Computer Speech & Language 65 (2021) 101130

Contents lists available at ScienceDirect

Computer Speech & Language

journal homepage: www.elsevier.com/locate/csl

Towards a speech therapy support system based on phonological processes early detection

Maria Helena Franciscatto[a], Marcos Didonet Del Fabro[a], João Carlos Damasceno Lima[b], Celio Trois[b], Augusto Moro[b], Vinícius Maran[c*], Marcia Keske-Soares[d]

[a] Departamento de Informática, Universidade Federal do Paraná, Curitiba, Brazil
[b] Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria, Brazil
[c] Laboratory of Ubiquitous, Mobile and Applied Computing (LUMAC), Universidade Federal de Santa Maria, Rua Júlio de Castilhos, 342, Sala 4, Centro, Cachoeira do Sul 96501-000, Brazil
[d] Centro de Ciências da Saúde, Universidade Federal de Santa Maria, Santa Maria, Brazil

ARTICLE INFO

Article History:
Received 5 September 2019
Revised 24 April 2020
Accepted 10 June 2020
Available online 24 June 2020

2010 MSC:
00-01
99-00

Keywords:
Speech therapy
Speech sound disorders
Situation-awareness
Case-based reasoning
Phonological processes
Machine learning

ABSTRACT

Phonological disorders are characterized by substitutions, insertion and/or deletions of sounds during the process of language acquisition, which are known as Phonological Processes (PPs). In the speech therapy domain, an early identification of PPs allows the diagnosis and treatment of various pathologies and may improve clinical tasks, however, there are few proposals that focus on the identification of PPs for supporting Speech-Language Pathologists (SLPs). Recent research applied Case-Based Reasoning (CBR) in medical domain to identify specific cases related to patients. Situation-Awareness (SA) is a technique that allows computing systems to adapt itself and respond to users or other systems according to environment information. Moreover, there is no indicative in related literature of CBR and SA being used for detecting PPs that may occur in pronunciation. In this paper, we introduce the union of SA and CBR, tied to machine learning algorithms for proposing a system to predict PPs, supporting specialists in their clinical decisions. To evaluate the system, we implemented it in a software architecture prototype and evaluated the prototypes using a knowledge base containing near one hundred thousand audio files, collected from more than 1,000 pronunciation assessments. The evaluation of the prototypes showed an accuracy over 93% in the prediction of PPs, resulting in a efficient tool for clinical decision support and therapeutic planning. We also presented a direct qualitative comparison between our approach and related work.

1. Introduction

The speech acquisition is a complex process where the production of phonemes is characterized by continuities and discontinuities in the childs path toward mature production of the segments and structures of the ambient language (Rvachew and Bernhardt, 2010). Some error patterns in speech, also known as *phonological processes* (PPs), are typical in the language acquisition and occur when the child tries to adapt his/her speech to adult speech (Ceron et al., 2017). The presence of these error patterns may indicate disordered or deviant development (Dodd, 2013), in which children often experience difficulties with literacy,

*Corresponding author.
E-mail address: vinicius.maran@ufsm.br (V. Maran).

https://doi.org/10.1016/j.csl.2020.101130
0885-2308/

including decoding, reading comprehension, and the production of written text (Joye et al., 2019). In this context, the monitoring of PPs in pronunciation is vital for the identification of delayed or disordered phonological development in children, which compromise biological, psychological, and social factors of the individual (Martín-Ruiz et al., 2013).

Knowledge of phonological development has great significance in the clinical population to determine whether a child has a phonological impairment and needs intervention (Abou-Elsaad et al., 2019). However, several barriers are found in speech therapy, especially in developing countries: the lack of resources, rehabilitation centers and qualified personnel to provide intervention services, the work overload of professionals, and the poor development or adaptation of assistive technologies according to patients needs (Robles-Bykbaev et al., 2016). In order to face these difficulties, many studies have proposed computer-based speech therapy systems or virtual speech therapists (VSTs) for people with speech disorders (Chen et al., 2016). Also, the related literature presents studies focused on the development of speech recognition systems (Abad et al., 2013; Grossinho et al., 2016; Caballero-Morales and Trujillo-Romero, 2014; Bolaños et al., 2011) for supporting SLPs in clinical environment.

Although there are many studies that focus on risk prevention and patient monitoring, there are few proposals in the speech therapy domain that use knowledge modeling to improve tasks such as diagnosis, therapy planning, and therapeutic intervention (Chuchuca-Méndez et al., 2016). Moreover, the proposals generally are not concerned with identifying and inferring error patterns in the speech. The identification of PPs in pronunciation may aid the Speech-Language Pathologists (SLPs) in extending traditional speech therapy, identifying relevant and recurrent patterns in the speech, and predicting behaviors according to these patterns. Consequently, it is possible to provide support to these professionals through optimization of therapeutic tasks, planning of therapy, and recommendation of necessary actions.

On these grounds, in this paper we present an architectural model that applies Situation-Awareness (SA) and Case-Based Reasoning (CBR) for automatically identifying PPs in children's speech. SA has been recognized as an important resource in many different domains, and involves collecting contextual information about the environment, making decisions based on this collection, acting according to decisions, and gathering feedback from the environment for making better decisions in the future (Kokar and Endsley, 2012). CBR can also be a favorable choice in speech and health contexts, since this methodology has good learning capabilities, and its ability to solve problems improves as new cases are stored in history log files or in databases (Husain and Pheng, 2010).

According to a recent published systematic review (Franciscatto et al., 2018a), there is not yet a proposal in the literature that applies CBR and SA for automatically identifying Phonological Processes (PPs) and suggesting them to the SLPs. These theories applied together may represent powerful tools to indicate PPs with accurate precision, since the knowledge base is constantly updated and acquires capabilities over time as the diagnosed cases are confirmed.

Thus, the architectural model presented in this paper applies these theories in two main modules (*Capture Module* and *Service Module*) for automatically identifying PPs in children's speech. The Capture Module is responsible for SA's *perception* phase, through speech data collection (audio recording), while the Service Module is responsible for SA's *comprehension* phase, through classification of pronunciation as correct or incorrect. In this module, the four activities of the CBR cycle (retrieval, reuse, review, and retain) are applied for prediction and recommendation of Phonological Processes, completing the SA's *projection* level in the approach.

We evaluated our proposal with a Phonological Knowledge Base containing speech samples collected from 1,114 evaluations performed with 84 Portuguese words in the last three years. These target words were chosen to facilitate the identification of PPs, where each word has a score associated with the possible PPs for the acquisition of the Portuguese language. The results showed an average accuracy of 92.5% for classifying pronunciations as correct or incorrect, in concordance with the therapist evaluation. The pronunciation results were used for predicting the PPs; in this step, our approach achieved an accuracy over 93%, showing that it is possible to predict PPs for clinical decision support.

The present paper is structured as follows. In Section 2, the main concepts related to this research are presented, as well as the motivation of this work. In Section 3 we report our situation-aware and case-based approach for speech therapy support, while Section 4 shows the evaluation, results and discussion. Lastly, in Section 5 we present the conclusions and final remarks.

2. Background and motivation

This section presents the main concepts and related work covering knowledge areas involved in this study, including Situation-Awareness, situation-aware systems in health care and speech-language domain, Case-Based Reasoning and Phonological Processes. Also, the motivation for this work is presented, with regard to promising theories and unexplored research possibilities.

2.1. Situation-awareness

Situation-Awareness (SA) is a term that expresses *"the perception of the elements of the environment within a volume of time and space, the understanding of its meaning and projection of its effects in the near future"* (Endsley, 1995). This definition suggests that through SA, applications and systems are able to understand surrounding events and, through them, design actions that can offer benefits to human life, from the simple task of providing a personalized service to make effective actions in risk scenarios.

Endsley (1995) proposed a model that illustrates how the elements of an environment are treated for situation recognition (Fig. 1). The model is widely accepted in the literature and it is based on three main levels to obtain Situation-Awareness. In the first level, ***Perception***, all spatial elements in the environment and their current states must be evaluated according to their

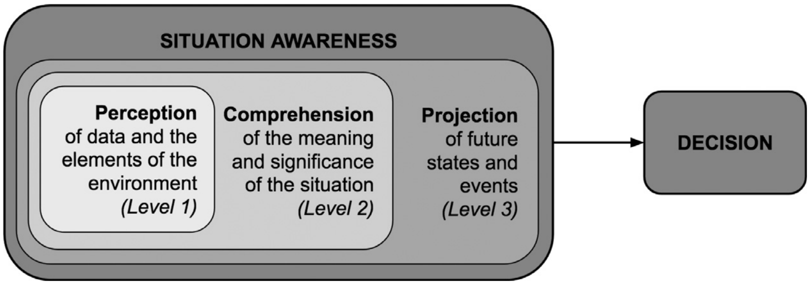

Fig. 1. SA model. Adapted from (Endsley, 1995).

relevance to decision-making process (Endsley and Garland, 2000). These elements are interpreted in a second level, ***Comprehension***, which establishes relationships among the elements and understands how these information will impact the user's goals. The comprehension level involves processes of pattern recognition, interpretation, and evaluation of acquired information, in order to create a world picture that is meaningful to user actions. Finally, the ***Projection*** level (the highest level in the SA model), is achieved from the understanding of the situation, which allows to anticipate or predict the near future, taking necessary actions to respond to it.

One of the main focus of SA researches is on the design of intelligent systems that can support independent living and human security, recognizing risk situations and taking appropriate actions proactively. We can mention, for example, studies that propose the monitoring of the patient's health, where the information acquired from sensor-based technologies can be useful for health promotion and disease prevention, such as obesity caused by sedentary habits (Goto et al., 2013).

Regarding the speech therapy domain, it is observed that, generally, pediatricians do not have enough time to properly monitor the children' s neuro-evolutive and social development (Martín-Ruiz et al., 2013). This is motivating many researchers to propose automated solutions to face the challenges found in traditional therapy: In Robles-Bykbaev et al. (2016), the authors present a specialist system for automatic generation of therapeutic guidelines, which suggests the best activities or intervention strategies for a specific patient profile. In Abad et al. (2013), an automatic speech recognition technology was proposed, based on a hybrid recognizer and intended for patients with aphasia. The authors in Iliya and Neri (2016) presented a technique based on neural system to segment speech utterances, in order to detect inconsistencies in speech signals. Finally, in Le et al. (2016), the authors proposed an automated system to evaluate clarity, fluency, and prosody in the aphasic speech, providing a meaningful feedback to therapists.

The projection capability of a SA system can be useful for both, the patient and the clinician, as it allows the therapist to have a good view on the patient's condition and thus, making better informed decisions. However, as demonstrated in Fig. 2, SA Projection still represents a major challenge in speech therapy, since most systems proposed for the area do not fully support the professional's actions.

Estimating future situations and providing decision support represent characteristics of Situation-Awareness that allow systems to deal with the evolution of complex environments. Although SA functionalities are needed in many application domains, they become more urgent in areas that involve healthcare and prevention, especially in the speech therapy domain, where there are several research opportunities. It is important to emphasize, however, that Situation-Awareness does not come to replace the professional, but to offer assistance in the problems observed in the area, such as the screening of language disorders and therapeutic planning.

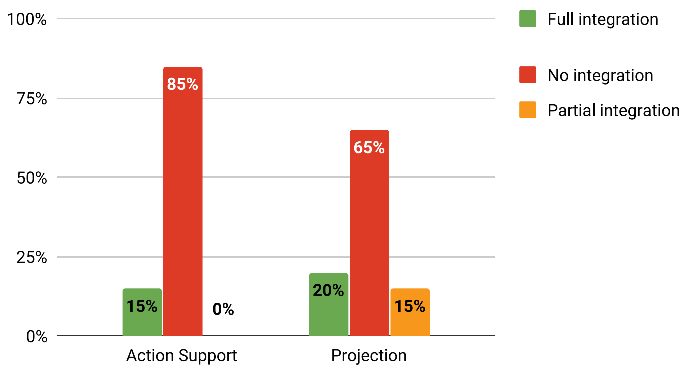

Fig. 2. SA criteria with low integration in speech therapy proposals (Franciscatto et al., 2018a).

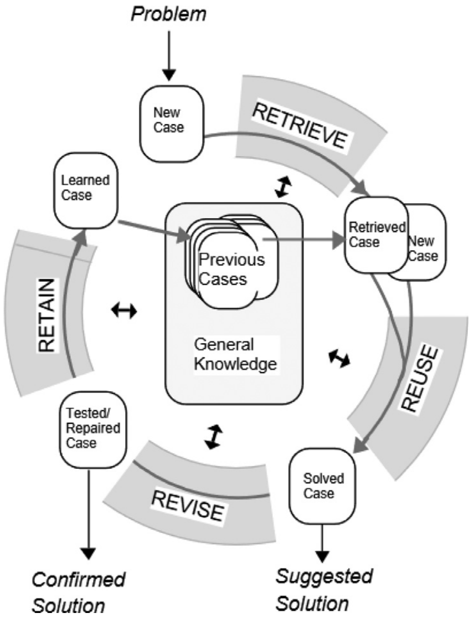

Fig. 3. CBR cycle (Aamodt and Plaza, 1994).

2.2. Case-based reasoning

Case-Based Reasoning (CBR) is a paradigm of Artificial Intelligence that solves problems using or adapting solutions to old problems (Riesbeck and Schank, 1989). This methodology draws on human reasoning for problem solving, since in real life, once we have learned a solution to a problem, we often try to reuse that solution in similar problems (Smiti and Elouedi, 2019).

The problem solving in the CBR methodology goes through a cycle of four activities (Fig. 3): *retrieving* cases that resemble the description of the problem, *reusing* an existing solution for a similar case, *reviewing* this solution in order to meet the new problem, and *retaining* this solution once it has been confirmed (Watson, 1999).

CBR has been applied in many different areas such as engineering, communication networks, manufacturing design, route finding, among others (Rao, 2017). Also, the clinical domain is one of the main areas of CBR application, where the problem is represented by the patient's symptoms and the solution is the diagnosis and/or treatment (Begum et al., 2011). In general, CBR has been studied as a knowledge-based systems approach to developing intelligent systems for facilitating knowledge maintenance and increasing problem-solving efficiency (Leake, 2015).

There are many advantages of applying CBR: the knowledge acquisition task is reduced, there is flexibility in knowledge modeling, it is possible to reasoning when domains are not completely understood or when data are imprecise, the cumulative experiences allow learning successes and failures over time, thus mistakes made in the past can be avoided (Rao, 2017). In addition, the methodology is compatible with other methods of Artificial Intelligence, i.e. Machine Learning, so that systems using CBR can take advantage of other techniques to improve their performance (Ahmed et al., 2012). We can mention, for example, the proposal in Yeow et al. (2014), that addresses the development of an Intelligent Forensic Autopsy Report System, which analyzes the case similarities by coupling CBR with a Naïve Bayes learner for feature-weights learning.

With respect to the speech therapy domain, the literature presents CBR for several tasks, e.g., generating natural language corpora (Fan and Kendall, 2005), performing intelligent processing for a speech-enabled e-Learning system (Azeta et al., 2009) and detecting depression in an individual from the properties of the pronunciation (Guillén and Usrey, 2005).

However, there is a lack of proposals that integrates Case-Based Reasoning as a method for solving problems and, specifically, identifying error patterns in speech, from which it is possible to optimize therapeutic tasks and the diagnosis of the specialist. In this sense, error patterns in speech will be addressed in the next section, also known as Phonological Processes.

2.3. Phonological processes

A phonological process (PP) is a mental operation that is applied in speech to substitute a class of sound or sound sequences that are difficult to the speech capacity of the individual, for an alternative class identical but lacking the difficulty property (Stampe, 1979).

Phonological acquisition research suggests that there is an underlying representation of the different speech sounds that needs to be acquired before proper articulation takes place, so during the phoneme acquisition process, children present omissions or substitutions for some phonemes that are difficult for them until the grammatical mapping of sounds gets settled, and these mispronunciations are called PPs (Fringi et al., 2015).

Table 1
Phonological processes in Portuguese acquisition. Adapted from (Yavas and Lamprecht, 1988).

Phonological Process (PP)	Definition	Phonemes (Ph)
Cluster reduction	Reduction of a consonant within the same syllable.	bruxa (witch) [brúʃa] → [búʃa] flor (flower) [flóχ] → [fó] ladrão (thief) [ladrə w] → [latə w] assoprei (I blew) [asopréj] → [asopéj]
Weak syllable deletion	Deletion of an unstressed syllable (pre- or post-tonically) from multisyllabic words, but also possible in disyllabics.	edifício (building) [edʒ ifísju] → [fíʃju] vejo (I see) [véʒ u] → [fé]
Final fricative[a] deletion	Deletion of phoneme /s/ in both syllable and word final positions.	escada (stair) [iskáda] → [ikáda] lápis (pencil) [lápis] → [ápi]
Final liquid[b] deletion	Deletion of a liquid in both syllable and word final position.	barco (boat) [bárku] → [páku] flor (flower) [flóχ] → [fó]
Intervocalic liquid deletion	Deletion of a liquid in intervocalic position.	tirou (took out) [tʃiró] → [tʃió] jacaré (alligator) [ʒ akarɛ] → [sakaɛ]
Word initial liquid deletion	Deletion of a liquid in word initial position.	roupa (clothes) [χópa] → [ópa] lápis (pencil) [lápis] → [ápi]
Labial assimilation	Change of a consonant to a labial when another labial occurred in the word.	nove (nine) [nóvi] → [mófi] girafa (giraffe) [ʒ iráfa] → [viráfa]
Backing	Substitution of a palato-alveolar fricative for a more anterior fricative.	segurando (holding) [sigurə ndu] → [ʃigurə du] soltou (let go) [soltó]-> [ʃotó]
Devoicing	Devoicing of any obstruent.	ovo (egg) [óvu] → [ófu] bico (beak) [bíku] → [píku] feijão (beans) [fejʒ ə w] → [feʃõ]
Fronting	Substitution of an alveolar fricative for a palato-alveolar fricative.	chamou (called) [ʃamó] → [samo] colégio (school) [kolɛ ʒ ju] → [kolɛ su]
Liquid substitution	Substitution of one liquid for another.	morreu (died) [moχéu] → [moléu] ovelha (sheep) [ovéʎa] → [avéra] zero (zero) [zə ru] → [sə lu]
Liquid gliding	Substitution of a liquid by a glide.	carteira (purse) [kartéra] → [katéja] amarela (yellow) [amarə la] → [amajə ja]
Plosivation	substitution of a fricative consonant for a plosive.	vaca (cow) [vaka] → [baka]
Intervocalic liquid nasalization	Substitution of a liquid by a nasal in intervocalic position.	cadeira (chair) [kadéra] → [katéna] carro (car) [káχu] → [kámu] eles (they) [élis] → [émis]
Intervocalic nasal gliding	Gliding of nasals in intervocalic position.	anos (years) [ə nus] → [ə ju] pequeno (small) [pikénu] → [pikéju]

[a] Fricative consonants are produced by narrowings of the air passage.
[b] Liquid consonants are characterized by obstruction of the expiratory current made with the tongue tip, and also by the production of vibrating sounds.

The identification of PPs is clinically important. A child who follows the normal course of development, albeit slowly, is less linguistically impaired than a child who produces deviant errors (Dodd and Lacano, 1989). In this context, phonological process analysis has a great influence on studies of phonological acquisition and phonological disorders, and it is currently the most widely used analytical procedure for the investigation of phonological development in children (Yavas and Lamprecht, 1988).

Automated methods of phonological processes analysis have also been investigated in the literature. In Sreedevi et al. (2013), the authors introduce an indigenous computerized software to assess the phonological processes in native Malayalam speaking children, which provides quick screening for speech sound error patterns. The authors in Fringi et al. (2015) investigate whether phonological processes are reflected in the performance of a baseline recognizer through systematic error patterns, so a statistical significance test is proposed to identify substitution errors in the childrens data that cannot be explained by the expected variation in the adult data.

Phonological processes have been found to occur in typically developing children in different languages, including Portuguese (Jesus et al., 2015). Considering the phonology of Brazilian Portuguese (BP), several differences can be identified between BP and other languages. For exemplifying, whereas English has 24 consonants, BP has 19 consonants, represented by /p/ , /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /ʃ/, /Z/, /m/, /n/, /η/, /χ/, /r/, /l/ and /L/ (Ceron et al., 2017). The phonological processes in BP can be grouped according to error patterns that are shared by most children. Some PPs described in Yavas and Lamprecht (1988) are presented in Table 1, along with substitutions examples in BP words.

2.4. Motivation

Expressive phonological difficulties can persist beyond the expected age, seriously impacting the life quality of the individual. People with developmental language disorders experience problems with the structural dimensions of language, including grammar and vocabulary (Joye et al., 2019). Similarly, correlations between impaired speech and orthographic skills have been found, since phonological awareness enables the child to understand and exploit the mappings between graphic symbols and sound structure of spoken language (Moll et al., 2014).

The combination of difficulties involving the perception and articulation of speech may impact even the individual's intelligibility and acceptability: the presence of a speech disorder degrades the person's role in society and precludes interactions in social activities (Chen et al., 2016). This may increase the risk of social, emotional, and/or academic challenges, especially in young people and children, who use communicative competence as a measure of peer popularity (Hitchcock et al., 2015).

Related studies have shown a wide range of speech disorders, including articulation disorder, phonological delay, chilhood apraxia of speech, among others (Dodd, 2014) that make up typical caseload for SLPs. Also, it was found that children with developmental language disorders produce more phonological processes in almost all age categories when compared to the performance of their typically developing age peers (Abou-Elsaad et al., 2019). The above-mentioned facts along with the speech disorders consequences on society motivate us to research how to identify phonological processes and use them to promote clinical support.

In this context, we propose the combination of Situation-Awareness and Case-Based Reasoning, which represent two concepts widely used in healthcare systems. While SA applies techniques of perception, comprehension, and projection for monitoring a situation and acting in advance, CBR is based on previous diagnostics to solve a problem and keep the knowledge base updated. Thus, the use of SA and CBR in speech therapy systems can offer great advances, from the perspectives of the user and the therapist.

Considering a patient with speech disorders, a system based on SA and CBR may be useful as an extension of traditional speech therapy, where mobile tools and automated speech analysis algorithms act as a virtual assistant, indicating personalized activities according to the characteristics and needs of the individual. A situation-aware and case-based system could monitor the pronunciation evolution over time and decide which exercises are best for the child to perform as a complement to the face-to-face therapy, based on effective solutions taken in the past. In addiction, if an unexpected situation is identified, the system could analyze the risk of future illnesses comparing the current case with past similar cases, thus sending alerts to parents or therapist.

Recent studies have focused on the development of speech recognition systems, assessment of pronunciation issues, remote administration of speech exercises, assessment of speech intelligibility, and prediction of speech disorders (Franciscatto et al., 2018a). However, the identification of phonological processes is little addressed in these studies, and involves non trivial tasks such as comprehension about phonemes acquisition, identification of error patterns and discovery of how these patterns can be used to infer speech disorders. Moreover, there have been few studies on phonological processes in Portuguese language (Jesus et al., 2015), which reinforces our motivation and justifies the development of a system for predicting PPs and supporting speech therapists.

From these considerations, the next section presents an architecture that integrates Situation-Awareness and Case-Based Reasoning for speech therapy support.

3. Situation-aware and case-based architecture for speech therapy support

As seen previously, traditional speech therapy presents some obstacles which include, mainly, the lack of specialists in the area and the difficulty of performing adequate patient monitoring. We believe that a situation-aware approach can mitigate these issues, thus the proposed software architecture aims to integrate aspects of the SA Model (Endsley, 1995) through Perception, Comprehension, and Projection levels. Moreover, the architecture includes the CBR methodology for expanding the knowledge base and acquiring capabilities over time. The proposed software architecture (Fig. 4) is composed by two modules: Capture Module and Service Module.

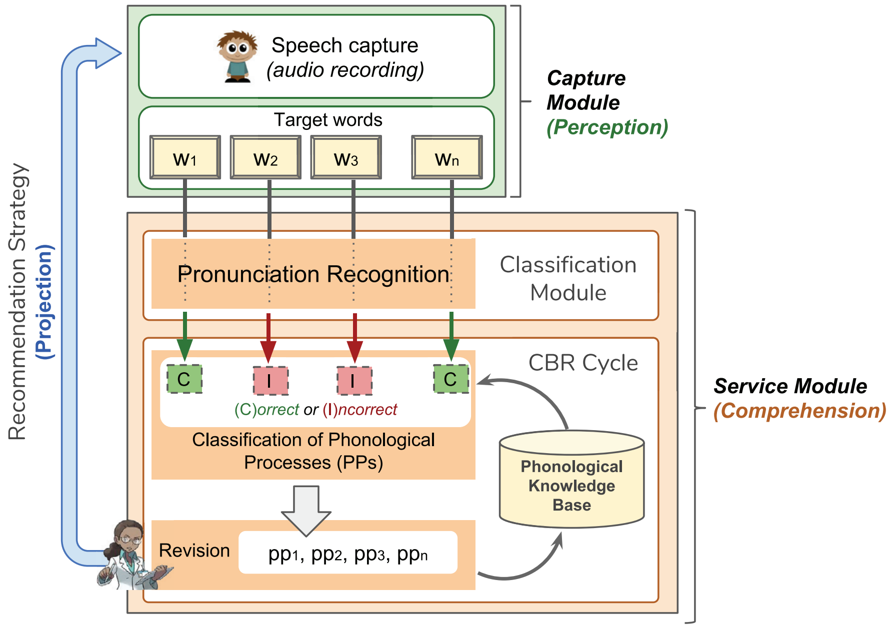

Fig. 4. Software architecture.

3.1. The capture module

Capture Module is responsible for collecting speech data from the target audience, it can be installed on any mobile device. When collecting data from the observed environment, the system is able to reach the first level of SA - *Perception* - and use the acquired information to classify the patient as a healthy individual or an individual with speech disorder, pinpointing his phonological processes. The captured data go through a series of steps, including preprocessing the speech signal, converting the audio to images (visual representations) for detecting speech patterns, feature extraction, and classification of the pronunciation as correct or incorrect. In this way, after extracting data from the Capture Module, the data are sent to the Service Module, discussed below.

3.2. The service module

The Service Module is responsible for taking each patient evaluation analyzed in the previous module as input for Case-Based Reasoning. This process occurs, firstly, through a classification of the collected speech as correct or incorrect. The results obtained in the classification are used to identify and group error patterns in order to form types of phonological cases. Thus, the CBR cycle begins, where the identified case types are *retrieved* and taken as the basis for evaluation of a new input case.

According to their suitability, the retrieved cases can be *reused* to categorize the new case based on the error pattern. The categorization is presented to the speech-language pathologist, who *reviews* the suggestions and confirms the identified situation. In the last step of the CBR cycle, the processed and classified information are stored in the server repository so that the knowledge base can be constantly updated with new cases. Further, these information are evaluated to establish appropriate recommendations, which will be returned to the user. The activities performed in the architecture will be minutely discussed in the specification of the Service Module's components.

3.2.1. The classification module

In the Classification Module (Fig. 5), the speech data collected in the Perception Module are preprocessed and the patient's pronunciation is classified as correct or incorrect. The preprocessing stage aims to optimize the quality of the collected sound, removing any background noise and making cuts in the audio as needed. This step is important as it influences the classification results and the classifiers performance.

After preprocessing, the audios are used to generate spectrograms, that is, they are converted into images or visual representations of the sound frequencies. By generating sound spectrograms, the Classification Module is able to distinguish patterns in the individual's speech and detect characteristics that define pronunciation errors. From the generation of spectrograms, feature extraction techniques are applied, which will be used as input to the classification process. In this process, a classifier is trained based on the correct and incorrect pronunciations of past cases, classifying the words according to their correctness. After classification, the processing in the CBR cycle of the architecture begins, which is responsible for understanding error patterns in the pronunciation.

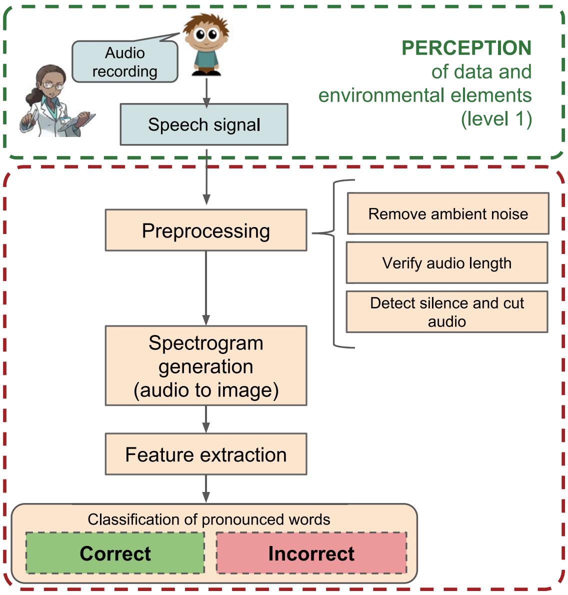

Fig. 5. The Classification Module.

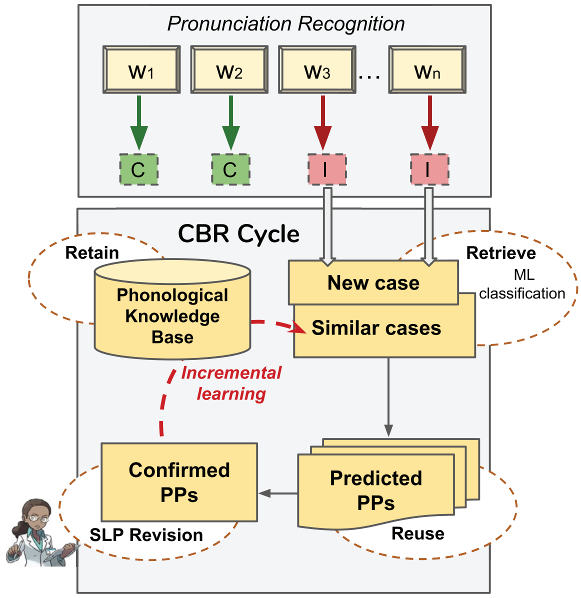

Fig. 6. CBR cycle.

3.2.2. The CBR cycle

The processing in the CBR cycle (Fig. 6) starts with the results of the Classification Module, where the incorrect pronunciation of one or more words in the speech evaluation are considered as a new case in the cycle, and it is assumed that this case needs investigation. For each new case, the objective is to verify in the incorrectly pronounced words if there is a pattern that has already been identified and diagnosed (which pronunciation errors were employed and how many occurrences for each). Therefore, the *retrieval* of existing cases from the system's knowledge base is performed.

Since the utility of a retrieved case cannot be evaluated directly *a priori*, the similarity between problem descriptions is used to estimate the expected utility of the cases under the new case (Kuo et al., 2015). Thus, after the retrieval process, the similarity between the new case and the retrieved cases is evaluated.

As shown in Fig. 7, the error patterns found in past cases receive scores calculated according to the probability of a word *W* occurs for each phonological process *pp*, also considering its possible phonemes (/*p*/, /*r*/, etc). This process begins with the identification of error patterns, i.e., each misspelled word is categorized into more than one PP, depending on the error patterns in the pronunciation. For example the portuguese word for dragon (D**r**agão), can be used for identifying five different PPs: devoicing

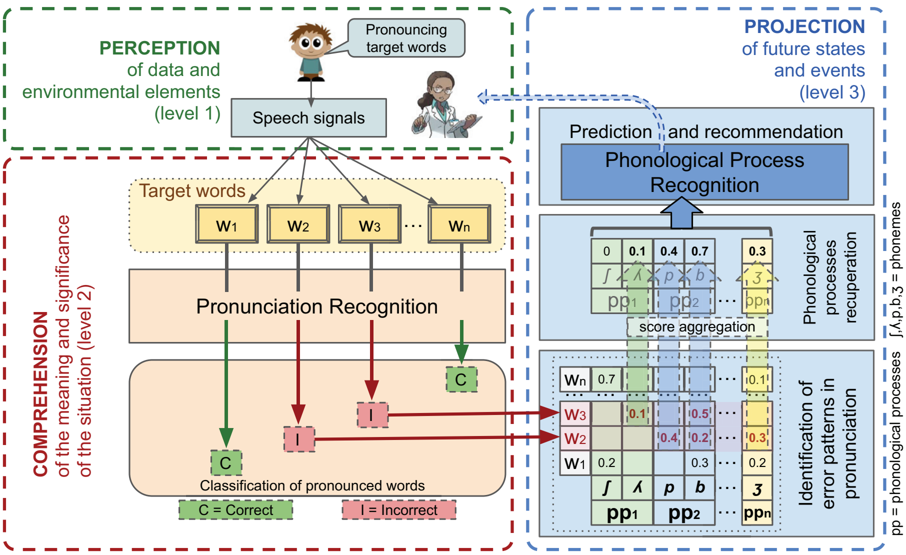

Fig. 7. Retrieval and Reuse steps of the CBR cycle.

Table 2
Identifying occurrences of PPs (*pph*) in the target words (*w*).

	w_1	w_2	w_3	w_4	...	w_n
pph_1		✓		✓		✓
pph_2	✓	✓				
pph_3			✓			✓
pph_4				✓		
...						
pph_m	✓	✓				

(/d/, /g/), fronting (/g/), backing (/d/), cluster reduction (/dr/), and liquid gliding (/**r**/). Similarly, the same PP can be found in different target words, e.g. liquid gliding (/**r**/) can be also observed in the portuguese word for witch (B**r**uxa). The Table 2 exemplifies the occurrences for each phonological process and its respective phonemes (*PPh*) in each target word.

Next, suppose that in most past cases the phonological process *liquid gliding* for phoneme */r/* (*PPh* − phonological process for a given phoneme) was identified in evaluations where the children misspelled the word *bruxa* (witch). Therefore, this word will receive a high score value for that *PPh*.

Upon a new case, the scores of misspelled words are aggregated for each *PPh*, creating the score vector for this evaluation. For example, considering a case where the child incorrectly pronounces the portuguese words *anel* (ring), *jornal* (newspaper), *fralda* (diaper), and *pastel* (pastry), the scores of these words are retrieved and aggregated to form the feature vector for this case. This feature vector is used as input to another classifier, which identifies the most similar patterns from past cases, thus indicating the phonological processes that may suit in the new case.

Lastly, the most similar patterns are *reused*, i.e., phonological processes are suggested as a resolution for the input case, completing the SA's comprehension level. These suggestions are *reviewed* by the human expert and the most appropriate PPs are confirmed for the new case, as a way of using human knowledge to guide the system in the learning process.

In the review step, the speech therapist analyzes the phonological processes and the probabilities (scores) with which they were inferred, chooses the most appropriate PPs and gives feedback to the system. This task is important since the specialist can observe the inferences scores and evaluate whether the characteristics or specific context of the case interfered in the prediction of PPs.

To elucidate this issue, can we illustrate a situation where a phonological process *A* was proposed to the SLP based on the error pattern *x*. The expert then reviews the case and realizes that *A* is not suitable for that particular case, since the error pattern *x* is common for the region where the patient resides (e.g. an individual's accent, which can determine the presence or absence of disorders in pronunciation, depending on the region of residence). This type of feedback given to the system can increase the reasoning and the variables analyzed in the processing.

In general, the evaluation of the cases made by the system and the suggestions review made by the human expert allow the increase of the comprehension level acquired in the previous module, since a comparison of different solutions occurs, as well as, the discovery of which is the most relevant to the current situation. From the confirmation of PP given by the SLP, it is possible to derive relationships between the information acquired and to interpret how the situation understood will impact on the user's goals. Considering this sequence of activities, the Recommendation Strategy presented below deals with the confirmed solution and its impact on the current system structure.

3.2.3. The recommendation strategy

The last level of the proposed approach (Fig. 8) is responsible for receiving the confirmed solution obtained in the CBR cycle and evaluating how it can be processed in order to generate appropriate recommendations. The confirmed solution, in this context, refers to the phonological processes that explain the patient's speech situation. Thus, once the patient's case has been assimilated (differentiated from other cases in history and properly categorized), it is necessary to perform a Recommendation Strategy, which has the main task of presenting the processing results and recommendations. Specifically, after PPs prediction the specialist should receive a results report in a mobile device. The results should include the error rate in the pronunciation, the patient's deviation degree, and therapeutic activities that the professional may perform according to the current situation. Through the mobile device that displays the results, the therapist can also confirm the usefulness of the recommendations, giving a feedback to the system.

After the solution is confirmed by the SLP, it is necessary to redefine the training data used as a basis for the reasoning. In other words, every suggestion of the system labeled by the professional is included in a next evaluation. This way, the recommendations of activities become more precise, since CBR handles cases more efficiently as new information are stored in the phonological database.

Also, it is desirable that the system maintains some statistics on previously evaluated cases, for example, in a given moment could be a fact that "cases of delayed speech development are observed in patients of a certain age when a given phoneme is omitted in 70% of the evaluations" but, after evaluating new cases and establishing appropriate solution for them, that fact could change. These statistics can be helpful for evaluating the composition of normal and abnormal cases in the Phonological Knowledge Base and are a part of future work.

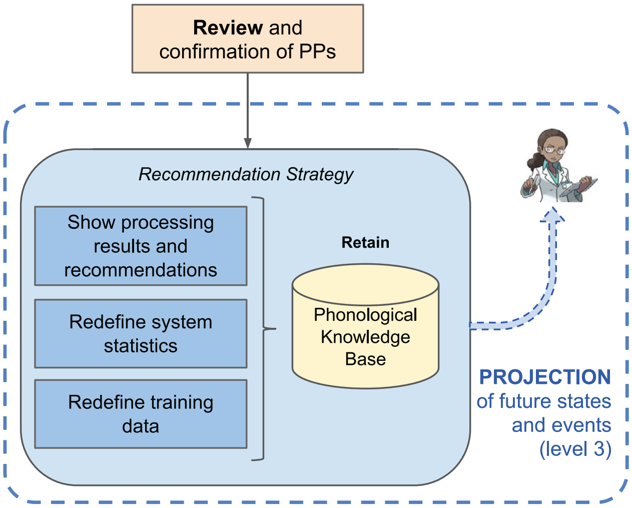

Fig. 8. The Recommendation Strategy.

Once the Recommendation Strategy is completed, the case, including all its characteristics, is stored (or *retained*) in the Phonological Knowledge Base, completing the CBR methodology cycle. From the recommendations managed by this module, the specialist can make better informed decisions based on the patient's profile and give feedback to the system, thus achieving the SA's Projection level.

4. Evaluation and results

In this section, the evaluation details of the architecture and the results obtained in each component module are presented. For this purpose, a prototype that implements the architecture was developed.

4.1. The capture module

To effectively understand a current situation, a situation-aware system must establish a perception phase, which involves capturing sensory cues from the surrounding environment. For this purpose, a set of 84 target words in Brazilian Portuguese was selected by a team of speech therapists from the Universidade Federal de Santa Maria (Brazil) in order to assess children's pronunciation skills. This team of SLPs performed a series of speech evaluations, which consisted of naming tasks. In these naming tasks, the child was presented to an image (referring to a target word) in a mobile application that was implemented as the Capture Module of the architecture (Fig. 9), and should pronounce the word corresponding to this visual stimulus.

In total, during the last three years, 1.114 evaluations were performed with 1.077 children aged 3 to 8 years and 11 months, forming a database containing 93.576 speech audios and their associated metadata (personal and contextual information, plus the word pronunciation transcription, performed by a SLP). From these collected data, the processing in the Service Module was started. Details on the conduction of the experiment are discussed in the following.

4.2. The classification module

For comprehending the correctness of pronunciation, every speech audio collected was submitted through a series of steps. Since the audios were noisy and they often contained a whole sentence and not just the desired word, we developed an automatic method for preprocessing the speech signal. To classify the pronunciations, a spectrogram of the speech signal was generated and LBP (Local Binary Pattern) was applied for generating a feature vector histogram; this histogram was used as input for the classifier. Results obtained from these steps are presented as following.

4.2.1. Preprocessing speech data

In previous work (Franciscatto et al., 2018b), the collected speech data were analyzed using Google Cloud Speech, an API that performs voice recognition converting speech into text (Mohamed et al., 2014). The tool demonstrated low accuracy when classifying speech, probably affected by the quality of the collected sound (noisy and of low volume) and its composition (often containing an entire sequence pronounced by the child, not just the target word).

From these results, it was necessary to perform preprocessing on the collected data, in order to improve speech classification according to the expert's assessment. The speech signals collected were preprocessed using two open source tools: SOX (*Sound*

Fig. 9. Screenshot of the mobile application running a naming task for the Portuguese word *cavalo* (horse).

eXchange, v14.4.1)[1] and FFmpeg (v3.2.12).[2] SOX was used for removing the noise and FFmpeg for detecting sentences and measuring the audio length. Whenever the length was larger than the average length for a given word, the method looked for moments of silence (which meant that the audio was a sentence), so the audio was cut for isolating just the desired word. The preprocessing method developed is discussed in detail in a previous research (Franciscatto et al., 2018c).

4.2.2. Spectrogram generation

After preprocessing the collected audios, there was the need to identify the word pronounced by the child at the moment of the evaluation. Consequently, it was necessary to distinguish error patterns in pronunciation, so we have used spectrograms for creating sound signatures that are used for detecting error patterns in the speech signal. The spectrograms were generated using *Matplotlib* (v2.2.2) Python package.[3] The Fig. 10 shows examples of spectrograms generated from the correct and incorrect pronunciation of the same Portuguese word "travesseiro" ("pillow", in English).

To reduce the computational complexity involved, the spectrograms were converted to gray scale. After this step, features were extracted from the images for differentiating patterns in pronounced words. The steps performed in the feature extraction are described in the sequence.

4.2.3. Feature extraction

Once the frequencies of the sounds were converted to spectrograms, the LBP method (Local Binary Pattern) (Ojala et al., 2002) was applied for identifying the differences in the visual patterns of words pronounced correctly and incorrectly. This method returns was a 256-dimensional feature vector histogram containing the textural information extracted from the spectrograms.

For this task, Python*Scikit-Learn* library (v0.19.1)[4] was used. Thus, the histograms are used as input to the classification process, described in the following topic.

4.2.4. Speech classification

Due to the dynamism required in situation-aware systems, machine learning techniques are commonly used for performing predictive tasks. In our proposal, we examined the accuracy of the following machine learning methods for classification: Neural Networks, K-Nearest Neighbors, Support Vector Machine, and Decision Tree algorithms. We chose the Decision Tree classifier (DT) as it presented a stable and satisfactory accuracy rate (Swain and Hauska, 1977).

The DT classifier was used in a total of 27,000 samples of the speech corpus. The corpus was randomly divided in 1/2 as the training set and 1/2 as testing set for preventing the classification from being biased, and the classifier was trained for recognizing the pronunciation of each word as correct and incorrect; these classification results are presented in Fig. 11.

---

[1] Available at: http://sox.sourceforge.net/ .
[2] Available at: https://ffmpeg.org/ .
[3] Available at: https://matplotlib.org .
[4] Available at: http://scikit-learn.org.

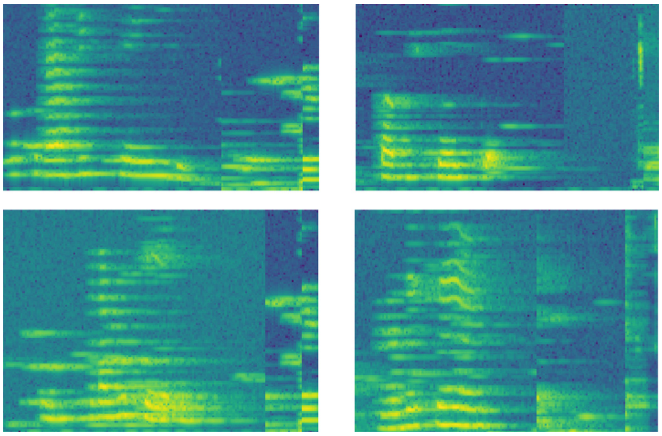

Fig. 10. Spectrograms generated from the Portuguese word “*travesseiro*” (pillow, in English) spoken correctly (to the left) and incorrectly (to the right).

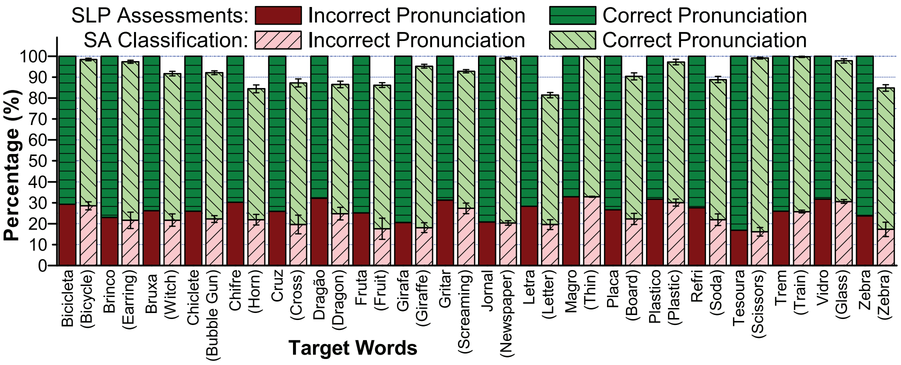

Fig. 11. Overall Concordance and Discordance rates in the target words.

The chart shows 20 randomly chosen target words written in Portuguese, and their English translations in brackets. For each word there is a darker column that provides information from SLPs’ assessments (showing the percentage of correct and incorrect pronunciations stored in the speech database), and a column of lighter colors that shows the average classification accuracy and standard deviation of our method.

Comparing the prediction of DT classifier against the SLP’s assessments, for all 84 target words, our method reached an accuracy of 92.5% for classifying the correctness of children’s pronunciation. By analyzing the set of words pronounced incorrectly, it is possible to infer which PPs the children used during their speech; thus, we used this information as input in the CBR cycle for augmenting our Phonological Knowledge Base.

4.3. The CBR cycle

In the CBR cycle, the patient’s pronunciations classified as correct and incorrect in the previous module are evaluated in order to identify patterns. These patterns allow the evaluation of possible PPs, favoring the anticipation of a future situation, so that the professional can make better informed decisions based on inferred results. This task is performed by scoring phonological processes that occurred in the target words and using these scores for feeding a ML classifier (*retrieve*), so it can *reuse* and suggest PPs for the SLP. Once the case is *reviewed* and confirmed by the SLP, it is incorporated in the Phonological Knowledge Base, the scores are recalculated, and the classifier’s training set is updated (*retain*). These steps for predicting PPs are described below.

4.3.1. Identification of speech error patterns

The process for identifying speech error patterns is based on PPs which, in the context of speech therapy, refers to the strategies employed by children to adapt the language spoken by adults to their phonological system, as presented in Section 2.3.

Since each misspelled word can be categorized into more than one PP, we analyzed exchanges, substitutions, and omissions of phonemes occurred in pronunciations, as well as the position in the word in which they occurred for all 84 target words, to categorize them according to PPs for acquisition of Portuguese language, as exemplified in Table 2 of Section 3.2.2. We used this information to feed our Phonological Knowledge Base by generating a **Phonological Matrix** (*PM*), which is formalized in Eq. 1.

$$
\begin{aligned}
TW &= tw_1 \cdots \mathrm{tw_m} \\
Ph &= ph_1 \cdots \mathrm{ph_j} \\
PPh &= PP(Ph) \\
&= [pph_1 \cdots \mathrm{pph_n}] \\
PM &= TW \times PPh \\
&= \begin{bmatrix} tw_1 \\ \vdots \\ tw_m \end{bmatrix} \times [\, pph_1 \quad \cdots \quad pph_n \,] \\
&= \begin{bmatrix} (tw_1, pph_1) & (tw_1, pph_2) & \cdots & (tw_1, pph_n) \\ (tw_2, pph_1) & (tw_2, pph_2) & \cdots & (tw_2, pph_n) \\ \vdots & \vdots & & \vdots \\ (tw_m, pph_1) & (tw_m, pph_2) & \cdots & (tw_m, pph_n) \end{bmatrix}
\end{aligned} \tag{1}
$$

where:

TW - vector of target words;
Ph - vector of phonemes;
PP - phonological processes;
PPh - vector of phonological processes, considering their specific phonemes; and
PM - Phonological Matrix.

The elements of *PM* are defined as presented in Eq. 2.

$$
(tw_i, pph_j) = \begin{cases} 0 & \text{if } pph_j \text{ does not apply } (\notin) \text{ in this } tw_i \\ 1 & \text{otherwise} \end{cases} \tag{2}
$$

It is important to emphasize that *PM* stores the definitions of which target words are subject to which *PPs*. These values are defined by speech therapists, and may be different for each region due to the cultural characteristics and the strong presence of immigrants in the social formation of the country. Also, this matrix can be modified to include new words and *PPs*, as necessary.

4.3.2. Calculating scores for phonological processes

The *PM* is given as input to generate a second matrix, PM_{score}, which stores the proportional scores of PPs for phonemes that the children had difficulty in pronouncing. The scores are automatically calculated considering the evaluations assessed by SLPs and stored in the phonological database. PM_{score} has the same structure as *PM*, but contains additional information (the scores), which are calculated as presented in Eq. 3. In this equation, n is the number of phonological processes (*PPh*), m is the number of cases, confirmed by a SLP, in which the pph_j was used by the children for pronouncing the tw_i, while l means all stored cases involving the pph_j. The relation (tw_i, pph_j) is 0 if $pph_j \notin tw_i$ and maintains the probability (0..1) of occurrence of pph_j for a given word tw_i.

$$
(tw_i, pph_j) = \begin{cases} 0 & \text{if } \ pph_j \notin tw_i \\ \dfrac{\sum_{k=1}^{m} PPh_k(tw_i)}{\sum_{l=1}^{n} PPh_l} & \text{otherwise} \end{cases} \tag{3}
$$

For example, as explained in Section 3.2.2, the Portuguese word for dragon (Dragão), can be used for identifying five different *PPhs*: devoicing (/d/, /g/), fronting (/g/), backing (/d/), cluster reduction (/dr/), and liquid gliding (/r/); so these words will receive scores on all these *PPh*.

To calculate the score, all cases confirmed by the SLPs are used. Suppose that, in all evaluations assessed by SLPs for PPh “liquid gliding” (/r/), all children incorrectly spelled the word “witch”. So, this PM_{score} cell (witch, liquid gliding (/r/)) is set to 1.0. Similarly, if in the previous evaluations SLPs identified that half of children spelled this word incorrectly, the score is set to 0.5.

4.3.3. Inferring phonological processes

The phonological processes of new child’s evaluation are inferred using the set of incorrectly spelled words, by summing their *PPh* scores stored in the PM_{score} matrix. The Table 3 exemplifies the scoring aggregation for a new case where the child incorrectly pronounced words w_3 and w_n. The resulting set for this evaluation is calculated by aggregating only the *PPh* scores for these misspelled words. In this case, aggregating the scores for pph_1 and pph_3, since they represent the error patterns found in the pronunciation. The zero value is assigned to all other *PPh* that did not occur in the misspelled words w_3 and w_n. Therefore, the resulting set of the entire pronunciation is [pph_1=**0.124**, pph_2=0, pph_3=**0.602**, pph_4=0, $\cdots$, pph_m=0], being the value 0.602 the sum of pph_3 scores in w_3 and w_n. This vector, which comprises the entire child speech evaluation, is used as input for a classifier that uses the information stored in Phonological Knowledge Base to identify the most likely cases. If confirmed by a SLP, this new case is incorporated in our base.

Table 3
Scores of occurrences for each phonological process.

	w_1	w_2	$\boldsymbol{w_3}$	w_4	$\cdots$	$\boldsymbol{w_n}$
pph_1		0.045		0.214		**0.124**
pph_2	0.104	0.291				
pph_3			**0.210**			**0.392**
pph_4				0.121		
$\vdots$						
pph_m	0.109	0.072				

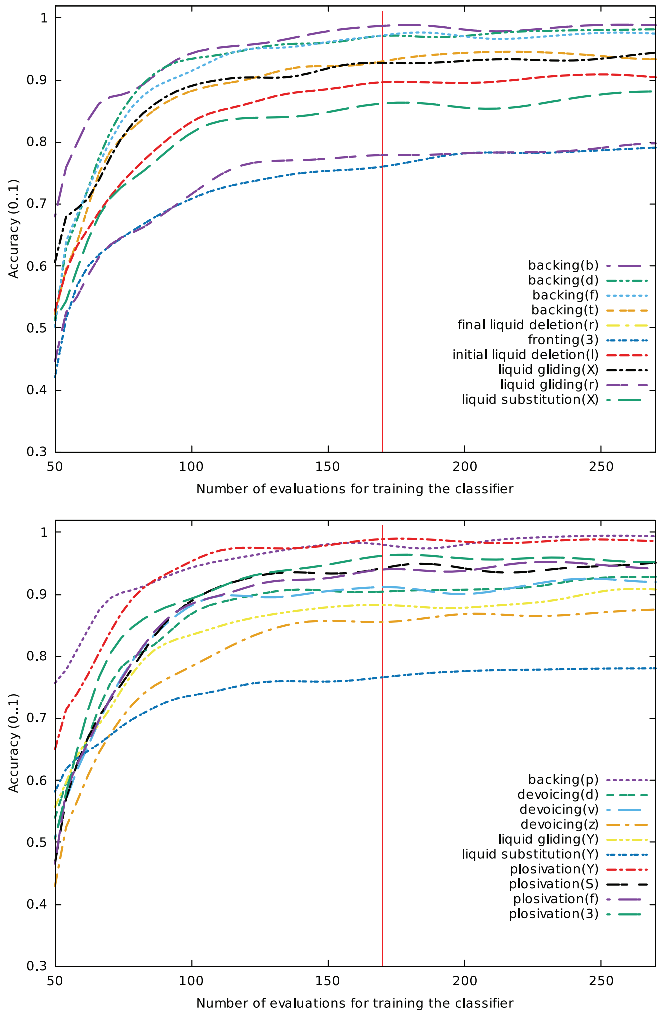

Fig. 12. Incremental learning for predicting the phonological processes.

As demonstrated in the x-axis of Fig. 12, we start evaluating the classification of phonological processes feeding the classifier with 50 speech assessments, confirmed by the SLPs and, with the arrival of new confirmed cases, the classifier was incrementally trained. Each assessment may contain from none to all the *PPs*, associated with their possible phonemes, as shown in Table 1. We have firstly used a Stochastic Gradient Descent (SGD)[5] as the classifier (Zhang, 2004) because it provides the necessary incremental learning. This classifier implements regularized linear models with SGD learning where the gradient of the loss is estimated each sample at a time and its model is updated on the fly. The training parameters were to their default values[6].

[5] Available at: https://scikit-learn.org/stable/modules/sgd.html .
[6] The description of SGD classifier's default parameters is available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html .

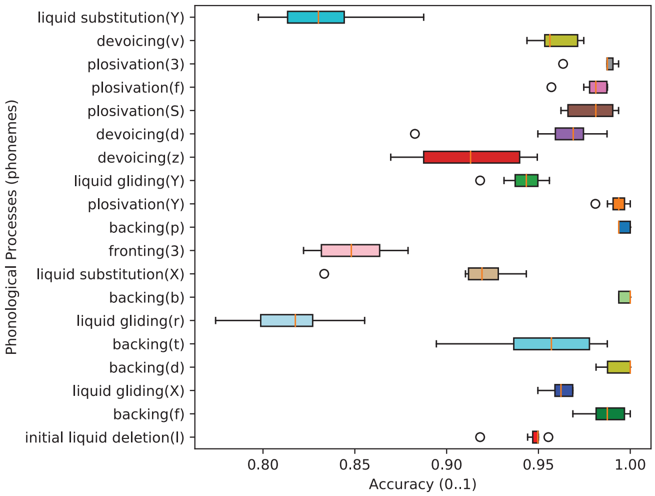

Fig. 13. SGD classifier accuracy for predicting phonological processes.

Fig. 12 shows the result in this step, presenting the accuracy of 20 *PPhs* randomly selected from a total of 51 *PPhs*; to ease the visualization, the *PPhs* were divided into two figures. As can be seen, in the beginning of charts, the accuracy of all *PPhs* were not satisfactory, ranging from 45% to 65%. As new evaluations were incorporated to the base, there was a significant improvement in the results and the prediction accuracy curve became flat, for most of *PPhs*, when the Phonological Knowledge Base had more than 170 confirmed speech assessments, asidentified in the charts with a vertical red line.

It can also be observed in Fig. 12 that some *PPhs* are still increasing their prediction accuracy, but not ascending as significant as in the beginning of the learning process. The figure also shows *PPhs* with a lower accuracy rate, as in the case of liquid substitution in phoneme /*Y*/. For both cases, we observed few occurrences, confirmed by SLPs, for children using these *PPhs* during their phonological assessments.

Fig. 13 shows the prediction accuracy for the selected *PPhs* when the SGD classifier was trained with the entire current phonological knowledge base. The average accuracy for predicting the selected *PPhs* was 94.55%, but when considering all 51 *PPhs*, the accuracy was 90.22%. We can also observe in Fig. 13 that some *PPhs* obtained low accuracy (such as *liquid gliding(r), fronting and liquid substitution(Y)*). This is explained by the fact that the knowledge base had few samples of these *PPhs*; however, their accuracy prediction rate was above 80%.

We firstly used SGD classifier because it simplified the incremental learning process; however, when a satisfactory accuracy rate was achieved we examined other classifiers for predicting the *PPhs*, such as Random Forest classifier (Breiman, 2001), Neural Network (Adams) (Kingma and Ba, 2014), Support Vector Machine (SVM) (Noble, 2006), Decision Tree (Swain and Hauska, 1977), and K-Nearest Neighbors (Zhang and Zhou, 2007).

In addition to accuracy, these classifiers were evaluated through other metrics, such as Precision, Recall, and F1-Score, which evaluates respectively, the agreement of the data labels with the positive labels given by the classifier, the effectiveness of a classifier to identify positive labels, and the weighted averaged combination between recall and precision (Raschka, 2014; Sokolova and Lapalme, 2009). These metrics were tested in addition to accuracy since the relation among different performance measures help resolving disagreements among performance evaluations, which frequently happen in experimental studies (Sokolova and Lapalme, 2009).

The evaluations were performed with the above-mentioned metrics, considering thirty trials by randomly splitting the data in 1/2 as the training set and 1/2 as testing set. With respect to the parameters used in the classifiers, Random Forest was evaluated considering fifty estimators, KNN with five neighbors, Adams Neural Network with maximum number of iterations = 400 and initial learning rate = 0.01, and SVM considering RBF kernel, penalty parameter C=0.031, gamma coefficient = 8.0, and probability = True.

Fig. 14 shows the results for all classifiers. The best average accuracy for predicting the *PPhs*, considering the whole Phonological Knowledge Base, was achieved by the Random Forest classifier (93.09%), followed by KNN (92.16%) and Adams Neural Network (91.49%). Although SVM showed good accuracy, the precision was not stable, so this classifier was less effective in comparison with the others. Also, SGD had greater variation in accuracy and precision than the other classifiers.

After classification, the *PPhs* of higher probability can be recommended to the SLP, as well as therapeutic activities suitable to the patient's situation. The therapist should review the suggestions and return a feedback to the system, accepting or rejecting the recommendations in order to keep the knowledge base up to date. The above-mentioned tasks are part of the Recommendation Strategy (Section 3.2.3), which aims to cover the last level of SA (Projection). The evaluation of the Recommendation Strategy is described in the next section.

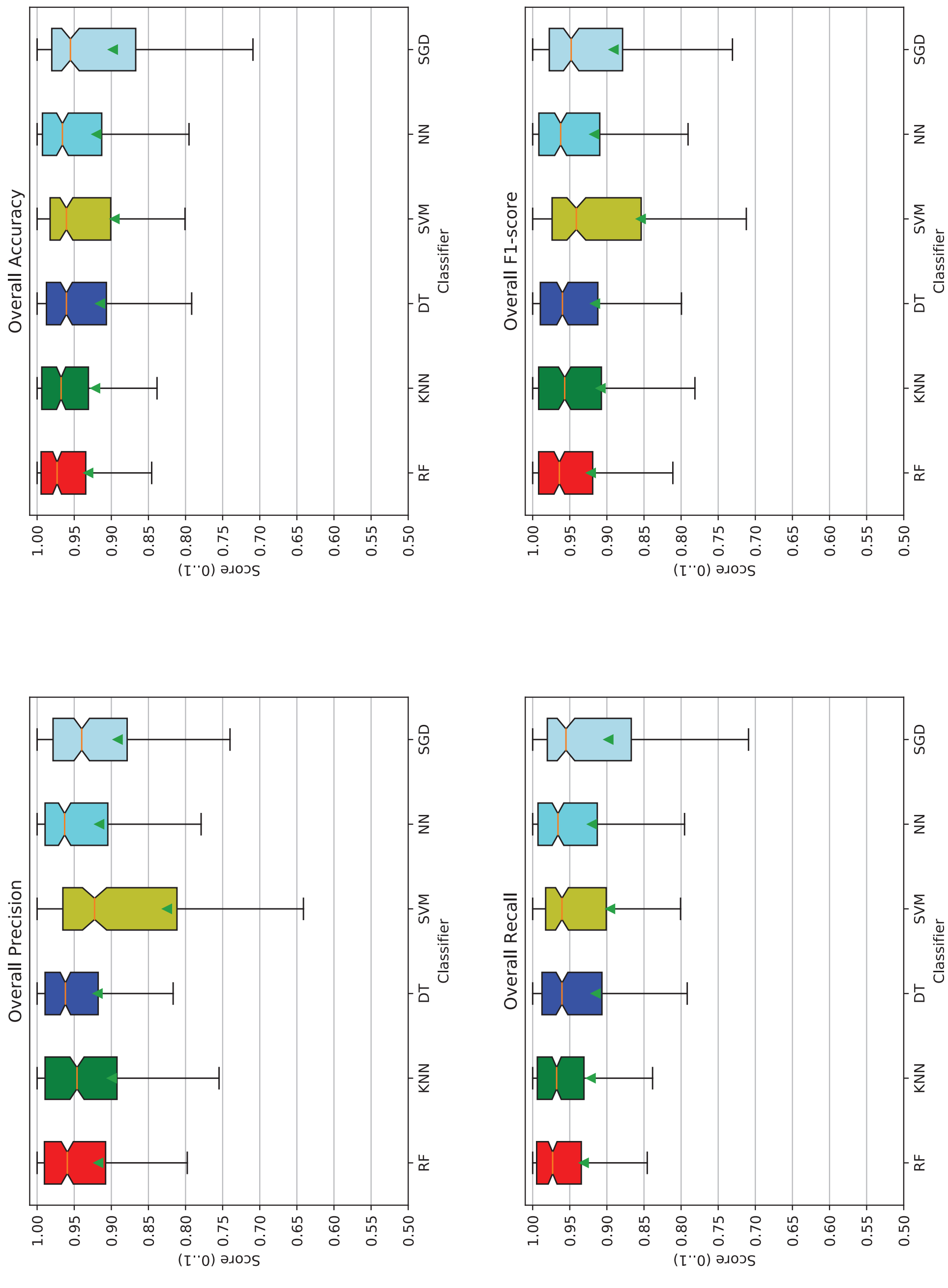

Fig. 14. Prediction results with different classifiers.

Table 4
Indication of speech disorder according to PCC-R value. Adapted from (Shriberg et al., 1997b).

PCC-R value	Deviation Level
Less than 50%	High
From 51% to 65%	Moderate-high
From 66% to 85%	Low-moderate
Greater than 85%	Low

4.4. The recommendation strategy

The results obtained in the CBR Cycle show that our approach was able to achieve an effective comprehension level, since the phonological processes in the pronunciation were predicted with an average accuracy over 90%. Considering that the understanding of a situation is fundamental for clinical decision making, we used the classification results for calculating the probability of a speech disorder in the child's pronunciation and projecting it to the speech therapist. This scenario represents a *Recommendation Strategy*, which is the last module of our situation-aware and case-based architecture.

Our Recommendation Strategy refers to activities suggested to the SLP, in order to provide assistance in the therapy planning according to the patient's situation. By responding to a situation, we achieve the *Projection* level (the last level of the SA model), which aims to create a view that is meaningful to decision-making.

We based our Projection strategy on a formula called **PCC-R** (*Percent Consonants Correct - Revised*, demonstrated in Equation 4), which determines the presence of speech disorder from the number of meaningful words (Shriberg et al., 1997b):

$$PCCR = \frac{CC}{TC} \tag{4}$$

where CC means the *number of Correct Consonants pronounced by the individual* and TC means the *Total number of Consonants pronounced, correct and incorrect*. The resulting value, multiplied by 100, expresses the percentage of intended consonant sounds in a conversational sample that were articulated correctly (Shriberg et al., 1997b; 1997a); this value indicates the presence of speech disorder in many severity levels, as shown in Table 4.

The PCC-R measure has been derived from the original one (PCC), differing in the way it scores speech-sound distortions. The original measure scores all omissions, substitutions, and distortions as errors, whereas the revised measure scores only omissions and substitutions as errors (Shriberg et al., 1997b; Shriberg and Kwiatkowski, 1982). In the most severe disorders *substitution* processes occur, but *omission* processes are more frequent since there is no awareness of the phoneme. In less severe cases, there is awareness of phonemes, however, children do not dominate their production, presenting more substitution processes.

In general, the disorder severity measured by the PCC-R indicates how much the communication capacity is below expectations, i.e., how much this capacity is not adequate to the speech standard for a given age. A study relating the occurrence of Phonological Processes to the phonological disorder severity (Ghisleni et al., 2010) found out that the greater the severity level, the greater the number of PPs present in the speech.

Thus, considering the close relationship between PCC and the occurrence of PPs, coupled with the fact that the measure has been widely used in the speech therapy domain (Klintö et al., 2016; Allen, 2013; Brancalioni and Keske-Soares, 2016; Ceron et al., 2017), we based our strategy on the PCC-R calculation, using the resulting value for making suggestions to the therapist. For presenting these suggestions to the expert and evaluate the recommendation strategy, we have improved the prototype developed in the Capture Module (Section 4.1), adding an interface for visualization of results, where the professional can review the recommendations and confirm actions that can be taken in clinical environment.

As observed in Fig. 15, the results interface is composed by a series of information including, firstly, the total number of correct pronunciations obtained in the speech assessment (*CC* rate, exemplified as the value "35"), followed by the total number of consonants pronounced (*TC* rate, exemplified as the value "50").

In the sequence, the interface shows the identified phonological processes (Backing and Final Fricative Deletion, both in the phoneme /s/), followed by the PCC-R calculation ("0.7") and the resulting deviation degree *baixo-moderado* (or low-moderate, in English). Finally, at the bottom of the interface, activities are recommended to the SLP, which can be translated as "Auditory discrimination exercises", "Target segment stimulation at the level of word and/or sentence" and "Target segment imitation training", respectively. Next to each suggested activity there is a checkbox, where the therapist can mark the appropriate activities and give a feedback, ignoring the suggestions ("*Ignorar*") or confirming them ("*Confirmar atividade(s)*").

The central idea is that the system learns about therapeutic activities from human feedback. Initially, the prototype should receive a list of intervention activities elaborated by a team of SLPs; at the moment of the speech assessment, as the professional evaluates the situation and opts for the most appropriate strategies, the system learns from this feedback. It is worth mentioning that the final decision belongs to the specialist, while the system supports this decision.

In addition to allowing the recommendation of PPs and therapeutic activities to the specialist, our approach may be useful to perform personalized speech evaluations according to the patient's needs. At first, for example, the child can be evaluated with all the target words. In case of errors in pronunciation errors, the PPs are retrieved. In a second moment, the system could

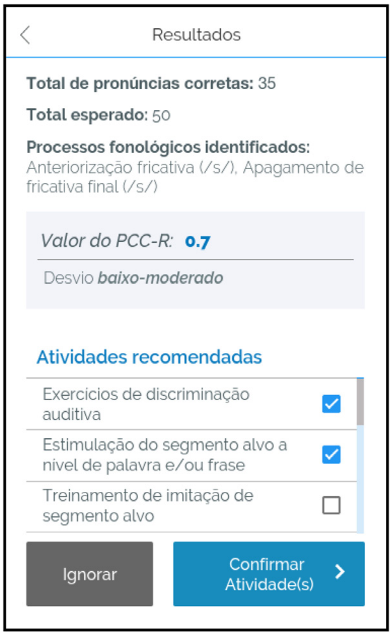

Fig. 15. Screenshot of the recommendation strategy presented to the professional.

customized for testing the patient's pronunciation for specific PPs. This allows more patient-oriented and adaptive assessments, just as it gives the therapist more options.

5. Conclusions

In this work, a situation-aware and case-based architecture was proposed to assist Speech-Language Pathologists in tasks involving screening of speech disorders and therapeutic planning. Analyzing related work, we observed that the identification of phonological processes is little addressed; furthermore, to the submission of this paper, we could not find any work applying Situation-Awareness and Case-Based Reasoning for problem solving in the speech therapy domain. Considering that an early identification of speech disorders may prevent negative consequences on the individual's development, we proposed an architecture combining the advantages of CBR and SA for predicting error patterns in children's speech and elaborating suitable recommendations, as a way to assist the therapist in clinical tasks.

We used machine learning methods to classify the pronunciation correctness from speech signals collected from more than 1,000 children aged 3 to 8 years performing naming tasks. The results show that a good level of comprehension about the current situation was established, since it was possible to derive relevant meanings from the perceived data.

The pronunciations classified as incorrect were used as input in a CBR cycle. The previously diagnosed cases were retrieved and compared with the input case, in order to identify recurrent error patterns, that is, phonological processes applied in the pronunciation. This task was performed by analyzing exchanges, substitutions, and omissions occurred in the speech samples, in which we obtained an average accuracy over 93% for predicting PPs, in concordance with the assessments performed by therapists. The satisfactory accuracy rate obtained shows that our approach is able to provide a faster indication of PPs applied in children's pronunciation.

From the identification of the individual's speech situation, as well as the Recommendation Strategy, it is possible to provide action support to SLP in clinical environment. Since our goal is not to substitute the SLP in diagnostic tasks, we believe the action support is the best way to improve the planning of therapeutic activities and help the therapist to make better decisions. However, the results found by merging CBR, SA, and machine learning are encouraging, showing that it is possible to developed an automated projection (a decision-making strategy) with little or no human intervention.

Our future work includes adapting the approach to cover other languages, that is, phonemes and PPs that may compose different phonetic structures. The speech development is affected by several factors, such as parental education, age, and type of school, so it is relevant to investigate how these factors constitute differences in the speech among individuals from many countries or regions, and perform the appropriate treatment of these peculiarities. That way our proposal could be applied in social politics for children's speech diagnosis, mainly those belonging to public schools, who currently make up the majority in the country. Secondly, we aim to include the management of statistics in the Recommendation Module, in order to understand which factors are related to speech disorders and can help characterize the normality of a case. Lastly, we want to perform tuning of the classifiers and inspect different approaches and machine learning algorithms aiming to improve the prediction results.

Finally, we conclude that our assessments presented encouraging results in understanding and projecting the speech situation of an individual. By integrating capabilities of Situation-Awareness and Case-Based Reasoning, the proposal is able to recommend PPs to professionals and support decision-making in speech therapy contexts, since the phonological knowledge base remains updated with information from assessed cases. Also, the developed method is highly configurable, that is, it can be modified to allow its application in a multidisciplinary way.

Declaration of Competing Interest

No other relationships/conditions/circumstances that present a potential conflict of interest Relationships.

Acknowledgements

This work has been supported by Fundação de Amparo a Pesquisa do Estado do RS (FAPERGS), grant number 17/2551-0000875-8, Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq Brasil, grant number 423518/2018-6 and UFSM/FATEC through project 041250-9.07.0025(100548). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)) - Finance Code 001.

References

Aamodt, A., Plaza, E., 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. 7 (1), 39–59.
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. Comput. Speech l Lang. 27 (6), 1235–1248.
Abou-Elsaad, T., Afsah, O., Rabea, M., 2019. Identification of phonological processes in arabic−speaking egyptian children by single-word test. J. Commun. disord. 77, 80–93.
Ahmed, M.U., Begum, S., Funk, P., 2012. Case studies on the clinical applications using case-based reasoning. Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on. IEEE, pp. 3–10.
Allen, M.M., 2013. Intervention efficacy and intensity for children with speech sound disorder. J. Speech Lang. Hear. Res. 1–13.
Azeta, A., Ayo, C., Atayero, A., Ikhu-Omoregbe, N., 2009. A case-based reasoning approach for speech-enabled e-learning system. 2009 2nd International Conference on Adaptive Science & Technology (ICAST). IEEE, pp. 211–217.
Begum, S., Ahmed, M.U., Funk, P., Xiong, N., Folke, M., 2011. Case-based reasoning systems in the health sciences: a survey of recent trends and developments. IEEE Trans. Syst. Man Cybern. Part C 41 (4), 421–434.
Bolaños, D., Cole, R.A., Ward, W., Borts, E., Svirsky, E., 2011. Flora: Fluent oral reading assessment of children's speech. ACM Trans. Speech Lang. Process. 7 (4), 16.
Brancalioni, A.R., Keske-Soares, M., 2016. Phonological disorders treatment effect with a stimulability and segment complexity strata model with speech intervention software (sifala). Revista CEFAC 18 (1), 298–308.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Caballero-Morales, S.-O., Trujillo-Romero, F., 2014. Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition. Expert Syst. Appl. 41 (3), 841–852.
Ceron, M.I., Gubiani, M.B., de Oliveira, C.R., Keske-Soares, M., 2017. Factors influencing consonant acquisition in brazilian portuguese−speaking children. J. Speech Lang. Hear. Res. 60 (4), 759–771.
Chen, Y.-P.P., Johnson, C., Lalbakhsh, P., Caelli, T., Deng, G., Tay, D., Erickson, S., Broadbridge, P., El Refaie, A., Doube, W., et al., 2016. Systematic review of virtual speech therapists for speech disorders. Comput. Speech Lang. 37, 98–128.
Chuchuca-Méndez, F., Robles-Bykbaev, V., Vanegas-Peralta, P., Lucero-Saldaña, J., López-Nores, M., Pazos-Arias, J., 2016. An educative environment based on ontologies and e-learning for training on design of speech-language therapy plans for children with disabilities and communication disorders. Ciencias de la Informática y Desarrollos de Investigación (CACIDI), IEEE Congreso Argentino de, pp. 1–6.
Dodd, B., 2013. Differential Diagnosis and Treatment of Children with Speech Disorder. John Wiley & Sons.
Dodd, B., 2014. Differential diagnosis of pediatric speech sound disorder. Curr. Dev. Disord. Rep. 1 (3), 189–196.
Dodd, B., Lacano, T., 1989. Phonological disorders in children: Changes in phonological process use during treatment. Int. J. Lang. Commun.Disord, 24 (3), 333–352.
Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Fact. 37 (1), 32–64.
Endsley, M.R., Garland, D., 2000. Theoretical underpinnings of situation awareness: acritical review. Situation Aware. Anal. Meas. 3–32.
Fan, Y., Kendall, E., 2005. A case-based reasoning approach for speech corpus generation. International Conference on Natural Language Processing. Springer, pp. 993–1003.
Franciscatto, M.H., Augustin, I., Lima, J.C.D., Maran, V., 2018. Situation awareness in the speech therapy domain: a systematic mapping study. Comput. Speech Lang. 53, 92–120.
Franciscatto, M.H., Lima, J.C., Moro, A., Maran, V., Augustin, I., Soares, M.K., Costa, C.C., 2018. A case-based system architecture based on situation-awareness for speech therapy. Enterprise Information Systems (ICEIS), 20th International Conference on.
Franciscatto, M.H., Trois, C., Lima, J.C.D., Augustin, I., 2018. Blending situation awareness with machine learning to identify children's speech disorders. 12th International Conference on Application of Information and Communication Technologies, AICT 2018, Almaty, Kazakhstan, October 17-19. IEEE.
Fringi, E., Lehman, J.F., Russell, M., 2015. Evidence of phonological processes in automatic recognition of children's speech. Sixteenth Annual Conference of the International Speech Communication Association.
Ghisleni, M.R.L., Keske-Soares, M., Mezzomo, C.L., 2010. O uso das estratégias de reparo, considerando a gravidade do desvio fonológico evolutivo. Revista CEFAC 12 (5), 766–771.
Goto, J., Kidokoro, T., Ogura, T., Suzuki, S., 2013. Activity recognition system for watching over infant children. RO-MAN, 2013 IEEE. IEEE, pp. 473–477.
Grossinho, A., Guimaraes, I., Magalhaes, J., Cavaco, S., 2016. Robust phoneme recognition for a speech therapy environment. Serious Games and Applications for Health (SeGAH), 2016 IEEE International Conference on, pp. 1–7.
Guillén, R., Usrey, R., 2005. Case based reasoning using speech data for clinical assessment. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, pp. 85–94.
Hitchcock, E.R., Harel, D., Byun, T.M., 2015. Social, emotional, and academic impact of residual speech errors in school-aged children: a survey study. Seminars in Speech and Language, 36. Thieme Medical Publishers, pp. 283–294.
Husain, W., Pheng, L.T., 2010. The development of personalized wellness therapy recommender system using hybrid case-based reasoning. Computer Technology and Development (ICCTD), 2010 2nd International Conference on. IEEE, pp. 85–89.
Iliya, S., Neri, F., 2016. Towards artificial speech therapy: a neural system for impaired speech segmentation. Int. J. Neural Syst. 26 (6), 1–16.
Jesus, L.M., Lousada, M., Domingues, D., Hall, A., Tomé, D., 2015. Phonological processes in portuguese children with speech sound disorders. Poznan Stud. Contemp. Linguist. 51 (1), 75–88.

Joye, N., Broc, L., Olive, T., Dockrell, J., 2019. Spelling performance in children with developmental language disorder: a meta-analysis across european languages. Sci. Stud. Read. 23 (2), 129–160.

Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. International Conference on Learning Representations, pp. 1–15.

Klintö, K., Salameh, E.-K., Lohmander, A., 2016. Phonology in swedish-speaking 5-year-olds born with unilateral cleft lip and palate and the relationship with consonant production at 3 years of age. Int. J. Speech-Lang. Pathol. 18 (2), 147–156.

Kokar, M.M., Endsley, M.R., 2012. Situation awareness and cognitive modeling. IEEE Intell. Syst. 27 (3), 91–96.

Kuo, J.Y., Yu, H.-F., Liu, K.F.-R., Lee, F.-W., 2015. Multiagent cooperative learning strategies for pursuit-evasion games. Math. Probl. Eng. 2015.

Le, D., Licata, K., Persad, C., Provost, E.M., 2016. Automatic assessment of speech intelligibility for individuals with aphasia. IEEE/ACM Trans. Audio SpeechLang. Process. 24 (11), 2187–2199.

Leake, D., 2015. Problem solving and reasoning: case-based. International Encyclopedia of the Social I& Behavioral Sciences (Second Edition), pp. 56–60.

Martín-Ruiz, M.L., Valero Duboy, M.A., Pau de la Cruz, I., 2013. Deployment and validation of a smart system for screening of language disorders in primary care. Sensors 13 (6), 7522–7545.

Mohamed, S.A.E., Hassanin, A.S., Othman, M.T.B., 2014. Educational system for the holy Quran and its sciences for blind and handicapped people based on google speech api. J. Softw. Eng. Appl. 7 (03), 150.

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P.H., Lohvansuu, K., et al., 2014. Cognitive mechanisms underlying reading and spelling development in five European orthographies. Learn. Instruct. 29, 65–77.

Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24 (12), 1565.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach.Intell. 24 (7), 971–987.

Rao, B., 2017. Case-based reasoning (cbr) in condition monitoring & diagnostic engineering management (comadem): a literature survey. Int. J. COMADEM 20 (1), 15–22.

Raschka, S., 2014. An overview of general performance metrics of binary classifier systems. arXiv preprint arXiv:1410.5330 1–5.

Riesbeck, C.K., Schank, R.C., 1989. Inside Case-Based Reasoning.Erlbaum, Northvale, NJ

Robles-Bykbaev, V.E., Guamán-Murillo, W., Quisi-Peralta, D., López-Nores, M., Pazos-Arias, J.J., García-Duque, J., 2016. An ontology-based expert system to generate therapy plans for children with disabilities and communication disorders. Ecuador Technical Chapters Meeting (ETCM), IEEE, 1, p. 6.

Rvachew, S., Bernhardt, B.M., 2010. Clinical implications of dynamic systems theory for phonological development. Am. J. Speech-Lang. Pathol. 34–50.

Shriberg, L.D., Austin, D., Lewis, B.A., McSweeny, J.L., Wilson, D.L., 1997. The percentage of consonants correct (PCC) metric: extensions and reliability data. J. Speech Lang. Hear. Res. 40 (4), 708–722.

Shriberg, L.D., Austin, D., Lewis, B.A., McSweeny, J.L., Wilson, D.L., 1997. The speech disorders classification system (SDCS) extensions and lifespan reference data. J. Speech Lang. Hear. Res. 40 (4), 723–740.

Shriberg, L.D., Kwiatkowski, J., 1982. Phonological disorders III: a procedure for assessing severity of involvement. J. Speech Hear. Disord. 47 (3), 256–270.

Smiti, A., Elouedi, Z., 2019. Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems. Theor. Comput. Sci. 1–9.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45 (4), 427–437.

Sreedevi, N., John, M., Chandran, S., 2013. Computerized assessment of phonological processes in malayalam (capp-m). Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013 IEEE. IEEE, pp. 264–269.

Stampe, D., 1979. A Dissertation on Natural Phonology. Indiana University Linguistics Club.

Swain, P.H., Hauska, H., 1977. The decision tree classifier: design and potential. IEEE Trans. Geosci. Electron. 15 (3), 142–147.

Watson, I., 1999. Case-based reasoning is a methodology not a technology. Knowl.-Based Syst. 12 (5), 303–308.

Yavas, M., Lamprecht, R., 1988. Processes and intelligibility in disordered phonology. Clin.Linguist. Phonet. 2 (4), 329–345.

Yeow, W.L., Mahmud, R., Raj, R.G., 2014. An application of case-based reasoning with machine learning for forensic autopsy. Expert Syst. Appl. 41 (7), 3497–3505.

Zhang, M.-L., Zhou, Z.-H., 2007. Ml-knn: a lazy learning approach to multi-label learning. Pattern Recognit. 40 (7), 2038–2048.

Zhang, T., 2004. Solving large scale linear prediction problems using stochastic gradient descent algorithms. ICML 2004: Proceedings of the 21st International Conference on Machine Learning. Omnipress, pp. 919–926.