



HAL
open science

Training speech-breathing coordination in computer-assisted reading

Delphine Charuau, Andrea Briglia, Erika Godde, Gérard Bailly

► **To cite this version:**

Delphine Charuau, Andrea Briglia, Erika Godde, Gérard Bailly. Training speech-breathing coordination in computer-assisted reading. Interspeech 2024 - 25th Annual Conference of the International Speech Communication Association, Sep 2024, Kos, Greece. pp.5128 - 5132, 10.21437/interspeech.2024-992 . hal-04700395

HAL Id: hal-04700395

<https://hal.science/hal-04700395v1>

Submitted on 17 Sep 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.



Distributed under a Creative Commons Attribution 4.0 International License



Training speech-breathing coordination in computer-assisted reading

Delphine Charuau¹, Andrea Briglia¹, Erika Godde², Gérard Bailly¹

¹GIPSA-Lab, Univ. Grenoble-Alpes, CNRS, Grenoble INP, France

²LEAD, CNRS, Université de Bourgogne, France

delphinecharuau1@gmail.com, erika.godde@u-bourgogne.fr, gerard.bailly@gipsa-lab.fr

Abstract

Aims of this study are: 1) identify respiratory features that could serve as objective markers of fluency of aloud reading; 2) investigate the effects of computer-assisted training on the improvement of speech-breathing coordination. Our training method combines the principles of repeated and assisted close-shadowed reading. Reading assistance takes over the principles of karaoke: highlighting text units in sync with pre-recorded adult performance. The aloud reading of 66 French young pupils is studied. They were divided into 3 groups: control (repeated reading only with no assistance), trained with word highlighting vs. trained with word and breath group highlighting. All children were recorded before and after 3 weeks of training while reading both a trained and an untrained text. The results indicated that respiratory planning and pauses management improved by computer-assisted reading training. However, there was no significant transfer of these improvements to the untrained text.

Index Terms: pauses, speech-breathing coordination, prosody, computer-assisted reading

1. Introduction

Reading fluency is traditionally measured as the number of words correctly read per minute, making decoding ability and speech rate the main indicators of reading proficiency [1]. However, this conception of fluency completely overlooks a crucial aspect of aloud reading: prosody. Prosody, which includes rhythm, phrasing, and intonation variations, plays a significant role in text comprehension. Studies have shown that prosody in aloud reading facilitates the listener's text comprehension, and reflects the reader's own comprehension of the text, underscoring the importance of how words are read [2].

Our study focuses on the development of speech-breathing coordination in children, and its impact on children's phrasing on a computer-assisted reading program based on the karaoke principle. This method combines the benefits of choral reading and repeated reading, which have already demonstrated its effectiveness in second language acquisition [3, 4].

The aim of our study is twofold: (a) analysing the speech-breathing coordination in this particular context of computer-assisted reading in order to identify respiratory features that could be considered as the hallmark of fluency improvement; (b) quantifying the impact of specific training on speech-breath coordination through computed-assisted reading. By investigating how children breathe during karaoke reading and how it influences their phrasing, we want to understand the benefits of this reading assistance on the development of respiratory patterns and their contribution in improving reading proficiency.

2. State of the art

Breathing contributes to the rhythm and phrasing of reading aloud. Reading aloud requires a specific control of breathing to coordinate respiratory needs with the discourse structure. Inhalations mainly occur at the boundaries of syntactic units [5, 6]. The text structure determines the distribution of pauses: breathing pauses occur at the end of paragraphs and sentences, and non-breathing pauses are more commonly placed between syntactic units [7, 8]. Texts provide visual cues (punctuation, spaces) that favour respiratory and pause planning. Therefore, reading relies on readers' ability to anticipate punctuation marks as indicators of pauses, and their mastery of the syntactic and semantic structure of the text to realise these pauses.

Given the importance of pause placement in reading, it is important to consider the development of these skills in learners. Pauses should mostly be grammatical, but children's reading is characterized by a high number of ungrammatical pauses, due to decoding errors and lack of pause planning [9, 10]. A recent cross-sectional study [11] reveals a significant evolution in the management of pauses during the early years of reading learning. Their results indicate a significant decrease in the ratio of ungrammatical pauses to the total number of pauses between 3rd and 4th grades. This proportion gradually decreases from grade 4 to grade 7, although it remains higher than that observed in adults, suggesting a continuous developmental process. Thus, pause management appears to be a dynamic aspect of reading learning, gradually evolving towards patterns similar to those of adults over time.

The distribution of pauses is highly reliant on respiratory control. Hence, the development of speech-breathing coordination is therefore an important issue in reading learning. This coordination develops gradually between 4 and 10 years old, with a significant transition towards more mature respiratory patterns between 7 and 10 years old [12, 13]. Children take a breath at more frequent intervals than adults, given their smaller lung capacities [14]. Furthermore, respiratory strategies for speech are not yet fully developed. Several studies indicate that younger children regularly rely on their expiratory reserve volume to complete their breath groups, demonstrating an immaturity in respiratory planning for speech [12, 15]. Their expiratory expenditure per syllable is higher than that of older children [13]. This large expenditure, combined with a slower speech rate, reduces the number of articulated syllables in a breath group. Immature respiratory control can be detrimental to reading fluency: frequent interruptions in speech flow compromise fluency. Children's respiratory strategies gradually refine as they grow, approaching patterns observed in adults between 10 and 14 years old [13, 15]. The development of speech-breathing coordination occurs parallel to the development of reading skills,

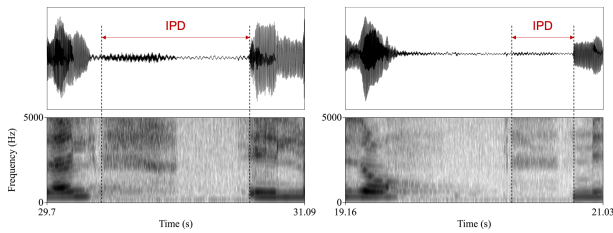


Figure 1: Illustrations of a typical IPD of a novice reader, at the beginning of the pause (left) vs. an IPD of a fluent pupil at the end of the pause, similar to the adult pattern (right).

suggesting that some respiratory features could serve as objective markers of fluency of aloud reading.

In this study, we focus on one of the indices of speech-breathing coordination: *inhalation-to-phonation delay* (IPD). This measure, representing the interval between the onset of inhalation noise and the onset of speech, evolves during growth. Based on the description of respiratory pauses by Grosjean & Collins [8], IPD corresponds to inhalation and the subsequent post-inspiratory phase before phonation. In adults, there is no post-inspiration: inhalation usually occurs at the end of the pause, *just in time* (JIT) before speech. The average IPD in adults is about 400 ms [11]. In contrast, children tend to take a breath immediately after speech, suggesting an *emergency or as-soon-as-possible* (ASAP) inhalation. Thus, in children, the pre-inspiration phase is very short compared to adults, while their post-inspiration phase is significantly longer (see figure 1). According to Godde et al [11], in 2nd grade, IPD is significantly greater compared to adults and older children. From the 3rd grade onwards, IPD decreases gradually with increasing grade level. One of the purpose of our computer-assisted reading program is to improve breathing planning, of which inhalation at the end of the pause is a sign.

3. Background

Godde et al [16] conducted a study on the subjective evaluation of fluency in a computer-assisted reading context, in French. Novice readers were recorded before and after 3 weeks computer-assisted reading training. Their fluency was evaluated by three researchers using the multidimensional fluency scale [17] adapted for French [18], for their first minute of reading. The results showed a significant improvement in the total fluency score, in particular for phrasing, *i.e.* pausing placement and logical grouping of words among them, with computer-assisted reading training. Based on this finding, we hypothesise that computer-assisted reading should have beneficial effects on the speech-breathing coordination. This should be evidenced by (a) a reduction in pause duration and number, (b) a reduction of IPD (as a sign of JIT inhalation) and (c) a decrease in the number of ungrammatical pauses after computer-assisted reading training. The second hypothesis is that IPD mean and variance are reliable objective markers of fluency in aloud reading. Our analysis is based on the data from the study by Godde et al [16]. In this paper, we compare the results of an objective analysis of speech-breathing coordination with their subjective analysis on the same dataset. The method employed in their study is detailed below.

4. Method

4.1. Computer-assisted reading

The original experiment was conducted using the karaoke reading application RAKE (Reading Assistance by KaraokeE), which provides audio-visual assistance for aloud reading and allows the user to read-while-listening (RwL)[19]. Equipped with headphones, the speaker hears an adult reading the text displayed on the screen while text units being highlighted in synchrony. Two levels of highlighting were studied : word-only vs. word with the preceding and following breath groups.

The novice reader benefits from explicit visual and auditory training, with gradual text highlighting to guide breathing and pause placement, and with a prosodic model provided by the expert reader. Highlighting by breath group aims to assist novice readers in anticipating inhalations and placing them at adequate syntactic boundaries. The model provided by the expert reader allows progression in terms of phrasing, expressiveness, and rate, by imitating the expert's way of distributing pauses as well as appropriate intonation.

This type of training furthers implicit learning through *close-shadowing* reading, where the novice reader closely follows the expert reader with a very short delay, typically around 100 ms [20]. This aims at an implicit transfer of motor skills, particularly regarding breathing. Furthermore, repeated and varied reading favours fluency improvement and the learning of reading skill such as rate and prosody [21].

4.2. Experimental protocol

97 pupils between 3rd and 5th grades, from 9 classes across 2 elementary schools, were selected to participate in this study. After pre-tests, 66 pupils were chosen according to their good reading profile. They demonstrated satisfactory decoding skills and all displayed a reading rate ranging from 90 to 130 words per minute. Their prosody was evaluated using multidimensional fluency scale from Rasinski [17], adapted to French [18], with a score of 2 out of 4 for speed, but less than 3 for phrasing. The syntactic skills of the children were assessed in pre- and post-tests. These assessments were not used for participant selection. Pupils who missed two or more training sessions, as well as those who missed the pre- or post-test phase, were excluded from the study. Finally 66 pupils were randomly assigned to 3 groups based on the training mode they followed:

- Group C (control): without training;
- Group W: training with word highlighting;
- Group B: training with word, breath group and pause highlighting.

The distribution of global scores and syntactic scores are homogeneous across the groups. Children in the experimental groups (W and B) participated in 9 sessions of 20 minutes on the computer-assisted reading using the corresponding highlighting type for their group. Training sessions occurred in small groups of 8 participants during class hours, under the supervision of a researcher. During each session, pupils read aloud the same text 3 times. The training program spanned 3 weeks, and each week, a new text (A, B, C) was introduced. In pre- and post-test, children were recorded on 4 texts, 3 of which they have been trained on (A, B, C), and one for which they had not (D). This protocol was approved by local ethical committee and French Ministry of National Education.

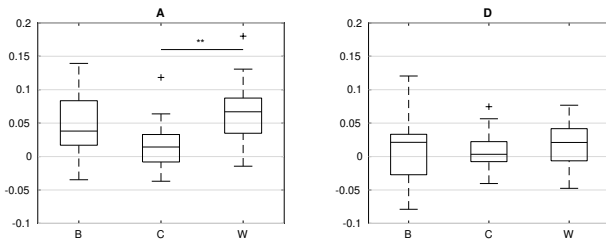


Figure 2: Variation in phonation rate between pre- and post-test for texts A (left) and D (right), according to training conditions. A positive variation indicates a higher proportion of speech.

4.3. Data processing

4.3.1. Data labelling and measurements

We hand-checked automatic alignments of recordings with text, adding repetitions, mispronunciations and deletions of words if any. Our analyses focused on the pre-test and post-test recordings of the trained text A and the untrained text D. Within breathing pauses, we manually segmented the IPD when inhalation was clearly audible by using Praat [22]. We did not use respiratory belts to measure variations in thoracic and abdominal movements. In absence of respiratory data, IPD onset coincides with the onset of breathing noise if any. We also added an annotation tier for pause grammaticality. Pauses occurring at the boundaries of syntactic units (clauses, syntactic groups) are considered grammatical (G). Pauses located outside these boundaries (as well as in-words) were considered ungrammatical (NG).

For each recording, we computed three features:

- phonation rate: ratio between duration of phonation vs. total reading time
- grammaticality rate: ratio NG pauses
- IPD mean and standard deviation

4.3.2. Statistical analyses

The significance of the results was tested using a multiple analysis of variance, followed by a Tukey-Kramer test (Multcompare on Matlab). Significant results were indicated with the usual significance thresholds on the figures, if any.

5. Results

5.1. Phonation rate

Figure 2 displays the progression of phonation rate between pre- and post-test readings of trained A vs. untrained D texts, according to training conditions.

The results for text A indicated that phonation rate was greater in group W compared to the control group. Group B showed a similar positive trend, but the differences with the control group were not significant. Note that we observed a high increase in variability for experimental groups (B and W). Computer-assisted reading training tends to both increase phonation rate and decrease pause numbers and durations. Concerning text D, none of the speaker groups showed any improvement of the phonation rate between the pre- and post-test readings.

We noticed an increase in variability for group B for both texts A and D. An increase of variability was also observed for group W, but only during reading of text A.

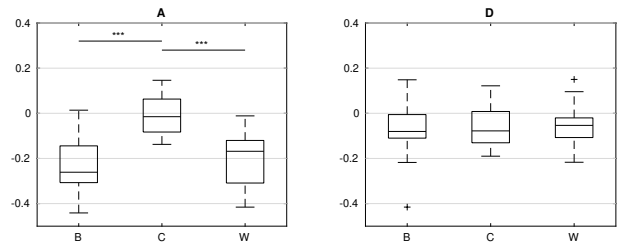


Figure 3: Variation in the percentage of ungrammatical pauses between pre- and post-test for texts A (left) and D (right), by training condition. A negative average indicates improvement.

5.2. Grammatical vs. ungrammatical pauses

The variation in the percentage of ungrammatical pauses relative to the total number of pauses is given in figure 3. Here the impact of repeated computer-assisted reading on pause placement is statistically significant.

For text A, the proportion of ungrammatical pauses significantly decreased for groups B and W, compared to group C. A slight improvement in pause placement was also observed during the reading of text D, but there was no significant difference between the three groups. However, group W showed higher variability in the proportion of ungrammatical pauses.

5.3. Inhalation-to-Phonation Delay

Figure 4 shows the evolution of the mean *Inhalation-to-Phonation Delay* between pre-test and post-test readings for text A and D, by training condition. With this measure, we verify whether computer-assisted reading promotes the shift of inhalation towards the end of the pause. To do this, we considered the IPD of pauses longer than 400 ms. These pauses are long enough for the IPD shift to be visible.

For text A, the IPD mean significantly decreased in group W compared to the control group. After the training, this group took a breath later, towards the end of the pause. The IPD mean also tended to decrease for group B, but the differences with control group were not significant.

Similarly to the previous results for text D, no significant improvement in mean IPD was observed between the speakers groups.

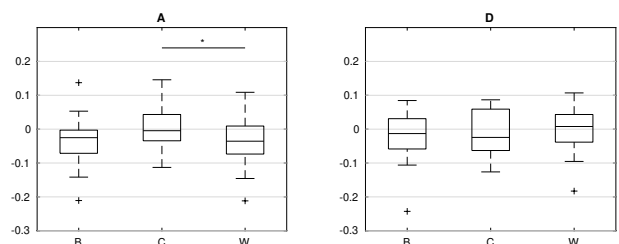


Figure 4: Variation in mean IPD between pre- and post-test for texts A (left) and D (right), according to training conditions. A negative value indicates improved speech-breathing coordination.

Figure 5 displays standard deviation (SD) of the IPD within all pauses. For this measure, no duration threshold was applied.

The SD of the IPD improved for groups B and W, compared to the control group. This indicated that the IPD was more steady during the post-test reading. Concerning text D,

the SD of IPD remained constant between the pre-test and post-test readings for all three training conditions.

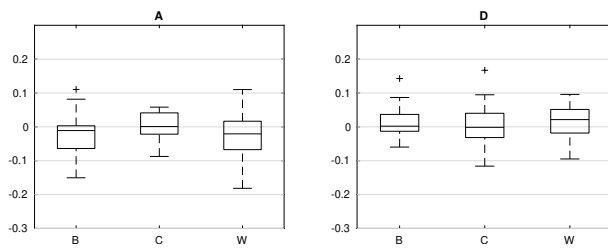


Figure 5: Variation in standard deviation of IPD between pre- and post-test for texts A (left) and D (right), according to training conditions. A negative value indicates improved speech-breathing coordination.

6. Discussion

Our study shows the positive effect of computer-assisted reading training on speech-breathing coordination during reading, specifically pause management. The increase in phonation rate in post-test phase reflects a decrease in pause duration during reading, which mainly result from a decrease in hesitations and decoding errors, and improved pause planning after three weeks of training. We also observed a significant improvement in the syntactic location of breathing pauses for both experimental groups (B and W). Thus, computer-assisted reading training helps children plan their pause placement according to the structure of the text.

A better respiratory planning is reached thanks to computer-assisted reading, which is shown by a decrease in IPD values between pre- and post-tests for groups W and B. During the pre-test reading, the high number of ungrammatical pauses and the high IPD indicate a deficit of respiratory planning. We can draw parallels with models of motor control in coarticulation. Adults anticipate the articulation of sounds when necessary, initiating movement *just-in-time* for the sound to be correctly articulated at the desired moment [23]. In children, these anticipatory actions are scarce, probably due to immature motor control [24]. This lack of anticipation, also evident in respiratory control, prompts the child to articulate or breathe ASAP rather than JIT.

Two highlighting modes during computer-assisted reading were tested. Unexpectedly, the sole word highlighting mode showed stronger effects on IPD and phonation rate than joint word/breath group and pause highlighting. The latter karaoke policy yet provides cues to help children to plan their breathing more effectively than highlighting only the word. The lower effectiveness of breath group highlighting may be explained by the surplus of information it provides to the child during reading. Indeed, this highlighting mode introduces competition between the smooth progression of reading (word) and the programmed fixation of the current and following breath group. Smooth pursuit of the reading by the eyes then competes with anticipatory saccades. This competition between the two eye controls perhaps overwhelms the novice reader, who has not yet acquired the skill and competence to manage this surplus of information. The question of the salience of breath intake on the computer screen during reading may also be questioned. Here, inhalation is cued only by the highlighting of the punctuation/space it is associated with. The highlighting of these small characters between the two large highlighted breath groups is

perhaps not salient enough and should be revisited.

We observed a progression in the respiratory features for the trained text A: the results for text D showed no significant improvement for the experimental groups compared to the control group. Therefore, improvements in speech-breathing coordination were not transferred when reading an untrained text. It should be made clear that the training lasted only three weeks. A longer training period may be necessary to facilitate transfer of skills developed to an untrained text. The larger complexity of text D could also be one of the possible reasons for the lack of transfer. Indeed, compared to text A, this text contains fewer punctuation marks to help identify the syntactic structure of sentences and to identify suitable locations for pauses. Furthermore, the presence of rare and difficult words for children has led to a significant number of reading errors and hesitations. To optimise this type of training, introducing more punctuation marks could reduce difficulties during reading.

In light of results from the subjective evaluation of fluency previously conducted on the same dataset [16], we expected to observe more significance in our results. The difference in significance between these two studies can be explained by the duration of reading used in the analyses: the subjective evaluation was conducted only on the first minute of reading, while our analysis of respiratory cues covered all the reading. At the beginning of reading, planning is usually easier, following by increasing difficulties in planning as fatigue sets in, which can impair speech-breathing coordination. Furthermore, text D contains several rare and difficult words towards the end, increasing the number of decoding errors and hesitations, which impacts pause placement.

Cues related to pause placement prove to be reliable objective markers of fluency improvement. IPD also appears to be a marker of fluency. Our results suggest that other factors may influence the position of inhalation within pauses. Further exploratory studies, including respiratory data, may be necessary to identify these factors and respiratory strategies employed during reading. In addition, we should extend our study to involve a larger number of pupils, and a greater variety of texts to test the robustness of these cues for objective fluency assessment.

Finally, it is worth noting that our study is limited by the lack of a specific respiratory data. The use of respiratory belts/vest would allow for more precise identification of respiratory pauses. Acoustic onsets of breath noises are sometimes faint or even inaudible, and therefore ignored during annotation. Nasal inhalations are also often missed. Moreover, additional respiratory data would allow for more precise delineation of IPDs. Furthermore, such data would allow for exploration of respiratory strategies deployed at the thoracic and abdominal levels during reading.

7. Conclusions and perspectives

This study confirms that computer-assisted reading training contributes to the improvement of pause and breathing planning. Training mode seems to influence the improvement of IPD with higher significance for word highlighting. Highlighting mode by word and breath group seems too complex for children to improve the speech-breathing coordination.

In future works, a new study, based on a similar protocol, will be conducted with a larger panel of pupils (≈ 1000). A more intensive training program will be conducted over a period of 10 weeks, three sessions per weeks and will focus on reading texts of increasing complexity.

8. Acknowledgements

This work was supported by the ANR TRANS3 (ANR-22-FRAN-0008) and the Fluence project (e-Fran, PIA2) funded by the CDC. We would like to thank the students and their teachers who participated in this study.

9. References

- [1] Z. Breznitz, *Fluency in Reading Synchronization of Processes*, 2nd ed. Psychology Press, 2012.
- [2] P. J. Schwanenflugel, A. M. Hamilton, J. M. Wisenbaker, M. R. Kuhn, and S. A. Stahl, "Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers," *Journal of Educational Psychology*, vol. 96, no. 1, pp. 119–129, 2004.
- [3] D. Luo, N. Minematsu, Y. Yamauchi, and K. Hirose, "Automatic assessment of language proficiency through shadowing," *2008 6th International Symposium on Chinese Spoken Language Processing*, pp. 1–4, 2008.
- [4] S. Webb and A. C.-S. Chang, "Vocabulary learning through assisted and unassisted repeated reading," *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, vol. 68, pp. 267 – 290, 2012.
- [5] A. Henderson, F. Goldman-Eisler, and A. Skarbak, "Temporal patterns of cognitive activity and breath control in speech," *Language and speech*, vol. 8, no. 4, pp. 236–242, 1965.
- [6] A. L. Winkworth, P. J. Davis, E. Ellis, and R. D. Adams, "Variability and consistency in speech breathing during reading," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 3, pp. 535–556, 1994.
- [7] B. Conrad, S. Thalacker, and P. Schönle, "Speech respiration as an indicator of integrative contextual processing," *Folia Phoniatrica*, vol. 35, no. 5, pp. 220–225, 1983.
- [8] F. Grosjean and M. Collins, "Breathing, pausing and reading," *Phonetica*, vol. 36, no. 2, pp. 98–114, 1979.
- [9] M. Lalain, L. Mendonca-Alves, R. Espesser, A. Ghio, C. D. Looze, and C. Reis, "Lecture et prosodie chez l'enfant dyslexique, le cas des pauses," in *Journées d'Études sur la Parole*, 2012, pp. 41–48.
- [10] D. Charuau, B. Vaxelaire, and R. Sock, "L'organisation spatio-temporelle de la respiration chez l'enfant," *SHS Web of Conferences*, vol. 138, p. 08005, 2022.
- [11] E. Godde, G. Bailly, and M.-L. Bosse, "Pausing and breathing while reading aloud: Development from 2nd to 7th grade in French speaking children," vol. 35, no. 1, pp. 1–27, 2022.
- [12] C. A. Boliek, T. J. Hixon, P. J. Watson, and P. B. Jones, "Refinement of speech breathing in healthy 4- to 6-year-old children," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 4, pp. 990–1007, 2009.
- [13] J. D. Hoit, T. J. Hixon, P. J. Watson, and W. J. Morgan, "Speech breathing in children and adolescents," *Journal of Speech, Language, and Hearing Research*, vol. 33, no. 1, pp. 51–69, 1990.
- [14] N. K. Russell and E. Stathopoulos, "Lung volume changes in children and adults during speech production," *Journal of Speech and Hearing Research*, vol. 31, no. 2, pp. 146–155, 1988.
- [15] E. T. Stathopoulos and C. M. Sapienza, "Developmental changes in laryngeal and respiratory function with variations in sound pressure level," *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 3, pp. 595–614, 1997.
- [16] E. Godde, G. Bailly, and M.-L. Bosse, "Un Karaoké pour Entraîner Prosodie et Compréhension en Lecture," in *EIAH 2019 - Environnements Informatiques pour l'Apprentissage Humain*, Paris, France, 2019.
- [17] T. V. Rasinski, "Assessing reading fluency." 2004.
- [18] E. Godde, M.-L. Bosse, and G. Bailly, "Échelle Multi-Dimensionnelle de Fluence : nouvel outil d'évaluation de la fluence en lecture prenant en compte la prosodie, étalonné du CE1 à la 5ème," *L'année Psychologique/Trends in Cognitive Psychology*, vol. 121, no. 2, pp. 19–43, 2021.
- [19] G. Bailly and W.-S. Barbour, "Synchronous reading: learning French orthography by audiovisual training," in *Interspeech 2011-12th Annual Conference of the International Speech Communication Association*, 2011, pp. 1153–1156.
- [20] G. Bailly, "Close shadowing natural vs. synthetic speech," in *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*, Perthshire - Scotland, 2001, pp. 87–90.
- [21] T. V. Rasinski, "Effects of repeated reading and listening-while-reading on reading fluency," *Journal of Educational Research*, vol. 83, no. 3, pp. 147–50, 1990.
- [22] P. Boersma and D. Weenink, "Praat : doing phonetics by computer." [Online]. Available: <http://www.praat.org/>
- [23] M. H. Christiansen and N. Chater, "The now-or-never bottleneck: A fundamental constraint on language," *Behavioral and Brain Sciences*, vol. 39, p. e62, 2016.
- [24] G. Barbier, P. Perrier, Y. Payan, M. K. Tiede, S. Gerber, J. S. Perkell, and L. Ménard, "What anticipatory coarticulation in children tells us about speech motor control maturity," *PLOS ONE*, vol. 15, no. 4, pp. 1–26, 2020.