Automatic scaffolding for L2 listeners by leveraging natural language processing

Maryam Sadat Mirzaei[1], Kourosh Meshgi[2], and Toyoaki Nishida[3]

Abstract. This paper introduces a new captioning tool, Partial and Synchronized Caption with Hints (PSCH), as a means to facilitate second language (L2) listening by providing cues for ambiguous and difficult words/phrases in the caption while filtering out the easy words. Each word in the caption is synchronized to the corresponding audio to enable text-to-speech mapping. The words to be shown in the caption are carefully selected by defining the features that lead to listening difficulty. The hints are generated in the form of short explanations/definitions of the words to allow for meaning construction and resolving difficulties on-the-fly. With the use of Natural Language Processing (NLP) tools and word sense disambiguation, we tried to generate appropriate hints for the selected words to provide instantaneous and minimally intrusive assistance. Experimental results revealed that learners' scores significantly increased when they used PSCH compared to having no hints. Furthermore, PSCH received positive learner feedback in providing appropriate and useful hints for improving listening comprehension.

Keywords: partial and synchronized caption, L2 listening difficulty, word ambiguity, instantaneous hints.

1. Introduction

L2 listening entails constant effort as the listeners need to process each part of the input quickly, without having the option to return to the earlier points. Learners must go through perception, recognition, comprehension, meaning construction, ambiguity resolution, and inferencing, etc. in a short time (Rost, 2005), which

1. Kyoto University, Kyoto, Japan; mirzaei@ii.ist.i.kyoto-u.ac.jp
2. Kyoto University, Kyoto, Japan; meshgi-k@sys.i.kyoto-u.ac.jp
3. Kyoto University, Kyoto, Japan; nishida@i.kyoto-u.ac.jp

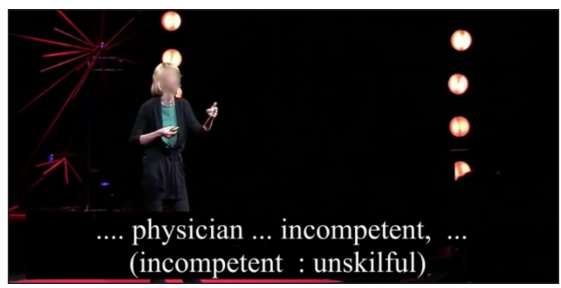
How to cite this article: Mirzaei, M. S., Meshgi, K., & Nishida, T. (2018). Automatic scaffolding for L2 listeners by leveraging natural language processing. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 200-206). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.837

imposes large working memory loads (Chang, 2009). To foster L2 listening, captioning is used as a popular tool that facilitates the comprehension of the input by allowing for reading the text along with listening to the audio. The use of captions, however, is subjected to some limitations as it promotes more reading than listening, thus inhibiting listening skill development (Pujolà, 2002).

To alleviate these problems, we introduced a Partial and Synchronized Captioning (PSC) system which selects limited numbers of words/phrases and presents them in the caption by synchronizing each word to its relevant speech segment (word-level text-to-speech alignment). The selection of words in PSC is based on the factors that cause difficulty for L2 listeners, such as word frequency, specificity, and speech rate. Additionally, more acoustic features were added by extracting the Automatic Speech Recognition (ASR) system's errors as a source that indicates perception difficulties in the given audio (Mirzaei, Meshgi, & Kawahara, 2018).

Through the experiments, Mirzaei et al. (2018) found that PSC can assist L2 listeners by successfully detecting and presenting difficult segments. However, further observations suggested that merely showing the words in the captions may not provide the optimal assistance, especially for words out of the learner's vocabulary reservoir. In such cases, learners' attention is confined to the ambiguous segment, which inhibits them from moving on to process the next input. This happens particularly for those who overemphasize on using bottom-up strategies and word-by-word decoding (Osada, 2004).

Figure 1. Screenshot of a TED talk with PSCH, which includes instantaneous hints



To address this issue, we augmented PSC with Hints (PSCH) to provide minimal assistance on the spot, when learners encounter difficult/ambiguous segments. This assistance is provided in the form of context-matching synonyms, short definitions, named entity tags, or co-references for such segments (Figure 1). Only 30% of

the words are shown in PSC and one-third of these words are accompanied with hints, to promote more listening than reading and using top-down strategies for comprehension.

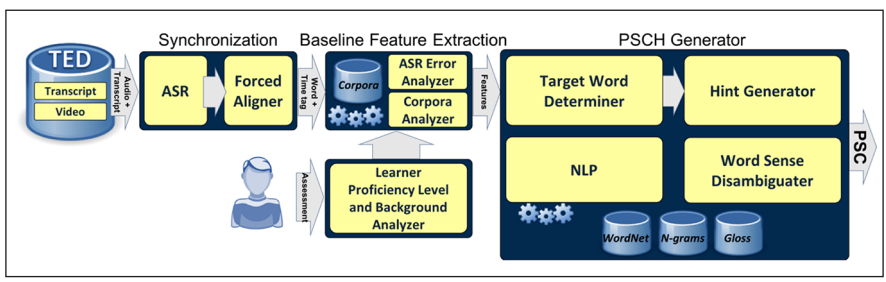
2. Instant scaffolding for L2 listeners

The process of generating PSCH involves: (1) determining the words/phrases that need supplementary description or clarification, and (2) providing useful description for them to assist learners.

First, we focus on the problematic categories that were specified by L2 listeners including: low-frequency, technical, ambiguous, and polysemous words, proper nouns, named entities, ambiguous references, and uncommon/multi-purpose abbreviations. Low frequency words were detected in PSC by referring to the corpora (BNC and COCA). For detecting other categories, however, we use NLP techniques for parsing, analyzing, extracting information (e.g. named entities and part-of-speech), and resolving ambiguities (e.g. co-references and word sense) of the transcript. We leverage the state-of-the-art open-source NLP libraries such as CoreNLP (Manning et al., 2014), NLTK, and spaCy.

We used the Term Frequency/Inverse Document Frequency (TF/IDF) index and/or domain-specific encyclopedias to find technical/specific jargon, WordNet synset size to detect polysemous/homonym words, named-entity recognizers and part-of-speech taggers to identify proper nouns, and Wikipedia and the Urban Dictionary to determine symbolic names and abbreviations (Figure 2).

Figure 2. PSCH system architecture



Next, the proper hints for each category are retrieved from in-house and online resources. A synonym is selected as the hint for low-frequency words (e.g.

cobble→put together). Wikipedia and glosses are consulted for word definitions and abbreviation expansions (e.g. neocortex→part of mammalian brain). Short descriptions are retrieved for proper nouns, named-entities, and symbolic names from Wikipedia and the Google search engine (e.g. Basel→city in Switzerland, big apple→New York City). The referent of references is displayed as a reminder hint if their co-reference was distant. For words with different meanings (e.g. polysemous words), we employed word sense disambiguation to find the most probable meanings from available synonyms/descriptions.

The hints provided to the learners should be short, helpful, and relevant. To this end, we seek the shortest description for the word or generate one by searching for the keywords in the retrieved description. Along with this, a filtering process assures that the final hint includes high-frequency words that are familiar for the learner. We carefully controlled the display of hints. Hints appeared in sync with the utterances and remained for a pre-defined duration, providing enough time for reading and processing the input.

3. Experimental evaluation

3.1. Participants

Our participants were 30 graduate and undergraduate students of intermediate English levels with TOEIC[4] scores ranging from 810-920.

3.2. Procedure

The participants were asked to watch a series of short segments (two to three minutes) from eight TED talks using Baseline PSC (no hint) and PSCH (with hints). We selected 40 words from the videos that appeared in PSC, among which only 15 words were supplemented by hints and were used as the target words of our experiment. The rest were used as distractors.

First, the participants watched all the videos with the baseline PSC followed by a listening test on 40 words. The listening test was made by extracting short audio clips (10~15 seconds) from the experimental videos. The participants were asked to write the meaning of each word in the given context.

4. Test of English for International Communication®

Next, participants watched the videos with PSCH in shuffled order. On average, each video segment included six hints. Learners were then asked to take the same listening test only on the 15 target words with hints. We aimed to see the participants' performance on the target words before and after receiving the hints. A questionnaire was designed to elicit learner feedback on the usefulness of PSCH.

4. Results and discussion

Figure 3(a) shows the result of participants' scores on defining the target words before and after receiving hints, which indicates a significant increase by the use of PSCH as compared to PSC. Figure 3(b) shows how the participants' answers changed before and after receiving hints. Results revealed that in most cases the answers were corrected after receiving hints (55%) but there were cases where participants could answer the questions correctly without having the hints (29%). Participants noted that in these cases hints were mostly used as a confirmation rather than assistance.

Figure 3. (a) Participants scores using PSC (with no hints) and PSCH (with hints); (b) Distribution of participants' answers before and after receiving hints

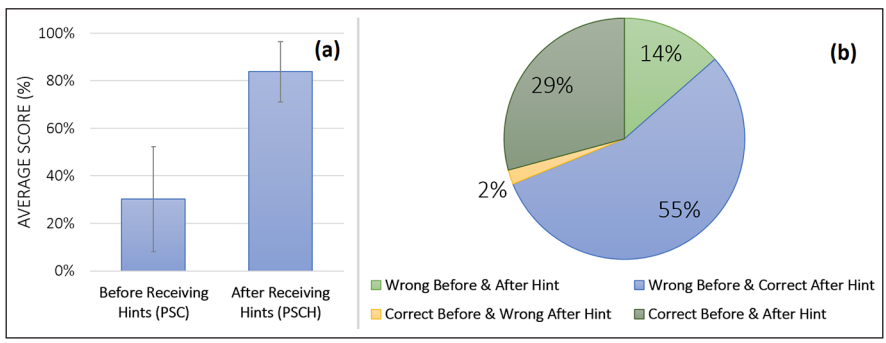
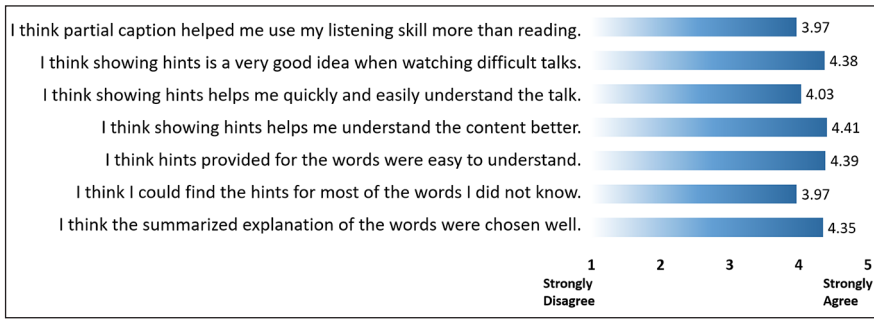


Figure 4 demonstrated learner feedback on PSCH and the type of hints provided during the experiment. The figure suggests that the majority of the participants found the hints easy to understand, useful to resolve difficulty, and helpful to comprehend the content on-the-fly, which highlights the benefit of instantaneous assistance and concurrent lexical support to facilitate L2 listening (Rost, 2005). It was noted, however, that at some points hints were not necessary as the words were easy or common, whereas, for some other words, participants preferred to receive hints. This emphasizes the importance of adjusting the type and frequency

of the hints with the learner's level and needs to provide more effective assistance (Durlach & Lesgold, 2012).

Figure 4. 5-point Likert-scale questionnaire on PSCH



5. Conclusions

This paper investigated the use of PSCH as a captioning system that detects difficult words in the listening material and presents them in the caption while supplementing some of them with hints in the form of short definitions when necessary. PSCH aims to provide necessary but minimal assistance to foster L2 listening by delivering instantaneous hints. Experimental results suggested that using PSCH has significantly assisted L2 listeners to disambiguate the content when listening on-the-fly and facilitated listening to the authentic materials. Findings also suggested that further improvement is needed by considering each individual's requirement to provide more learner-specific assistance.

6. Acknowledgements

This study was supported by Japan's MEXT Kakenhi grant No.17K02925. Baseline PSC was developed under supervision of Professor Tatsuya Kawahara, to whom we are always grateful.

References

Chang, A. C. S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, *37*(4), 652-663.

Durlach, P. J., & Lesgold, A. M. (2012). *Adaptive technologies for training and education*. Cambridge University Press. https://doi.org/10.1017/CBO9781139049580

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of ACL* (pp. 55-60).

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

Osada, N. (2004). Listening comprehension research: a brief review of the past thirty years. *Dialogue, 3*, 53-66.

Pujolà, J. T. (2002). CALLing for help: researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, *14*(2), 235-262. https://doi.org/10.1017/S0958344002000423

Rost, M. (2005). L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 503-527). Erlbaum.

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

Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018
Edited by Peppi Taalas, Juha Jalkanen, Linda Bradley, and Sylvie Thouësny

Publication date: 2018/12/08

Typeset by Research-publishing.net
Cover theme by © 2018 Antti Myöhänen (antti.myohanen@gmail.com)
Cover layout by © 2018 Raphaël Savina (raphael@savina.net)
Drawings by © 2018 Linda Saukko-Rauta (linda@redanredan.fi)

ISBN13: 978-2-490057-22-1 (Ebook, PDF, colour)
ISBN13: 978-2-490057-23-8 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-21-4 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

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

Legal deposit, UK: British Library.
Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: Décembre 2018.