



HAL
open science

Study's System: A Three-Stage System for Children Handwriting Recognition and Spelling Error Detection

Sofiane Medjram, Véronique Eglin, Stéphane Bres

► To cite this version:

Sofiane Medjram, Véronique Eglin, Stéphane Bres. Study's System: A Three-Stage System for Children Handwriting Recognition and Spelling Error Detection. *Multimedia Tools and Applications*, 2024. hal-04519952

HAL Id: hal-04519952

<https://hal.science/hal-04519952>

Submitted on 25 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Study's System: A Three-Stage System for Children Handwriting Recognition and Spelling Error Detection

Sofiane Medjram^{1*}, Véronique Eglin¹ and Stéphane Bres¹

^{1*}Univ Lyon, INSA Lyon, CNRS, LIRIS, UCBL, UMR5205,
F-69621, Villeurbanne, 69100, France.

*Corresponding author(s). E-mail(s):

medjram.sofiane@gmail.com;

Contributing authors: veronique.eglin@insa-lyon.fr;

stephane.bres@insa-lyon.fr;

Abstract

In this paper, we present the first version of Study's system, a system that modernizes learning to spell based on dictation for elementary school children. Study's system configuration is composed of an ISKN Repaper tablet connected to a computer with a screen. The ISKN Repaper tablet allows children to write on an ordinary paper attached to the tablet slate using an ordinary pen. Study's system uses handwriting recognition HTR models on three stages to recognize children handwriting and localize different spelling errors. Study's system is based on dictation academic method; it enables better learning for children, where they receive feedback at moments of adequate concentration. Furthermore, Study's system allows teachers to gain a better understanding of children, track their learning progression more effectively, and adapt to their profiles more efficiently.

Keywords: Offline HTR model, Children Handwriting Recognition, ISKN Repaper, Dictation, Elementary Schools

1 Introduction

At elementary schools, children learn languages by acquiring skills in reading, writing, and spelling words. The classical academic method uses dictation for learning, where the teacher dictates words, children listen and write on paper what they hear (using a pencil/pen), and finally, the teacher reviews and addresses their spelling errors.

The academic method is reliable when the number of children is low. However, when the number of children becomes significant, the task of reviewing children's handwritten words and addressing their spelling errors becomes sluggish. The delayed return of errors to children after dictation can impact their learning due to the disparity in their state of concentration while writing and receiving corrections.

To enhance the efficiency of teacher reviewing tasks and children's spelling learning, we have modernized teaching and learning through the utilization of an ISKN Repaper tablet [1]. We have also studied and experimented several Handwriting Recognition methods [2][3]. This modernization enables teachers to track children's learning progression and adapt appropriate dictations based on individual profiles. Additionally, it allows children to concentrate on dictation learning and receive feedback on their spelling errors at suitable moments.

The ISKN Repaper tablet [1] utilizes an ordinary pen and paper for writing and drawing. In the context of learning to spell, this tablet preserves the natural handwriting aspect and enables children to receive feedback on their spelling errors in real-time or after dictation through external AI machine-based with external screen.

Nowadays, handwriting recognition methods have significantly improved, largely due to the utilization of deep learning techniques [2][4][5][6]. For the Study's System, we investigated and experimented several handwriting recognition models using real and synthetic data of adult and children's handwriting [3]. The result of our investigation showed us how to efficiently build our system. In addressing the issue of spelling errors in children's dictation, we have found that a word-level handwriting recognition approach is the viable solution. Using the ISKN Repaper tablet [1], we employ a BLSTM-CTC encoder-decoder model [2] to modernize spelling learning for elementary school children.

The paper is structured as follows: In section 2, we present various drawing tablets currently available in the market. In section 3, we discuss recent deep learning models for offline handwriting recognition. In section 4, we describe the configuration of Study's system, along with the distinct stages employed for handwriting recognition and spelling error detection. Section 5 presents the results of Study's system, showcasing various case scenarios involving dictation and spelling errors. Finally, section 6 provides the concluding remarks for this study.

2 Drawing tablets

Modernizing teaching methods in elementary schools, particularly concerning the academic dictation approach, primarily involves incorporating interactive techniques that retain the natural movements of handwriting.

Drawing tablets represent a novel form of interaction where human hand movements can be accurately tracked. These tablets can be categorized into two types: tablets that directly draw and display on the tablet screen, and tablets that utilize an external computer to visualize the drawings.

2.1 Drawing tablets with a digital screen

Drawing tablets with digital screens are tablets that do not require an external computer for communication and generating user-drawn renderings. These tablets track the pressure and movement of the user's hand on an integrated screen using a specialized stylus. The emulation of the user's hand on this type of tablet involves three technologies: pressure sensitivity, software pen brush, and refresh rate.

Several companies have designed drawing tablets with digital screens, including notable examples like Wacom Cintiq [7] and iPad Pro .

Although drawing tablets with digital screens and styluses are highly developed, they are more geared towards professional artists rather than children's handwriting recognition and spelling learning (Figure 1).



Fig. 1 Wacom Cintiq Tablet [7].

2.2 Drawing tablets with an external screen

Drawing tablets with an external screen are devices that are connected to an external computer for visualizing and generating user-drawn renderings.

These tablets can be used in two ways: drawing directly on the tablet's surface or on a physical sheet of paper affixed to the tablet slate.

4 *Study system*

Examples of tablets with an external screen include the Wacom Intuos Pro Paper edition and the ISKN Repaper.

The Wacom Intuos tablet consists of a slate surface and a stylus/pen for interaction. To draw using this tablet without attaching a sheet of paper, the user must use the stylus and software for drawing on an external computer to visualize and track the stylus's position and rendering. To draw on paper, a sheet of paper can be attached to the slate's surface, and drawing can be done directly using a Wacom pencil pen (Figure 2).



Fig. 2 Wacom Intuos Pro Paper edition (Stylus and Wacom pencil) [7].

The ISKN Repaper tablet offers a device that functions similarly to a Wacom tablet. It can be used with or without a sheet of paper for drawing.

The ISKN tablet comprises a slate area where an A5-sized sheet of paper can be attached. It utilizes a standard pen (such as a pencil, felt-tip, or ball-point pen) for drawing. The pen interacts with the tablet slate using a magnetic ring attached to it, without requiring a battery or electronics (Figure 3).



Fig. 3 ISKN Repaper Tablet [1].

Although drawing tablets with an external screen are highly advanced and primarily designed for professional artists, their feature of utilizing an ordinary pen and paper sheet makes them highly suitable for children's handwriting recognition and spelling learning. For our proposed system, we have opted for the ISKN Repaper tablet.

3 Deep learning-based handwriting recognition

Handwriting recognition represents a challenging problem. The high variance in handwriting styles, both within the same person and across different individuals, presents significant obstacles in transcribing it into machine-readable text.

Handwriting recognition methods are divided into two types: online and offline. Online methods involve tracking a digital pen to gather information about the position and strokes as the text is being written. Offline methods transcribe text from an image into digital text.

We are situated here in a word recognition context, and the models we are presenting are focused on handwritten sequences. The advantage of this type of model is that sequence-level annotated data is sufficient for training, so it is not necessary to know the precise position of each character in order to learn the model, this represents a significant time-saving in the data annotation process.

Deep learning has been applied to offline handwriting recognition and has successfully produced models with high performance.

In [4], the authors proposed a simple model consisting of an end-to-end encoder-decoder architecture that begins with CNN layers, followed by RNN layers for encoding, and finally a CTC layer for decoding (refer to Figure 4).

A CNN model with 5 layers is employed to extract spatial feature maps from word images, which are then passed to a recurrent network to create contextual feature representations using the BLSTM mechanism. The output of the RNN layers is subsequently fed into a Connectionist Temporal Classification (CTC) layer for predicting sequential digital text (Figure 4).

In [5], an attention-based model for end-to-end handwriting recognition has been proposed.

The attention model proposed is composed of four stages: transformation, feature extraction, sequence modeling, and prediction (refer to Figure 5). The model commences by transforming and normalizing input images from varying irregular shapes. Subsequently, spatial feature maps are extracted from the normalization results using the CNN Resnet model. Then, employing the BLSTM mechanism, a sequence model of the spatial feature maps is generated. Finally, an attention-based decoder receives the sequence model for predicting sequential digital text (Figure 5).

Although the models in [4] and [5] achieve good performances in offline handwriting recognition, they become slow in training when the text sequences become long (as in databases based on lines) due to the use of the LSTM

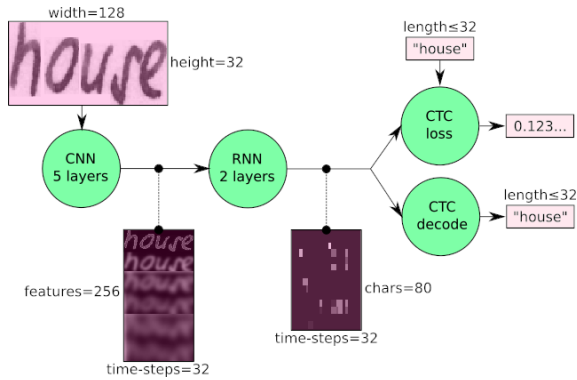


Fig. 4 Offline BLSTM-CTC HTR model proposed in [4].

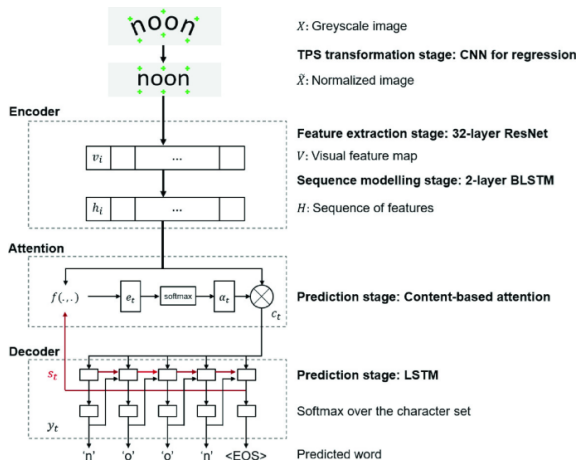


Fig. 5 Offline Attention HTR model proposed in [5].

mechanism. To overcome this drawback, Transformers have been introduced as an efficient alternative solution to LSTM and Attention mechanisms.

In [6], the authors studied the original Transformers architecture model for HTR and proposed a lighter version of it. Their new model contained fewer layers and parameters, yet it demonstrated nearly the same performance as the original HTR transformers (Figure 6).

The HTR transformers model begins by encoding visual features using CNN and Transformer encoder layers. The result of the encoder is then passed to the decoder to sequentially predict the character sequence written within the image. The Transformers encoder relies on multi-head attention, followed by a position-wise feed-forward layer. The Transformers decoder incorporates positional encoding and leverages both the output from the encoder section and the sequence of predicted characters.

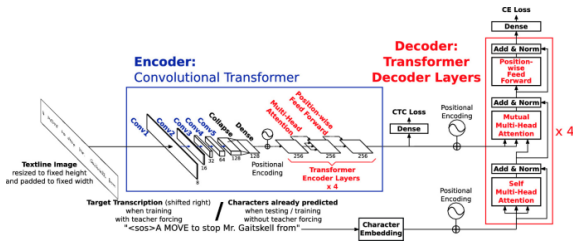


Fig. 6 Offline Light Transformers HTR model proposed in [6].

4 Study's System for Learning to Spell

In this section, we provide a comprehensive overview of our proposed system for learning to spell. We will discuss Study's system configuration, various spelling error scenarios within Study's system, the stages involved in Study's system, and ultimately, the handwriting recognition model integrated into Study's system.

4.1 Study's System presentation

Study's system introduces the modernization of the traditional method of dictation for learning spelling in elementary school. The configuration of Study's system includes an ISKN Repaper Tablet connected to a computer (such as a tablet, laptop, etc.), see Figure 7.

Once children are comfortably seated in front of their computers and tablets, the teacher dictates the text, and the children write the words they hear. Study's system detects various spelling errors and displays them on the screen.

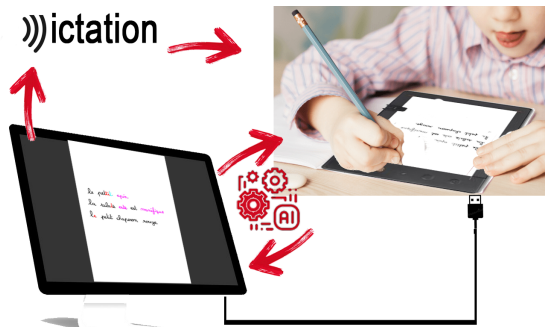


Fig. 7 Study's system first version presentation.

4.2 Study's System scenarios

Our system for detecting spelling errors based on dictation primarily relies on offline handwriting recognition, the quality of interaction with the tablet, and the text image rendering quality of the tablet being used.

Thanks to the ISKN Repaper tablet, the rendering of handwritten text images is consistently clean and of high quality, making the task of offline handwriting recognition more achievable. In terms of interaction quality, the ISKN Repaper tablet employs a specialized ring attached to the pen, maintaining a defined distance from the pen's tip (neither too far nor too close). The daily use of the pen and the manner in which it is used can impact the ring's position, necessitating periodic verification to ensure optimal interaction and rendering.

At the handwriting recognition system level, the system involves the comparison, word by word and character by character, of a dictated text (the ground-truth text) with the tablet-rendered text.

Five scenarios comparison are expected by Study's system, these scenarios are as follow:

1. The handwritten text aligns with the dictation text when no words are missing and no spelling errors are detected.
2. The handwritten text aligns with the dictation text's word count, may vary in terms of word vocabularies from the dictation, and words that share likelihood with dictation word vocabularies may or may not contain spelling errors.
3. The handwritten text does not align with the word count of the dictation text, may vary in terms of word vocabularies from the dictation, and words that share likelihood with dictation word vocabularies may or may not contain spelling errors.

4.3 Study's system stages

Based on the comparison scenarios anticipated by Study's system, we are proposing a three-stage system for handwriting recognition and spelling error detection.

4.3.1 Study's system first stage

The first stage of Study's system consists of an offline handwriting recognition model that categorizes detected words into accepted or rejected vocabulary. Utilizing the knowledge that the words in the dictation text are familiar, we classify each transcribed word produced by the model. It is classified as part of the accepted vocabulary if it bears a 50% likelihood resemblance to one of the dictation words; otherwise, it is rejected (for further details, refer to Algorithm-1 below).

4.3.2 Study's system second stage

At this stage, the system will focus solely on the accepted word vocabularies identified in the first stage.

The second stage of Study's system also involves an offline handwriting recognition model. For each word written by the child, its transcription as generated by the model will be compared to the dictation words in terms of word position and spelling errors. However, the repetition or missing cases (such as repeated accepted words or missing accepted words) are not yet addressed in this initial version of Study's system (for more details, please refer to Algorithm-2 below).

4.3.3 Study's system third stage

The final stage of the Study system involves visualizing and presenting all correctly recognized words, identifying various spelling errors detected within accepted words, and listing different rejected words.

This last stage can be achieved through a reversed offline handwriting recognition model. Utilizing the transcription results obtained from the model used in stage 2, a reverse operation is applied to identify spelling errors character by character across various segments of the original accepted word image.

For lowercase words, an enhanced approach involves dividing the accepted word image into segments corresponding to the number of characters in the model's transcription (refer to Algorithm 3 for details).

4.4 Study's system handwriting recognition model

The Study's system presented above utilizes an offline handwriting recognition model at each composing stage. Thanks to the result of our research investigation on HTR models in [3], we employed the offline BLSTM-CTC word level model [4] for classifying detected words into accepted or rejected word vocabularies, transcribing accepted words, and visualizing spelling errors.

The BLSTM-CTC model represents the lightest HTR encoder-decoder model in the literature, with only 6.1 million parameters and 7 layers.

We applied different training weights to the BLSTM-CTC model across system stages. While using the same weights for all stages is possible, employing separate weight models proves more suitable for the dictation case. Here, the words are already known, and the model at each stage is more specialized and requires a smaller training dataset.

Based on our experiments using real children's data [3], we discovered that:

1. BLSTM-CTC effectively recognizes and handles children's handwriting images, similar to adult handwriting images.
2. Transfer learning of BLSTM-CTC from an adult database (such as IAM database) and synthetic data enhances performance on children's handwriting images.

3. Training the model on limited vocabularies enables achieving human-level performance (100%) on the same word vocabularies, regardless of new handwriting styles.

4.4.1 Study's system offline HTR model first stage

The first stage of Study's system consists of classifying words into accepted and rejected categories based on the dictation vocabularies. For each ASCII transcription of the detected handwritten word, the transcription is compared with all ASCII words from the dictation. If the comparison yields a 50% likelihood, the word is accepted; otherwise, it is rejected (refer to Algorithm 1).

For this stage, we used the BLSTM-CTC model weights that resulted from the IAM transfer learning conducted on real children handwriting images, as presented in [3] (Table 1).

Table 1 BLSTM-CTC model performances on children handwriting database with and without IAM transfer learning (bold text represents best result: low CER and high WAR). Children handwriting database presents 3832 real words collected from Scoledit [8] (3640 train, 192 val, 192 test). Results extracted from [3]

Transfer Learning	Val CER %	Val WAR %	Test CER %	Test WAR %
No	9.50%	77.08%	31.09%	54.68%
Yes	06.38%	82.81%	28.87%	56.25%

The core concept of this stage is to prioritize a stronger likelihood with dictation words, even if the performance of the used model is not perfect. This stage addresses situations where children may partially write words, cross them out, write completely incorrect words, or misunderstand due to hearing confusion.

4.4.2 Study's system offline HTR model second stage

From the list of accepted words resulting from Study's system's first stage, we utilize BLSTM-CTC model weights derived from dictation words, both with and without anticipated spelling errors.

Based on our study conducted on real children's data gathered from an elementary school [8], we determined that most spelling errors manifest in combinations of phonetic characters, character omissions, or instances of character mishearing. These discrepancies, when contrasted with correct words, can alter the word's length by preserving, removing, or extending characters.

The second HTR model employed in Study's system comprises a specialized BLSTM-CTC model intended for recognizing dictation words, both with and without spelling errors, based on phonetic character combinations, character omissions, and character mishearings. The specialized BLSTM-CTC model utilized in stage 2 employs Beam Search-based decoding [2].

Algorithm 1 Study's System First Stage

Require: BLSTM-CTC Model 1, Word Detector Model

```

1: dictation = Read words dictation
2: img = Read ISKN Reapaper image rendering
3: words = Detect words localisation on img
4: HTR1 = load BLSTM-CTC Model 1
5: for word ∈ words do
6:   wordR = HTR1(word) transcribing word image with Model 1
7:   for dic ∈ dictation do
8:     if wordR is 50% dic then
9:       AcceptedWords ← word
10:      continue
11:    else
12:      RejectedWords ← word
13:    end if
14:  end for
15: end for

```

Some pioneering works are founded upon phonetic guidance from the Phonetisaurus model [9], which is a stochastic Grapheme-to-Phoneme (G2P) model based on a Weighted Finite State Transducer (WFST). The WFST employs the principle of joint sequences [10] to align grapheme sequences with their corresponding phonemes. An N-Gram model is derived from the aligned joint sequences and transformed into a WFST model. Subsequently, the G2P model can predict the pronunciation of a new word. In Corbillé's research [11], the author harnesses the G2P model in conjunction with a Phoneme-to-Grapheme (P2G) model to generate a set of phonetically similar pseudo-words. In the scope of our study, phonetic variations were meticulously crafted through a collaborative effort involving a team of educators and linguists who possess specialized expertise in children's spelling and writing acquisition.

Data preparation is a crucial process that significantly influences model training and performance. In our case, data preparation was rooted in the actual collection of data from several adult users (10 users) utilizing the ISKN tablet (in lowercase). These acquisitions were subsequently normalized to IAM database annotations. Each user wrote three sentences, six times each, both with and without spelling errors, to ensure intrinsic variability in handwriting.

We defined three dictation sentences in the French language for data acquisition:

1. *le lapin est magnifique*
2. *la tortue mange de la salade verte*
3. *le petit chaperon rouge*

By utilizing phonetic character combinations, character omissions, or instances of character mishearing, we expand the list of dictation words to

include expected spelling errors (Table 2). In comparison to a correct word, spelling errors can maintain, remove, or extend the number of characters within it (refer to Algorithm 2). A selection of spelling errors is presented in the table.

Table 2 List of some spelling errors including phonetic character combination, character missing and confusion hearing character

	spelling error (-)	spelling error (=)	spelling error (+)
lapin	apin	lapun	lapain
tortue	tortu	tordue	torttue
chaperon	chapron	shaperon	chapperon

During our experiment, we prepared 2335 word images, both with and without spelling errors. We divided the data into 95% for training and 5% for validation, followed by testing the model on several new users.

Like any offline HTR model, the Character Error Rate (CER) and Word Error Rate (WER) are utilized for performance evaluation. The CER is calculated as the Levenshtein distance, which aggregates character substitutions (S_c), insertions (I_c), and deletions (D_c) required to transform one string into another, and then divides by the total number of characters in the ground truth (N_c). Formally refers to Equation 1,

$$CER = \frac{S_c + I_c + D_c}{N_c} \quad (1)$$

Similarly, the Word Accuracy (WA) metric calculates the sum of word substitutions (S_w), insertions (I_w), and deletions (D_w) needed to transform one string into another. It is then divided by the total number of words in the ground truth (N_w). Formally refers to Equation 2,

$$WA = 1 - \frac{S_w + I_w + D_w}{N_w} \quad (2)$$

The performances obtained are presented in Table 3.

Table 3 BLSTM-CTC model performances on our database based dictation

BLSTM-CTC	Word CER %	Word WAR %
Validation set	0.04%	99.06%
Test set	0.12%	98.47%

At this stage, the Word Accuracy Rates (WAR) are excellent, and the Handwritten Text Recognition (HTR) is capable of recognizing words and their phonetic variants with a very high degree of accuracy. The inclusion of comprehensive examples of phonetic compositions in the dataset allows for diversifying each word through controlled variation. Word variants that exhibit

non-phonetic construction or entirely aberrant character compositions are rejected due to their non-recognition. These will be identified at the end of the dictation without pinpointing the exact error locations (whether orthographic or grammatical).

Algorithm 2 Study's System Second Stage

Require: BLSTM-CTC Model 2, AcceptedWords

```

dictation = Read words dictation
2: img = Read ISKN Reapaper image rendering
   HTR2 = load BLSTM-CTC Model 2
4: wordR = HTR2(word) transcribing word image with Model 2
   i = 0
6: j = 0
   indexSpellErrors = []
8: while i < len(AcceptedWords) do
   if AcceptedWords[i] == dic[i] then
10:     Fully recognized word
       continue
12: else if len(AcceptedWords[i]) == len(dictation[i]) then
   Spelling error with same number of character than word dictation
14:     for char ∈ dictation[i] do
       if char! = AcceptedWords[i][j] then
16:         indexSpellErrors.append(j)
       end if
18:     j = j + 1
       end for
20: else if len(AcceptedWords[i]) > len(dic[i]) then
   Spelling errors with more number of characters than word dictation
22:     for char ∈ dictation[i] do
       if char! = AcceptedWords[i][j] then
24:         indexSpellErrors.append(j)
       end if
26:     j = j + 1
       end for
28: else
   Spelling errors with less number of characters than word dictation
30:     for char ∈ dictation[i] do
       if char! = AcceptedWords[i][j] then
32:         indexSpellErrors.append(j)
       end if
34:     j = j + 1
       end for
36:     end if
       i = i + 1
38: end while

```

4.4.3 Study's system offline HTR model third stage

The final stage of Study's system involves the localization of identified spelling errors within word images.

As previously mentioned, a reverse HTR process can be employed here, utilizing the same weights as Study's system's second stage. Starting from the ASCII transcription containing spelling errors, we identify the characters with spelling errors. Subsequently, using a sliding window approach, we search for the corresponding erroneous character within the image.

In the first version of Study's system, where we established a simple database in lowercase, we implemented a boosted method for the localization of spelling errors. Leveraging the transcription results from the BLSTM-CTC model in stage 2 and the positions of spelling errors, we divide the word image into n slices, corresponding to the length of the transcription. We then highlight the slices containing spelling errors (refer to Algorithm3).

Algorithm 3 Study's System Third Stage

Require: BLSTM-CTC Model 2, indexSpellErrors

```

dictation = Read words dictation
img = Read ISKN Reapaper image rendering
3: HTR2 = load BLSTM-CTC Model 2
   wordR = HTR2(word) transcribing word image with Model 2
   slices = []
6: lenWordR = len(wordR)
   slices = divideWordImageIntoSlices(lenWordR)
   for  $i \in \text{indexSpellErrors}$  do
9:   highlight(slices [i])
   end for

```

5 Experiments and results.

In this section, we showcase the results of Study's system using real data acquired from the ISKN Repaper tablet. The Study system underwent testing with multiple users, yielding highly promising and noteworthy outcomes.

5.1 Handwriting acquisition using ISKN Tablet

Here's an example of user acquisition for dictation sentences. The user's acquisition includes several spelling errors and encompasses all the various expected scenarios, such as correct handwriting words, handwriting words with spelling errors, and handwriting words different from the dictation vocabulary (Figure8).

le petit apin
 la salate est est magnifique
 le petit chapeau rouge
 le petit apin stop rouge

Fig. 8 ISKN User handwriting text.

5.2 Accepted and rejected word vocabularies

In order to avoid providing incorrect spelling feedback to the child during dictation, the rejection of unrecognized words is crucial. When no recognition occurs (due to an inadequate recognition score from the proposed word to the BLSTM-CTC model), or when alignment between the ASCII transcriptions of recognized items and those of the dictated words is unachievable, the decision is made to reject the word currently being processed. This is the case for the word 'stop,' which gets rejected and for which no feedback will be given to the children. The implementation of rejection leads to a 6.5% enhancement in the system's word recognition rate (compared to our previous best results in [3]), thereby providing a more precise feedback to the child. Please refer to Figure 9 for visualization.

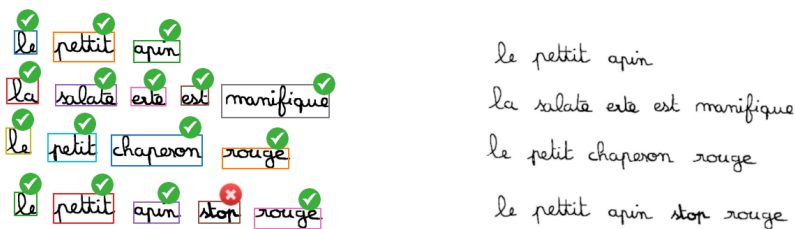


Fig. 9 Study's system accepted and rejected words visualization.

5.3 Recognition and spelling errors localization

Building upon the approved words extracted from the prior stage of the study, the specialized BLSTM-CTC model will be employed for both detecting and pinpointing spelling errors. The localization of spelling errors is depicted in (Figure 10).

This stage facilitates an in-depth analysis of the handwriting, allowing for instant spelling feedback on a range of errors that the child might have

le petit apin
 la salate est est magnifique
 le petit chapeau rouge
 le petit apin ~~apin~~ rouge

Fig. 10 Study's system spelling errors visualization.

committed. The overall method relies on a three-step process: verifying the presence of dictated words, fine word recognition along with potential phonetic or orthographic errors, and visualization of corrections. The nuanced analysis of each word is guided through a progressive and step-by-step approach, which either offers the child a point of error localization or dismisses the word as incorrect (either an extra word or severely misspelled).

The effectiveness of this approach hinges on the quality of the knowledge incorporated into the guidance, which could pose a limitation when adapting the method to another language where phonetic guidance needs updating or is not feasible. Furthermore, employing guidance extends the analysis duration, potentially impeding the ability to offer prompt feedback to the child for words that are excessively segmented (words composed with inaccurate hyphenation or combinations of words).

6 Acknowledgments

We warmly acknowledge the Auvergne-Rhone-Alpes Region for supporting the *Study* project on the Research and Development Booster 2020-2023 Program involving several teams of researchers and industrialists including the two French partner companies *AMI* and *SuperExtraLab*.

7 Conclusion

In this paper, we present the first version of Study's system: an artificial intelligence system designed for children's handwriting recognition and spelling error detection. We demonstrate how the integration of the developed ISKN Repaper tablet with an efficient BLSTM-CTC deep learning model contributes to the modernization of the dictation-based school method for both teachers and children.

With the use of the ISKN Repaper tablet, the traditional academic approach of learning to spell with pen and paper is maintained. Children receive feedback on spelling errors at moments conducive to concentration, and teachers are better able to monitor children's learning progression.

We illustrate how our three-stage system for dictation-based learning to spell effectively manages various spelling error scenarios. By employing specialized offline HTR models at each stage of the system, we achieve improved performance, reduced data requirements, and shortened training times.

Our work has demonstrated the value of the encoding and decoding method for words using a simple recurrent network that incorporates Connectionist Temporal Classification (CTC) for providing feedback on the quality of writing's spelling. In its current version, The system can identify and pinpoint spelling errors based on a prompt. Looking ahead, there is potential to enrich this feedback by introducing the concept of error types (such as phoneme confusion, dyslexia, etc.) in order to focus on specific areas of improvement for each student. This diagnostic approach could also be beneficial for teachers in hindsight, assisting in identifying potential specific written language disorders.

References

- [1] ISKN: Tablette graphique Repaper: tablette papier crayon originale iskn" (2020). <https://www.iskn.co/fr>
- [2] Scheidl, H., Fiel, S., Sablatnig, R.: Word beam search: A connectionist temporal classification decoding algorithm. Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR **2018-Augus**, 253–258 (2018). <https://doi.org/10.1109/ICFHR-2018.2018.00052>
- [3] Medjram, S., Eglin, V., Bres, S.: Challenging Children Handwriting Recognition Study Exploiting Synthetic, Mixed and Real Data. In: Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13237 LNCS, pp. 538–552 (2022). https://doi.org/10.1007/978-3-031-06555-2_36
- [4] Scheidl, H.: Build a Handwritten Text Recognition System using TensorFlow (2018). <https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5>
- [5] Kass, D., Vats, E.: AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **13237 LNCS**, 507–522 (2022) <https://arxiv.org/abs/2201.09390>. https://doi.org/10.1007/978-3-031-06555-2_34
- [6] Barrere, K., Soullard, Y., Lemaitre, A., Couïasnon, B.: A Light Transformer-Based Architecture for Handwritten Text Recognition, pp.

- 275–290. Springer (2022). https://doi.org/10.1007/978-3-031-06555-2_19.
https://link.springer.com/chapter/10.1007/978-3-031-06555-2_19
- [7] Wacom: Cintiq 22HD Creative Pen Display (2015). <https://www.wacom.com/fr-fr/products/wacom-cintiq>
- [8] Université Grenoble Alpes: SCOLEDIT - Etat du Corpus (2013). <http://scoledit.org/scoledition/corpus.php>
- [9] Novak, J.R., Dixon, P.R., Minematsu, N., Hirose, K., Hori, C., Kashioka, H.: Improving WFST-based G2P conversion with alignment constraints and RNNLM N-best rescoring. 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012 **3**(August 2015), 2525–2528 (2012). <https://doi.org/10.21437/interspeech.2012-654>
- [10] Bisani, M., Ney, H.: Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication* **50**(5), 434–451 (2008). <https://doi.org/10.1016/j.specom.2008.01.002>
- [11] Krichen, O., Corbillé, S., Anquetil, E., Girard, N., Nerdeux, P.: Online Analysis of Children Handwritten Words in Dictation Context. In: *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12916 LNCS, pp. 125–140 (2021). https://doi.org/10.1007/978-3-030-86198-8_10. https://www.researchgate.net/publication/354357100_Online_Analysis_of_Children_Handwritten_Words_in_Dictation_Context