

Design and Evaluation of a Spanish Language Therapy Support System

Karla I. Arce-Ruelas

Faculty of Sciences, Autonomous University of Baja California, Mexico

Omar Alvarez-Xochihua

Faculty of Sciences, Autonomous University of Baja California, Mexico

José A. Gonzalez-Fraga

Faculty of Sciences, Autonomous University of Baja California, Mexico

Evelio Martinez-Martinez

Faculty of Sciences, Autonomous University of Baja California, Mexico

Patricia Paez-Manjarrez

Faculty of Administrative and Social Sciences, Autonomous University of Baja California, Mexico

Abstract—*Specific Language Impairment (SLI)* is a language disorder that delays progress in mastering speech-language skills, and typically occurs in childhood. Most speech-language pathologists commonly use paper-based instruments to diagnose and treat this problem. This article describes the design, implementation, and evaluation of SATEL, an ontology-based system used both in diagnosing this condition and as part of speech therapy for children with SLI. With the help of a Kinect sensor, SATEL is able to recognize and classify pronounced words. The proposed system was designed and evaluated by a team of four speech-language pathologists and 26 children diagnosed with SLI. Results showed an accuracy rate of 94.42% and 97.75% in recognizing syllables and words correctly and incorrectly pronounced in the diagnostic and treatment modules, respectively.

Index Terms—specific language impairment, speech therapies, ontologies, assistive technology

I. INTRODUCTION

Some language disorders affect a significant number of children in early school years (Jackson-Maldonado, 2011). Particularly, *Specific Language Impairment* is associated with problems of grammaticality, poor vocabulary, and speech fluency. Children with SLI do not suffer from any physical condition, such as hearing loss or neurological disease (Jackson-Maldonado, 2011), but do not develop ordinary language skills as expected in early childhood, and face specific difficulties encoding and decoding in the communication process.

SLI negatively influences children's participation in classroom activities, since they struggle to learn new words and communicate ideas and thoughts; this affects their academic development as the reading fluency has greater effects on their performance in later years (Lange, 2019). Additionally, this type of disorder is a significant cause of undesired child behaviour such as anxiety, frustration, and insecurity, which has repercussions on academic performance and social interaction (Jackson-Maldonado, 2011; Woodman et al., 2018).

Technically speaking, children facing SLI may have difficulties with some or all dimensions of language, such as phonology, the lexicon, semantics, morphology, and syntax. Thus, SLI is not a homogeneous disorder. Indeed, there are many subtypes and combinations (Vázquez et al., 2011), this is why it is essential that children with this condition receive appropriate and personalized therapy. On the other hand, although SLI is an international and recurrent problem, in practice, speech-language pathologists rarely leverage the full potential of computer-based technology. Indeed, only a few specific computer-based systems exist to support SLI therapy. Moreover, it was found that current systems are not considering adaptability with a view to personalizing care.

This paper describes the design, development, and evaluation of an automated tool called *Language Therapy Support System* (referred to as its Spanish acronym SATEL). SATEL aims to provide support for SLI diagnosis and language therapy for native Spanish-speaking children. Specifically, the speech therapy provided by the system focuses on helping children to overcome problems as they learn to pronounce certain words, syllables, and phonemes. This study involved language specialists and children treated at the *Psychopedagogical Care Centre for Preschool Education* (referred to as its Spanish acronym CAPEP), in Ensenada, Mexico.

II. TECHNOLOGY SUPPORTING SLI CARE

Due to the high influence and rapid spread of computer-based technology and the Internet, organizations in all fields (e.g. industry, tourism and education), have achieved increased productivity. However, there remain some specific fields or subfields in which this possibility should be addressed or improved. Specifically, in the field of education, the last two decades have seen several theoretical and empirical studies on language learning systems (Schoelles & Hamburger, 1996; Segalowitz & Gatbonton, 1995; Sun & Gao, 2019; Xu et al., 2019).

Recently, research by Sharifi (2018) compared computer-based learning and traditional learning by students of English, and found better language learning performance among students using computer tools, while Sydorenko, Daurio, and Thorne (2018) generated computer simulations based on real-world scenarios to teach students a new language, leading to better performance by students learning to master a new language. However, despite extensive research in educational technology and language pedagogy, schools and teachers have not fully appropriated the technology to apply it to language teaching (Kuure et al., 2015). Particularly, little research was found in SLI. Examples of studies in this field are outlined below.

One of the first tools to address SLI was the BALDI system (Massaro, 2003). BALDI is a computer animation tool that uses a talking head to illustrate the necessary articulatory movements to pronounce a word. Years later, in Toki and Pange (2010), an improved environment was implemented based on the BALDI project. In this system, children have access to the 'Articulation Gym' module, where they can practice phoneme utterance by completing specific tasks.

More recently, Di Pretoro (2016) proposed visual programming languages as a mean to develop SLI treatment. In particular, she recommended the Scratch platform, an environment that uses building blocks of different colours and shapes that resemble Lego bricks, and promotes the structural development of cognition, which is associated with improvement in children's linguistic capabilities.

One of the most recent contributions is presented in Nasiri, Shirmohammadi, and Rashed (2017), who conducted a literature review on serious games used in therapy for children with language problems. Based on the outcomes of this study, the group proposed a game in which the child has to give spoken instructions to move an avatar, explore an environment, and find objects.

Although some progress has been made, none of the above systems provide support for both the diagnosis and treatment of SLI. All the cited tools offer support in one way or another, either by assisting the therapist in diagnosing the SLI or by complementing the child's therapy, disregarding the symbiosis required between these two interconnected activities. Furthermore, they do not consider personalizing the learning environment with features such as a customized system interface, a specific sequence of instructions, or personalized feedback. The rest of this article describes how personalization is addressed in the implementation of SATEL.

III. SATEL IMPLEMENTATION

SATEL was developed to support the diagnosis and treatment of SLI, explicitly addressing phonetics, phonology, and semantics. The concept of SLI diagnosis refers to any medical intervention or examination on an individual who may have any language disorder not caused by any physical or evident condition, in order to identify the nature of the problem (Jackson-Maldonado, 2011). On the other hand, the concept of therapy implies the required post-diagnosis treatment to help a person to overcome a disease or condition (Cambridge Dictionary, 2019).

Ideally, such a system should 1) evaluate incorrect pronunciations produced by the child when speaking, and 2) provide audible and visual feedback. In implementing SATEL, engaging characters were used based on the age of the intended user, using conventional technology like the Kinect motion sensor from the Xbox video game console. As far as system functionality is concerned, open source libraries were used to recognize and evaluate children's pronunciation.

The methodology for developing the proposed system consisted of two main stages: a) *the design and implementation of the environment*, and b) *the evaluation of the proposed environment*.

Design and Implementation of the Environment

In this phase, the structures and algorithms to conduct the diagnosis and treatment of children with SLI were created, as well as the system architecture design and required user interfaces. The design and development of SATEL was structured as follows: 1) implementation of the system architecture; 2) knowledge representation and algorithms; and 3) design of the user interfaces.

A. Implementation of the System Architecture

Figure 1 shows the design of the SATEL architecture, enabling interaction between the components of the environment. The architecture consists of four layers: *User, Application, Database, and Expert Knowledge*.

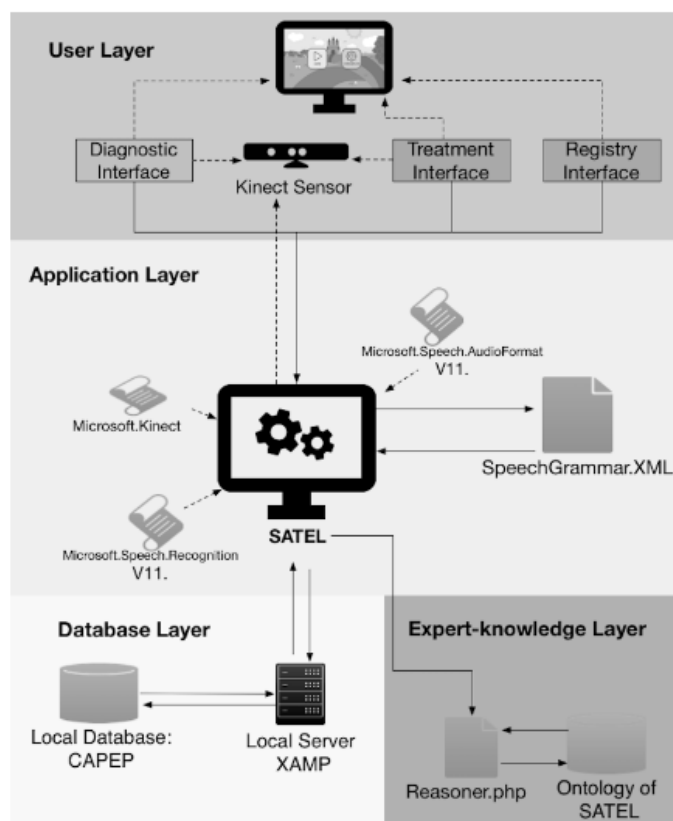


Figure 1. SATEL architecture: a) The User layer presents the devices and interfaces that interact directly with the user; b) The Application layer is the central core of SATEL including the libraries used; c) The Database Layer includes the local database; and lastly, d) The Expert Knowledge Layer incorporates the ontology that represents the expert knowledge.

Within the *User layer*, three components interact directly with the user: 1) the user registration module; 2) the user interfaces, in order to conduct the diagnosis and language therapy; and 3) the Kinect sensor, which captures the user's pronunciations.

In the *Application layer*, through the use of Microsoft® speech recognition libraries, the algorithms implemented in SATEL are used to perform the necessary diagnostic and treatment procedures. These libraries are used to process the information provided by users as they use the system. This information is then compared with a previously defined Spanish grammar included in the *SpeechGrammar.xml* file.

The *Database layer* includes the components that make it possible to store the information provided by users as they interact with SATEL, and the XAMPP server, which enables communication between the database and SATEL.

Finally, within the *Expert Knowledge layer*, the knowledge of the speech-language pathologists is represented by an ontology and formalized using the Web Ontology Language (OWL) (Flórez-Fernández, 2007). According to the research carried out in Liaw et al., (2013) and Liaw et al., (2014), the representation of knowledge through ontologies is suitable for use and application in information systems for health support. There is previous work where ontologies are used to model expert knowledge, such as the work presented in Alloni et al., (2018), where an ontology-based system is developed to support cognitive rehabilitation therapies, which includes exercises similar to those used in speech therapies. Similarly, in Greenbaum et al., (2019), ontologies are used to manage a considerable amount of medical terms and concepts. In SATEL this knowledge base is used through a reasoner to diagnose and determine personalized treatment for children with SLI.

B. Knowledge Representation and Algorithms

Nowadays, adaptive learning has become a relevant research topic. Adaptive learning aims to personalize instruction based on several factors, such as devices, context, and individual student characteristics, customizing aspects like the instructional content and sequence, the look and feel of the material, and the type of feedback (Liu, 2007). Primarily, the design of the domain model of SATEL seeks to adapt and integrate language therapy sequences and instructional content, in order to provide instruction and support according to the specific language requirements of each child using the system.

Considering the heavy reliance on expert knowledge to understand and validate the proposed model, it was decided to explore the use of ontologies. Through this knowledge representation technique, a knowledge structure was obtained that was easy for therapists and computers to understand (Gascuena et al., 2006).

IV. EXPERT KNOWLEDGE BASE

Below is included some of the information obtained from an interview conducted with therapists from CAPEP. The instrument outcomes were used to design and implement the SATEL ontology, which represents the expert knowledge base.

1. The therapists from CAPEP divide Spanish phonemes into four complexity levels:
 - Level I. Vowel phonemes: {a, e, i, o, u}.
 - Level II. Consonant phonemes {b, k, d, f, g, j, l, m, n, p, s, t, y}.
 - Level III. Consonant phonemes {r, err, ch, ñ, z}.
 - Level IV. Consonant phonemes {bra, kra, dra, fra, gra, pra, tra, bla, kla, fla, glo, pla, tla, rra}.
2. Child age and complexity levels:
 - (a) Children between 2 and 2.5 years of age must be competent using words including phonemes from complexity levels I and II.
 - (b) Children between 2.5 and 5 years of age must be competent using words including phonemes from complexity levels I to III.
 - (c) From age 5 and up, children must be competent using words including phonemes from complexity levels I to IV.

A. Description of the SATEL Ontology

The above information was incorporated into the proposed ontology as shown in Figure 2. The ontology consists of three layers of classes: high-level (*Diagnostic* and *Treatment*), core-level (*Word* and *Vocabulary*), and symbolic-level (*Syllable* and *Phoneme*).

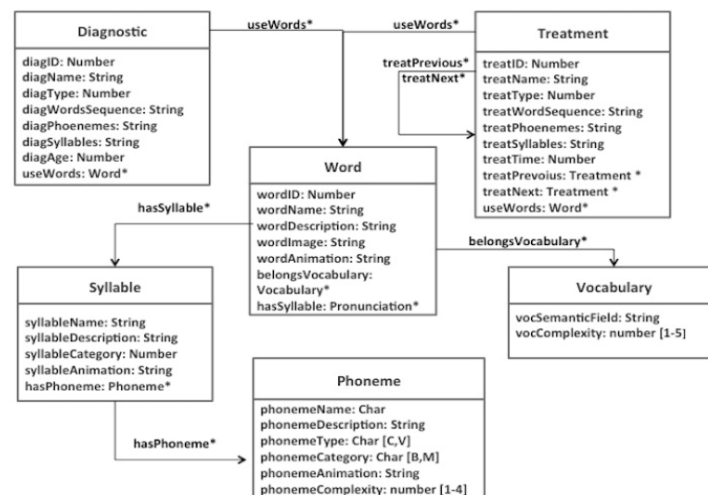


Figure 2. The SATEL ontology. The *diagnosis* and *treatment* classes use the *word* class, composed of syllables and phonemes.

The first layer is intended to cover the top elements within the domain ontology. In this case, it includes the high-level activities that serve to address SLI: *Diagnosis* and *Treatment*. The *diagnostic* class contains descriptive attributes and also some variables to indicate the phonemes, syllables, and words to consider for children of each specific age. Similarly, the *treatment* class represents the phonemes, syllables, and words to be strengthened during the intervention, as well as the previous and next treatment in the event of success, partial success, or failure in reaching the expected pronunciation performance. Both classes are connected to the *word* class through the *useWords* attribute, which makes it possible to establish a relationship between the high-level classes and their used words.

The *core-level* layer includes the words selected for SLI therapy purposes. The *word* class consists of the core elements of the knowledge base, including all the words selected for diagnosis, treatment, or both. The *wordAnimation* attribute is used to indicate the animation file for each specific word. This animation provides the feedback for the correct pronunciation of words, through an avatar representing a girl. Each word can be used for improving pronunciation or vocabulary skills; words are connected to the corresponding classes through the *hasSyllable* and *belongsVocabulary* attributes.

Finally, the *symbolic-level* layer includes the basic classes *Syllable* and *Phoneme*, which are used to compose the set of entities within the *word* class. These classes are also linked through the *hasPhoneme* attribute. The *phonemeAnimation* and *syllableAnimation* attributes are used to indicate the multimedia file.

B. Speech Processing Algorithms

Traditionally, CAPEP therapists use 52 printed images to conduct diagnostic and treatment activities, and commonly show children some or all of these images, as required, hoping that they spontaneously pronounce the corresponding

word. These images are based on the full set of words from the standardized language test “Linguistic Exploration in Preschool Children” (Exploración lingüística del niño preescolar - ELNP) (Ballester, 2019). If therapists are diagnosing SLI, they evaluate the vocabulary level and the correct pronunciation of the words. On the other hand, if they are treating SLI, the therapists provide feedback on the pronunciation mistakes.

The implemented algorithms consistently use the *SpeechGrammar.xml* file, which stores the set of words defined for the diagnosis procedure. Following the recommendations of the CAPEP therapists who participated in the study, the SATEL knowledge base includes the 52 words used in the ELNP language test. This instrument includes categories of words that make it possible to identify phonetic and phonological problems present in children, representing all Spanish phonemes described in the four levels of complexity outlined in the expert knowledge base section. The words in the ELNP instrument are distributed as follows, based on the complexity grading scale for Spanish phonemes: 25 level II, 12 level III, and 15 level IV (all also considered level I). A set of these words was considered in SATEL for the diagnosis of SLI. In addition, in order to avoid duplicating words, the treatment process was supplemented with 38 additional words recommended and provided by the therapists, and distributed as follows: 24 level II, 8 level III, and 6 level IV. The results of this study showed that, by using the Kinect sensor and its development libraries, it is possible to identify the correct or incorrect pronunciation of this group of words and phonemes with a high confidence level. The implemented algorithms are described below.

1. Diagnostic Module

This module shows a set of images, then records and classifies the words pronounced by the child. *Algorithm 1* summarizes the activities in this module.

<p>Algorithm 1. Diagnosis</p> <p>Input: Set of images that represent words.</p> <p>For each image in the diagnostic dataset:</p> <ol style="list-style-type: none"> 1. The image is shown. 2. The child’s pronunciation is recorded. 3. The audio is converted into a character token. 4. The pronunciation is classified as correct or incorrect. <p>Output: The number of correct and incorrect words, and the set of incorrect words.</p>
--

Each image in the dataset has its corresponding word, defined in the grammar file *SpeechGrammar.xml*. Through the *Microsoft.Speech.Recognition* library the system converts the audio received by the Kinect sensor to a character token named *Result.Semantics.Value*. A confidence index is also obtained, indicating the perceived clarity in the pronunciation.

To determine the correct pronunciation of a word, two factors are considered: (1) whether children vocalize the word corresponding to the image presented, and (2) the clarity and precision with which they pronounce each phoneme. Once all the images have been presented to the child, the outputs of the diagnostic module are the number of correct and incorrect words, and a list of incorrect words.

2. Expert Reasoning

Expert knowledge is represented by the ontology depicted in Figure 2. This ontology is used by the reasoner module to conduct the diagnosis and treatment of children with SLI. *Algorithm 2* summarizes the activities in this intermediate process.

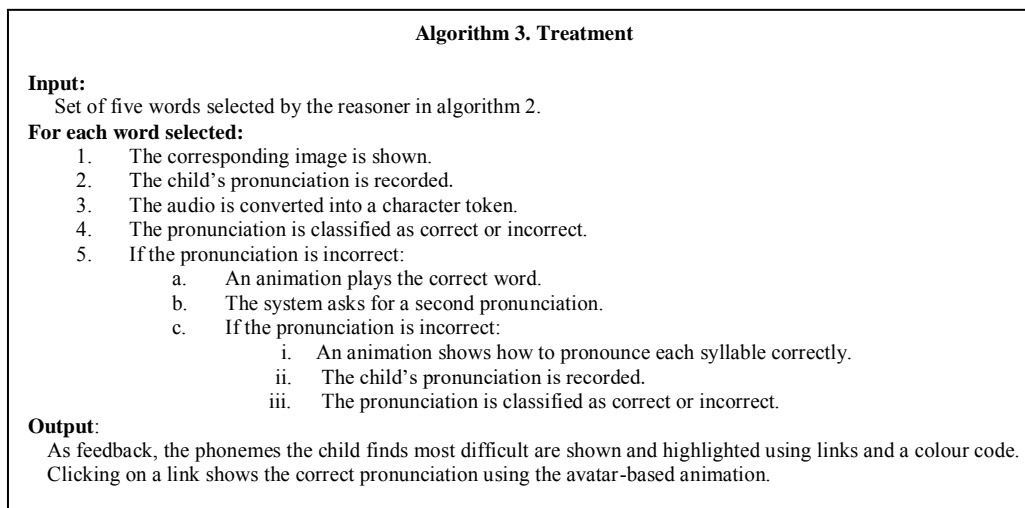
<p>Algorithm 2. Expert reasoning</p> <p>Input: Set of mispronounced words detected in the diagnostic step.</p> <p>For each mispronounced word:</p> <ol style="list-style-type: none"> 1. The word is broken down into syllables and phonemes. 2. The frequency of each phoneme is determined. 3. The most difficult phonemes are detected based on frequency. 4. Five words containing the difficult phonemes are randomly selected. <p>Output: Set of five words.</p>
--

The expert reasoning algorithm uses as an input the set of incorrect words from the child’s last diagnosis. This is sent to the *reasoner.php* file, which uses *algorithm 2* to break down the received words into syllables and phonemes. Each word is separated into syllables, using the knowledge base represented by the ontology. Then, it is separated into phonemes to identify those that recur most frequently; this process allows the system to detect which phonemes are the most difficult for each child.

Once the most difficult phonemes are identified, a new query is sent to the ontology, using the list of detected phonemes as a parameter. This makes it possible to obtain words containing the specific phonemes. Finally, five words that include the problematic phonemes are selected randomly for consideration in the treatment process. Thus, the system is able to provide personalized treatment to children.

3. Treatment Module

In a similar manner to the diagnosis activity, *algorithm 3* shows the required procedure for the treatment activity.



A list of five words is received from the reasoner module. Then, for each word, the corresponding image is displayed to the child (see Figure 3). Like in the diagnostic module, the Kinect sensor detects the words pronounced by the child and then, using the Microsoft® *Speech* libraries, the pronounced word is classified as correct or incorrect.

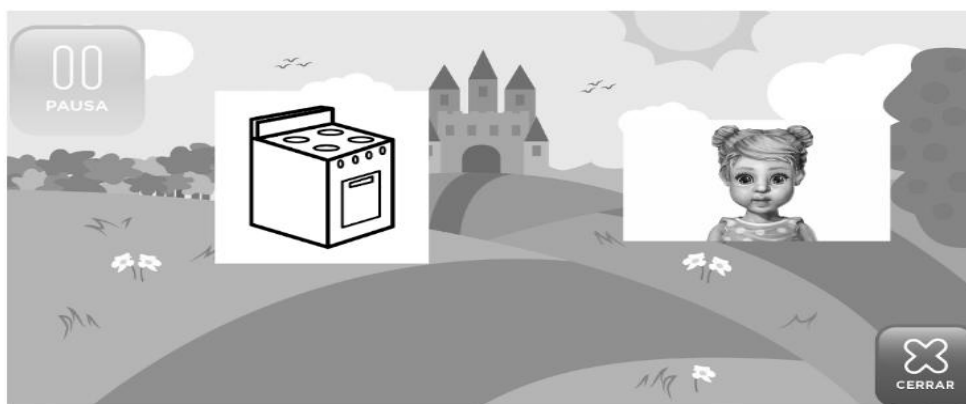


Figure 3. SATEL treatment interface. This interface presents the set of images used during the treatment activity, and the avatar used for feedback.

If the detected pronunciation does not match the displayed image, it will be saved in the database as an incorrect answer, and an animation will be played showing how to say that word. Then, the system asks for a second pronunciation of the same word. If the system detects an incorrect pronunciation again, that word is divided into syllables and an avatar-based animation of how to pronounce each syllable is shown.

V. SATEL EVALUATION

This section describes the results of the evaluation of the system's performance and the user experience provided by the environment.

A. Study Methodology

1. Participants

In general, the implemented system was evaluated by four female therapists with more than 20 years of experience each in the field of SLI, and 26 children. The language therapists were professionals working at CAPEP with expertise in areas involving language development. The children were 20 boys and 6 girls aged between 4 and 7 years. All were from different public schools and received language therapy at CAPEP.

CAPEP receives children identified as having a language problem by teachers from different schools in the town, in

order to diagnose the possible existence of SLI professionally. If CAPEP confirms SLI, the language therapy service is provided. Therapists determined which children participated in the evaluation of SATEL, mainly on the basis of which children displayed a favourable attitude towards interacting with technology and had parental consent.

2. Data Collection and Sources

The data used in this study was collected from multiple sources: interviews, observation, questionnaires, videos, and system logs.

First, during the therapy sessions, each child's pronunciation was logged in the system database, and the whole intervention was recorded. Then, using the logs and videos, precision and accuracy were evaluated to estimate the system's ability to identify correct and incorrect pronunciation of words and syllables by children.

This activity was carried out by the first author of this article. Following the therapists' instruction that a word should be correctly articulated to be considered a satisfactory pronunciation; word pronunciation was evaluated one by one. Then, in order to estimate the precision and accuracy of SATEL in classifying correct and incorrect pronunciations, a comparison was made between the results of the video evaluation and the classification data in the SATEL system logs.

3. SATEL Performance Evaluation

This phase evaluated the *precision* and *accuracy* of SATEL in categorizing correct and incorrect pronunciations.

Considering that SATEL performs a binary classification of the children's pronunciations, *precision*, *inverse precision*, and *accuracy* were estimated using the equations for binary classifiers (Powers, 2011). In this context, *precision* implies the system's confidence level in classifying the pronunciation of a word or syllable as correct (positive pronunciation), and is quantified by equation (1), where *TP* is the number of *true positive pronunciations* and *FP* the number of *false positive pronunciations*:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

Inverse precision measures the proportion of incorrect cases that are indeed negative pronunciations (equation 2), where *TN* is the number of true negative pronunciations and *FN* is the number of false negatives:

$$\text{Inverse precision} = \frac{TN}{TN+FN} \quad (2)$$

Finally, the accuracy of the classifier indicates the proportion of correct classifications with respect to the full set of positive and negative pronunciations, and is defined by equation 3 as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

The results of the performance evaluation of SATEL are given below.

B. Diagnostic Module

A total of 26 children from CAPEP participated in evaluating the SATEL diagnostic module. The purpose of this evaluation was to assess the system's ability to properly identify correct and incorrect pronunciations using the implemented algorithms. Based on the therapists' recommendations, participants were asked to pronounce the first 20 words of the 52 initially included in SATEL. These words were considered by the researchers and therapists sufficient to evaluate the children's speech and language skills and the system's performance. The complexity level of the selected words is distributed as follows: 9 level II, 9 level III, and 2 level IV. These included words with two and three syllables. The data shown in Table 1 was obtained from this evaluation.

TABLE 1
PRECISION, INVERSE PRECISION, AND ACCURACY OF THE SATEL'S DIAGNOSTIC MODULE.
The User Identifier Consists Of A Unique Id Followed By The User's Gender – M (Male) Or F (Female) – And Age.

User	Proper identification of correct pronunciations	Precision	Proper identification of incorrect pronunciations	Inverse Precision	Accuracy
1-M7	4 out of 4	100%	16 out of 16	100%	100%
2-M5	6 out of 7	85.71%	11 out of 13	84.61%	85%
3-M5	2 out of 2	100%	18 out of 18	100%	100%
4-M5	5 out of 5	100%	15 out of 15	100%	100%
5-M4	7 out of 8	87.50%	11 out of 12	91.60%	90.00%
6-F6	0 out of 0	—	20 out of 20	100%	100%
7-M5	5 out of 6	83.30%	14 out of 14	100%	95.00%
8-M6	4 out of 5	80.00%	15 out of 15	100%	95.00%
9-M6	3 out of 4	75.00%	15 out of 16	93.75%	90.00%
10-M6	0 out of 0	—	19 out of 20	95.00%	95.00%
11-F6	4 out of 5	80.00%	14 out of 15	93.33%	90.00%
12-M6	6 out of 6	100%	12 out of 14	85.71%	90.00%
13-M6	2 out of 3	66.66%	17 out of 17	100%	95.00%
14-M6	7 out of 7	100%	11 out of 13	84.61%	90.00%
15-M5	6 out of 7	85.71%	13 out of 13	100%	95.00%
16-M7	10 out of 11	90.90%	8 out of 9	88.88%	90.00%
17-F6	6 out of 7	85.71%	12 out of 13	92.30%	90.00%
18-M6	3 out of 4	75.00%	16 out of 16	100%	95.00%
19-F6	6 out of 6	100%	14 out of 14	100%	100%
20-M6	5 out of 5	100%	15 out of 15	100%	100%
21-M4	0 out of 0	—	20 out of 20	100%	100%
22-M4	7 out of 8	87.50%	12 out of 12	100%	95.00%
23-F6	4 out of 4	100%	14 out of 16	87.50%	90.00%
24-F6	7 out of 7	100%	13 out of 13	100%	100%
25-M6	11 out of 11	100%	8 out of 9	88.88%	95.00%
26-M5	9 out of 9	100%	9 out of 11	81.81%	90.00%
	129 out of 141	91.65%	362 out of 379	94.92%	94.42%

The precision of SATEL varies depending on the detection of correct or incorrect pronunciations. A better performance was observed in identifying words pronounced incorrectly (*TN*), with 94.92% correct identifications. The precision of the system was lower in detecting correct pronunciations. However, an acceptable performance was observed, with an average precision of 91.65%. It should be noted that there were nearly three times more incorrect pronunciations than correct ones; this behaviour is considered normal because the test was conducted with children facing SLI. Overall, for the diagnostic module, the accuracy of SATEL in the detection of both correct and incorrect pronunciations was 94.42%.

The first column in Table 1 (User) uses a letter and number combination to indicate the gender (female or male) and age of children participating in the study. SATEL's performance is consistent regardless of age and gender. The accuracy of the system was 94.44% and 94.41% for children 4-5 and 6-7 years old, respectively; and 95.00% and 94.25% for girls and boys, respectively.

C. Treatment Module

The performance evaluation for the treatment module was carried out with 15 of the 26 students who participated in the previous evaluation: 11 boys and 4 girls. The objective of this evaluation was to assess SATEL's precision and accuracy in identifying correct and incorrect pronunciations of words and syllables. As mentioned in algorithm 2, children were asked to pronounce 5 previously selected words on the basis of their diagnosis results. Whenever they said a word incorrectly twice, they were asked to pronounce the syllables in that word. Our evaluation of SATEL's ability to classify pronunciations takes into account each individual child's pronunciation of words and syllables, resulting in a different number of evaluations for each child (see Table 2).

TABLE 2
PRECISION, INVERSE PRECISION, AND ACCURACY OF THE SATEL'S TREATMENT MODULE.
The User Identifier Consists Of A Unique ID Followed By The User's Gender – M (Male) Or F (Female) – And Age.

User	Proper identification of correct pronunciations	Precision	Proper identification of incorrect pronunciations	Inverse Precision	Accuracy
1-M6	2 out of 2 (0)*	100%	15 out of 16 (9)*	93.75%	94.44%
2-M5	0 out of 0 (0)*	100%	21 out of 21 (11)*	100%	100%
3-M6	2 out of 2 (1)*	100%	19 out of 19 (10)*	100%	100%
4-F6	5 out of 5 (4)*	100%	12 out of 13 (5)*	92.30%	94.44%
5-M4	4 out of 4 (2)*	100%	10 out of 10 (4)*	100%	100%
6-M5	1 out of 1 (1)*	100%	20 out of 20 (10)*	100%	100%
7-F6	1 out of 1 (0)*	100%	18 out of 19 (10)*	94.73%	95.00%
8-M6	4 out of 4 (3)*	100%	12 out of 13 (4)*	92.30%	94.11%
9-M5	1 out of 1 (1)*	100%	21 out of 21 (11)*	100%	100%
10-M4	2 out of 2 (2)*	100%	15 out of 16 (7)*	93.75%	94.44%
11-M4	5 out of 5 (2)*	100%	10 out of 10 (4)*	100%	100%
12-M5	2 out of 2 (0)*	100%	13 out of 14 (7)*	92.85%	93.75%
13-M6	3 out of 3 (2)*	100%	19 out of 19 (10)*	100%	100%
14-F6	3 out of 3 (2)*	100%	16 out of 16 (8)*	100%	100%
15-F6	3 out of 3 (0)*	100%	10 out of 10 (4)*	100%	100%
	38 out of 38	100%	231 out of 237	97.31%	97.75%

*(number of evaluated syllables)

From these results, a fairly similar level of precision was observed in SATEL's classification of incorrect pronunciations. A slight increase of 2.39% was observed in the treatment module. As for correct pronunciations, the results of the experiment showed that 100% precision was achieved. This excellent performance was obtained as a result of considering the evaluation of syllable pronunciation (the number of syllables evaluated is given in brackets). The treatment module exhibited 97.75% accuracy in the detection of pronunciations, both correct and incorrect, which the creators of the system consider a remarkable performance for this version of SATEL. Again, SATEL's performance is consistent regardless of age and gender, exhibiting only a slight variation. The accuracy of the system was 98.31% and 97.25% for children 4-5 and 6-7 years old, respectively; and 97.36% and 97.89% for girls and boys, respectively.

In general, SATEL's ability to identify the correct and incorrect pronunciation of words is outstanding. The use of the confidence index to identify incorrect pronunciations, implemented through the classification algorithms, allowed the system to perform its work effectively, achieving an error rate in accuracy of 5% or less. Furthermore, system performance improves considerably when evaluating the pronunciation of syllables. In this case, when classifying words and syllables, SATEL exceeded the researchers' expectations with around 98% accuracy. However, these outcomes indicate that SATEL performance may be affected by several factors, including word length and the combination of certain syllables.

Another important point to note is the performance of SATEL at an individual level. From this perspective, the lowest classification accuracy during the diagnostic process was 85%; this only occurred in one case. On the other hand, classification accuracy for 57% of children was 95% or more. Similarly, for the treatment module, individual accuracy was 95% or more for 67% of users. These outcomes indicate an adequate performance in general and in individual cases.

Finally, the use of SATEL's ontology was adequate to provide personalized assistance within the treatment module. Once the SLI diagnostic process had been conducted, SATEL was able to personalize treatment therapies by asking for the pronunciation of specific words and syllables according to the children's language deficiencies.

VI. CONCLUSIONS AND FUTURE WORK

This paper described the SATEL system, a computer-based tool intended to complement and reinforce SLI therapies. Unlike current computer-based systems supporting SLI, SATEL integrates language therapists' two main activities: diagnosis and treatment. This paper proposes an ontology to represent the knowledge of speech-language specialists and presents the results of an evaluation of SATEL performance conducted with children facing SLI.

An outstanding system performance was achieved in terms of identifying correct and incorrect words and syllables spoken by children facing SLI. The outcomes indicate that the system, supported by the Kinect sensor, is able to evaluate children's pronunciations with accuracy of 94.42% and 97.75% for the diagnostic and treatment modules, respectively. Using the Kinect sensor and the information provided by its development libraries, SATEL is able to identify correct or incorrect pronunciations of words and phonemes with a high confidence level.

Previous studies on the use of computer-assisted tools to support language learning, specifically in the case of SLI, have emphasized the use of specific multimedia elements to support children in mastering pronunciation skills. Different technologies and learning strategies have been proposed and implemented, such as the use of computer animations, computer games, and mobile technologies. While this interactive approach is fully supported by the authors of this study, it is also complemented by two main characteristics: the design and implementation of an ontology-based knowledge base and a set of algorithms to personalize instruction.

SATEL's threefold ontological scheme enables the integration and linkage of the syllables and phonemes of existing

words. Even though the words currently available in the SATEL ontology are limited to the system's needs, this knowledge representation is scalable and versatile; it can be easily adapted to many different language learning activities. On the other hand, the implemented algorithms provide the capacity to personalize instruction.

REFERENCES

- [1] Alloni, A., Quaglini, S., Panzarasa, S., Sinfioriani, E., & Bernini, S. (2018). Evaluation of an ontology-based system for computerized cognitive rehabilitation. *International journal of medical informatics*, *115*, 64-72. <https://doi.org/10.1016/j.ijmedinf.2018.04.005>
- [2] Ballesteros, F. (2019). *Exploración Lingüística Del Niño Preescolar*. Retrieved April 30, 2019, from https://www.academia.edu/35832501/EXPLORACION_LINGUISTICA_DEL_NINO_PREEESCOLAR
- [3] Cambridge University Press. (2019). *Cambridge online dictionary*, Cambridge Dictionary online. Retrieved at April 30, 2019, from the website <https://dictionary.cambridge.org/es/diccionario/ingles/therapy>
- [4] Di Pretoro, N. (2016). Linguistic and social treatment of Specific Language Impairment (SLI) using Scratch. *Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16*, 541-547. <https://doi.org/10.1145/2930674.2935985>
- [5] Flórez-Fernández, H.A. (2007). Construcción de ontologías OWL. *Revista v ínculos*, *4*(1), 19-34. <https://doi.org/10.14483/2322939X.4112>
- [6] Gascuena, J.M., Fernández-Caballero, A., & González, P. (2006). Domain Ontology for Personalized E- Learning in Educational Systems. *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. <https://doi.org/10.1109/icalt.2006.1652470>
- [7] Greenbaum, N. R., Jernite, Y., Halpern, Y., Calder, S., Nathanson, L. A., Sontag, D. A., & Horng, S. (2019). Improving documentation of presenting problems in the emergency department using a domain-specific ontology and machine learning-driven user interfaces. *International Journal of Medical Informatics*, *132*, 103981. <https://doi.org/10.1016/j.ijmedinf.2019.103981>
- [8] Jackson-Maldonado, D. (2011). La identificación del Trastorno Específico de Lenguaje en Niños Hispano-hablantes por medio de Pruebas Formales e Informales. *Revista Neuropsicología y Neuropsiquiatría y Neurociencias*, *11*(1), 33-50.
- [9] Kuure, L., Molin-Juustila, T., Keisanen, T., Riekkki, M., Iivari, N., & Kinnula, M. (2015). Switching perspectives: from a language teacher to a designer of language learning with new technologies. *Computer Assisted Language Learning*, *29*(5), 925–941. <https://doi.org/10.1080/09588221.2015.1068815>
- [10] Lange, A. A. (2019). Technology, instructional methods, and the systemic messiness of innovation: improving reading fluency for low socio-economic elementary school students. *Educational Technology Research and Development*, *67*(5), 1333-1350.
- [11] Liaw, S. T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S. & Talaei-Khoei, A. (2013). Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International journal of medical informatics*, *82*(1), 10-24. <https://doi.org/10.1016/j.ijmedinf.2012.10.001>
- [12] Liaw, S. T., Taggart, J., Yu, H., de Lusignan, S., Kuziemsky, C., & Hayen, A. (2014). Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. *Journal of Biomedical Informatics*, *52*, 364-372. <https://doi.org/10.1016/j.jbi.2014.07.016>
- [13] Liu, F. (2007). Personalized learning using adapted content modality design for science students. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore! - ECCE '07*. <https://doi.org/10.1145/1362550.1362612>
- [14] Massaro, D. W. (2003). A computer-animated tutor for spoken and written language learning. *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*. <https://doi.org/10.1145/958432.958466>
- [15] Nasiri, N., Shirmohammadi, S., & Rashed, A. (2017). A serious game for children with speech disorders and hearing problems. *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. <https://doi.org/10.1109/segah.2017.7939296>
- [16] Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37-63.
- [17] Schoelles, M., & Hamburger, H., (1996). Cognitive tools for language pedagogy. *Computer Assisted Language Learning*, *9*(2-3), 213-234. <https://doi.org/10.1080/0958822960090210>
- [18] Segalowitz, N., & Gathbonton, E., (1995). Automaticity and lexical skills in second language fluency: Implications for computer assisted language learning. *Computer Assisted Language Learning*, *8*(2-3), 129-149. <https://doi.org/10.1080/0958822940080203>
- [19] Sharifi, M., Rostami AbuSaedi, A., Jafarigohar, M., & Zandi, B., (2018). Retrospect and prospect of computer assisted English language learning: a meta-analysis of the empirical literature. *Computer Assisted Language Learning*, *31*(4), 413-436. <https://doi.org/10.1080/09588221.2017.1412325>
- [20] Sun, Y., & Gao, F. (2020). An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research and Development*, *68*, 1181–1198. <https://doi.org/10.1007/s11423-019-09733-9>
- [21] Sydorenko, T., Daurio, P., & L. Thorne, S. (2017). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, *31*(1-2), 157-180. <https://doi.org/10.1080/09588221.2017.1394326>
- [22] Toki, E. I., & Pange, J. (2010). E-learning activities for articulation in speech language therapy and learning for preschool children. *Procedia Social and Behavioral Sciences*, *2*(2), 4274-4278. <https://doi.org/10.1016/j.sbspro.2010.03.678>
- [23] Vázquez, M.V., Méndez, M.R., Hernández, X., & Bustos, R. (2011). Alteraciones del lenguaje en niños de occidente de México. *E-journal de Medicina Familiar y Atención Primaria Internacional*, *5*(1), 13-20.
- [24] Woodman, A. C., Demers, L., Crossman, M. K., Warfield, M. E., & Hauser-Cram, P. (2018). Part C Early Intervention dosage

and growth in adaptive skills from early childhood through adolescence. *Early Childhood Research Quarterly*, 43, 73-82. <https://doi.org/10.1016/j.ecresq.2018.01.007>

- [25] Xu, Z., Chen, Z., Eutsler, L., Geng, Z., & Kogut, A. (2020). A scoping review of digital game-based technology on English language learning. *Educational Technology Research and Development*, 68, 877–904.

Karla I. Arce-Ruelas is a graduate student from Universidad Autónoma de Baja California (UABC), where received her bachelor's degree in computer science in 2014 and her master's degree in computer science in 2017. She is currently involved in doctoral research studies at UABC. Her research interests focus on the use of technology in support of early childhood education and knowledge modelling for the development of intelligent agents.

Omar Alvarez-Xochihua is a professor of computer science at Universidad Autónoma de Baja California (UABC). His main research interest focuses on the investigation of educational technology, mainly implementing intelligent tutoring systems, modelling students and expert knowledge, and conducting research in the field of natural language processing.

José Ángel González-Fraga received his BSc degree in electrical engineering from Universidad Autónoma de San Luis Potosí (UASLP), México, in 2002 and his MSc and Ph.D. degrees in computer science from Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), México, in 2004 and 2007, respectively. He is currently a full-time professor at Universidad Autónoma de Baja California (UABC). His research interests include pattern recognition, adaptive image processing and also educational robotics and technology.

Evelio Martínez is a graduate of the bachelor's degree in Computer Science from the Autonomous University of Baja California (UABC) since 1991. In 2001 he completed a master's degree in Telecommunications and Information Networks. He has participated in various Conferences, Symposiums and International Forums as a speaker. It has more than 150 dissemination articles published in various print and electronic journals.

Patricia Paez-Manjarrez is a professor of psychology and educational sciences at Universidad Autónoma de Baja California (UABC). Her research interest is in the training of university professors, personnel management in educational institutions and the inclusion of students with Asperger syndrome at the university.