Multimedia Appendix 1) clustered into 3 categories:

- Fixes or bugs: reliability and alignment issues (failed uploads, shifting regions, and missed words).
- Usability improvements: clearer wording in the user interface, simultaneous listening and text entry, and time markers.
- Feature requests: phonetic equivalence in intelligibility scoring, and pause analysis.

These results confirmed high usability while identifying areas for refinement before the second cycle.

Table 2. End-user testing and refinement—cycle 1.

Survey question	SLP ^a score (out of 5), mean (SD)
The web application met my expectations.	3.88 (1.13)
The navigation within the app was easy and logical.	4.50 (0.76)
The design of the app (colors, fonts, layout) was appealing and professional.	4.88 (0.35)
The session upload process was straightforward	4.63 (0.74)
Selecting audio segments was intuitive	4.63 (0.52)
The time required for upload and selection was acceptable	4.50 (0.53)
How easy was it to select audio segments and upload your video or audio session?	4.50 (1.41)
The automatic intelligibility score seemed accurate	3.13 (1.25)

Survey question	SLP ^a score (out of 5), mean (SD)
The intelligibility score was easy to understand	4.13 (0.99)
The side-by-side view of patients' transcription versus reference was clear	4.50 (0.76)
The phoneme-level error highlighting was intuitive	3.88 (1.25)
How easy was the intelligibility score to understand?	4.00 (2.51)
Overall mean across questions	4.26 (0.58)

^aSLP: speech and language pathologist.

Table 3. End-user testing and refinement—cycle 1.

ID ^a	Representative SLP ^b feedback	Category
I01-002	First files did not upload.	Fix or bug
I01-013	The selected region shifted right when I extended the clip.	Fix or bug
I01-004	The term “misspelled” is confusing because it suggests the SLP made the error.	Improvement
I01-005	I want to listen to the segment while I enter the reference words.	Improvement
I01-007	Show minutes/seconds under the clip to guide word and sentence selection.	Improvement
I01-019	Count phonetic equivalents as correct in the intelligibility calculation.	Feature
I01-018	Analyze the duration of pauses between words to reflect prosody issues.	Feature

^aID: feedback identifier.

^bSLP: speech and language pathologist.

End-User Testing and Refinement-Cycle 2

Five SLPs participated in the second usability cycle (mean clinical experience 4.3, SD 2.0 y; [Table 1](#)). The testing environments and browsers were the same as those used in cycle 1.

The mean SUS score increased to 91.7 (SD 4.1), which is again considered excellent [38]. Survey results ([Table](#)

4) showed consistently high ratings. Navigation and the new acoustic analysis page both received a score of 4.75, indicating that participants found the workflow intuitive and the additional analysis features useful. Frequency and intensity graphs were rated positively (4.33 each), while the formants graph was rated lower (mean 3.66, SD 0.65), suggesting limited clarity or perceived clinical utility. Overall expectations were rated at a mean of 4.20 (SD 0.42).

Table 4. End-user testing and refinement—cycle 2.

Survey question	SLP ^a score (out of 5), mean (SD)
I found the fundamental frequency graph useful.	4.33 (0.57)
I found the intensity graph useful.	4.33 (0.89)
I found the formants graph useful.	3.66 (0.65)
How easy was the acoustic analysis page to understand?	4.75 (0.41)
The web application met my expectations	4.20 (0.42)
The navigation within the app was easy and logical.	4.75 (0.78)
Overall mean across questions	4.34 (0.40)

^aSLP: speech and language pathologist.

Representative feedback ([Table 5](#)) emphasized advanced feature requests. Suggestions included the option to add notes to assessment results, integration of advanced voice-quality indices (eg, Cepstral Peak Prominence [CPPS] and Acoustic

Voice Quality Index [AVQI]), and audio playback directly from result pages. These comments indicate readiness for clinical refinement, with priorities shifting toward advanced analysis rather than fundamental usability.

Table 5. End-user testing and refinement—cycle 2.

ID ^a	Representative SLP ^b feedback	Category
I02-002	Ability to add notes to an assessment result. It should be possible to comment on patient behavior (eg, “struggled with breath here”).	Feature
I01-013	Inclusion of voice-quality indices (eg, CPPS ^c and AVQI ^d) was suggested to provide additional clinical insight.	Feature
I01-004	Playback of the analyzed audio directly from the Results page was requested to allow comparison between perceived audio and computed metrics.	Feature
I01-005	Ability to listen to the uploaded audio on the Results page was requested, especially to review the alignment between the reference text and the transcription	Feature

^a ID: feedback identifier.

^bSLP: speech and language pathologist.

^cCPPS: Cepstral Peak Prominence.

^dAVQI: Acoustic Voice Quality Index.

Summary of Results

Across all 3 substudies, the iSpeak prototype demonstrated promising performance and usability. The pilot validation confirmed that ASR achieved accuracy comparable to or greater than that of untrained human listeners across varying severities of dysarthria. Both usability cycles with SLPs yielded excellent SUS scores (>88), with cycle 2 reaching a mean of 91.7 (SD 4.1) after refinement. Core workflows were consistently rated highly, and feedback evolved from identifying technical issues in cycle 1 to requesting advanced analytical features in cycle 2. Together, these results indicate that the system is both usable and potentially clinically relevant, with future development focused on enhanced analytic capabilities and broader validation.

Discussion

Principal Findings

This study evaluated a web-based application designed to support SLPs in assessing dysarthric speech in patients with poststroke. Developed using a user-centered design framework in accordance with the International Organization for Standardization (ISO) 9241-210, the prototype integrates ASR and acoustic analysis into a clinical dashboard. A pilot validation compared ASR with lay listeners, followed by 2 iterative usability cycles with SLPs. The findings demonstrate that ASR can approximate human performance in intelligibility assessment and that the application was consistently rated as highly usable, with SUS scores above 85 in both cycles. Feedback confirmed the value of automatic intelligibility scoring and acoustic analysis while identifying priorities for further development.

Preliminary Testing

The pilot validation provided initial evidence of how ASR compares with human listeners on dysarthric speech. Across tasks, word lists produced higher WERs than sentence lists, reflecting the contextual advantage of sentences [27]. Importantly, lay listeners were not native English speakers, which may have introduced bias, as subtle dysarthric articulations could be more difficult to interpret for nonnative listeners [40]. This may have led to an underestimation of intelligibility relative to trained or native listeners [41].

Patient-level differences were evident. For “Patient 001,” ASR outperformed most lay listeners across tasks, suggesting that the system could handle this speech relatively well. For “Patient 003,” ASR performance was competitive, especially in the sentences task. “Patient 002” posed the greatest challenge: lay listeners showed high variability in WERs for the word task, and ASR performed worse than all listeners in the sentences task. These findings highlight how severity

and individual variability in dysarthria strongly affect ASR performance [27,40,41].

Although Whisper achieves state-of-the-art performance on typical speech, its accuracy declines with dysarthric input. Stroke-related dysarthria alters pronunciation, speech rate, and voice quality in diverse ways. Because such data are underrepresented in training corpora, ASR robustness remains limited [32,41,42]. Nonetheless, pilot results indicate that Whisper can sometimes match or surpass untrained human listeners, suggesting promise as a baseline for automatic intelligibility scoring. Similar findings are reported in recent validation efforts of automatic intelligibility measures for motor speech disorders [43]. Future validation with larger patient cohorts and SLP raters is needed to confirm these observations.

End-User Testing and Refinement—Cycle 1

The first usability cycle demonstrated that the prototype was already perceived as highly usable. The mean SUS score of 88.4 (SD 4.6) falls in the “excellent” range, confirming that the design was well-adapted to SLPs’ needs [44]. Importantly, the application was tested both in clinical and home environments, as well as across different browsers, indicating technical robustness in varied contexts of use.

Survey ratings averaged 4.26/5. Core workflow elements—navigation, uploading, segment selection, and processing time—scored above 4.5, and visual design was rated highest (mean 4.88, SD 0.35). These results align with reviews of existing speech therapy apps, which similarly show that visual design and general usability are often rated highly while output clarity (such as the accuracy of automated scores or error highlighting) tends to receive more mixed feedback [38,45].

Three items scored lower: perceived accuracy of the intelligibility score (mean 3.13, SD 1.25), phoneme-level error highlighting (mean 3.88, SD 1.25), and meeting expectations (mean 3.88, SD 1.13). These reflected a technical alignment bug in this prototype version, which caused intelligibility scores and error highlights to miss clearly spoken words. This limitation directly influenced perceptions of accuracy, a pattern also reported in empirical studies of online SLP tools, where users frequently note misalignments or ambiguous output explanations [39].

Open feedback aligned with these quantitative findings. Fixes focused on reliability (eg, failed uploads and region shifting). Usability improvements included clearer terminology, simultaneous listening and transcription, and visible time markers. Feature requests suggested phonetic equivalence in scoring and pause analysis. Similar issues of clinician-perceived usability and workflow integration have been noted in

broader reviews of automated speech therapy tools [45,46]. Fourteen priority fixes and improvements were implemented for cycle 2, balancing feasibility with the need to proceed rapidly to the next evaluation.

End-User Testing and Refinement–Cycle 2

The second cycle reinforced and extended these findings. The mean SUS score increased to 91.7 (SD 4.1), again in the “excellent” range, suggesting that refinements made after cycle 1 successfully improved usability, as SLP tools evolve from basic usability to richer features [44,45]. As before, testing across home and clinical settings and multiple browsers confirmed robustness. This pattern mirrors findings in reviews of eHealth speech-language therapy applications, where iterative improvements based on clinician feedback tend to yield measurable increases in satisfaction and functionality [38,45].

Survey ratings averaged 4.34/5. Navigation and the new acoustic analysis page received the highest scores (mean 4.75, SD 0.78). Frequency and intensity graphs were positively rated (4.33 each), confirming clinical relevance. The formants graph was rated lower (mean 3.66, SD 0.65), indicating that its value was less evident in practice. The “meeting expectations” item improved to a mean of 4.20 (SD 0.42). These improvements in visual clarity and interactive analytics reflect the user experience findings from telepractice tools, where dashboards and visualization features are increasingly valued as maturity grows [47].

Qualitative feedback shifted from identifying bugs to requesting advanced features. Suggestions included adding notes to assessment results, incorporating voice-quality indices such as CPPS and AVQI, and enabling audio playback directly from results pages. This shift is consistent with empirical studies of speech therapy platforms, where, once baseline usability is achieved, user requests focus more on analytic depth and interactivity [20,48]. Broader surveys of online and AI-enabled speech therapy systems echo this pattern, with early iterations focusing on usability and later development emphasizing advanced functionality [45,46].

The Prototype

The final prototype can be considered a simple, efficient, and potentially clinically relevant dashboard for assessing speech clarity and phonetic accuracy. Usage analytics confirmed the intuitive design: task times decreased after initial use, and only 3 dead clicks were recorded. The modular architecture enables extension to multilingual contexts, although current phoneme segmentation is restricted to English. While Whisper is multilingual, the phoneme module relies on English-only resources, limiting its applicability across languages.

Occasional ASR hallucinations were observed in cases of severe dysarthria or poor microphone quality. These issues were mitigated through design decisions (eg, suppressing insertions in alignment), ensuring that phoneme analysis

was not distorted. The current implementation also assumes one speaker per session; while speaker diarization could be added to handle overlapping speech, it introduces additional complexity.

Future Work

Several directions for development emerged. First, multilingual phoneme support is required to reflect the diversity of clinical populations. Second, advanced acoustic indices (eg, jitter, shimmer, CPPS, and AVQI) should be integrated to capture additional aspects of dysarthria severity [35–37]. Third, diarization would allow for the analysis of multispeaker sessions, accounting for overlaps or caregiver contributions. Finally, larger-scale validation with patients and practicing SLPs is essential to confirm accuracy and usability across broader contexts. An important next step is to systematically evaluate the generalizability of the system across different neurological populations, as initial testing across mixed etiologies suggests feasibility but does not yet establish condition-specific validity. Future work should also consider implementation within the broader framework of telerehabilitation, where evidence is growing for the effectiveness of remote interventions in speech-language therapy [49].

Limitations

At present, the system is not compliant with data protection regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH), which poses a barrier to clinical deployment over the internet. Use within closed clinical networks is feasible, but full compliance will be essential for wider adoption. A limitation of this study is the small sample size in the usability testing cycles, which may limit the generalizability of the usability findings and warrants further testing in larger cohorts.

Conclusion

This study demonstrates the feasibility of integrating ASR and acoustic analysis into a web-based application to support SLPs in assessing dysarthric speech. A pilot validation confirmed that ASR performance was comparable to that of untrained human listeners, while 2 usability cycles with SLPs yielded consistently excellent SUS scores, indicating that the system was perceived as highly usable and clinically relevant. Iterative refinements improved navigation and workflow, and feedback evolved from bug reports to requests for advanced analytic features, underscoring both the robustness of the core design and the demand for deeper functionality. Although current limitations include reliance on English-only phoneme segmentation, the prototype establishes a solid foundation for scalable digital assessment. Future work should extend validation to larger and more diverse patient groups, expand multilingual support, and integrate additional advanced outcome measures to further enhance clinical adoption and impact.

Acknowledgments

Generative artificial intelligence (ChatGPT 5.5 high) assistance was limited to the wording and language editing of the manuscript; all scientific content, analysis, and interpretation were developed by the authors.

Funding

This project did not receive external funding.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to data privacy considerations but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: PV, PD, AS, SG, MB, CEA, CMB

Data curation: PV, CMB

Formal analysis: PV

Funding acquisition: CMB

Investigation: PV, JC

Methodology: PV, PD, AS, CMB

Project administration: CMB

Resources: MB, CMB

Software: PV, SS, ED

Supervision: PD, CMB

Validation: PV, PD, CMB

Visualization: PV

Writing—original draft: PV, CMB

Writing—review & editing: PD, AS, SS, SG, ED, JC, MB, CEA

Conflicts of Interest

None declared.

Multimedia Appendix 1

Needs, requirements, feedback, and style guide.

[\[DOCX File \(Microsoft Word File\), 234 KB-Multimedia Appendix 1\]](#)

References

1. Jayaraman DK, Das JM. Dysarthria. In: StatPearls. StatPearls Publishing; 2023. URL: <https://www.ncbi.nlm.nih.gov/books/NBK592453/> [Accessed 2026-06-06]
2. Dysarthria in adults. American Speech-Language-Hearing Association. URL: <https://www.asha.org/practice-portal/clinical-topics/dysarthria-in-adults/> [Accessed 2026-06-06]
3. Vogel AP, Graf L, Weiß M, Chan CSJ, Hepworth G, Synofzik M. Development and validation of the dysarthria impact scale: a patient-reported outcome for motor speech disorders. *J Neurol*. Mar 10, 2026;273(3):195. [doi: [10.1007/s00415-026-13740-1](https://doi.org/10.1007/s00415-026-13740-1)] [Medline: [41805906](https://pubmed.ncbi.nlm.nih.gov/41805906/)]
4. Atkinson-Clement C, Letanneux A, Baille G, et al. Psychosocial impact of dysarthria: the patient-reported outcome as part of the clinical management. *Neurodegener Dis*. 2019;19(1):12-21. [doi: [10.1159/000499627](https://doi.org/10.1159/000499627)] [Medline: [31112944](https://pubmed.ncbi.nlm.nih.gov/31112944/)]
5. Enderby P. Frenchay Dysarthria Assessment. *Int J Lang Commun Disord*. Jan 1980;15(3):165-173. [doi: [10.3109/13682828009112541](https://doi.org/10.3109/13682828009112541)]
6. Riolo V, Pizzorni N, Guanziroli E, et al. Cross-cultural adaptation into Italian and validation of the Frenchay Dysarthria Assessment - 2. *Eur J Phys Rehabil Med*. Jun 2022;58(3):342-351. [doi: [10.23736/S1973-9087.21.07029-5](https://doi.org/10.23736/S1973-9087.21.07029-5)] [Medline: [34498832](https://pubmed.ncbi.nlm.nih.gov/34498832/)]
7. Cardoso R, Guimarães I, Santos H, et al. Frenchay Dysarthria Assessment (FDA-2) in Parkinson's disease: cross-cultural adaptation and psychometric properties of the European Portuguese version. *J Neurol*. Jan 2017;264(1):21-31. [doi: [10.1007/s00415-016-8298-6](https://doi.org/10.1007/s00415-016-8298-6)] [Medline: [27747392](https://pubmed.ncbi.nlm.nih.gov/27747392/)]
8. Icht M, Bergerzon-Biton O, Ben-David BM. Validation and cross-linguistic adaptation of the Frenchay Dysarthria Assessment (FDA-2) speech intelligibility tests: hebrew version. *Int J Lang Commun Disord*. Sep 2022;57(5):1023-1049. [doi: [10.1111/1460-6984.12737](https://doi.org/10.1111/1460-6984.12737)] [Medline: [35714104](https://pubmed.ncbi.nlm.nih.gov/35714104/)]
9. Telepractice. American Speech-Language-Hearing Association. URL: https://www.asha.org/practice-portal/professional-issues/telepractice/#collapse_3 [Accessed 2025-02-10]

10. Telehealth guidance. Royal College of Speech and Language Therapists. URL: <https://www.rcslt.org/members/delivering-quality-services/telehealth-guidance/> [Accessed 2026-06-06]
11. What is speech? What is language? American Speech-Language-Hearing Association. URL: <https://www.asha.org/public/speech/development/speech-and-language/> [Accessed 2026-06-06]
12. Scott AM, Clark J, Cardona M, et al. Telehealth versus face-to-face delivery of speech language pathology services: a systematic review and meta-analysis. *J Telemed Telecare*. Oct 2025;31(9):1203-1215. [doi: [10.1177/1357633X241272976](https://doi.org/10.1177/1357633X241272976)] [Medline: [39387166](https://pubmed.ncbi.nlm.nih.gov/39387166/)]
13. Yorkston KM, Strand EA, Kennedy MRT. Comprehensibility of dysarthric speech: implications for assessment and treatment planning. *Am J Speech Lang Pathol*. Feb 1996;5(1):55-66. [doi: [10.1044/1058-0360.0501.55](https://doi.org/10.1044/1058-0360.0501.55)]
14. Kent RD, Kim YJ. Toward an acoustic typology of motor speech disorders. *Clin Linguist Phon*. Sep 2003;17(6):427-445. [doi: [10.1080/0269920031000086248](https://doi.org/10.1080/0269920031000086248)] [Medline: [14564830](https://pubmed.ncbi.nlm.nih.gov/14564830/)]
15. Kim Y, Kim M, Kim J, Song TJ. Efficacy and feasibility of a digital speech therapy for post-stroke dysarthria: protocol for a randomized controlled trial. *Front Neurol*. 2024;15:1305297. [doi: [10.3389/fneur.2024.1305297](https://doi.org/10.3389/fneur.2024.1305297)] [Medline: [38356882](https://pubmed.ncbi.nlm.nih.gov/38356882/)]
16. Baevski A, Zhou H, Mohamed A, Auli M. Wav2vec 2.0: a framework for self-supervised learning of speech representations. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020:12449-12460. URL: <https://dl.acm.org/doi/abs/10.5555/3495724.3496768> [Accessed 2026-06-06]
17. Tucker JK. Perspectives of speech-language pathologists on the use of telepractice in schools: the qualitative view. *Int J Telerehabil*. 2012;4(2):47-60. [doi: [10.5195/ijt.2012.6102](https://doi.org/10.5195/ijt.2012.6102)] [Medline: [25945203](https://pubmed.ncbi.nlm.nih.gov/25945203/)]
18. Hsu WN, Bolte B, Tsai YHH, Lakhota K, Salakhutdinov R, Mohamed A. HuBERT: self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Trans Audio Speech Lang Process*. 2021;29:3451-3460. [doi: [10.1109/TASLP.2021.3122291](https://doi.org/10.1109/TASLP.2021.3122291)] [Medline: [25079929](https://pubmed.ncbi.nlm.nih.gov/25079929/)]
19. Chen S, Wang C, Chen Z, et al. WavLM: large-scale self-supervised pre-training for full stack speech processing. *IEEE J Sel Top Signal Process*. Oct 2022;16(6):1505-1518. [doi: [10.1109/JSTSP.2022.3188113](https://doi.org/10.1109/JSTSP.2022.3188113)] [Medline: [25079929](https://pubmed.ncbi.nlm.nih.gov/25079929/)]
20. Molini-Avejonas DR, Rondon-Melo S, Amato CA, Samelli AG. A systematic review of the use of telehealth in speech, language and hearing sciences. *J Telemed Telecare*. Oct 2015;21(7):367-376. [doi: [10.1177/1357633X15583215](https://doi.org/10.1177/1357633X15583215)] [Medline: [26026181](https://pubmed.ncbi.nlm.nih.gov/26026181/)]
21. Mitchell J, Shirota C, Clanchy K. Factors that influence the adoption of rehabilitation technologies: a multi-disciplinary qualitative exploration. *J Neuroeng Rehabil*. Jun 20, 2023;20(1):80. [doi: [10.1186/s12984-023-01194-9](https://doi.org/10.1186/s12984-023-01194-9)] [Medline: [37340496](https://pubmed.ncbi.nlm.nih.gov/37340496/)]
22. Selvanayakam S, Giovanoli S, Slot A, et al. I speak Tele outlines the design of a digitized dysarthria assessment. *Sci Rep*. 2025;15(1):35903. [doi: [10.1038/s41598-025-19726-9](https://doi.org/10.1038/s41598-025-19726-9)] [Medline: [41087405](https://pubmed.ncbi.nlm.nih.gov/41087405/)]
23. Chandran S, Al-Sa'di A, Ahmad E. Exploring user centered design in healthcare: a literature review. Presented at: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); Oct 22-24, 2020; Istanbul, Turkey. [doi: [10.1109/ISMSIT50672.2020.9255313](https://doi.org/10.1109/ISMSIT50672.2020.9255313)]
24. Good A, Omisade O. Linking activity theory with user centred design: a human computer interaction framework for the design and evaluation of mHealth interventions. *Stud Health Technol Inform*. Jul 30, 2019;263:49-63. [doi: [10.3233/SHTI190110](https://doi.org/10.3233/SHTI190110)] [Medline: [31411152](https://pubmed.ncbi.nlm.nih.gov/31411152/)]
25. Rudzicz F, Namasivayam AK, Wolff T. The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Lang Resources & Evaluation*. Dec 2012;46(4):523-541. [doi: [10.1007/s10579-011-9145-0](https://doi.org/10.1007/s10579-011-9145-0)]
26. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. *Am J Theor Appl Stat*. 2016;5(1):1-4. [doi: [10.11648/j.ajtas.20160501.11](https://doi.org/10.11648/j.ajtas.20160501.11)]
27. Hustad KC. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatri Logop*. 2007;59(6):306-317. [doi: [10.1159/000108337](https://doi.org/10.1159/000108337)] [Medline: [17965573](https://pubmed.ncbi.nlm.nih.gov/17965573/)]
28. ggml-org/whisper.cpp. GitHub. URL: <https://github.com/ggml-org/whisper.cpp> [Accessed 2026-06-06]
29. Grier RA, Bangor A, Kortum P, Peres SC. The system usability scale: beyond standard usability testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. ;57(1):187-191. [doi: [10.1177/1541931213571042](https://doi.org/10.1177/1541931213571042)]
30. Konstantinidis S. Computing the edit distance of a regular language. *Inf Comput*. Sep 2007;205(9):1307-1316. [doi: [10.1016/j.ic.2007.06.001](https://doi.org/10.1016/j.ic.2007.06.001)]
31. Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. *Sov Phys Dokl*. 1966;10:707-710. URL: <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf> [Accessed 2026-06-06]
32. Xue W, van Hout R, Cuccharini C, Strik H. Assessing speech intelligibility of pathological speech in sentences and word lists: the contribution of phoneme-level measures. *J Commun Disord*. 2023;102:106301. [doi: [10.1016/j.jcomdis.2023.106301](https://doi.org/10.1016/j.jcomdis.2023.106301)] [Medline: [36709701](https://pubmed.ncbi.nlm.nih.gov/36709701/)]

33. Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J Speech Hear Disord.* Nov 1989;54(4):482-499. [doi: [10.1044/jshd.5404.482](https://doi.org/10.1044/jshd.5404.482)] [Medline: [2811329](https://pubmed.ncbi.nlm.nih.gov/2811329/)]
34. Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatr Logop.* 2001;53(1):1-18. [doi: [10.1159/000052649](https://doi.org/10.1159/000052649)] [Medline: [11125256](https://pubmed.ncbi.nlm.nih.gov/11125256/)]
35. Nylén F. An acoustic model of speech dysprosody in patients with Parkinson's disease. *Front Hum Neurosci.* 2025;19:1566274. [doi: [10.3389/fnhum.2025.1566274](https://doi.org/10.3389/fnhum.2025.1566274)] [Medline: [40356883](https://pubmed.ncbi.nlm.nih.gov/40356883/)]
36. Villain M, Cosin C, Glize B, et al. Affective prosody and depression after stroke: a pilot study. *Stroke.* Sep 2016;47(9):2397-2400. [doi: [10.1161/STROKEAHA.116.013852](https://doi.org/10.1161/STROKEAHA.116.013852)] [Medline: [27507865](https://pubmed.ncbi.nlm.nih.gov/27507865/)]
37. Ross ED. Affective prosody and its impact on the neurology of language, depression, memory and emotions. *Brain Sci.* Nov 9, 2023;13(11):1572. [doi: [10.3390/brainsci13111572](https://doi.org/10.3390/brainsci13111572)] [Medline: [38002532](https://pubmed.ncbi.nlm.nih.gov/38002532/)]
38. Vaezipour A, Campbell J, Theodoros D, Russell T. Mobile apps for speech-language therapy in adults with communication disorders: review of content and quality. *JMIR mHealth uHealth.* Oct 29, 2020;8(10):e18858. [doi: [10.2196/18858](https://doi.org/10.2196/18858)] [Medline: [33118953](https://pubmed.ncbi.nlm.nih.gov/33118953/)]
39. Wouda L, Boerma T, Gerrits E, Blom E. First steps toward implementation of the online test battery LITMUS-NL: a usability and feasibility study. *Perspect ASHA Spec Interest Groups.* Oct 3, 2024;9(5):1439-1455. [doi: [10.1044/2024.PERSP-23-00308](https://doi.org/10.1044/2024.PERSP-23-00308)]
40. Kim Y, Thompson A, Lee SJ. Does native language matter in perceptual ratings of dysarthria? *J Speech Lang Hear Res.* Sep 12, 2024;67(9):2842-2855. [doi: [10.1044/2024.JSLHR-23-00668](https://doi.org/10.1044/2024.JSLHR-23-00668)] [Medline: [38662924](https://pubmed.ncbi.nlm.nih.gov/38662924/)]
41. Qian Z, Xiao K, Yu C. A survey of technologies for automatic dysarthric speech recognition. *EURASIP J Audio Speech Music Process.* Nov 11, 2023;2023(1):48. [doi: [10.1186/s13636-023-00318-2](https://doi.org/10.1186/s13636-023-00318-2)]
42. Qian Z, Xiao K. A survey of automatic speech recognition for dysarthric speech. *Electronics (Basel).* Oct 16, 2023;12(20):4278. [doi: [10.3390/electronics12204278](https://doi.org/10.3390/electronics12204278)]
43. Tröger J, Dörr F, Schwed L, et al. An automatic measure for speech intelligibility in dysarthrias-validation across multiple languages and neurological disorders. *Front Digit Health.* 2024;6:1440986. [doi: [10.3389/fdgth.2024.1440986](https://doi.org/10.3389/fdgth.2024.1440986)] [Medline: [39108340](https://pubmed.ncbi.nlm.nih.gov/39108340/)]
44. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum-Comput Interact.* Jul 2008;24(6):574-594. [doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)]
45. Attwell GA, Bennin KE, Tekinerdogan B. A systematic review of online speech therapy systems for intervention in childhood speech communication disorders. *Sensors (Basel).* Dec 11, 2022;22(24):9713. [doi: [10.3390/s22249713](https://doi.org/10.3390/s22249713)] [Medline: [36560082](https://pubmed.ncbi.nlm.nih.gov/36560082/)]
46. Green JR. Artificial intelligence in communication sciences and disorders: introduction to the forum. *J Speech Lang Hear Res.* Nov 7, 2024;67(11):4157-4161. [doi: [10.1044/2024.JSLHR-24-00594](https://doi.org/10.1044/2024.JSLHR-24-00594)] [Medline: [39418586](https://pubmed.ncbi.nlm.nih.gov/39418586/)]
47. Shankar V, Ramkumar V, Kumar S. Understanding the implementation of telepractice in speech and language services using a mixed-methods approach. *Wellcome Open Res.* 2022;7:46. [doi: [10.12688/wellcomeopenres.17622.2](https://doi.org/10.12688/wellcomeopenres.17622.2)] [Medline: [36158869](https://pubmed.ncbi.nlm.nih.gov/36158869/)]
48. Weidner K, Lowman J. Telepractice for adult speech-language pathology services: a systematic review. *Perspect ASHA Spec Interest Groups.* Feb 21, 2020;5(1):326-338. [doi: [10.1044/2019.PERSP-19-00146](https://doi.org/10.1044/2019.PERSP-19-00146)]
49. Cetinkaya B, Twomey K, Bullard B, EL Kouaissi S, Conroy P. Telerehabilitation of aphasia: a systematic review of the literature. *Aphasiology.* 2024;38(7):1271-1302. [doi: [10.1080/02687038.2023.2274621](https://doi.org/10.1080/02687038.2023.2274621)]

Abbreviations

- ASR:** automatic speech recognition
- AVQI:** Acoustic Voice Quality Index
- CPPS:** Cepstral Peak Prominence
- FDA-2:** Frenchay Dysarthria Assessment, Second Edition
- HIPAA:** Health Insurance Portability and Accountability Act
- HITECH:** Health Information Technology for Economic and Clinical Health Act
- ISO:** International Organization for Standardization
- SLP:** speech and language pathologist
- SUS:** System Usability Scale
- WER:** word error rate

Edited by Alicia Stone; peer-reviewed by Clarion Mendes, Rizwana Irfan; submitted 03.Oct.2025; final revised version received 08.May.2026; accepted 15.May.2026; published 03.Jul.2026

Please cite as:

Vinet P, Dillenbourg P, Slot A, Selvanayakam S, Giovanoli S, Du E, Cardoso J, Branscheidt M, Easthope Awai C, Bauer CM

Automatic Speech Recognition and Acoustic Analysis for Dysarthria Assessment in Telerehabilitation: User-Centered Design and Usability Study

JMIR Form Res 2026;10:e85230

URL: <https://formative.jmir.org/2026/1/e85230>

doi: [10.2196/85230](https://doi.org/10.2196/85230)

© Pierre Vinet, Pierre Dillenbourg, Amelieke Slot, Sharmila Selvanayakam, Sandra Giovanoli, Elisa Du, Julia Cardoso, Meret Branscheidt, Chris Easthope Awai, Christoph Michael Bauer. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 03.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.