Automatic assessment of verbal responses in a speech and language therapy platform

Evanthia Faliagka
Electrical & Computer Engineering Department University of Peloponnese
Patras, Greece
e.faliagka@go.uop.gr

Manos Plitsis
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
manos.plitsis@athenarc.gr

Kosmas Palios
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
kosmas.palios@gmail.com

Nassos Katsamanis
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
nkatsam@athenarc.gr

Spyridoula Stamouli
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
pstam@athenarc.gr

Athanasia-Lida Dimou
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
ndimou@athenarc.gr

Christos Panagiotou
PIKEI New Technologies
Patras, Greece
christos@pikei.io

Dimitris Karadimas
PIKEI New Technologies
Patras, Greece
dimitris@pikei.io

Christos Antonopoulos
University of Peloponnese
Patras, Greece
ch.antonop@uop.gr

Nikolaos Voros
University of Peloponnese
Patras, Greece
vorosn@gmail.com

Vassilis Katsouros
Institute for Language and Speech Processing Athena Research Center
Athens, Greece
vsk@athenarc.gr

Abstract

The aim of this paper is to propose a software subsystem for the automatic evaluation of patients' verbal responses in word and sentence production exercises, integrated into a comprehensive speech and language therapy intervention platform. This subsystem serves both the needs of patients by enabling easy and remote practice on speech production activities, and the needs of speech and language therapists by providing immediate and detailed feedback on their patients' performance. More specifically, speech therapy exercises require the patient to produce a word or a sentence that corresponds to an image displayed on screen. To evaluate the performance of patients in these tasks, a system was developed, which allows the recognition and automatic evaluation of the patients' speech, providing feedback to the therapist regarding the degree of success in producing the target word or sentence. For the purposes of the system development, a Greek custom speech recognition model as well as an open-source multilingual speech recognition system were utilized to transcribe the patient's speech into text. At the word level, accuracy is evaluated by calculating the Levenshtein distance between the transcription and the target word, while at the sentence level, the Word Error Rate (WER) is used to assess similarity to the target sentence. Evaluation of the subsystem's ability in recognizing mispronounced words—which represent variations of target words containing phonological errors—produced by neurotypical adults showed a word recognition accuracy of 71,4% for perfectly recognized words, while a substantial proportion of words were recognized with only one or two incorrect characters (8,8% and 7,6%, respectively), indicating that the system's effective accuracy in capturing acceptable responses in clinical settings can be considered higher.

This work is licensed under a Creative Commons Attribution International 4.0 License.

PETRA '25, Corfu Island, Greece
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1402-3/2025/06
https://doi.org/10.1145/3733155.3733223

CCS Concepts

• **Software and its engineering**;

Keywords

speech and language therapy, automatic speech recognition, aphasia

ACM Reference Format:
Evanthia Faliagka, Manos Plitsis, Kosmas Palios, Nassos Katsamanis, Spyridoula Stamouli, Athanasia-Lida Dimou, Christos Panagiotou, Dimitris Karadimas, Christos Antonopoulos, Nikolaos Voros, and Vassilis Katsouros. 2025. Automatic assessment of verbal responses in a speech and language therapy platform. In *The PErvasive Technologies Related to Assistive Environments (PETRA '25), June 25–27, 2025, Corfu Island, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3733155.3733223

1 Introduction

Serious and chronic speech and language impairments, as a result of acquired neurological disorders, affect a large percentage of the adult population, which is increasing due to population aging, and have serious consequences for the quality of life of sufferers, as they limit their ability to communicate on a daily basis. The medical causes of these disorders vary, with stroke, brain tumor and traumatic brain injury as the most common. In addition, certain neurodegenerative diseases, many of which are serious and chronic conditions increasingly affecting the elderly population, such as dementia, Alzheimer's disease, Primary Progressive Aphasia, etc., are often associated with speech and language disorders. According to the World Health Organization, 15,000,000 people worldwide suffer a stroke annually. In Greece, according to the latest data from the Hellenic Statistical Authority[1], 2.9% of the Greek population aged 65-74 and 5.7% aged 75+ suffered a stroke or continued to suffer the chronic consequences of a previous stroke, 6.5% of those aged 75+ suffered dementia or Alzheimer's disease, while it is estimated that of those who have suffered a stroke, approximately one third acquire aphasia, a chronic language disorder [1].

Effective support for these individuals requires individualized and systematic speech and language therapy intervention, the quality and frequency of which significantly affect the final outcome, i.e. the improvement of communication skills. However, while evidence indicates that the frequency and quantity of therapy is directly linked to faster patient recovery [1], resource and infrastructure shortages often make it impossible to provide sustained clinical therapy—whether publicly funded, insurance-supported, or private—for a long period of time. Furthermore, in addition to speech and language impairments, patients, who are mostly elderly, often face mobility issues that further limit their access to therapy.

Technology-assisted speech and language therapy applications help bridge this gap by providing accessible and cost-effective solutions for individuals with communication disorders. Technology-supported applications enable patients to engage in self-managed therapy at their own pace and convenience, reducing dependence on in-person clinical sessions. Research has shown that long-term and consistent use of such applications can lead to measurable improvements in language and cognitive function. For example, Liu et al [2] demonstrated that patients who regularly practiced with a digital therapeutic platform experienced sustained gains over a 30-week period. Similarly, studies [3–5] highlight the effectiveness of app-based interventions in supporting recovery for stroke patients with aphasia.

In recent decades, significant progress has been made in the field of automatic speech recognition (ASR), and in particular towards the development and improvement of pathological speech recognition systems, with aphasia being the most common pathology. More specifically, algorithms are being developed using machine learning techniques, which are trained on large volumes of spoken speech data from people with aphasia. These systems are applied in the automatic assessment and classification of speech into types of aphasia or degrees of severity [6–9], while in recent years they have begun to be applied in the context of technologically assisted speech and language therapy intervention. In this context, patients are given the opportunity to practice activities that require speech production through their computer, receiving feedback from an automated speech assessment system on their degree of success in producing target words or sentences. The advancements in the field of pathological speech processing have a direct impact on the faster and more effective rehabilitation of individuals with speech and language disorders, since technologically assisted therapy increases both the quantity and frequency of patients' practice.

This paper presents a subsystem, integrated in a comprehensive speech and language therapy platform (planv-project.gr), designed for the automated assessment of patients' speech. This subsystem supports therapy exercises that target individuals with language disorders and require speech production at both the word and sentence levels. The rest of the paper is organized as follows: In Section 2 related research is presented while in Section 3 the subsystem for automated evaluation of speech production is described. Section 4 outlines the architecture of the subsystem and in Section 5 its evaluation and results are discussed. Finally, Section 6 presents the conclusions and future directions.

2 Background

Most speech recognition systems that have been applied to pathological speech are mainly aimed at people with dysarthria. While dysarthria can be caused by a stroke and coexist with aphasia, the nature and characteristics of speech disorders in dysarthria and aphasia are quite different [10]. Dysarthria is a speech disorder characterized by weakness, paralysis, or poor coordination of the muscles involved in speech articulation. Patients with dysarthria, while knowing a word and being able to retrieve it as a lexical item, present articulation problems when producing it. In contrast, aphasia does not involve motor problems in speech articulation but rather disturbances in comprehension, word finding, or grammar. Problems with word retrieval occur, as well as semantic errors (paraphasias), i.e. production of words semantically related to the target word (e.g. "coffee" instead of "tea"), or phonological paraphasias, i.e. production of phonologically related words, or production of neologisms, i.e. pseudowords. Furthermore, there is a great deal of variation in the types of errors both among individuals with aphasia who share the same type of disorder, and within the same individual over time. This variation in speech production poses a challenge for speech recognition systems and explains the lack of extensive research on ASR in individuals with aphasia to date. The earliest studies in the field of automatic speech recognition of individuals with aphasia used commercial ASR applications developed for healthy speakers. These tools were used to provide automatic feedback on patient performance during treatment with mixed results. Specifically, Linebarger [11] reported positive treatment outcomes without providing a detailed presentation of the impact and performance of the ASR tool. Wade & Petheram [12], investigated the effectiveness of ASR in individuals with aphasia, reporting an accuracy rate of 80% in word and sentence production. The tool required extensive training sessions to allow the system to adapt to the patient's speech, while a limited vocabulary of 50 words and 24 sentences was used.

[1]Hellenic Statistical Authority (2020). Press Release: Health Survey 2019. available at https://www.statistics.gr/documents/20181/5856c8ff-59cd-aef0-2883-f90e4b265706 [in Greek]

More recent studies have focused on the development of new ASR technologies for aphasia. Technological developments in the field of ASR combined with advances in the field of deep neural networks (DNN) have been applied by research groups with the following results: (a) automatic transcription of oral narratives from individuals with aphasia into English and Cantonese [8, 13], (b) detection of phonological paraphasias in the spontaneous speech of individuals with aphasia [14], (c) quantitative analysis of the spontaneous speech of individuals with aphasia [15], and (d) assessment of aphasia severity in Cantonese speakers [9]. However, current ASR solutions are still affected by large individual differences between patients in the type of errors produced or even the dialect and pronunciation of each speaker. While large corpora of aphasic speech data are available in free databases, such as the AphasiaBank [16], which are constantly being enriched and provide important training data for the development of new models, they mainly represent widely spoken languages.

Regarding the automatic assessment of word production in picture naming tasks, the challenge of ASR is to confirm that a specific target word is produced and to provide feedback on the correctness of response. In this context, the Vithea research team [17] developed a speech and language therapy intervention application for Portuguese speakers with aphasia. They presented the results of an in-house ASR system, called AUDIMUS [18], which assesses picture naming using the keyword spotting technique. The researchers report an ASR accuracy of 82% on average, with a range between 69% and 93% across patients [10]. Another research group [19] evaluated a technology-assisted intervention application involving naming tasks in English speakers with apraxia and aphasia. They used the freely available ASR tool CMU PocketSphinx (Cmusphinx/Pocketsphinx, 2014/2020) to provide patients with "correct" or "incorrect" response feedback for each naming attempt. They reported an overall system accuracy performance of 80% and a range between 65.1% and 82.8% among patients depending on the severity of the deficit. Another study [20] performed word recognition in naming tests, utilizing a deep learning algorithm developed to identify pronunciation errors by non-native English speakers, compared to native speakers. In the NUVA system (Naming Utterance Verifier for Aphasia treatment), each word pronounced by a healthy speaker is compared to the same word pronounced by a speaker with aphasia. For this purpose, auditory models were utilized that were trained on a large volume of data from healthy English speakers (WSJCAM corpus of British healthy English speakers, [21]). The system uses two recordings from healthy speakers for each target word, which are converted into posteriograms based on their neural network. The system was evaluated on a sample of 8 patients, who made 1760 naming attempts. Accuracy ranging from 83.6 to 93.6% is reported, a performance significantly higher than the baseline (Google STT service) and in many cases comparable to the assessment of naming accuracy by speech therapists. Finally, Google Research [22] has developed personalized ASR systems that achieve high accuracy in recognizing pathological speech. Their study reports that these personalized models outperformed state-of-the-art, speaker-independent ASR systems and even human transcribers, especially in cases involving severe speech disorders.

3 INTEGRATED SUBSYSTEM FOR AutomatED EvALUATION OF SPEECH PRODUCTION: APPLICATION SCENARIOS

The subsystem presented in this paper is integrated into a comprehensive speech and language therapy platform (planv-project.gr), which offers patients the opportunity for remote speech and language intervention in their own environment, without the physical presence of a Speech and Language Therapist (SLT). The platform is a valuable tool for SLTs, who plan and oversee the intervention process remotely. It allows them to design and administer individualized speech and language therapy programs to their patients, which act complementarily to in-person sessions.

Exercises which involve speech production are part of a comprehensive therapy protocol for people with chronic, neurogenic language disorders, such as aphasia. The protocol focuses on two fundamental levels: The word level and the sentence level. The exercises related to the word level include nouns and verbs. These exercises require naming of nouns and verbs based on a visual stimulus displayed on the screen. Patients who perform the naming exercises are provided with multiple aids for word production, such as semantic help (e.g., "We use this object to brush our teeth"), phonological help (e.g., "The word begins with ca"), as well as orthographic help (written representation of the target word). Regarding the production exercises at the sentence level, the goal is to spontaneously produce sentences based on an image, to compose sentences based on some given words, or to place the arguments of sentences in the correct order.

Figure 1 presents the user interface in the word production exercise. The user records their answer, i.e. the word corresponding to the image displayed, which is then evaluated by the automatic evaluation system.

4 Architecture

The services and the subsystems that support the speech and language therapy platform have been developed using a microservices development model, with the entire functionality organized into five service groups as we can see in Figure 2 [23]. More specifically:

- Graphical interfaces: Graphical interfaces are the point of interaction between the user and the system. They are responsible for transferring all the functionalities of the platform to patients and SLTs. The interfaces use both synchronous and asynchronous communication models as the data sent to and from the user present great heterogeneity.
- User Identification/Authorization: This subsystem is responsible for registering the users and granting them access to the functions that concern them. This is where the main security and privacy mechanisms of the platform are implemented. Access to the subsystem is provided through a modern application programming interface (API).
- Platform content management: The content management subsystem consists of a set of services related to the storage and provision of user data, intervention plans/exercises and usage statistics. It is accessible through a modern API.
- Gateway: Receives all requests from the graphical interfaces to the identification and content management services. The

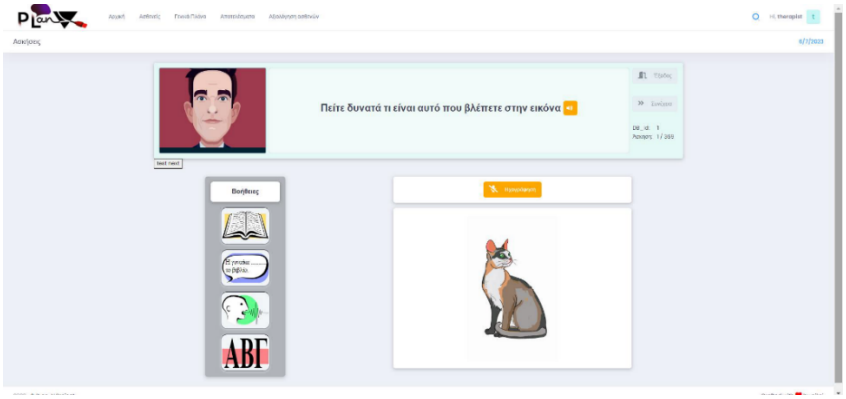

Figure 1: User interface in a word production exercise (Naming)

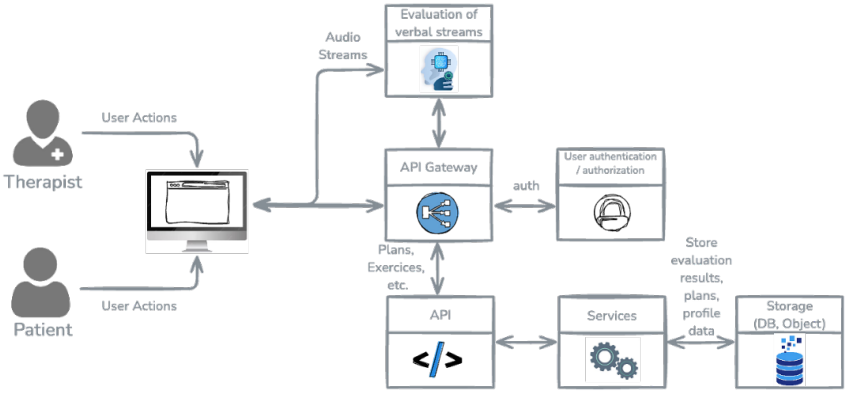

Figure 2: Platform's Architecture

purpose of this service is to properly manage and route the requests to the appropriate services.

- Verbal response evaluation subsystem: This subsystem processes and analyzes the users' verbal responses by interacting with and consuming data from other services on the platform.

In the verbal response evaluation subsystem, the evaluation process is carried out in three stages corresponding to three individual microservices (Figure 3):

- Initially, the exercise data sent by the frontend is used to retrieve the target words or phrases from a database.
- Next, the audio data, together with the target phrases, are forwarded to the central evaluation subsystem, which forwards them directly to the ASR subsystem for transcription into text.
- Finally, the central evaluation subsystem receives the transcription of the user's speech and gives it as input to the transcription evaluation subsystem. The latter generates a score on a scale from 1 to 100, which is finally made available to the SLTs.

The entry point of the system is the preprocessing subsystem. It is a lightweight microservice implemented in FastAPI, which receives the target words or sentences for each individual query from the exercise database via HTTPS GET requests, making them available for the next processing stage. It also receives audio data from the user's application, and forwards it to the subsequent components.

In the second stage, the ASR subsystem performs the transcription of the user's voice input into text. The voice input is streamed in real time via websocket, while the text is also returned in real time, without requiring the user to complete the entire sentence. The data is transferred through the central evaluation system. The speech recognition subsystem supports interconnection with cloud services (e.g., Google speech), as well as with internal Greek speech recognition models developed by the Institute for Language and Speech Processing (ILSP) in Kaldi (kaldi-asr.org), and with the Whisper multilingual ASR system (https://github.com/openai/whisper).

For the evaluation of speech production at word level, custom ASR models developed internally by ILSP using Kaldi were employed. To improve ASR accuracy, a list of possible mispronunciations of all target words used in naming exercises was created by specialized speech-language therapists. Specifically, a dataset of 4,567 phonological variations, derived from 369 target words, was generated. These variations were automatically produced by applying predefined linguistic rules based on common phonological

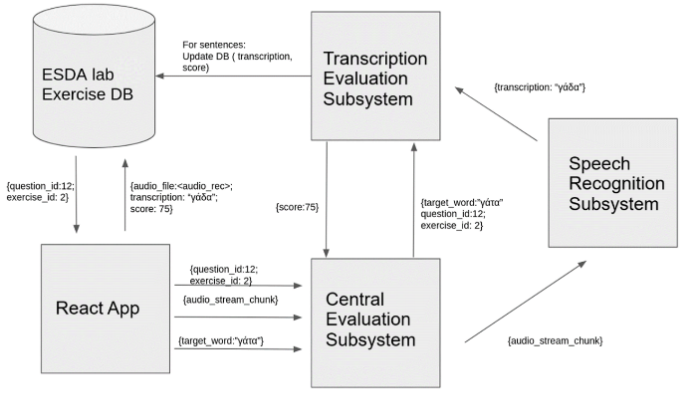

Figure 3: Architecture of the verbal response evaluation subsystem

error patterns observed in individuals with phonological impairments. This list of word variations was incorporated into the ASR system's dictionary, enabling the system to detect common word mispronunciations made by speech-impaired patients.

For the sentence recognition task, we used the OpenAI Whisper model, which, despite its increased computational complexity, has proven to perform much better in recognizing free-form speech. To reduce response time and computational cost, we employed an implementation of this system written entirely in C++ (https://github.com/ggerganov/whisper.cpp), which can run on any standard processor (CPU) within a reasonable time. However, since sentence-level processing may still require significantly more time than word-level recognition, the system was designed to first send an initial message back to the central evaluation system indicating that the processing is ongoing, and this status is also recorded in the database. Once transcription is completed, the result is sent back to the central evaluation system, which directly updates the correct value in the database through an update message.

The system is written in FastAPI and uses multiple processing threads (multithreading) for the simultaneous recognition of audio messages and sending of the results. In the third stage, the central evaluation subsystem uses the transcription evaluation subsystem to assess the user's responses at both the word and sentence levels. This subsystem receives as input the transcription of the user's speech (e.g., «*γάδα*»[target: «*γάτα*», Eng. "cat"], or «Η*μητέραπλέκει*»Eng. "The mother is knitting") along with the target word or sentence.

The evaluation is performed through textual similarity measurement with the reference (target) answers. More specifically, for word-level evaluation, the system calculates the Levenshtein distance between the transcription and the target word. A score of 100 is assigned when there are no differences (distance = 0), and 15 points are subtracted for each additional difference (+1 in distance), with a minimum score of 0 when the distance is greater than 4 (i.e., when 4 or more characters differ between transcription and target word). For sentence-level evaluation, the score is based on the Word Error Rate (WER) between the transcribed sentence and the target sentence, reflecting the proportion of errors relative to the total number of words in the target sentence. This metric accounts for word substitutions, omissions, and insertions. To normalize the score and ensure it lies within [0,1], even when there are significant differences in sentence length, the WER is bounded within this range, and the final score is computed as (1 - bounded WER) $\times$ 100. Thus, a score of 100 reflects a perfectly correct sentence, while 0 corresponds to a maximally erroneous one.

The evaluation subsystem produces two outputs: a) The final score of the user's response on a scale from 1 to 100, b) the automatic transcription of the user's speech. Both the score and the automatic transcription, along with the recording of the response, are available to the SLT, who can review and edit them, if necessary. Importantly, the patient doesn't receive any score on their pronunciation accuracy; instead, they always receive positive feedback from the system's front end, encouraging them to continue, regardless of the score.

5 System Evaluation

To evaluate the accuracy of the word-level recognition system, a subset of 1,781 word variants was selected from the larger dataset of 4,567 phonological variations (derived from 369 target words, see Section 4: *Architecture*). These word variants included common phonological errors and were designed to simulate mispronunciations typically produced by individuals with speech production impairments. All words were pronounced by six neurotypical speakers, each contributing approximately 300 words, following a protocol intended to reflect realistic error patterns observed in clinical populations.

It should be noted that this evaluation was conducted for word-level recognition, as this subsystem was specifically enriched with phonological variants of the target words to enable accurate detection of mispronunciations expected by individuals with speech and language disorders. Sentence-level recognition, handled by a general-purpose multilingual ASR system, was not evaluated in this study and remains a subject for future work.

Our word-level ASR system was custom-designed for the specialized purpose of detecting phonological errors, and not optimized for general-purpose recognition. For this reason, we could not rely on large end-to-end ASR models such as OpenAI's Whisper, which lack the capacity to handle phonological variants without explicit modeling. Instead, we used simpler acoustic models trained with Greek data, which are significantly smaller in scale than typical English corpora. The limited availability of Greek speech data for ASR training, compared to the extensive resources available for English, is a well-known limitation that affects system performance. For instance, even Whisper's large end-to-end model, which is substantially more powerful than our prototype, reports approximately 13% in Greek under general-purpose recognition tasks (OpenAIWhisperGitHub). Furthermore, Greek ASR models incorporating language models report even higher WER scores in general ASR tasks [24].

To assess word-level recognition accuracy, we measured how many of the produced words were correctly identified by the system exactly as pronounced, including when phonological errors were present. Out of the 1,781 recorded word variants, 325 words (18.2%) resulted in no recognition output, meaning the system failed to identify any word. Among the remaining recognized words, the following breakdown was observed: (a) 71.4% (1,040 words) were recognized perfectly, without any errors, (b) 8.8% (128 words) were recognized with a single incorrect character, (c) 7.6% (110 words) were recognized with two incorrect characters, (d) the remaining 12.2% were recognized with three or more incorrect characters.

These results demonstrate that the system achieves an overall moderately good performance for phonologically altered words under controlled conditions. However, while 71.4% perfect recognition is a promising result for words with phonological errors, the 18.2% no-recognition rate points to the system's sensitivity to both acoustic variability and/or the limitations of the underlying models and dictionary. While part of this outcome may stem from the complexity of recognizing atypical phonological patterns, it also likely reflects a need to improve both the robustness of the ASR system and the consistency of recording procedures. Although all recordings were made in quiet environments with good-quality microphones, variability in speaker behavior, microphone distance, and room acoustics may have contributed to the observed no recognition results.

On the other hand, it is worth noting that a considerable proportion of recognized words (8.8%) were identified with only a single incorrect character or with two incorrect characters (7,6%). Given that the mean word length in our dataset is approximately 6.88 characters, such minor deviations indicate that in many cases the system successfully captures the intended word form with high precision. In clinical contexts, and more specifically, in naming tasks where productions containing small phonological errors may still be considered acceptable (e.g. "olange" or "olance" instead of "orange"), this performance can be seen as highly functional. If we consider that in standardized procedures for the assessment of linguistic skills of individuals with aphasia the production of a word in naming tasks is typically regarded as correct if fewer than half of its phonemes are incorrect, the system's ability to limit deviations to just one or two characters in a significant proportion of cases suggests that its effective accuracy in capturing acceptable responses could be considered even higher than the strict 71.4% perfect recognition rate.

However, it is important to note that these tests were conducted using neurotypical speakers simulating phonologically erroneous word productions, thus representing a best-case scenario for system performance. We expect that recognition accuracy may be lower when working directly with individuals who have speech production difficulties, due to the greater variability and severity of errors in their speech.

6 Conclusions and future directions

In this paper, we presented a software subsystem for the automatic evaluation of verbal responses, integrated into a comprehensive speech and language therapy platform. This subsystem supports word and sentence production exercises, which are essential yet challenging components of therapy. The system employs custom Greek speech recognition models, enriched with phonological variants, to detect common mispronunciations and assess speech production accuracy using specialized similarity measures. The evaluation of the word-level recognition component demonstrated 71.4% perfect recognition of phonologically altered words pronounced by neurotypical adults. Although this is a moderately good result for such a challenging task, it should be noted that when words recognized with only minor deviations (e.g., one or two incorrect characters) are considered, the system's effective accuracy could be regarded as even higher, from a clinical perspective. Nevertheless, this evaluation represents a best-case scenario, and future evaluations on real patient data are crucial to assess the system's robustness under more demanding realistic conditions.

Although the system does not provide explicit scores to patients, it facilitates independent, home-based practice, by allowing patients to engage in speech production exercises assigned and controlled by their SLTs. At the same time, it offers SLTs detailed feedback on each patient's performance, including: (1) an objective, fine-grained score (1–100), (2) the automatic transcription of the response, and (3) the audio recording itself. The availability of such detailed data provides SLTs with a valuable tool to support and monitor therapy remotely, especially in exercises involving speech production, which are inherently more difficult to evaluate automatically compared to comprehension tasks. The system allows SLTs to track patient progress over time, identify persistent error patterns, and adjust therapy plans accordingly, thus enhancing the overall effectiveness and personalization of intervention. Moreover, by enabling asynchronous supervision of therapy, the subsystem helps maximize therapy intensity and frequency, even when direct, in-person sessions are limited, facilitating better long-term outcomes. Furthermore, the longitudinal analysis of speech data could support the development of predictive models for recovery trajectories, identifying patterns associated with better or worse outcomes. This could lead to personalized and adaptive therapy programs, adjusted dynamically to each patient's evolving needs.

Nevertheless, several areas remain open for future work. A key next step is to evaluate the system with actual speech-impaired users, as testing with neurotypical speakers simulating errors, though useful for initial validation, cannot fully capture the variability and complexity of real pathological speech. Additionally,

expanding Greek speech corpora, particularly with pathological speech data, is critical for improving ASR robustness and reducing recognition failures. The sentence-level recognition component, while technically integrated, also requires formal evaluation to verify its effectiveness in real therapeutic scenarios. Finally, clinical validation through user studies will be essential to assess the platform's practical benefits for patients and STLs, and to guide its further integration into clinical practice. In summary, the proposed subsystem lays the groundwork for scalable, accessible, and individualized remote speech and language therapy, with the potential to improve treatment consistency, enhance progress tracking, and support better long-term outcomes for individuals with speech and language disorders.

Acknowledgments

This work has been co-funded by PASS4Autism project, funded by the European Union. Project Nr. 2023-1-ES01-KA220-ADU-000154910.

References

[1] Brady, M.C., Kelly, H., Godwin, J., Enderby, P., Campbell, P. (2016) Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev.* 2016 Jun 1;2016(6):CD000425.

[2] Liu, H., Cordella, C., Ishwar, P., Betke, M. & Kiran, S. Consistent long-term practice leads to consistent improvement: Benefits of self-managed therapy for language and cognitive deficits using a digital therapeutic. Front. Digit. Heal. Sect. Pers. Medicine 5:1095110 (2023). Fraser, K. C., Rudzicz, F. & Rochon, E. (2013α) Using text and acoustic features to diagnose progressive aphasia and its subtypes. Proc. INTERSPEECH, 2177–2181.

[3] Des Roches, C. A., Balachandran, I., Ascenso, E. M., Tripodis, Y. & Kiran, S. Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. Front. Hum. Neurosci. 8, DOI: 10.3389/fnhum.2014.01015 (2015).

[4] Mallet, K. *et al.* RecoverNow: A patient perspective on the delivery of mobile tablet-based stroke rehabilitation in the acute care setting. Int J Stroke 14, 174–179, DOI: 10.1177/1747493018790031 (2019).

[5] Nichol, L., Pitt, R., Wallace, S. J., Rodriguez, A. D. & Hill, A. J. "there are endless areas that they can use it for": speech language pathologist perspectives of technology support for aphasia self-management. Disabil. Rehabil. Assist. Technol. 18, 1473–1488, DOI: 10.1080/17483107.2022.2037758.

[6] Fraser, K. C., Rudzicz, F. & Rochon, E. (2013α) Using text and acoustic features to diagnose progressive aphasia and its subtypes. Proc. INTERSPEECH, 2177–2181.

[7] Le, D., Licata, K., Mercado, E., Persad, C. & Provost, E. M. (2014) Automatic Analysis of Speech Quality for Aphasia Treatment, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014.

[8] Le, D. & Provost, E. M. (2016) Improving Automatic Recognition of Aphasic Speech with AphasiaBank. Interspeech 2016, 2681–2685.

[9] Qin, Y., Wu, Y., Lee, T. & Hin, A. P. (2020) An End-to-End Approach to Automatic Speech Assessment for Cantonese-speaking People with Aphasia. J Sign Process Syst 92: 819–830.

[10] Abad, A., Pompili, A., Costa, A., Trancoso, I., Fonseca, J., Leal, G., Farrajota, L., & Martins, I. P. (2013). Automatic word naming recognition for an on-line aphasia treatment system. Computer Speech & Language, 27(6), 1235–1248.

[11] Linebarger, M. C., Schwartz, M. F., & Kohn, S. E. (2001). Computer-based training of language production: An exploratory study. *Neuropsychological Rehabilitation*, *11*(1), 57–96.

[12] Wade, J. R. C., B. Petheram. (2001). Voice recognition and aphasia: Can computers understand aphasic speech? *Disability and Rehabilitation*, *23*(14), 604–613.

[13] Lee, T., Liu, Y., Huang, P.W., Chien, J.T., Lam, W.K., Yeung, Y.T., Law, T.K., Lee, K.Y., Kong, A.P.H., Law, S.P. (2016). Automatic speech recognition for acoustical analysis and assessment of cantonese pathological voice and speech. In Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6475–6479). Shanghai: IEEE.

[14] Le, D., Licata, K., & Provost, E. M. (2017). Automatic Paraphasia Detection from Aphasic Speech: A Preliminary Study. *Interspeech 2017*, 294–298.

[15] Le, D., Licata, K., Provost, E.M. (2018). Automatic quantitative analysis of spontaneous aphasic speech. Speech Communication, 100, 1–12.

[16] MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25: 1286-1307.

[17] Pompili, A., Abad, A., Trancoso, I., Fonseca, J., Martins, I. P., Leal, G., & Farrajota, L. (2011). An on-line system for remote treatment of aphasia. *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, 1–10.

[18] Meinedo, H., Caseiro, D., Neto, J., & Trancoso, I. (2003). AUDIMUS.MEDIA: A Broadcast News Speech Recognition System for the European Portuguese Language. In N. J. Mamede, I. Trancoso, J. Baptista, & M. das Graças Volpe Nunes (Eds.), *Computational Processing of the Portuguese Language* (Vol. 2721, pp. 9–17). Springer Berlin Heidelberg.

[19] Ballard, K. J., Etter, N. M., Shen, S., Monroe, P., & Tan, C. T. (2019). Feasibility of Automatic Speech Recognition for Providing Feedback During Tablet-Based Treatment for Apraxia of Speech Plus Aphasia. *American Journal of Speech - Language Pathology (Online); Rockville*, *28*(2S), 818–834.

[20] Barbera, D.S/, Huckvale, M., Fleming, V., Upton, E., Coley-Fisher, H., Doogan, C., Shaw, I., Latham, W., Leff, A.P. & Crinion, J. (2021). NUVA: A Naming Utterance Verifier for Aphasia Treatment. *Computer Speech & Language*, 69, 101221.

[21] Robinson, T., Fransen, J., Pye, D., Foote, J., & Renals, S. (1995). Wsjcam0: A British English Speech Corpus For Large Vocabulary Continuous Speech Recognition. In Proc. ICASSP 95, 81–84.

[22] Green, J.R., MacDonald, R.L., Jiang, P.-P. Cattiau, J., Heywood, R., Cave, R., Seaver, K., Ladewig, M.A., Tobin, J., Brenner, M.P., Nelson, P.C. Tomanek, K. (2021) Automatic Speech Recognition of Disordered Speech: Personalized models outperforming human listeners on short phrases. *INTERSPEECH 2021*, 30 August – 3 September, 2021, Brno, Czechia

[23] Faliagka, E., Panagiotou, C., Antonopoulos, K., Stavropoulos, V., Karadimas, D., Keramidas, G., ... & Voros, N. (2022, June). An architecture for handling heterogenous services in a speech and language therapy platform. In 2022 11th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-6). IEEE.

[24] Rouvalis, G. (2023). GREC: Multi-domain Speech Recognition for the Greek Language. Unpublished Master Thesis.