World Journal of Information Systems 2(1) 70-77
DOI: https://doi.org/10.17013/wjis.v2i1.34

Article

Toward the development of software to classify speech sound disorders

Pere Porcuna [1*], David Bañeres [1] and Llorenç Andreu [2]

[1] Internet Interdisciplinary Institute

[2] eHealth Center

Universitat Oberta de Catalunya (UOC) Barcelona, Spain

***** Correspondence: pporcuna@uoc.edu

(Article history: Received December 29, 2024; Received in revised form January 28, 2025; Accepted February 27, 2025; Available online March 15, 2025)

Abstract: The detection of children with speech sound disorders is a research area widely explored in many languages. Early detection and treatment positively affect the child's learning development. This work aims to present a part of a project called ROMULO, which allows the evaluation, diagnosis, and treatment of children with speech sound disorders in Catalan where standard techniques based on signal processing cannot be applied due to the lack of public corpus. Specifically, we present an error classifier developed with a rule-based method, to detect speech problems codified in text at the word, syllable, and segmental levels. The approach can efficiently detect and classify speech sound errors, helping the speech therapist automate this process. Preliminary results show promising results on the detection, reaching an average success rate of 74.6%.

Keywords: Speech Sound Disorders, Speech sound errors, Phonological processes, Rule-based method, Error Classifier.

1. Introduction

Speech sound disorders (SSD) encompass a range of difficulties related to the perception, production, and organization of speech sounds and segments. It is prevalent in the pediatric population, impacting approximately 10% to 15% of preschoolers and 6% of school-aged children [1].

There are several types of speech sound errors that children with SSD produce. These errors can be classified differently [2]. One of them categorizes speech sound errors according to typical and atypical sound changes, often denoted as phonological processes [3], [4], [5].

Phonological processes are patterns of speech sound errors that children typically use to simplify speech while they learn to talk. An SSD occurs when phonological processes persist beyond the age when most typically developing children have stopped using them or when the processes used are much different than expected [6].

Different classifications of phonological processes have been created in diverse languages (e.g. [3], [4], [5], [7]). In Catalan, a romance language spoken in the western part of Spain by about 9 million people [8], there is one standardized assessment test [9] that includes an exhaustive classification and that is very used among speech and language therapists. Their

© 2025 ITMA. All rights reserved.

classification categorizes phonological processes at three levels: word, syllable and phoneme. At the word level, phonological processes are organized related to structure changes (syllable omissions or inversions) and assimilation (one sound becomes the same or similar to another sound in the word). At the syllable level, phonological processes can affect the syllable structure (reductions, deletions, epenthesis, coalescence, and metathesis). Finally, at the phoneme level, phonological processes can generate substitutions (one sound is systematically substituted for another sound) that can affect three different sound traits (place, manner, and voice), generating different errors (fronting, backing, stopping, affrication, voicing, devoicing, among others).

In [10], we presented a new technology-based solution, denoted ROMULO (RObot MUltifunctional for LOgopaedics), as a support tool to assist speech and language therapists in identifying and treating children with Speech Sound Disorders. The proposed solution consisted of different modules: a) a module to assess speech development, b) a module to classify and diagnose speech sound errors, and c) a module to recommend a treatment for this patient. Module a) was developed and comprised an assessment test based on 33 words to analyze the phonetic and phonological development in Catalan. This work focuses on module b), that is, the development of the system to detect and classify speech sound errors.

This paper is structured as follows. Section II introduces related work on computer-based speech therapy and speech disorder recognition. The technology-based solution, ROMULO, is described in Section III. The error classifier is presented in Section IV, while Section V summarizes its preliminary results. Finally, Section VI discusses the conclusions and future work.

2. Preliminaries

2.1. Speech Sound Disorders (SSD)

SSD refers to a collection of communication disorders in which children consistently struggle to articulate words or sounds accurately. Children with speech sound disorder may encounter challenges in either their understanding of speech sounds or their ability to coordinate the necessary movements for speech. These communication difficulties can hinder children's development by restricting their capacity to actively engage in social, academic, or occupational activities [2].

SSD often coexist with other neurodevelopmental conditions, making assessing, diagnosing, and selecting effective interventions a complex, time-consuming, and expensive process [11]. Typically, speech and language therapists provide intervention for these disorders through face-to-face sessions in their offices or private consultations. Due to the resistant nature of speech and language disorders, numerous therapy sessions over an extended period are required. Consequently, families bear significant financial burdens as public coverage for these services is limited in most countries.

2.2. Speech Sound Disorders and the Technology

Technology implementation has enabled the detection, assessment, monitoring, diagnosis, and treatment of various neurodevelopmental disorders in recent years [11], [12]. Additionally, technology can reduce inefficiencies, improve access to information, reduce costs, increase quality, and personalize health for users and relatives [13], [14].

Digital tools have been utilized to make assessment and treatment more convenient. Mobile/tablet apps, robots, serious games, computerized tests, videos, and virtual reality are the most used technologies. Many of these digital tools, encompassed in the field of

© 2025 ITMA. All rights reserved.

Computer-Aided Speech Therapy [15], are computer applications that focus on some of the previously described speech disorders.

One of the technologies being applied in the field of SSD is Speech Technology (ST), which allows the processing of spoken words. Inside the ST area, we have the Speech Recognition area where we find the Automatic Speech Recognition (ASR) techniques.

Automatic Speech Recognition focuses on converting voice into text. Multiple approaches to performing this transformation exist, but one of the most remarkable is Speech-to-Text (STT) techniques [16]. ASRs can be used to evaluate and recognize SSDs based on algorithms, patterns, or models by identifying known speech errors and making the detection of these speech problems possible [17].

The ASRs can detect speech problems by using frequency parameters of the speech. Combining parameters such as Mel Frequency Cepstral Coefficients (MFCC) of the speech with a Gaussian Mixed Model (GMM), a speech error classifier can be developed to detect speech errors [18], [19]. Also, other techniques have been developed by combining MFCCs with the two derivatives of the coefficients and the least squares support vector machine [20].

Using an ASR system with an acoustic model of a deep neural network (DNN) also allows classifying and quantifying the SSDs using constituent speech phonemes [21]. This is possible by checking the mismatches between the speech parameters with and without pathologies. However, background noise plays a decisive role in this type of technique to perform a reliable recognition.

These techniques have been used as tools to help people with SSD. Among the solutions for these children, *the Vocaliza* app [22] is a tool that allows to train speech in concrete articulation habits employing words or short sentences. The tool offers four activities with three language levels: phonological, semantic, and syntactic. ASR techniques are used to accept or reject the word pronunciation and evaluate it through a word-level Pronunciation Verification algorithm.

Another tool [23] belonging to the ST field allows the improvement of the articulation of words and sounds. In particular, it uses speech signal processing and animations. The tool detects speech segments using a Voice Activity Detector (VAD). This signal is converted into lines on the screen, and the children can draw lines on the screen through their voices.

Although these works obtained successful results in error detection, they are developed in languages with extensive research on voice signal recognition. Minority languages like Catalan lack public corpora to develop such recognition models. Additionally, this problem is even worse *when* detecting children's SSDs. In the following sections, we present our proposal for *detecting SSDs and, specifically, developing* an error classifier of the speech codified in the text as a part of this new detection approach.

3. Approach

This work focuses on functional disorders such as articulation and phonological disorders. Articulation disorders focus on errors (e.g., distortions and substitutions) in producing individual speech sounds. Phonological disorders focus on predictable and rule-based errors (e.g., fronting, stopping, and final consonant deletion) that affect more than one sound. As mentioned, ROMULO aims to detect, diagnose, and recommend treatments for such disorders. To meet such objectives, the system has been divided into three modules to deal with these features:

- Detection: The system will evaluate the patient through the patient's pronunciation of a series of 33 words. The module will analyze the pronunciation of each word with the correct pronunciation.
- Diagnose: Once the detection has been performed, this module will assess the gathered information to diagnose the patient's speech disorder.

© 2025 ITMA. All rights reserved.

- Treatment: The system should be capable of recommending a personalized treatment that can be applied by the therapist. This recommendation module will decide on a treatment based on the evidence provided by the previous modules.

As described in Section II.B, different approaches based on ASR recognition have been successfully applied to detect SSDs. However, the ASR techniques have some drawbacks [9], mainly in minority languages: 1) the inexistence of public corpora makes the creation of identification models difficult, and 2) the utilization of adult speech for children is ineffective mainly on younger children due to the divergence between the speech characteristics.

To surpass these drawbacks, a new approach has been designed for ROMULO. The detection module architecture is depicted in Figure 1. The configuration and diagnosis stages share some components. Both stages use the recording application designed in [10] to capture new children's speech models. The recording application has two modes: configuration and diagnosis. The first mode is used to create a new children's speech corpus; meanwhile, the second one detects SSDs for new patients.

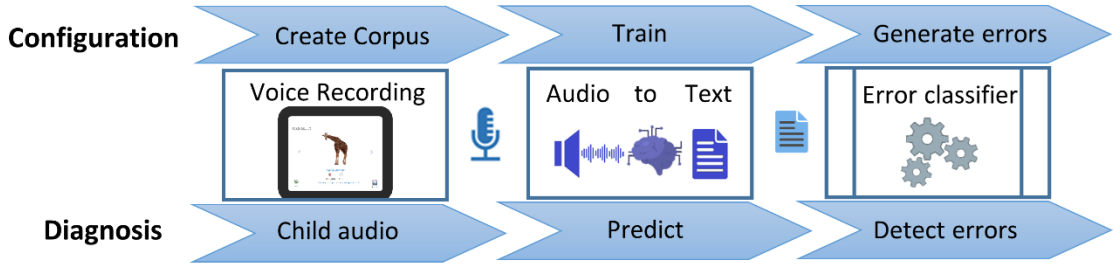


Figure 1. Detection module architecture

The second component will use a neural network to obtain the audio's text transcription, which will be trained using the developed speech corpus. In the diagnose mode, this neural network will be used to predict the transcript text of the child's audio.

Due to the complexity of this second task and while the corpus is being built, we started developing the third component to detect and diagnose the potential SSD based on the transcript text. A generic error classifier component has been designed to identify all possible errors for a word and detect speech errors on this word.

4. Error Classifier Component

The error classifier based on the rule method component, focuses on identifying an articulation and phonological disorder from the transcript text. The component expects as input, on the one hand, the word that the therapist is working with the patient, for example, "melo"[melon] and on the other hand, the pronounced word by the patient is "lelo". As output, the component determines whether the second word is correct or incorrect depending on the first word entered. If the second word, the one pronounced by the patient, is incorrect, the component provides the type of error detected. E.g. If the patient pronounces “lelo” instead of “melo”, the component will report the error of type "Assimilation of articulation mode of the liquid letters".

The classifier can identify 63 articulation-type errors at the word, syllable and phoneme levels. Specifically, the component will classify the speech problems at the level of word (word structure changes and assimilations), syllable (reductions, omissions, epenthesis, coalescence, and syllable metathesis) and phoneme (omissions, additions, distortions and substitutions).

© 2025 ITMA. All rights reserved.

The classifier also has two functional modes. The set of words being handled is introduced during the configuration mode, seeking all potential errors for each word. As a result, the diagnosis mode gives the error in the input word (if it exists).

Since the articulation errors are based on fixed identification rules (i.e., one letter changes to another, a letter is omitted), a systematic algorithm has been designed to address each type of error. In the configuration mode, the errors are artificially generated on each word by applying mathematical operations such as permutation, combination, and the Cartesian product. Thus, the classifier applies a set of mathematical operations for each known error to generate the potential erroneous word. Then, each generated word is tagged with the error (or errors) which produced the erroneous word. All these new words are created through the rules and patterns defining each type of error.

For instance, at the word level, in articulation mode, an error occurs in people with speech problems associated with the assimilation of occlusive phonemes. These difficulties can create changes in words such as "popa" instead of "sopa" [soup] or "pupa" instead of "lupa" [magnifying glass]. In this case, the occlusive letter "p" replaces the non-occlusive ones, i.e., "s" for the word "sopa" [soup] and "l" for the word "lupa" [magnifying glass], respectively.

For small words, exploring the possible erroneous words is straightforward. However, the complexity may increase on larger words depending on the letters used. Thus, we designed a dynamic programming approach to explore all possible errors efficiently. The word is initially segmented at the syllabic and phoneme levels. Then, the operation set for each error is applied to the segments. Note that, by the nature of the dynamic programming approach, applying different operation sets may lead to the same erroneous word but tagging the word with different error types. Additionally, further processing is avoided when reaching an already processed word since the algorithm has already applied the error operation sets on such words.

Note that the complexity of the classifier is during the configuration process since the diagnosis only queries for a word, whether it is an erroneous typing, and which error/errors are associated with it.

5. Preliminary Results

The rules and patterns defining each of the 63 errors have been codified with the Python programming language to develop the error classifier. The mathematical operations have been used on each word to generate all the possible diction errors.

Thus, to verify the correct operation of the developed error classifier, we have undergone a test by introducing five words of the 33 possible words discussed above. Note that we performed this experiment to see the initial performance of the error classifier. These words have been taken from the *A-RE-PA assessment test* [9] and from the book: *Avaluació del desenvolupament fonològic en nens catalanoparlants de 3 a 7 anys* [24] where different errors are also provided for each of the explored words.

The algorithm generated all possible diction errors in the configuration mode. The diagnosis mode was used to check whether the algorithm could find the correct error. The results are summarized for the five words in Table 1 where the explored words, the number of generated erroneous words, the number of tested errors, and the success rate are detailed.

As we can observe, the error classifier cannot find all tested errors. On the one hand, we identified a mismatch problem between the codified letter and some associated sounds (i.e., the coding of some phonemes). The solution involves refining the word generation for the correct codification of some Catalan phonemes.

© 2025 ITMA. All rights reserved.

Table 1. Preliminary Results

Words	Nº of generated erroneous words	Nº of tested errors	Success rate (%)
Cadira	1484	5	80
Casa	665	4	80
Melo	348	5	80
Reina	349	3	33
Sopa	688	5	100

On the other hand, we spotted some misinterpretations of the rules and patterns that generated some false negatives. The solution requires changes in the set of mathematical operations for some rules that generate the words with these errors.

6. Conclusion and Future work

In this article, we have presented the development of an error classifier that allows us to detect and diagnose articulation mode speech errors codified in text.

We have presented preliminary results on the success rate concerning a series of words. Although our method can be improved, we expect to refine the approach with the remaining words in successive tests.

As future work, we will focus on creating a corpus of pronunciations and using an ASR with a DNN to convert the speech to text more efficiently than what we can get through basic ASR. This will provide a complete approach to evaluating and diagnosing the potential SSD for a child based on voice recordings.

Acknowledgments: This research was supported by Proof of Concept Call 2023 from the eHealth Center (UOC).

Conflicts of Interest: The authors declare no conflict of interest

References

[1] R. Waring and R. Knight, “How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems,” International Journal of Language and Communication Disorders, vol. 48, no. 1. 2013. doi: 10.1111/j.1460-6984.2012.00195.x.

[2] R. Waring and R. Knight, “How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems,” International Journal of Language and Communication Disorders, vol. 48, no. 1. 2013. doi: 10.1111/j.1460-6984.2012.00195.x.

[3] M. L. Edwards, “In Support of Phonological Processes,” Lang Speech Hear Serv Sch, vol. 23, no. 3, 1992, doi: 10.1044/0161-1461.2303.233.

© 2025 ITMA. All rights reserved.

[4] L. Menn, “Phonology: Applications in Communicative Disorders. Mary Louise Edwards & Lawrence D. Shriberg. San Diego, Calif.: College-Hill Press, Inc., 1983. Pp. xx + 413 1 ,” Appl Psycholinguist, vol. 5, no. 3, 1984, doi: 10.1017/s0142716400005154.

[5] L. M. Laila Khan, “A Review of 16 Major Phonological Processes,” Lang Speech Hear Serv Sch, vol. 13, no. 2, 1982, doi: 10.1044/0161-1461.1302.77.

[6] F. Brosseau-Lapréa and E. Roepke, “Speech errors and phonological awareness in children ages 4 and 5 years with and without speech sound disorder,” Journal of Speech, Language, and Hearing Research, vol. 62, no. 9, 2019, doi: 10.1044/2019_JSLHR-S-17-0461.

[7] P. Grunwell, “David Ingram, Phonological disability in children. London: Edward Arnold, 1976. Pp. xvii + 167.,” J Child Lang, vol. 4, no. 1, 1977, doi: 10.1017/s0305000900000544.

[8] David M. Eberhard, G. F. Simons, and C. D. Fennig, “Ethnologue: Languages of the World. Twenty-fourth edition.,” Dallas, Texas : SIL International.

[9] M. S. Raventós and E. M. A. Mediavilla, A-RE-PA. Anàlisi del retard de la parla. in BIBLIOTECA UNIVERSITÀRIA. Publicacions i Edicions de la Universitat de Barcelona, 2004. [Online]. Available: https://books.google.es/books?id=RiBNael1pEwC

[10] P. Porcuna, L. Andreu, and D. Baneres, “TOWARDS AN INTELLIGENT SYSTEM FOR ANALYSING CHILDREN’S SPEECH DISORDERS,” in INTED2023 Proceedings, 2023. doi: 10.21125/inted.2023.1593.

[11] A. Z. Valentine, B. J. Brown, M. J. Groom, E. Young, C. Hollis, and C. L. Hall, “A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: A map of the current evidence,” Clinical Psychology Review, vol. 80. 2020. doi: 10.1016/j.cpr.2020.101870.

[12] G. Dawson and G. Sapiro, “Potential for Digital Behavioral Measurement Tools to Transform the Detection and Diagnosis of Autism Spectrum Disorder,” JAMA Pediatrics, vol. 173, no. 4. 2019. doi: 10.1001/jamapediatrics.2018.5269.

[13] WHO, “Building foundations for ehealth Progress of Member States,” …, Report of the WHO Global Observatory for ehealth, 2006.

[14] H. Oh, C. Rizo, M. Enkin, and A. Jadad, “What is eHealth (3): A systematic review of published definitions,” Journal of Medical Internet Research, vol. 7, no. 1. 2005. doi: 10.2196/jmir.7.1.e1.

[15] M. Shahin, U. Zafar, and B. Ahmed, “The Automatic Detection of Speech Disorders in Children: Challenges, Opportunities, and Preliminary Results,” IEEE Journal on Selected Topics in Signal Processing, vol. 14, no. 2, 2020, doi: 10.1109/JSTSP.2019.2959393.

[16] V. Bhardwaj et al., “Automatic Speech Recognition (ASR) Systems for Children: A Systematic Literature Review,” Applied Sciences (Switzerland), vol. 12, no. 9. 2022. doi: 10.3390/app12094419.

[17] J. Levis and R. Suvorov, “Automatic speech recognition,” The encyclopedia of applied linguistics, 2012.

© 2025 ITMA. All rights reserved.

[18] L. Verde, G. De Pietro, and G. Sannino, “Voice Disorder Identification by Using Machine Learning Techniques,” IEEE Access, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2816338.

[19] J. Liu et al., “Speech Disorders Classification in Phonetic Exams with MFCC and DTW,” in Proceedings - 2021 IEEE 7th International Conference on Collaboration and Internet Computing, CIC 2021, 2021. doi: 10.1109/CIC52973.2021.00015.

[20] N. Souissi and A. Cherif, “Dimensionality reduction for voice disorders identification system based on Mel Frequency Cepstral Coefficients and Support Vector Machine,” in Proceedings of 2015 7th International Conference on Modelling, Identification and Control, ICMIC 2015, 2016. doi: 10.1109/ICMIC.2015.7409479.

[21] Y. Liu, T. Lee, T. Law, and K. Y. S. Lee, “Acoustical assessment of voice disorder with continuous speech using asr posterior features,” IEEE/ACM Trans Audio Speech Lang Process, vol. 27, no. 6, 2019, doi: 10.1109/TASLP.2019.2905778.

[22] O. Saz, S. C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodríguez, “Tools and Technologies for Computer-Aided Speech and Language Therapy,” Speech Commun, vol. 51, no. 10, 2009, doi: 10.1016/j.specom.2009.04.006.

[23] W. R. Rodríguez Dueñas, C. Vaquero, O. Saz, and E. Lleida, “Speech technology applied to children with speech disorders,” in IFMBE Proceedings, 2008. doi: 10.1007/978-3-540-69139-6_64.

[24] Laura. Bosch Galceran, Avaluació del desenvolupament fonològic en nens catalanoparlants de 3 a 7 anys. in Documents (Universitat de Barcelona. Institut de Ciències de l’Educació) ; A 69. Barcelona: Promociones Publicaciones Universitarias, 1987.

© 2025 ITMA. All rights reserved.