Teaching EFL pronunciation with audio-synchronised textual enhancement and audiovisual activities: Examining questionnaire data

Valeria Galimberti, Joan C Mora, Roger Gilabert

Galimberti, V., Mora, J. C., and Gilabert, R. (2023). Teaching EFL pronunciation with audio-synchronised textual enhancement and audiovisual activities: Examining questionnaire data. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 70–82). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8174014

Teaching EFL pronunciation with audio-synchronised textual enhancement and audiovisual activities: Examining questionnaire data

Valeria Galimberti
University of Barcelona

Joan C. Mora
University of Barcelona

Roger Gilabert
University of Barcelona

Synchronising the enhancement of target words to their auditory onset has been found to promote a focus on their phonetic form (Stenton, 2013). In the case of L2 subtitled video, post-viewing activities involving interpretation and repetition of speech from the video offer further opportunities for noticing target pronunciation features and incorporating them into the learners' developing L2 system. This study investigated three groups of high-school EFL learners. Two intervention groups watched TV series clips with or without audio-synchronised textual enhancement of words containing past tense <-ed> endings in the subtitles and performed pronunciation-focused audiovisual activities such as revoicing and subtitling, whereas the control group was not exposed to the learning materials and thus provided a baseline of past tense <-ed> pronunciation rule knowledge. Questionnaire data provided information on the participants' language learning profiles, their perception of the enhancement and their ability to describe the past tense <-ed> pronunciation rule, and their impressions of the intervention. The participants' perceptions of the intervention were favourable, and the enhancement seemed to positively impact self-reported noticing of the target verbs, although not the internalisation of the pronunciation rule. We outline ideas for future research involving the implementation of these pedagogical tools in the language classroom.

Keywords: input enhancement, multimodal input, audiovisual activities, pronunciation teaching, English regular past

1 Introduction

Watching TV in a second language is an activity with a strong pronunciation learning potential, as it provides exposure to large amounts of L2 speech even in instructional contexts where the L2 is not spoken outside the classroom. This extensive listening practice can be supported and further enhanced by the simultaneous processing of subtitles that contain a verbatim transcription of each utterance. While research has shown that exposure to video with L2 subtitles facilitates speech segmentation (Charles & Trenkic, 2015) and promotes the development of speech perception skills (Mitterer & McQueen, 2009), we know very little about its effects on the development of L2 pronunciation (see Wisniewska & Mora, 2020 for a pioneering study) and how the characteristics of subtitles in L2 videos can be manipulated to further promote pronunciation learning. In particular, it is of interest whether visually highlighting words that contain fossilised L1 sound/symbol correspondences would disrupt automatic reading behaviours and direct learners' attention to the soundtrack containing the target-like realisation of those words (Stenton, 2013). In this study, we audio-synchronised the highlighting of target words in the subtitles and combined them with video-based activities to teach the pronunciation of English regular past tense <-ed> ending. This paper analyses participants' perceptions of the intervention, their self-reported noticing of the enhanced target verbs and their acquisition of the past tense <-ed> pronunciation rule.

2 Literature review

Watching TV in a target language is not only a fun extracurricular activity but also an effective way to practise L2 reading and listening. TV series in particular tend to keep viewers engaged for many hours, increasing the total amount of exposure to the foreign language (Pujadas & Muñoz, 2019). Moreover, L2 video represents an accessible source of L2 auditory input even for learners at lower proficiency levels, thanks to the possibility to rewind the video and listen again as many times as needed, and to the widespread availability of subtitles providing a verbatim transcription of the speech (Vanderplank, 2015).

Watching subtitled video involves exposure to large amounts of visual and auditory information, including language-related visual cues such as facial expressions and gestures, written text, and natural monologic or conversational L2 speech. However, the extent to which various grammatical, lexical and phonological features of the input are attended to and effectively processed may vary greatly, depending on viewing context (leisure vs. classroom activity), as well as learners' proficiency and attitudes towards the use of recreational materials in language learning (Vanderplank, 2015). To improve learners' noticing of target vocabulary or grammatical constructions during the viewing, a number of studies have used textual enhancement by typographically enhancing (e.g., highlighting or underlining) target words in the subtitles (see for example, Lee & Révész, 2020; Montero Perez et al., 2015).

Previous research on subtitled video enhancement in L2 pronunciation teaching has found that synchronising the enhancement of target words with the corresponding auditory onset in the soundtrack may promote a focus on the pronunciation of those words (Galimberti et al., 2023). The timely noticing of target words' pronunciation may, in turn, enhance awareness of any differences between the phonetic form of words as perceived through target-like auditory input and the learners' stored representation of the word. Further support for the synchronised enhancement of auditory and written word forms comes from research on reading-while-listening,

where it has been used to promote the update of L2 lexical stress patterns (Stenton, 2013) and the development of L2 and L1 reading skills (Bailly & Barbour, 2011; Gerbier et al., 2018).

Input enhancement in L2 video can promote, in the context of a primarily meaning-focused activity, the noticing of language form, which is a necessary step in the conversion of L2 input into intake (Schmidt, 1990; Sharwood Smith, 1991). In the presence of sufficient depth of processing, noticing may result, over time, in the development of rule-based representations, which may be more or less generalisable and accessible for testing depending on the explicitness of the learning conditions (Robinson, 1997). In Leow's (2015) L2 processing model, the initial stages of learning involve moving from L2 input processing to intake by engaging in memory-based processing (item learning), and/or rule-based processing (system learning). While restructuring of the learner's L2 system can be triggered by the conceptually-driven processing required to formulate a rule, the cognitive effort required to process the data can be reduced by the automatisation of linguistic data through repeated exposure and meaningful practice (Leow, 2015). This assumption is in line with Han et al.'s (2008) recommendation, based on a meta-analysis of 21 studies on textual enhancement, to combine textual enhancement with other strategies such as explicit instruction and/or interactional tasks involving the target feature.

To test the efficacy of this recommendation, we designed a pronunciation teaching intervention that combined TV series clips containing audio-synchronised enhancement with audiovisual (AV) activities, such as revoicing a silent clip or writing the L2 subtitles for an unsubtitled clip (see the AV framework in Zabalbeascoa et al., 2012). These activities, traditionally used to train translators, have been recently introduced into the language classroom to promote the development of listening and speaking skills (Danan, 2010; Zhang, 2016). Of particular relevance to this study are the studies that implemented AV activities to teach pronunciation (e.g., Chiu, 2012), including those with a broader focus on fluency (Sanchez-Requena, 2018) and comprehensibility (Lima, 2020).

The pronunciation of the English regular past tense <-ed> ending was selected as a target structure since the choice among the three allomorphs /d/, /t/ and /əd/ or /ɪd/ is derived from a morphophonological rule which depends on the phonemic environment. The main aspects of this rule can be explained in terms of spelling (Brutten et al., 1986); verbs ending in <-t> and <-d> in their present form take the /əd/ or /ɪd/ pronunciation, while other spelling endings take either /d/ or /t/. Learning this was expected to reduce the most common mispronunciations involving the erroneous addition of epenthetic vowels (e.g., *worked* pronounced */wɜrkəd/). These mispronunciations are as critical as the deletion of inflectional endings because the addition of an unexpected extra syllable through epenthesis may affect comprehensibility, i.e., the listener's perception of how difficult it is to understand a message, and intelligibility, which is the listener's actual understanding of the message (Levis, 2018). Finally, targeting the accurate pronunciation of <-ed> endings was expected to be appropriate and beneficial for high school learners, because regular past tense <-ed> endings are hard to perceive and produce accurately even at advanced proficiency levels, due to their low perceptual salience and redundancy with time adverbials (Strachan & Trofimovich, 2019).

3 Research methodology

3.1 Research questions

The study aimed to answer the following research questions:

RQ1: After a teaching intervention based on audio-synchronised textual enhancement and audiovisual activities:
a) do learners report noticing the target L2 pronunciation feature?
b) can learners successfully describe the target L2 pronunciation rule?

RQ2: What are the learners' perceptions of:
a) videos with audio-synchronised textual enhancement in subtitles?
b) pronunciation-focused audiovisual activities?

3.2 Participants

The intervention was implemented with three intact classes of L1 Spanish and Catalan 15-year-old students learning English as a foreign language. Out of 78 students, 53 completed a survey after obtaining their parents' written consent. The students' English proficiency level was estimated to be intermediate, based on the textbook used in class and on the participants' vocabulary size (m = 2715.09, SD = 592.87) as assessed by the X-Lex test (Milton, 2010). The groups were not significantly different in terms of vocabulary size ($F(2, 50) = .52$, $p = .60$), time spent in an English-speaking country ($F(2, 50) = .09$, $p = .92$), total time spent on English extracurricular classes ($F(2, 50) = .90$, $p = .41$), and weekly exposure to L2 TV shows ($F(2, 50) = .04$, $p = .96$). In order to ensure participant anonymity, a unique identifier was generated using a combination of alphanumeric characters. For instance, participant 2 in intervention group A was assigned the code A02.

3.3 Intervention materials

Students watched five video clips in which, under the enhancement condition only, a selected number of target words (past tense regular verbs) were enhanced in the subtitles 500 ms before the corresponding auditory onset by highlighting the whole word in yellow and underlining the <-ed> ending together with the orthographical representation of its phonological context, i.e., the vowel or consonant preceding it. Low frequency words and words not clearly audible in the soundtrack were not enhanced, in order to avoid interference with comprehension. In each AV activity, the participants in the intervention groups (see Figure 1) re-watched a clip and were instructed to either: 1) complete subtitles in which some words, including the target words, were missing; 2) order and label excerpts of the clip containing the target words; 3) identify muted target words in shorter unsubtitled excerpts and repeat the whole sentence out loud; 4) revoice a muted clip with the help of the subtitles; and 5) order unsubtitled excerpts containing the target words and revoice the obtained sequence. Therefore, in the first session learners could self-test their perception of the target feature through subtitling, whereas in the second session they needed to pay close attention to L2 speech, although in the context of a meaning-focused comprehension task. Finally, the three sessions that involved revoicing aimed at the automatisation of accurate and fluent production, with the support of (part of) the target utterances spoken by native L2 speakers (the characters). After each AV activity, participants did an awareness raising activity in which they read or listened to a list of verbs and were asked, for example, to underline the letter preceding the <–ed> ending of some verbs and decide if the sound was voiced or voiceless; group the verbs based on how the <-ed> ending sounded; or decide whether the vowel representing letter <e> in the <–ed> ending

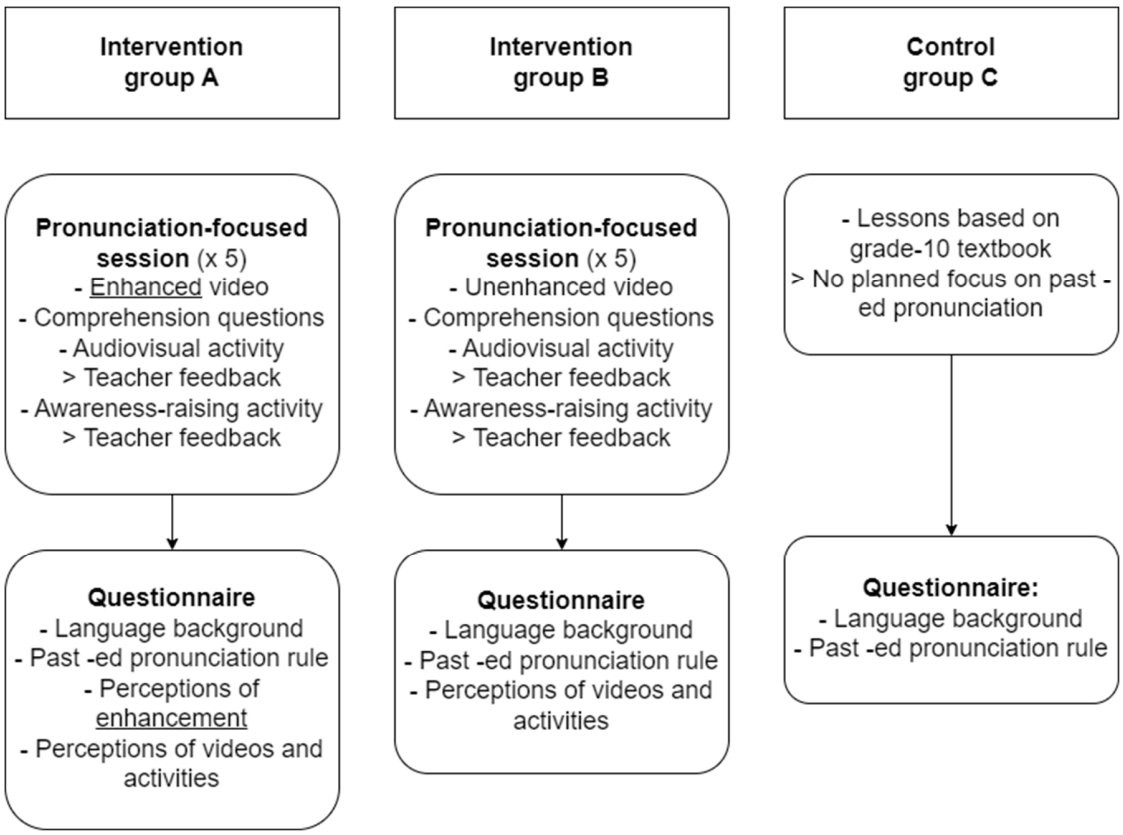
was pronounced or remained silent. These activities aimed at explicitly directing learners' attention to some aspects of regular past tense <-ed> pronunciation, such as the existence of different allomorphs, the difference between voiced and voiceless consonants, and the effects of the phonetic context preceding and following the <-ed> ending (Strachan & Trofimovich, 2019).

3.4 Procedure

The intervention lasted six weeks, with each group receiving fifty minutes of instruction per week (Figure 1). Intervention group A was exposed to audio-synchronised textual enhancement and carried out the AV activities; Intervention group B did the same activities but watched the clips without enhancement; Control group C followed their conventional textbook-based classes, and no planned or reactive focus on past tense <-ed> pronunciation was implemented by the teacher. The control group provided a baseline of past tense <-ed> pronunciation rule knowledge among learners who belonged to the same population as the intervention groups but had not received focused instruction.

Figure 1

Lesson Plan and Data Collection Procedure by Group



After a mock session, in each of the five sessions groups A and B watched a video containing enhanced and unenhanced target words respectively, worked on an AV activity in pairs while the teacher[1] walked around the classroom offering support, and they received feedback. The feedback phase involved having two or three pairs report on the activity in front of the whole class and asking other students if they agreed with the solution or performance proposed until the correct answer was provided. To conclude, each student answered ten comprehension questions and did an awareness-raising activity individually before receiving group feedback. A written questionnaire was administered in a quiet classroom the week after the intervention. Participants could choose between the Spanish and Catalan version and were asked to complete it within 20 minutes.

3.5 Questionnaire

After the language background section, which provided information on the participants' L1(s) and extracurricular exposure to English, all participants (groups A, B and C) were asked to describe the rule about the pronunciation of the <-ed> ending of past tense regular verbs, including examples if possible. Participants in group A and group B expressed their perceptions of the intervention by indicating to what extent they agreed with statements about the videos and the activities. The statements were mostly adapted from Sokoli's (2018) survey of learners' perceptions of AV activities, with a few novel items included. The item on peer collaboration was added as an initial (albeit very limited) measure of social interaction, a construct that has been related to active learning and to a higher focus on the task (Zabalbeascoa et al., 2012). In addition, participants were asked to indicate if they read the subtitles during the viewing, which provided a tentative measure of audiovisual processing under the circumstances (as collecting eye-tracking data during the implementation of whole-class activities was impossible). Sokoli's (2018) questions on participants' feeling of learning were adapted to assess whether the learners' general focus was primarily on grammar or pronunciation, since the intervention may have increased awareness of both the grammatical function and phonological form of the verbs. As a measure of reported noticing, participants in group A were also asked whether any letters were enhanced in the subtitles, what those letters had in common, and whether the participants believed that the enhancement was useful or distracting.

3.6 Data analysis

The data reported in this paper were collected via written questionnaire and analysed quantitatively. Yes/No questions resulted in binary variables (0 or 1), whereas the five-point Likert items on learner perceptions resulted in categorical variables with five levels from 1 (totally disagree) to 5 (totally agree). Knowledge of the past tense <-ed> rule was also operationalised as a categorical variable with four levels, with value range 0 (no response) to 3 (completely correct response). The rating of the responses was conducted by the author/teacher. Partially correct answers (value 1) mentioned some relevant elements but missed other important ones, e.g., "it is pronounced like a *t*" (A05). Answers were considered mostly correct (value 2) if they mentioned the existence of three allomorphs and/or the presence or absence of a vowel sound depending on the context, e.g., "in *walked* <e> makes no sound, in *provided* it sounds like /ed/ because the word ends in <e> (sic), other times it sounds like /t/ and others it makes no sound" (B46). An example

[1] The first author, Valeria Galimberti, was the teacher during the intervention.

of a response considered completely correct is: "There are verbs that in the past are pronounced as if they ended with /t/ (e.g., *walked*), others with /id/ (e.g., *waited*) or with /d/ (e.g., *turned*)" (A02).

Count data is presented for RQ1a, due to the small sample. To answer RQ1b, between group differences were also explored through Fisher's exact tests with Monte Carlo method (1e4 sampled tables), due to the low expected frequencies per variable level. When reporting in-text the participants' responses to the statements in RQ2, the response values 5 (agree) and 4 (somewhat agree) were collapsed into one category 5 (agree). Similarly, the response values 1 (disagree) and 2 (somewhat disagree) were collapsed into the category 1 (disagree). To offer a complementary picture of the data, the mean value and standard deviation on the original five-point scale of the responses to each statement were reported in Table 1 and Table 2 aggregated by group.

4 Results

4.1 Noticing and describing the target L2 pronunciation feature (RQ 1a, 1b)

All participants watching the videos with audio-synchronised enhancement (n = 18) reported noticing the enhanced words in the subtitles, with 16 finding the enhancement useful. Fourteen participants correctly identified that the enhanced words were regular past verbs and/or verbs ending in <-ed>, only one mentioned that the words had been enhanced because of their pronunciation, and three mentioned that the enhancement was related to pronunciation without further specification.

When asked to describe the rule about how to pronounce regular past tense <-ed> endings, 50% of the participants in group A, 12% of the participants in group B, and 22% of the participants in group C did not attempt to answer. The proportion of incorrect answers was 17% in group A, 71% in group B and 56% in group C. Of the twelve acceptable answers, only two for group A and one for group B were rated as mostly correct (11% and 6%, respectively), and two as completely correct (one in group A and one in group B). Fisher's exact tests with Monte Carlo method did not find significant differences between the responses of the three groups (two-tailed p = .66).

4.2 Learners' perceptions of videos with audio-synchronised textual enhancement and pronunciation-focused audiovisual activities (RQ 2a, 2b)

All participants reported understanding the videos, and around 80% in each group thought they were fun (Table 1). Two thirds of group A (videos with enhanced subtitles), but only half of group B (unenhanced subtitles), reported reading the subtitles. Around 70% of the participants in each intervention group believed that they had learned some English pronunciation from the video. While 65% also felt that they had learned some grammar or vocabulary from the videos in group B, only 50% of group A agreed.

Table 1

Responses (1–5) to Statements about the Enhanced Videos

	Intervention group A (Enhancement)			Intervention group B (No enhancement)		
	M	*SD*	95% CI	*M*	*SD*	95% CI
I understood the videos	4.83	.38	[4.64; 5.02]	4.53	.62	[4.21; 4.85]
The videos were fun	4.22	1.17	[3.64; 4.80]	4.12	.93	[3.64; 4.59]
I read the subtitles	3.72	1.02	[3.22; 4.23]	3.53	1.18	[2.92; 4.14]
I learned some English pronunciation from the videos	3.72	.57	[3.44; 4.01]	3.82	1.07	[3.27; 4.38]
I learned some English grammar or vocabulary from the videos	3.50	.71	[3.15; 3.85]	3.65	.99	[3.13; 4.16]

Note. 1 = totally disagree, 5 = totally agree

Almost all participants reported understanding the instructions of the AV activities and two thirds used the clues offered within each activity to complete them (Table 2). Eighty percent of the participants in group A and 60% of the participants in group B indicated that the activities were fun. Only one third of the participants in group A indicated that the activities were challenging, but in group B almost two thirds of participants found them challenging. Ninety percent of participants in group A and 75% in group B responded that both partners had contributed equally to the activity. Similar to the responses for the enhanced videos, around 70% of the participants in each group reported learning some pronunciation from the activities, but only half in each group reported learning some grammar and vocabulary.

Table 2

Responses (1-5) to Statements about the Audiovisual Activities

	Intervention group A (Enhancement)			Intervention group B (No enhancement)		
	M	*SD*	95% CI	*M*	*SD*	95% CI
I understood the instructions	4.39	.78	[4.00; 4.78]	4.29	.85	[3.86; 4.73]
We used the clues to do the activities	3.72	.89	[3.28; 4.17]	3.76	.90	[3.30; 4.23]
The activities were fun	3.89	1.23	[3.28; 4.50]	3.53	1.01	[3.01; 4.05]
The activities were challenging	3.00	1.19	[2.41; 3.59]	3.41	1.12	[2.84; 3.99]
My partner and I contributed equally to the activities	4.39	1.24	[3.77; 5.01]	4.06	1.20	[3.44; 4.67]
I learned some English pronunciation from the activities	3.89	.68	[3.55; 4.23]	4.06	.97	[3.56; 4.56]
I learned some English grammar or vocabulary from the activities	3.50	.71	[3.15; 3.85]	3.65	1.11	[3.07; 4.22]

Note. 1 = totally disagree, 5 = totally agree

5 Discussion

In relation to RQ1, participants who watched L2 videos with audio-synchronised, textually enhanced subtitles and did pronunciation-focused activities, reported noticing the enhanced words and connected the enhancement to the target feature. This suggests that they were able to move past the stage of *input processing* to that of *intake processing*, as the enhanced exemplars seemed to have been cognitively registered with some level of awareness (Leow, 2015, p. 17). However, even after five weeks of intervention, most participants were unable to describe the rule relative to regular past tense <-ed> pronunciation better than the control group. Despite adopting a sequential design in which learners were encouraged to process input for meaning first and then focus on form through subsequent activities, as recommended by Han et al. (2008), the intervention did not seem to promote the type of conceptually-driven processing necessary to extract abstract rules from the exemplars encountered in the input (Leow, 2015). One possible explanation is that the participants may have struggled to integrate visual and auditory input due to the low salience of the target phonological forms (Strachan & Trofimovich, 2019), and the salience created externally by highlighting the target words and creating activities that revolve around these words may not have aligned with the learners' internally created salience (Sharwood Smith, 1991). In line with this hypothesis, very few participants indicated the *pronunciation* of the regular past endings as the reason for their enhancement, which suggested that the processing of regular past verbs may have primarily focused on their grammatical or semantic properties rather than their phonological realisation. Other possible explanations for the null or negative findings associated with input enhancement typically involve the shortness and implicitness of the treatment (Han et al., 2008). However, in studies of similar length that assessed L2 speech production rather than rule acquisition, significant pronunciation gains have been observed from exposure to enhanced input (Stenton, 2013) as well as the implementation of AV activities (Sanchez-Requena, 2018). Considering that almost all participants in this study perceived the enhancement as useful and that the feeling of pronunciation learning was generally very high, exposure to audio-synchronised enhancement and AV activities may have benefited other, more implicit, dimensions of L2 pronunciation learning.

Regarding RQ2, learners' responses to the questionnaire seemed to indicate an overall positive perception of the videos and activities, in line with previous studies on AV activities (Danan, 2010; Sanchez-Requena, 2018; Sokoli, 2018). The learners indicated that they had understood the videos and the instructions of the activities, and that the videos were fun, confirming that the materials were appropriate for the target population. According to their responses to the questionnaire, participants in group A reported that they relied on subtitles more than participants in group B. If group A had been primarily processing the written input, the appearance of an enhanced word may have interrupted the automatic reading process and successfully redirected their attention to the corresponding auditory form (Stenton, 2013). However, this explanation is only tentative in the absence of online measures of attention allocation from eye-tracking and offline stimulated recall protocols. Moreover, participants in group A, who had already focused on the target words during the first exposure to the enhanced video and may have developed a stronger episodic memory of those auditory forms, were less likely to find the activities challenging than participants in group B. Almost all participants reported that both partners had contributed equally during the activities, suggesting that working in pairs may have helped learners overcome the challenges presented by the dual processing of meaning and form, fostering social interaction with positive effects on language learning (Zabalbeascoa, 2012). Finally, learners reported a higher feeling of learning for

pronunciation than for grammar and vocabulary, especially in relation to the AV activities. The lack of speech perception and production tests, which would have allowed us to draw more robust conclusions regarding pronunciation learning, is a major limitation of this paper, and will be addressed in future publications.

6 Conclusion

After an intervention featuring L2 video with audio-synchronised textual enhancement and video-based activities, our participants reported noticing the enhanced verbs in the subtitles but were unable to infer the past <-ed> pronunciation rule and describe it in writing. However, the intervention was well-received, and the participants' feeling of learning was high, in line with the hypothesis that incorporating these materials into the EFL classroom may foster active and collaborative learning (Zabalbeascoa, 2012). To ensure the successful implementation of AV activities, teachers should carefully select target features and video clips at the appropriate difficulty level and provide clear instructions before each activity. Prefacing the activities with explicit instruction may help direct learners' attention to the phonological properties of the target words, especially with a morphophonemic feature like the regular past <-ed>. To ensure active participation, the teacher should monitor the learners' execution of each stage, provide individualised feedback during pair work and foster a safe learning environment in which learners may be willing to perform the revoicing activities in front of the whole class.

This study focused on the participants' perceptions of audio-synchronised enhancement and AV activities and was therefore very limited in scope. The main limitation was the lack of objective measures of phonological development tapping into the learners' perception and production of the target feature. Recommendations for future research include the analysis of L2 pronunciation development through pre- and post-tests involving L2 speech production tasks, as well as the investigation of learners' attention allocation to audio-synchronised subtitle enhancement. Although collecting eye-tracking data in a classroom setting may not be feasible, a lab-based study featuring a comparable sample may provide valuable insights on the processing of audio-synchronised enhancement. Finally, due to our small and homogenous sample, further research is needed to assess how participants of different ages, proficiencies and mother tongues would respond to audio-synchronised enhancement and AV activities.

Acknowledgments

This work was supported by the Spanish Ministry of Science, Innovation and Universities [grant PID2019-107814GB-I00] and by the Secretary of Universities and Research of the Government of Catalonia and the European Social Fund [grant FI_2019].

References

Bailly, G., & Barbour, W. (2011). Synchronous reading: Learning French orthography by audiovisual training. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 1153–1156). ISCA. http://dx.doi.org/10.21437/interspeech.2011-342

Brutten, S. R., Mouw, J. T., & Perkins, K. (1986). The Effects of language group, proficiency level, and instruction on ESL subjects' control of the {D} and {Z} morphemes. *TESOL Quarterly*, *20*(3), 553–559. https://doi.org/10.2307/3586301

Charles, T., & Trenkic, D. (2015). The effect of bi-modal input presentation on second language listening: The focus on speech segmentation. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning* (pp. 173–197). Peter Lang.

Chiu, Y. (2012). Can film dubbing projects facilitate EFL learners' acquisition of English pronunciation? *British Journal of Educational Technology*, *43*(1), 24–27. https://doi.org/10.1111/j.1467-8535.2011.01252.x

Danan, M., (2010). Dubbing projects for the language learner: A framework for integrating audiovisual translation into task-based instruction. *Computer Assisted Language Learning*, *23*(5), 441–456. https://doi.org/10.1080/09588221.2010.522528

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System, 116,* 103078. https://doi.org/10.1016/j.system.2023.103078

Gerbier, E., Bailly, G., & Bosse, M. L. (2018). Audio–visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning. *Computer Speech and Language, 47*, 74–92. https://doi.org/10.1016.j.csl.2017.07.003

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4), 597–618. https://doi.org/10.1093/applin/amn010

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42*(3), 1–27. https://doi.org/10.1017/S0272263120000108

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge. https://doi.org/10.4324/9781315887074

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Lima, E. F. (2020). The Supra Tutor: Improving speaker comprehensibility through a fully online pronunciation course. *Journal of Second Language Pronunciation*, *6*(1), 39–67. https://doi.org/10.1075/jslp.18031.lim

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211-232). Second Language Acquisition and Testing in Europe Monograph Series 1. EUROSLA.

Mitterer H., & McQueen J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785

Montero Pérez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–47. https://www.jstor.org/stable/44488684

Sánchez-Requena, A. (2018). Intralingual dubbing as a tool for developing speaking skills. *Translation and Translanguaging in Multilingual Contexts*, *4*, 101–128. https://doi.org/10.1075/ttmc.00006.san

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*, 118–132. https://www.jstor.org/stable/43104426

Sokoli, S. (2018). Exploring the possibilities of interactive audiovisual activities for language learning. *Translation and Translanguaging in Multilingual Contexts*, *4*(1), 77–100. https://doi.org/10.1075/ttmc.00005.sok

Stenton, A. (2013). The role of the syllable in foreign language learning: Improving oral production through dual-coded, sound-synchronized, typographic annotations. *Language Learning in Higher Education*, *2* (1), 145–161. https://doi.org/10.1515/cercles-2012-0009

Strachan, L., & Trofimovich, P. (2019). Now you hear it, now you don't: Perception of English regular past –ed in naturalistic input. *Canadian Modern Language Review*, *75*(1), 84–104. https://doi.org/10.3138/cmlr.2017-0082

Vanderplank, R. (2015). Thirty years of research into captions/same language subtitles and second/foreign language learning: Distinguishing between 'effects of' subtitles and 'effects with' subtitles for future research. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language Learning* (pp. 19–40). Peter Lang.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599–624. https://doi.org/10.1017/S0272263120000029

Zabalbeascoa, P., Sokoli, S., & Torres, O. (2012). *Clipflair: Foreign language learning through interactive revoicing and captioning of clips* (Work Package 2, Lifelong Learning Programme 519085-LLP-1-2011-1-ES-KA2-KA2MP). Education and Culture DG, European Union. http://clipflair.net/wp-content/uploads/2014/06/D2.1ConceptualFramework.pdf

Zhang, S. (2016). Mobile English learning: An empirical study on an APP, English Fun Dubbing. *International Journal of Emerging Technologies in Learning, 11*, 4-8. https://doi.org/10.3991/ijet.v11i12.6314

About the authors

Valeria Galimberti is a PhD candidate in Applied Linguistics at the University of Barcelona, and a member of the GRAL (Language Acquisition Research Group). She has conducted research with high school and university students on L2 captioned video processing, vocabulary learning from TV series and production tasks in second language acquisition. Her PhD thesis explores the effects of input enhancement and video-based activities on the acquisition of L2 pronunciation.

Email: galimberti@ub.edu

Joan C. Mora is associate professor in the Department of Modern Languages and Literatures and English Studies and a member of GRAL (Language Acquisition Research Group) at the University of Barcelona (UB) in Spain. He is interested in understanding how contextual and individual factors shape L2 speech learning. His current research interests focus on the role of cognitive and emotional individual differences in the development of L2 pronunciation and speaking fluency, phonological learning in the mental lexicon, phonetic training methods, multimodal pronunciation training, and task-based pronunciation teaching and learning in instructed SLA.

Email: mora@ub.edu

Roger Gilabert is a lecturer and researcher at the University of Barcelona, and he is a member of the GRAL (Language Acquisition Research Group). His research has revolved around task and

syllabus design for the last 25 years. The focus of his research has also been L2 oral and written production, multimodal input processing through caption videos (genres), game-based learning and reading skills. He is currently leading the taskGen project on the automation of second and foreign language task design.

Email: rogergilabert@ub.edu