



Self-regulated listening experience with smart captioning

Maryam Sadat Mirzaei¹ and Kourosh Meshgi²

Abstract. This article introduces self-regulation features with Partial and Synchronized Caption (PSC) for practicing listening skills and fostering self-paced and autonomous learning. PSC generates a partial caption by focusing on acoustically and lexically difficult words and synchronizes each word's appearance with the speaker's speech. The aim is to reduce textual density in the caption and encourage more listening than reading. Given that difficulty translates differently for individual learners, we created a user-friendly interface that allows a more individualized experience with our generated caption. We collected the user-configuration log data of 33 intermediate English learners to analyze learner behavior during self-paced practice, along with the comprehension scores of the subsequent tests and learner feedback. The data showed differences in learners' strategies with a relative tendency to reduce their reliance on captions. Self-paced practice was received positively by most but not all learners, indicating the importance of individual differences in self-regulated listening.

Keywords: smart captioning, L2 listening, self-regulated learning, listening difficulties.

1. Introduction

In the sphere of second language acquisition, technologies have changed the way we train and develop skills. However, technologies for teaching and learning listening skills have been limited for many years to the use of captions and multimedia controllers (Gass, Winke, Isbell, & Ahn, 2019). Numerous research

1. RIKEN AIP, Tokyo, Japan; maryam.mirzaei@riken.jp; <https://orcid.org/0000-0002-0715-1624>

2. RIKEN AIP, Tokyo, Japan; kourosh.meshgi@riken.jp; <https://orcid.org/0000-0001-7734-6104>

How to cite this article: Mirzaei, M. S., & Meshgi, K. (2022). Self-regulated listening experience with smart captioning. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouéšny, & M. J. Whelpton (Eds), *Intelligent CALL, granular systems, and learner data: short papers from EUROCALL 2022* (pp. 273-278). Research-publishing.net. <https://doi.org/10.14705/rpnet.2022.61.1470>

studies have been conducted on the benefits of captions in classrooms, such as increased comprehension and vocabulary acquisition (Fievez, Perez, Cornillie, & Desmet, 2020), but studies also indicate that captions could promote reading over listening with increasing reliance on the text (Leveridge & Yang, 2013). Captions also ignore the requirements of individual learners by providing full text to everyone, regardless of level. Therefore, there is a need to open a window onto the development of contemporary tools for listening and pedagogy with particular attention to different learners' needs.

In this context, self-directed learning plays a critical role. It emphasizes that individuals take the initiative in their learning process by diagnosing their needs, adapting the appropriate learning strategies, and evaluating learning outcomes (Knowles, 1975). In the case of listening, learners are dealing with fleeting speech and need to process the input quickly before the next piece of information is received. Putting the control of listening input in the hands of the listener will allow for self-paced learning, which personalizes the learning practice, supports the self-regulatory processes (Kitsantas & Dabbagh, 2011), fosters the sense of agency, motivation, and positive learning experience (McBride, 2011), reduces cognitive overload (Ozcelik, Van den Branden, & Van Steendam, 2019), and improves information processing and performance (Roussel, 2011).

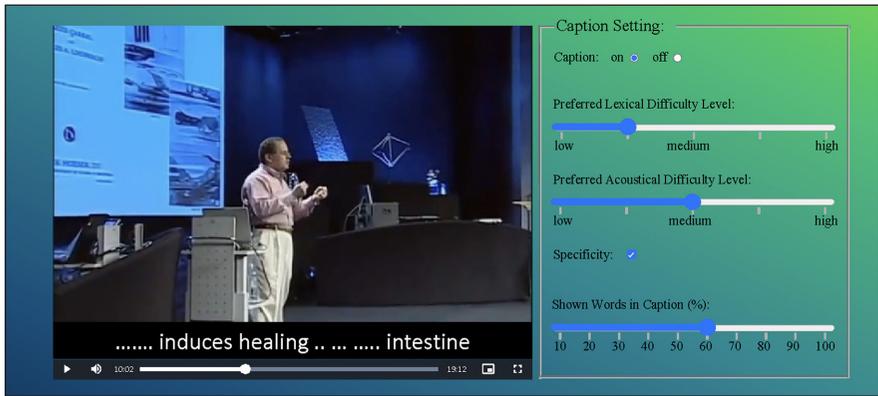
Learning from the ample research available on the L2 listening, with a particular focus on the individual learner's strengths and weaknesses, we introduce PSC with enhanced self-regulating features that allow for generating captions for different language learners. We investigate how learners use self-regulated captioning and how it affects their performance.

2. Self-regulation using PSC

PSC provides for a limited number of words by introducing features that enable the detection of difficult words/phrases. It benefits from automatic speech recognition systems and corpora to identify challenging acoustic and lexical video segments. It further provides word-level synchronization for smoother audio-to-text matching. We considered a combination of acoustic and lexical features, ranging from acoustic neighbors, speech rate, and word frequency, to conduct a rigorous calculation that distinguishes easy from problematic words (see Mirzaei, Meshgi, & Kawahara, 2018). The system prepares captions for different levels – beginners, intermediates, and advanced – by adjusting the parameters when generating the caption. Given that some learners have a better

vocabulary reservoir but less tolerance in coping with fast speech or vice versa, we have incorporated a user-friendly interface to adjust the system's features and override the default parameters. This enables increasing or decreasing the number of words shown for personalization and promotes autonomous and self-regulated learning (Figure 1).

Figure 1. PSC with self-regulation features to adjust the caption based on preference



3. Preliminary experiment

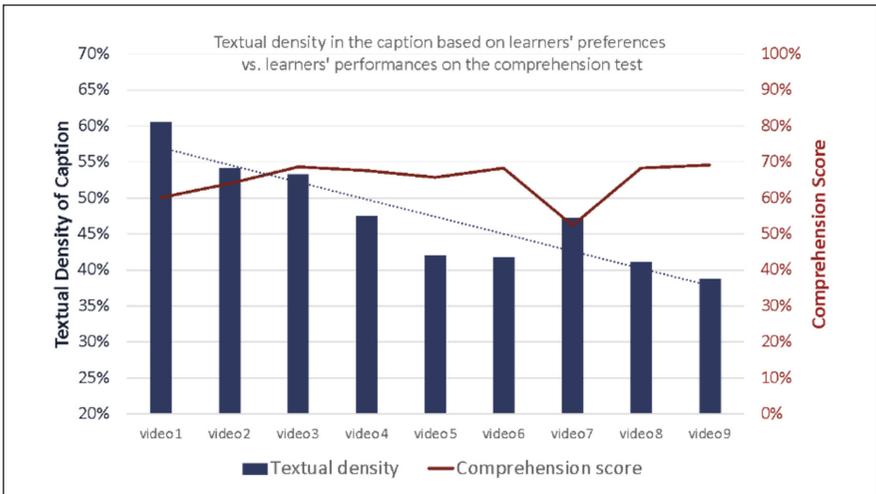
Our participants were 33 intermediate English learners (520~725 TOEIC scores), undergraduates (19~20 years old, 15 males and 18 females, from engineering and humanities fields) whose native language was Japanese. The whole experiment was conducted online. The participants connected to the experimental sessions using a web browser, and the entire sessions and log data were recorded. Two pilot sessions were conducted to help learners familiarize themselves with the system and suggest parameters for their level. Before watching videos, the learners were tested for vocabulary size and tolerable speech rate to adjust the system's default parameters accordingly. They watched a series of TED talks (15~20 minutes each) on various topics and answered multiple-choice comprehension questions after each. During the watching phase, learners could play with the system to generate the appropriate amount of captions or keep the suggested parameters. The first five minutes of each video were used to find the best setting through self-control, and the log data of the learner's preferable setting was analyzed once the setting had been decided. We also elicited learner feedback through an open-ended questionnaire to complement our data with the actual learner experience.

4. Results and discussions

Most participants (18) took a conservative approach by starting with a high-density caption, keeping it for a while, then gradually decreasing it. We observed that a few learners (5) took a more challenging path by starting with less than 30% of the text. Three learners quickly switched to a higher density caption, and only two continued with less than 30% text throughout the experiment. Ten participants leaned toward a trial-and-error approach, mostly starting with a medium-density caption and a slight increase or decrease over time. Two of them kept a steady strategy and barely changed the amount of the text.

As Figure 2 shows, the overall trend analysis shows that, on average, the learners tended to choose fewer words in the caption while striving to gain an acceptable performance. We also observed that such a decrease did not significantly affect the overall comprehension, and the learners’ performances showed a relatively steady increase.

Figure 2. Experimental results



Video #7 was notably harder than other videos in terms of topic (medical), vocabulary (terminologies), and speech rate. For this particular video, most learners considered receiving more captions to deal with the video’s difficulty; however, given that unknown technical terms were frequently presented in this video, the average performance showed a decrease. This result indicates that when the input is high above the learner’s level or the topic is difficult, using more text in the caption

will not necessarily lead to better performance. This finding is essential in that when presenting an input to the learner, it is necessary to consider the learner's background knowledge and the difficulty level of the video relative to the learner's level.

The analysis of self-regulation strategies used by the learners shows that they were primarily successful in picking the appropriate amount of words needed to attain a relatively acceptable performance level. Meanwhile, for some learners, adapting the wrong settings led to low performance, while for others, it was hard to find the best setting. Even though the overall performance was not significantly high, data from the learners' self-disclosure showed that most learners enjoyed self-regulation. They felt motivated to control their learning process, which may promote autonomous learning, highlighting the effectiveness of using self-regulated methodologies for practicing listening. However, it is crucial to consider that this approach may not work for all learners, especially those who find it hard to self-regulate (Roussel, 2011). It is anticipated that through practice, learners can become more familiar with the system, the types of listening challenges they face, and the areas they need to improve. Additionally, learner feedback revealed that for those overly reliant on the caption, the idea of using partial text is not easy to adapt. For this group, it may take significantly more time to feel confident enough to decrease the dependence on the caption slowly.

5. Conclusions

This article introduced a smart caption with adjustable parameters to meet the requirement of individual listeners. Several experiments were conducted with self-regulation listening episodes. The findings revealed a positive learning experience while gleaning valuable insights into differences in learner preferences and indicating the importance of incorporating learner differences into computer assisted language learning systems to provide personalized L2 listening practices. The current study focused on intermediate learners; however, the findings indicate a need to fortify the results by including different proficiency levels and backgrounds in future studies.

References

- Fievez, I., Perez, M. M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of learner-and word-related factors. *Calico Journal*, 37(3), 233-253. <https://doi.org/10.1558/cj.39370>
-

- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: a working-memory, eye-tracking study. *Language Learning & Technology*, 23(22), 85-104. <https://doi.org/10.125/44684>
- Kitsantas, A., & Dabbagh, N. (2011). The role of Web 2.0 technologies in self-regulated learning. *New directions for teaching and learning*, 2011(126), 99-106. <https://doi.org/10.1002/tl.448>
- Knowles, M. S. (1975). *Self-directed learning: a guide for learners and teachers*. Association Press. <https://doi.org/10.1177/105960117700200220>
- Leveridge, A. N., & Yang, J. C. (2013). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL*, 25(2), 199-214. <https://doi.org/10.1017/S0958344013000074>
- McBride, K. (2011). The effect of rate of speech and distributed practice on the development of listening comprehension. *Computer Assisted Language Learning*, 24(2), 131-154. <https://doi.org/10.1080/09588221.2010.528777>
- Mirzaei, M. S., Meshgi, K., & Kawahara, T. (2018). Exploiting automatic speech recognition errors to enhance partial and synchronized caption for facilitating second language listening. *Computer Speech & Language*, 49, 17-36. <https://doi.org/10.1016/j.csl.2017.11.001>
- Ozcelik, H. N., Van den Branden, K., & Van Steendam, E. (2019). Listening comprehension problems of FL learners in a peer interactive, self-regulated listening task. *International Journal of Listening*, 1-14. <https://doi.org/10.1080/10904018.2019.1659141>
- Roussel, S. (2011). A computer assisted method to track listening strategies in second language learning. *ReCALL*, 23(2), 98-116. <https://doi.org/10.1017/S0958344011000036>



Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2022 by Editors (collective work)
© 2022 by Authors (individual work)

Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022
Edited by Birna Arnbjörnsdóttir, Branislav Bédi, Linda Bradley, Kolbrún Friðriksdóttir, Hólmfríður Garðarsdóttir, Sylvie Thoučsny, and Matthew James Whelpton

Publication date: 2022/12/12

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2022.61.9782383720157>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover photo by © 2022 Kristinn Ingvarsson (photo is taken inside Veröld – House of Vigdís)
Cover layout by © 2022 Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-38372-015-7 (PDF, colour)

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2022.