CrossMark

*Studies in Second Language Acquisition* **42** (2020), 599–624
doi:10.1017/S0272263120000029

# *Research Article*

## CAN CAPTIONED VIDEO BENEFIT SECOND LANGUAGE PRONUNCIATION?

***Natalia Wisniewska*** *

*University of Barcelona, Barcelona*

***Joan C. Mora***

*University of Barcelona, Barcelona*

**Abstract**

The present study investigated the potential benefits of extended exposure to captioned videos for second language pronunciation. We tested 90 L2 adult learners of English on speech processing skills (segmentation, speed of lexical access, and sentence processing) and phonological accuracy in perception (ABX discrimination) and production (accentedness ratings) before and after an 8-week treatment consisting of regular exposure to audiovisual materials. Participants were randomly assigned to four experimental conditions involving two viewing modes (captioned or uncaptioned) and two task focus conditions (focus on phonetic form or focus on meaning). Results showed benefits in speech segmentation and speech processing skills irrespective of viewing mode. No significant benefits were found for phonological accuracy in perception. In production, a focus on phonetic form improved pronunciation only in the absence of captions, whereas captioned viewing led to pronunciation gains as long as there was no focus on phonetic form. These findings suggest that pronunciation improvement can take place with the help of captions or, in the absence of captions, when learners' attention is directed to pronunciation. Cognitive overload might explain why no benefits were obtained when attention was directed to pronunciation in a captioned viewing mode.

### INTRODUCTION

Learning a second language (L2) is a challenging task involving the mastery of different linguistic dimensions (e.g., vocabulary, grammar, pronunciation). The diverse nature of language acquisition contexts makes some of these dimensions more easily learnable than

We would like to thank the editors of this special volume and two anonymous reviewers for their useful suggestions and comments on earlier versions of this article. The study reported in this article was supported by grants FFI 2016-80564R from MINECO and FI-DGR 2017 from AGAUR.

* Correspondence concerning this article should be addressed to Natalia Wisniewska, Departament de Llengues i Literatures Modernes i Estudis Anglesos, Facultat de Filolgia, Universitat de Barcelona. Gran Via de les Corts Casatalanes, 558.08007 Barcelona, Spain. E-mail: nathalie.wisniewska@gmail.com.

© *Cambridge University Press 2020*

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

others. For example, in the context of instructed second language acquisition (SLA), pronunciation learning is particularly challenging, mainly due to the limited amount of input learners receive, the fact that the input learners receive from instructors is often accented, and the scarcity of opportunities to use the L2 in meaningful communication outside the classroom (Piske, 2007). Improving perceptual sensitivity to cross-language phonetic differences would help learners overcome L1-based perception (Best & Tyler, 2007), it would enhance accuracy in speech segmentation and processing through more effective decoding of the speech signal at the acoustic, phonetic, and phonological levels (Ramus et al., 2010), and it would have positive effects for higher-order levels of linguistic processing (e.g., lexical, semantic) and for listening comprehension skills. However, pronunciation does not often receive sufficient attention in instructed SLA settings (Thomson & Derwing, 2014). This is partly due to the lack of clear pedagogical guidelines and coherent goals (Darcy et al., 2012) and the shift in methodological focus from audio-lingualism to communicative approaches to language teaching prioritizing attention to meaning and interaction (Levis, 2005). Moreover, pronunciation training in the language classroom is very challenging because it is hard to integrate a focus on phonetic form into current task-based communicative approaches to language teaching (Darcy et al., 2012; Gurzynski-Weiss et al., 2017; Mora & Levkina, 2017).

The scarcity of L2 input in school settings (2–4 50-minute sessions per week) and teachers' highly variable oral fluency and proficiency (Muñoz, 2008) can make instructional settings nonfavorable for pronunciation development. Aided by easy access to technology, out-of-classroom sources of exposure have become a common supply of contact with the target language. Watching movies and TV series in the L2, for example, is a fundamental resource in providing particularly rich forms of authentic L2 input. Not surprisingly, for the last decade SLA research has been investigating the potential benefits of audiovisual sources of language input for second language learning (Vanderplank, 2010, 2016).

Studies in this research strand are grounded in the notion of bimodal reinforcement from Paivio's (1986) *Dual Coding Theory* and Mayer's (2009) principles of multimedia learning, claiming that the dual processing of auditory and visual information promotes learning by strengthening mental representations of perceived objects. A substantial amount of this research has focused on different sources of audiovisual input, such as subtitled video (Danan, 2004) and television programs (Rodgers, 2013; Webb & Rodgers, 2009), examining its effects on a variety of language learning outcomes (Webb, 2015), especially vocabulary acquisition. This research has also investigated the various conditions under which L2 learners watch audiovisual materials (Peters et al., 2016), such as with L1 subtitles or L2 subtitles (captions), and the role of learners' individual differences in age and proficiency (Muñoz, 2017). The overall picture emerging from this research suggests that audiovisual input in its various modalities is beneficial for foreign language learning in the domains of listening comprehension (Kruger & Steyn, 2014; Markham, 2001; Montero Perez et al., 2013; Vanderplank, 1988) and the incidental acquisition of vocabulary (Bird & Williams, 2002; Bisson et al., 2014; d'Ydewalle & Van de Poel, 1999; Peters & Webb, 2018; see Montero Perez et al., 2013 for an overview of research into captions). However, research examining the potential benefits of audiovisual input for the development of L2 learners' speech perception and production skills is still very scarce and has not directly examined its effects on L2 pronunciation (but see Borrás & Lafayette, 1994). To the best of our knowledge, the present study is the first to do so.

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

### *PREVIOUS RESEARCH ON SUBTITLES AND SPEECH PERCEPTION*

Charles and Trenkic (2015) showed, through the use of a shadowing task performed under laboratory listening conditions consisting in repeating short phrases, that L2 English university students having spent more than one year in the United Kingdom failed to repeat back 30% of the spoken utterances they heard. Such poor performance was attributed to problems of lexical segmentation originating in the inefficient use of low-level listening processes such as the inability to use L2-specific word boundary cues to identify words (Norris et al., 2001) or to a likely mismatch between the incoming auditory input and the inaccurate phono-lexical representations in the mental lexicon (Broersma, 2012). In a follow-up experiment, Charles and Trenkic (2015) administered a similar shadowing task to an analogous population of L2 English university-level students before and after exposure to audiovisual materials. The students watched documentaries in four viewing sessions under two experimental viewing conditions, with and without captions, and a third control viewing condition with captions but no soundtrack. The shadowing task contained test items from the training audiovisual materials, as well as from comparable test trials to which viewers had not been exposed. The bimodal presentation mode (captioned video) obtained the largest gain scores, suggesting that captioned video was superior to uncaptioned video in developing L2 learners' speech segmentation skills. However, Charles and Trenkic (2015) did not directly test whether the benefits obtained in L2 speech segmentation were related to or could extend to enhanced efficiency in L2 phonological processing.

In an earlier study, Mitterer and McQueen (2009) showed that at the perceptual level, captions can aid the phonological decoding and segmentation of speech by helping listeners map auditory input to linguistic form in running speech. In their study, L1-Dutch advanced learners of English watched a 25-minute-long video either in unfamiliar Scottish or Australian English accents and under one of three viewing conditions: with captions, with L1 Dutch subtitles, or without subtitles. After watching the video, they were tested on their ability to repeat excerpts of the auditory materials they had been exposed to (old) and comparable ones they had not been exposed to (new) in a shadowing task. L2 learners watching the video with captions outperformed those who had watched the videos with L1 Dutch subtitles and without subtitles in both old and new test trials. Thus, captions enhanced perceptual adaptation to unfamiliar accents when these were in the same language as the soundtrack, but not otherwise. The authors interpreted these adaptation effects as an instance of perceptual learning guided by the lexical forms provided orthographically during the reading of subtitles in the language of the soundtrack. Perceptual learning occurred because listeners were able to use lexical knowledge to retune perception at the phonetic prelexical level and this improved the recognition of words in the new test trials of the shadowing task. This notion of lexically guided perceptual learning (see also Clarke-Davidson et al., 2008) and the potential benefits of captioned videos for L2 perceptual learning is dependent on a speech-processing mechanism activating phonological representations from orthographic input when readers are simultaneously exposed to auditory and written word forms. Congruency of the language in both modalities would facilitate perceptual learning, that is, it would help the updating of phonological representations to more closely match the

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

auditory input associated with the written form of words, whereas bimodal input with language-incongruent auditory and orthographic word forms (such as L2 soundtrack with L1 subtitles) would hinder perceptual development.

In a similar study, Birulés-Muntané and Soto-Faraco (2016) exposed L1-Spanish intermediate learners of English to a 1-hour-long TV drama in English with subtitles either in Spanish, English, or without either of them. Before and after watching the video, participants took a listening test, a vocabulary test, and a content comprehension test. The listening test consisted of a 1-minute-long excerpt of a conversation between two of the characters learners to which had been exposed, but extracted from a different episode. The conversation was presented auditorily (twice) and in a written transcript that contained 24 gaps to be filled in by the learners. The percentage of correctly identified missing words was used as an index of the impact of viewing mode on L2 learners' speech perception. They found significantly stronger exposure effects for learners watching the audiovisual materials with English captions than for learners exposed to Spanish subtitles or no subtitles.

Birulés-Muntané and Soto-Faraco (2016) thus replicated Mitterer and McQueen's (2009) findings in a listening task. They also interpreted their results as supporting the hypothesis that captions help retune L2 learners' phonetic categories by strengthening the link between the auditory and written form of words during the processing of bimodal input. However, unlike the present study, none of the other studies here reviewed have directly tested whether phonetic categories are in fact retuned as a consequence of lexically guided perceptual learning, or whether L2 speech perception at the phonetic or phonological level improves through the exposure of multimodal input.

The findings from previous research clearly identify captions as being superior to L1 subtitles and viewing without subtitles in effecting benefits in L2 speech processing. However, subtitles may be cognitively overloading, making it difficult for L2 learners to synchronize reading with online auditory speech processing. Although the impact of subtitles on cognitive load in multimodal input environments has yet to be determined (Kruger & Doherty, 2016; Kruger et al., 2014, 2015), it is already known that factors such as type of audiovisual material, level of information redundancy between text and audio, speed of subtitle presentation, and different visual display affect dynamic text processing. Evidence from research on the impact of cognitive load on speech perception suggests that lexically guided perceptual learning can be seriously compromised by cognitive load during audiovisual speech processing, especially at lower levels of proficiency. For example, Mattys et al. (2009) found that informational masking induced by a competing task dividing attention reduced reliance on acoustic detail and increased reliance on lexical-semantic knowledge. Mitterer and Mattys (2017) found that the visual processing of faces while listening was detrimental to speech perception acuity in an oddity discrimination task, suggesting that cognitive load might hinder listeners' ability to rely on fine phonetic detail to perform sublexical tasks negatively affecting perceptual discrimination (Mattys & Wiget, 2011). Thus, lexically guided perceptual learning leading to retuning of phonetic categories may be facilitated by the bimodal input in captioned video, but it may just as well be hindered by L2 learners' proficiency level or some other factor generating cognitive overload. Without cognitive overload L2 bimodal input could serve as a form of perceptual training by reinforcing the link between the orthographic and auditory forms of words. This might lead to the updating

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

of previously established phonological representations toward target phonological forms, effectively making speech segmentation faster and less effortful, and overall improving L2 speech processing, segmental speech perception, and eventually, speech production. However, other factors moderating perceptual learning may be at play. For example, perceptual retuning may take place only as long as L2 learners manage to effectively synchronize and integrate the auditory and written form of words during bimodal input processing. Lexically guided perceptual learning may be more likely when the reading of the written form of words in the captions occurs *after* its auditory form has been processed (Wisniewska & Mora, 2018), whereas when reading occurs before auditory processing, learners' activation of preestablished phonological forms are likely to interfere with the processing of word forms in the auditory input, thus hindering pronunciation learning.

In the current study we set out to investigate the potential benefits of extended exposure to L2 captioned videos for L2 pronunciation. Our aim is to extend previous research in this area by examining both L2 speech processing skills, including speech segmentation, and perceptual and productive accuracy in the learners' L2 phonology. This study represents a first attempt at exploring whether extended exposure to multimodal input through captioned video can benefit L2 pronunciation development.[1]

## *THE PRESENT STUDY*

L2 learners were administered an 8-week exposure treatment to 16 L2 videos from a British TV series and benefits for L2 pronunciation, including L2 speech processing skills and phonological accuracy, were assessed in a pretest/posttest design. In addition, effects of viewing mode and task differences were assessed by randomly assigning participants to one of two *viewing conditions* (with captions or without captions) and one of two *task focus* conditions (directing attention either to phonetic form or meaning).

Manipulating viewing mode was motivated by findings from previous studies showing that captioned video was superior to uncaptioned video in effecting gains in speech perception. Previous studies had not included an extended treatment (Birulés-Muntané & Soto-Faraco, 2016; Mitterer & McQueen, 2009), or if they had (e.g., Charles & Trenkic, 2015), only speech segmentation skills had been tested. Therefore the present study not only includes an extended treatment but also assesses both L2 speech processing skills and L2 phonological accuracy in perception and production. Manipulating task focus was motivated by previous studies showing that L2 pronunciation development in instructional settings with limited L2 exposure, contrary to what might happen in other learning environments, are unlikely to occur without a focus on phonetic form, such as providing corrective feedback on mispronunciations (Saito, 2015) or providing explicit instruction on difficult L2 pronunciation features (e.g., Kissling, 2013; Saito, 2013). We therefore decided to introduce two different task focus conditions to compare the effect of directing learners' attention either to the phonetic form of words, or to the understanding of the plot. This was achieved by introducing attention-directing questions that had to be answered while watching the videos. Although the students knew what type of questions they would be answering, no indication was given for when the question would appear, forcing them to remain attentive throughout each viewing session.

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

The present study addressed the following three research questions (RQs):

RQ1: Do learners' L2 speech processing skills and phonological accuracy benefit from exposure to L2 video clips?
RQ2: Does *viewing mode* (i.e., captioned vs. uncaptioned) lead to gains in L2 speech processing and phonological accuracy?
RQ3: Do *task focus differences* (i.e., inducing attention to phonetic form vs. meaning) lead to gains in phonological accuracy under different viewing modes?

To address these research questions participants were assigned to experimental groups undergoing a viewing treatment lasting 8 weeks, during which they watched two episodes of the British BBC TV series *Luther* per week (2 episodes × 8 weeks = 16 episodes). They watched the episodes under two different viewing modes and task focus conditions: with captions and focusing on phonetic form (Captions + Form), with captions and focusing on meaning (Captions + Meaning), without captions and focusing on phonetic form (No Captions + Form), and without captions and focusing on meaning (No Captions + Meaning). The audiovisual materials were identical for all four experimental conditions. A group of participants from the same pool were recruited as controls and did the pretest and posttest assessment tasks without getting exposure to the audiovisual materials.

## METHODS

### *PARTICIPANTS*

Ninety L1-Spanish/Catalan bilingual learners of English (76 females, 14 males) participated in the study for course credit. They were first year undergraduate university students aged 18–52 ($M$ = 21.88, $SD$ = 8.2) of an upper-intermediate/advanced level, having acquired English mainly in an instructional foreign language setting. They were randomly assigned to one of the four experimental conditions described previously or to a control group. To ensure that participant groups were comparable in terms of their exposure to L2 audiovisual materials, we collected self-reports on frequency of viewing movies and series in English with original soundtrack and on viewing mode (with or without captions) through a questionnaire (see Table 1). Viewing Frequency was assessed through a 5-point Likert scale (1 = never; 2 = 1–3 times in a year; 3 = once a month; 4 = every week; 5 = every day), as was Viewing Mode (with captions: 1 = never; 2 = 1–3 times in a month; 3 = 1–3 times in a week; 4 = 4–6 times in a week; 5 = every day). We also ensured that participant groups were comparable in L2 proficiency by administering an elicited imitation task (Ortega et al., 2002).

## VIEWING TREATMENT AND VIDEO MATERIALS

Participants watched the first season of the British series *Luther* (BBC One, 2010), divided into chronological excerpts of comparable length (approximately 15–20 minutes) twice a week. They watched each session through an online platform *PlayPosit* (www.playposit.com). This platform allowed the researchers to embed question into each viewing sessions, collect students' answers and provide immediate feedback after each answer. The sessions were made available to learners for watching online only for a

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press



TABLE 1. Participants' demographics, viewing habits, and proficiency level

| | | Group | *N* | *M* | *SD* | *SE* | *95%CI* |
|---|---|---|---|---|---|---|---|
| Age (years) | | Captions + Form | 19 | 24.80 | 11.15 | 2.49 | 19.58–30.02 |
| | | Captions + Meaning | 19 | 20.32 | 8.97 | 1.91 | 16.34–24.30 |
| | | NoCaptions + Form | 19 | 20.63 | 4.66 | 1.07 | 18.39–22.88 |
| | | NoCaptions + Meaning | 20 | 19.89 | 2.45 | 0.58 | 18.67–21.11 |
| | | Control | 13 | 24.43 | 9.77 | 2.61 | 18.79–30.07 |
| L2 Proficiency (%) | | Captions + Form | 19 | 78.42 | 15.85 | 3.64 | 70.78–86.05 |
| | | Captions + Meaning | 19 | 75.39 | 14.36 | 3.29 | 68.47–82.31 |
| | | NoCaptions + Form | 19 | 76.21 | 11.64 | 2.67 | 70.59–81.82 |
| | | NoCaptions + Meaning | 20 | 75.75 | 15.70 | 3.51 | 68.40–83.09 |
| | | Control | 13 | 75.32 | 15.20 | 4.22 | 66.13–84.50 |
| Viewing (1–5) | Frequency | Captions + Form | 19 | 3.60 | 0.88 | 0.20 | 3.18–4.01 |
| | | Captions + Meaning | 19 | 3.95 | 1.05 | 0.22 | 3.49–4.41 |
| | | NoCaptions + Form | 19 | 3.63 | 1.34 | 0.31 | 2.98–4.27 |
| | | NoCaptions + Meaning | 20 | 3.17 | 0.92 | 0.22 | 2.70–3.62 |
| | | Control | 13 | 3.93 | 1.07 | 0.29 | 3.30–4.54 |
| | Captioned | Captions + Form | 19 | 2.35 | 0.88 | 0.20 | 1.94–2.75 |
| | | Captions + Meaning | 19 | 2.86 | 1.36 | 0.29 | 2.26–3.46 |
| | | NoCaptions + Form | 19 | 3.11 | 1.59 | 0.37 | 2.33–3.87 |
| | | NoCaptions + Meaning | 20 | 1.83 | 0.62 | 0.15 | 1.52–2.14 |
| | | Control | 13 | 2.71 | 1.49 | 0.40 | 1.85–3.57 |
| | Uncaptioned | Captions + Form | 19 | 2.70 | 1.42 | 0.32 | 2.03–3.36 |
| | | Captions + Meaning | 19 | 3.14 | 1.25 | 0.27 | 2.58–3.68 |
| | | NoCaptions + Form | 19 | 2.63 | 1.38 | 0.32 | 1.96–3.29 |
| | | NoCaptions + Meaning | 20 | 2.33 | 1.28 | 0.30 | 1.69–2.97 |
| | | Control | 13 | 3.29 | 1.59 | 0.42 | 2.36–4.20 |

limited period of time (3 days for each excerpt) to control for regularity of exposure (two excerpts per week) as well as to keep control of learners' log times. All learners watched the excerpts in the same order. They received detailed instructions on how to use the platform and, after the treatment, were asked to report under what conditions the viewing had taken place. All the experimental groups turned out to be comparable in terms of the device used ($\chi^2(12) = 10.14$, $p = 0.606$, $V = 0.21$), reporting having watched the series mostly on a computer and using headphones ($\chi^2(6) = 7.32$, $p = 0.29$, $V = 0.22$). The precise record of their log times and the registered time spent on the platform allowed the researchers to control that everyone followed the task procedure correctly. Lack of regularity in viewing or not completing all the viewing sessions resulted in exclusion from the study. For this reason we excluded three participants.

Viewing mode was operationalized between subjects by assigning participant groups to either a viewing condition with captions or without captions. As shown in Table 1, within each one of the viewing conditions, participants' attention was directed either to the phonetic form of words or to the understanding of the plot. This was achieved by providing participants with two different types of questions (4–5 per video excerpt) that had to be answered while watching. Questions on pronunciation served the purpose of directing learners' attention to the speech input focusing on the phonological representation of words and included a variety of prompts about segmental and suprasegmental features such as "*Do the words **mistake** and **mean** share the same initial vowel?*" or "*Which of the following words is pronounced differently? caught, thought, **though**, ought*" (see Appendix A for examples). The answer to each question was always related to the auditory input the participants had just heard. Participants viewing the excerpts under this task focus condition were expected to generally orient their attention to pronunciation while watching, as correctly answering the question was dependent on how much attention was paid to pronunciation. Questions on meaning were similar in structure and type to those on pronunciation but focused exclusively on the comprehension of the plot, such as "*What did the comment about Luther's 'merciless approach' refer to?*" For both types of attention-directing questions we used a variety of formats, including true/false, multiple choice, pick the odd-one-out, and short written answers. Participants received two types of feedback: immediate feedback on accuracy through a visual prompt (correct! or wrong!) and a percentage score after each one of the viewing sessions. In the present experimental design, watching video excerpts therefore had an interactive component not present in normal video viewing.

## TESTING L2 SPEECH PROCESSING AND PHONOLOGICAL ACCURACY

To assess the effects of extended exposure to the L2 video materials under different viewing conditions all participants performed a battery of pronunciation-related tests in the L2 before and after exposure to the video excerpts. Pronunciation-related development in speech processing was tested through a shadowing task that provided a measure of speech segmentation skill (Mitterer & McQueen, 2009), an animacy judgment task that measured speed of lexical access (Segalowitz & Frenkiel-Fishman, 2005), and a sentence verification task that measured the efficiency of L2 speech processing (Munro & Derwing, 1995). Phonological accuracy was tested in perception using an ABX categorical discrimination task (Flege, 2003; Gottfried, 1984) measuring accuracy in the perception of a

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

TABLE 2. Experimental design of the study

| Week 1 | → | Week 8 |
|---|---|---|
| Pretest | Treatment of video watching (16 sessions) twice a week | Posttest |
| **L2 speech processing measures**<br>• Shadowing task<br>• Animacy judgment task<br>• Sentence verification task<br>**L2 phonological accuracy**<br>• ABX discrimination task<br>• Foreign accent rating task | **(1) Viewing mode**<br>with captions<br>or<br>without captions<br>**(2) Instructional focus**<br>phonetic form<br>or<br>meaning | **L2 speech processing measures**<br>• Shadowing task<br>• Animacy judgment task<br>• Sentence verification task<br>**L2 phonological accuracy**<br>• ABX discrimination task<br>• Foreign accent rating task |
| **Proficiency measure (EIT)** | | |
| **Pretest questionnaire** | | **Posttest questionnaire** |

difficult vowel contrast, and a foreign accent rating task provided a holistic assessment of accuracy in participants' L2 speech production (see Table 2).

All the tests were administered individually in a quiet lab environment. Perception tasks (lexical decision, sentence verification, ABX discrimination, native speaker rating task) were performed using Beyerdynamic DT-770 noise-canceling headphones at self-adjusted volume levels. Production tests were performed using Beyerdynamic DT-990 open headphones and speech production was recorded onto a PMD-661 mkii solid-state digital recorder through a Shure-SM58 voice microphone at a 41.1Kz sampling rate. The following subsections provide a detailed description of the tasks.

## L2 SPEECH PROCESSING AND PROFICIENCY

### *SHADOWING*

We followed Mitterer and McQueen (2009) and Charles and Trenkic (2015) in using a shadowing task to measure speech segmentation. In this task participants listened to short excerpts of speech and repeated as many words as they could from them. As in previous research, trials for repetition included speech samples taken from the video excerpts participants had been exposed to (*Luther*) during the treatment (*old* items) as well as speech samples taken from a TV series (*Sherlock*) participants had not been exposed to (*new* items). The shadowing task included *old* and *new* items to assess whether potential gains in speech segmentation obtained through the treatment could be generalized to new items not included in the treatment, thus showing evidence of knowledge transfer and learning. All selected excerpts were produced by the lead characters in both series. Participants had to repeat back 40 sentences in total; 20 *old* items taken from the British series *Luther* and 20 *new* items taken from the British series *Sherlock.* Speech samples were selected to vary in difficulty through word length between 4 and 15 words, but average word length was comparable across series (*Luther: M* = 7.85, *SD* = 1.98, *95%CI* = 8.07–8.23; *Sherlock: M* = 8.15, *SD* = 2.53, *95%CI* = 7.78–7.92). The speech samples from the series participants had not been exposed to (*Sherlock*) were slightly longer and thus

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

slightly more difficult than those from the series they had been exposed to (*Luther*). A measure of speech segmentation was obtained by counting the total number of function and lexical words repeated (Mitterer & McQueen, 2009) and a proportion score measure was computed for every item and participant.

### *ANIMACY JUDGMENT*

To assess speed of lexical access we administered an animacy judgment task in English (Segalowitz & Frenkiel-Fishman, 2005). Participants were presented with 80 word stimuli and were prompted to decide, as fast as possible, whether words identified either a *living* (e.g., *dog*) or *nonliving* (e.g., *chair*) item by pressing either the right or left shift key on the computer keyboard. The presentation of the stimuli was visual (orthographic representation of words). All the test items were of high lexical frequency to ensure lexical activation by our L2 English learners. Only words from the first of the nine frequency word lists on the BNC/COCA (Nation, 2012) were used. Most of them (97%) belong to the 1–6k word families and the rest 3% were off-list compound words (e.g., *hairdresser*). Half the items qualified as *living* and the other half as *nonliving*. The presentation of test items was preceded by eight practice trials providing feedback on accuracy and reaction time. Our dependent measure for this task was the mean response latency (in milliseconds), which represents the processing time it took participant to access the lexical representation of the word appearing on the screen.

### *SENTENCE VERIFICATION TASK*

This task measured speed of processing of L2 speech. Following Munro and Derwing (1995), processing times were estimated by measuring the time needed for listeners to assign true/false values to a number of short sentences. Participants were presented aurally with 40 sentences spoken by a male native speaker of Southern British English and were instructed to answer, as fast as possible, whether the sentence they heard was *true* or *false* by pressing either the right or left shift key on the computer keyboard. Sentences classified as *true* were semantically coherent (e.g., *Hot and cold are opposite*) and those classified as *false* lacked semantic coherence (e.g., *Gasoline is an excellent drink*). Each item was a single-clause sentence of four to eight common words (mean word length = 5.9). The speech samples were normalized for peak and mean amplitude and high-pass filtered (50Hz). The British speaker was instructed to produce the sentences at conversational speed. The mean speech rate was 270 syllables per minute and the mean sentence length was 1.8 seconds. Our dependent measure for this task was the mean response latency (in milliseconds) representing the processing time it took participants to assign a truth value to each sentence.

### *ELICITED IMITATION*

An elicited imitation task (Ortega et al., 2002) was used to obtain an estimate of interlearner differences in L2 proficiency because it has been shown to discriminate well between different proficiency levels (Yan et al., 2016). In addition, scores rely heavily on processing and automaticity, as learners are required to auditorily process and repeat back

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

sentences of increasing grammatical complexity. This allowed us to determine proficiency differences between groups through their performance on a task that is related to what learners do when processing multimodal input. The elicited imitation task consisted of 30 sentences of increasing grammatical complexity ranging 7–19 syllables in length and containing high-frequency vocabulary items. Participants heard each sentence only once through headphones and were asked to repeat the sentences as accurately as they could after a 2-second delay signaled by a beep sound. Following the scoring rubric of Ortega et al. (2002) available in the IRIS digital repository (Marsden et al., 2016), each sentence was assigned 0, 1, 2, 3, or 4 points, depending on how much of the sentence could be repeated and the kind of errors produced (if any), to a maximum raw score of 120 for the 30 sentences presented, which was then converted to percentage scores.

## L2 PHONOLOGICAL ACCURACY

### *ABX DISCRIMINATION*

To assess L2 perceptual learning in terms of phonological accuracy we administered a speeded ABX categorical discrimination task (Flege, 2003; Gottfried, 1984) based on Darcy et al. (2016). Participants heard three stimuli in a row and had to choose if the last one (X) was more similar to the first one (A) or to the second one (B). Interstimulus interval was 500ms. Trials were presented 1000ms after response or automatically 2500ms after the previous trial if no response was recorded. To increase task demands, the stimuli consisted of trisyllabic nonwords following the rules of English phonotactics and a CV.ˈCV.CV(C) structure (e.g., *fadittick*). We tested the lax-tense /iː/-/ɪ/ vowel contrast, which is phonemically contrastive in English, but not in Spanish or Catalan, and thus problematic for English L2 learners having these languages as their L1. There were 32 difficult /iː/-/ɪ/ test trials and 16 easy control trials (/a/-/iː/). The percentage of correct responses on test trials was used as an accuracy measure at pre- and posttest. We also used a response latency measure (RT) based on correct responses screened for extreme values at 2.5 standard deviations below and above each participant's mean RT. The task was administered on a PC through headphones using the presentation software DMDX (Forster & Forster, 2003). Our dependent measures for this task were accuracy of correct responses and response latency (RT).

### *DELAYED SENTENCE PRODUCTION FOR FOREIGN ACCENT RATING*

In this task participants were asked to silently read and remember English sentences (5–7 words long) appearing in standard orthography on the computer screen for 2.5 seconds and to pronounce them as accurately as they could after a beep sound presented after the sentence had disappeared from the screen. Sentences had not been available on the screen for at least 1.5 seconds when produced and were not presented auditorily to avoid direct imitation and the influence of a model voice in L2 learners' production. In this way, we hoped to elicit samples of speech from memory reflecting learners' L2 phonological representations.

Given the number of participants in the study, three representative sentences containing rich phonological features were selected for presentation to native speaker judges in a

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

foreign accent rating task. Twelve British native speakers (mean age = 29.2, SD = 5.60); 7 females, 5 males) participated as paid judges in the foreign accent rating task. All of them reported familiarity with Spanish-accented speech. Their self-reported average Spanish proficiency level was 4.09 (*SD* = 1.78) on a 9-point Likert scale (1 = no knowledge of Spanish; 9 = proficient knowledge of Spanish), indicating a lower-intermediate proficiency level in Spanish. The raters listened to six sentences produced by each participant (the same three sentences from pretest and posttest). The sentences were presented in pretest-posttest pairs in randomized trials. In every trial, raters provided two ratings on a continuous foreign accent scale, one for each one of the two sentences in the pair. In half the trials the pretest sentence appeared first and in the other half the posttest sentence appeared first. Each sentence in each pair was rated separately on a continuous 0–100 scale by placing the cursor anywhere on the scale (see Appendix A). The two ends of the scale were labeled as 0 = "No Foreign Accent" and as 100 = "Heavy Foreign Accent." Due to the number of participants, we created two versions of the rating task so that 50% of the speech samples were assigned to one group of raters and the other 50% to another group. All 12 native speakers rated 10% of the speech samples to calculate an interrater reliability index. The Cronbach's alpha index for these ratings (360 items) resulted in a reliability coefficient of $\alpha = 0.862$, suggesting consistency in the ratings. Consequently, we calculated an average foreign accent rating score (1–100) across judges for each sentence by each participant that was used as a measure of phonological production accuracy. On this measure, the lower the percentage score, the less foreign accented the sentence was.

## RESULTS

The main aim of the present study was to assess the effects of a treatment based on the extended exposure to L2 bimodal input through video materials (TV series) on speech processing skills and phonological accuracy (RQ1). The treatment was provided under two different *Viewing* conditions (*with* or *without captions*, RQ2) and two *Task Focus* conditions (*form-focused* or *meaning-focused*, RQ3). We first assessed the between-group differences in L2 proficiency and exposure to media reported in Table 1. The scores obtained in the elicited imitation task ranged from 54 to 119, that is 45–99% accuracy ($M = 76.27$, $SD = 14.31$, *95%CI* = 73.27–79.27), suggesting an overall relatively advanced level of proficiency with noticeable interlearner variability. As expected, L2 proficiency was related to L2 learners' ability to repeat sentences in the shadowing task ($r = .75$, $p < .001$) and to measures of L2 speech processing (animacy judgment: $r = –.26$, $p = .015$; sentence verification: $r = –.33$, $p = .002$) and L2 phonological accuracy (accentedness: $r = –.30$, $p = .005$; but not ABX discrimination: $r = .15$, $p = .151$). However, L2 proficiency was unrelated to any of the gains obtained for either speech processing (shadowing: $r = .028$, $p = .800$; animacy judgment: $r = .054$, $p = .620$; sentence verification: $r = .03$, $p = .726$) or phonological accuracy (ABX: $r = .06$, $p = .534$; accentedness: $r = –.078$, $p = .470$), suggesting that interlearner differences in proficiency did not contribute fundamentally in explaining gains. In addition, a one-way ANOVA with the elicited imitation score as a dependent measure and participant group as the independent between-subjects factor confirmed that learner groups did not vary significantly on this proficiency measure ($F(4,85) = 0.14$; $p = 0.967$, $\eta^2 = .007$).[2] Similarly,

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

according to Kruskal-Wallis tests,[3] the effect of participant group failed to reach significance for age ($H(4) = 4.35$, $p = .361$, $\varepsilon^2 = .047$) viewing frequency ($H(4) = 7.15$, $p = .128$, $\varepsilon^2 = .078$), and viewing mode with ($H(4) = 8.97$, $p = .062$, $\varepsilon^2 = .098$) and without ($H(4) = 5.37$, $p = .251$, $\varepsilon^2 = .058$) captions. These analyses suggest that the four experimental groups and the control group were comparable in terms of L2 proficiency and self-reported viewing habits.

We next compared how accurate experimental participants were in answering treatment questions on either plot comprehension or pronunciation under the two viewing conditions. A one-way factorial ANOVA on response accuracy scores with viewing mode and task focus as between-subjects independent variables revealed main effects of *Viewing Mode* ($F(3,72) = 5.07$, $p = .027$, $\eta^2 = .066$), *Task Focus* ($F(3,72) = 4.98$, $p = .029$, $\eta^2 = .065$), and a significant *Viewing Mode* × *Task Focus* interaction ($F(3,72) = 4.11$, $p = .046$, $\eta^2 = .054$). The interaction arose because answering questions on plot comprehension in the absence of captions (no captions + meaning) yield significantly lower accuracy scores ($M = 67.42$; $SD = 8.41$; *95%CI* = 63.49–71.36) compared to all other experimental groups: captions + focus on phonetic form ($M = 77.72$; $SD = 12.12$; *95%CI* = 71.69–83.76), no captions + focus on phonetic form ($M = 77.21$; $SD = 9.15$; *95%CI* = 72.80–81.63), captions + focus on meaning ($M = 77.25$; $SD = 10.13$; *95%CI* = 72.37–82.14).

In the following subsections we assess the effect of the treatment, first on L2 speech processing skills under different viewing modes (*with* or *without captions*) and then on L2 phonological accuracy as a function of task focus conditions, that is, whether learners' attention was directed to the phonetic form of words (*form-focused*) or to the comprehension of the plot (*meaning-focused*) within each of the two viewing modes.

### *TREATMENT EFFECTS ON L2 SPEECH PROCESSING SKILLS UNDER DIFFERENT VIEWING CONDITIONS*

Treatment effects on L2 speech processing skills were assessed through pre/posttest changes in L2 learners' performance on the *Shadowing*, *Animacy Judgment*, and *Sentence Verification* tasks.

Overall, the descriptive statistics in Table 3 indicate positive gains between pretest and posttest in L2 learners' segmentation skills (*Shadowing*), speed of lexical access (*Animacy Judgment*), and speed of speech processing at a sentence level (*Sentence Verification*). Gain sizes appear to be substantially superior for the experimental participants compared to those of controls in the tasks involving the processing of sentence-long stretches of speech (*Shadowing* and *Sentence Verification*). Average gains in speed of lexical access did not appear to differ between experimental and control groups (*M gain* = 18 for both). As regards viewing conditions, the group average gain for speed of lexical access was twice as fast for participants exposed to captioned video (*M gain* = 24) than for those exposed to video without captions (*M gain* = 12). A speeded decision on the true value of a sentence presented auditorily (*Sentence Verification*) appears to have become faster after the treatment for those participants in the uncaptioned viewing condition (*M gain* = 120) than for those in the captioned viewing condition (*M gain* = 93.7).

It is worth noting that the *Shadowing* task included test items from the TV series participants had been exposed to (*Luther*) as well as test items from a series they had not

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press



TABLE 3. Descriptive statistics for L2 speech processing tasks

| | | *Shadowing* (% words repeated) | | | *Animacy Judgment* RT (ms) visual | | | *Sentence Verification* RT (ms) auditory | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Mean Scores* | Test | *M* | *SE* | *95%CI* | *M* | *SE* | *95%CI* | *M* | *SE* | *95%CI* |
| *Experimental* | T1 | 58.05 | 1.9 | 54–62 | 747 | 12 | 724–772 | 2221 | 15 | 2191–2251 |
| (n = 77) | T2 | 69.66 | 1.9 | 66–73 | 730 | 12 | 707–753 | 2114 | 15 | 2085–2143 |
| *Captions* | T1 | 59.42 | 2.5 | 54–64 | 760 | 18 | 726–795 | 2239 | 22 | 2196–2282 |
| (n = 38) | T2 | 70.93 | 2.5 | 66–76 | 736 | 17 | 704–770 | 2145 | 21 | 2104–2186 |
| *No Captions* | T1 | 56.62 | 2.5 | 52–62 | 736 | 17 | 704–770 | 2204 | 21 | 2163–2246 |
| (n = 39) | T2 | 68.42 | 2.5 | 63–73 | 724 | 17 | 693–757 | 2084 | 20 | 2045–2124 |
| *Control* | T1 | 52.29 | 4.7 | 43–62 | 757 | 30 | 702–818 | 2183 | 36 | 2112–2256 |
| (n = 13) | T2 | 58.02 | 4.7 | 49–67 | 739 | 29 | 684–797 | 2111 | 35 | 2043–2181 |
| *Mean Gain Scores* | | *Gain* | *SE* | *% Gain* | *Gain* | *SE* | *% Gain* | *Gain* | *SE* | *% Gain* |
| *Experimental* | T2–T1 | 11.6 | 0.7 | 19.98 | 18 | 3 | 2.41 | 107 | 8 | 4.82 |
| *Captions* | T2–T1 | 11.5 | 1.1 | 19.35 | 24 | 4 | 3.16 | 93.7 | 12 | 4.18 |
| *No Captions* | T2–T1 | 11.7 | 1.1 | 20.66 | 12 | 4 | 1.63 | 120 | 11 | 5.44 |
| *Control* | T2–T1 | 5.7 | 2.0 | 10.90 | 18 | 7 | 2.38 | 72 | 19 | 3.30 |

*Note*: M = mean, SE = standard error, CI = confidence interval, ms = milliseconds, T1 = pre–test, T2 = post–test, Gain = T2 minus T1 difference.

been exposed to (*Sherlock*) to test for generalization of treatment gains to a new context. Although on average L2 learners were able to repeat slightly more words in test items at pretest from the untrained series *Sherlock* ($M$ = 61.9, $SE$ = 1.8, *95%CI* = 58.4–65.5) compared to the trained series *Luther* ($M$ = 56.0, $SE$ = 1.9, *95%CI* = 52.2–59.8; see mean scores by group in Table 3), gain sizes at posttest reached comparable scores for test items from both series (*Sherlock*: $M$ = 70.5, $SE$ = 1.7, *95%CI* = 67.3–73.8 vs. *Luther*: $M$ = 70.3, $SE$ = 1.7, *95%CI* = 67.0–73.7). We ran a linear mixed-model analysis on the experimental participants' data only where *Series*, *Time*, *Viewing*, and their interactions were included as fixed factors and *Subject* and *Item* as random intercepts. In this and all subsequent statistical analyses for this task, we also included *MaxScore*, the maximum number of words a participant could repeat back on a given test trial as a covariate. This was done to control for word length, as shorter test trials were much more likely to obtain higher scores than longer ones. This analysis revealed significant main effects of *Time* ($F(1,5911)$ = 216.8, $p < .001$; $\beta = 10.4$, $SE = 1.6$, $t = 6.58$), but neither *Series* ($F(1,5911) = 3.3$, $p < .078$; $\beta = 0.74$, $SE = 1.6$, $t = .047$) nor any of the interactions reached significance. Thus, learners could extend the benefits in speech segmentation skill obtained through extended exposure to a TV series (*Luther*) to help them segment speech in a TV series to which they had not been exposed. In subsequent analyses only the data from the trained series *Luther* were used.

To assess the effect of exposure to the TV series *Luther* on L2 speech processing skills, we ran, for each one of the speech processing tasks, a series of mixed-model analyses with *Time* (T1, T2), *Viewing* (Captions, NoCaptions, Control), and their interaction as fixed factors, including random intercepts for *Subject* and *Item* (see model coefficients in Table 1 in Appendix C). For the *Shadowing* task, the main effect of *Viewing* did not reach significance ($F(2, 83) = 2.03$, $p = .138$), but both the main effect of *Time* ($F(1,3332)$ = 113.6, $p < .001$) and the *Time* × *Viewing* interaction did ($F(2, 3332) = 6.92$, $p = .001$). The interaction arose because both experimental groups significantly improved between testing times ($p < .001$), whereas the control group did not ($p = .059$), and because whereas one of the experimental groups differed significantly from the control group at T2 (Captions: $p = .020$; NoCaptions: $p = .073$), they did not differ from the control group at T1, nor did the two experimental groups differ from one another ($p = 1.0$). Thus, no differential effects were found on speech segmentation skills as a function of whether learners were watching the TV series with or without captions, both viewing modes appear to have affected speech segmentation skills positively. In the *Animacy Judgment* task there was a significant main effect of *Time* ($F(1,12423) = 34.88$, $p < .001$) and a nonsignificant main effect of *Viewing* condition ($F(2,85) = .292$, $p = .748$). The *Time* × *Viewing* interaction did not reach significance ($F(2,12424) = 2.14$, $p = .117$). Thus, although L2 learners exposed to captioned videos obtained larger gains (3.16%) than those exposed to uncaptioned videos (1.63%), the group difference did not reach significance at either T1 or T2 ($p = 1.0$). Finally, the linear mixed-model analysis on the *Sentence Verification* scores showed a main effect of *Time* ($F(1,5813) = 236.6$, $p <$ .001), a nonsignificant main effect of *Viewing* condition ($F(2, 84) = 1.29$, $p = .279$), and a significant *Time* × *Viewing* interaction ($F(2, 5813) = 3.98$, $p = .019$). This is due to the fact that although all groups significantly improved in speed of sentence processing from pretest to posttest, both experimental groups improved more than the control group did (4.18% and 5.44% vs. 3.30%). L2 learners exposed to the TV series *without* captions

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

improved to a larger extent (5.44%) than L2 learners exposed to the TV series *with* captions did (4.18%), but the two experimental groups did not differ from one another at either pretest ($p = .739$) or posttest ($p = .109$).

Overall, the benefits in L2 speech processing skills identified suggest that extended viewing of TV series in the L2, both with and without captions, might extend to L2 phonological processing skills, leading to improvement in L2 speech perception and production. In the following section we examine such potential effects by assessing the effect of viewing mode and task focus conditions on L2 phonological accuracy in perception (L2 sound discrimination) and production (accentedness ratings).

### *TREATMENT EFFECTS ON L2 PHONOLOGICAL ACCURACY UNDER DIFFERENT TASK FOCUS CONDITIONS*

Treatment effects on L2 phonological accuracy in perception were assessed through an ABX categorical discrimination task. The results of the ABX discrimination task correspond to the average scores L2 learners obtained by group and experimental condition on the ABX task test items only (Table 4). Control trials were much easier to discriminate ($M = 0.80$, $SD = 0.39$, 95% $CI = 0.79$–$0.82$ vs. $M = 0.65$, $SD = 0.47$, 95% $CI = 0.64$–$0.67$) and were also discriminated faster ($M = 1182$, $SD = 339$, 95% $CI = 1168$–$1196$ vs. $M = 1221$, $SD = 340$, 95% $CI = 1210$–$1232$), as expected, than the test items targeting the difficult English vowel contrast /iː/-/ɪ/. Test and control items differed significantly both in terms of accuracy ($t(8670) = -14.74$, $p < .001$) and speed ($t(5993) = 4.39$, $p < .001$).

As the descriptive data in Table 4 show, for ABX discrimination accuracy and RT gains ranged considerably between experimental groups, which were not always larger than those of the control group. For the accent rating scores there was overall improvement between testing times, but gains in accentedness were much larger for learners exposed to captions while focusing on meaning (5.18%) than for those focusing on phonetic form (1.09%), whereas learners exposed to the TV series without captions focusing on meaning obtained much smaller gains (0.96%) than those focusing on phonetic form (5.28%). Thus, for accent ratings, only two of the experimental groups appeared to outperform the control group: those who, while being exposed to captions were focusing on meaning, and those who, while not being exposed to captions, were focusing on phonetic form. Therefore, incidental learning of pronunciation would seem to occur only through exposure to captioned viewing, whereas in the absence of captions gains would seem to be driven by an intentional focus on pronunciation.

To assess the effect of exposure to the TV series *Luther* on L2 phonological accuracy, we fit the ABX discrimination and accent rating data to a set of linear mixed models with *Viewing* (Caption, NoCaption, Control), *Task Focus* (Form-Focused, Meaning-Focused, Control), *Time* (T1, T2), and their interactions as fixed factors, including random intercepts for *Subject* and *Item* (see model coefficients in Table 3 in Appendix C). These analyses showed that none of the experimental conditions significantly affected ABX accuracy scores (all main effects and interactions failed to reach significance), except for *Time* ($F(1,5750) = 4.01$, $p = .045$), suggesting that the slight overall improvement observed in ABX accuracy was unrelated to the treatment. However, only the experimental group exposed to captions and focusing on meaning obtained larger gains (9.73%) than those of the control group (6.18%), which in turn obtained larger gains than the other experimental

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press



TABLE 4. Descriptive statistics for L2 phonological accuracy tasks

| | | | *ABX Discrimination* | | | | | | *Accent Rating* (1–100) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Accuracy | | | RT (ms) | | | | | |
| *Mean Scores* | | | *M* | *SE* | *95%CI* | *M* | *SE* | *95%CI* | *M* | *SE* | *95%CI* |
| *Experimental (n = 77)* | | T1 | .649 | .013 | .62–.67 | 1244 | 25 | 1197–1293 | 53.9 | 1.6 | 51–57 |
| | | T2 | .672 | .013 | .65–.70 | 1163 | 23 | 1119–1209 | 52.2 | 1.6 | 49–55 |
| *Captions* | *Form* | T1 | .660 | .026 | .61–.71 | 1271 | 50 | 1177–1372 | 53.0 | 3.3 | 46–60 |
| *(n = 38)* | *(n = 19)* | T2 | .679 | .026 | .63–.73 | 1197 | 47 | 1108–1293 | 52.4 | 3.3 | 46–59 |
| | *Meaning* | T1 | .586 | .027 | .53–.64 | 1291 | 51 | 1195–1394 | 54.3 | 3.2 | 48–61 |
| | *(n = 19)* | T2 | .643 | .027 | .59–.69 | 1178 | 46 | 1090–1272 | 51.5 | 3.2 | 45–58 |
| *NoCaptions* | *Form* | T1 | .668 | .026 | .62–.72 | 1267 | 51 | 1171–1370 | 53.3 | 3.3 | 47–60 |
| *(n = 39)* | *(n = 19)* | T2 | .704 | .025 | .65–.75 | 1218 | 49 | 1126–1317 | 50.4 | 3.3 | 44–57 |
| | *Meaning* | T1 | .674 | .025 | .62–.72 | 1173 | 45 | 1089–1264 | 54.9 | 3.2 | 49–61 |
| | *(n = 20)* | T2 | .657 | .026 | .60–.71 | 1080 | 42 | 1002–1165 | 54.3 | 3.2 | 48–61 |
| *Control (n = 13)* | | T1 | .599 | .033 | .53–.66 | 1260 | 62 | 1144–1389 | 58.4 | 4.0 | 50–66 |
| | | T2 | .636 | .033 | .57–.70 | 1207 | 59 | 1096–1329 | 56.4 | 4.0 | 48–64 |
| *Mean Gain Scores* | | | *Gain* | *SE* | *% Gain* | *Gain* | *SE* | *% Gain* | *Gain* | *SE* | *% Gain* |
| *Experimental* | | T2–T1 | .023 | .034 | 3.54 | 81 | 23 | 6.96 | 1.70 | 0.8 | 3.21 |
| *Captions* | *Form* | T2–T1 | .019 | .014 | 2.88 | 74 | 9 | 6.18 | 0.59 | 1.7 | 1.09 |
| | *Meaning* | T2–T1 | .057 | .027 | 9.73 | 113 | 19 | 9.59 | 2.76 | 1.7 | 5.18 |
| *NoCaptions* | *Form* | T2–T1 | .036 | .028 | 5.39 | 49 | 20 | 4.02 | 2.90 | 1.7 | 5.28 |
| | *Meaning* | T2–T1 | –.017 | .027 | –2.52 | 93 | 18 | 8.61 | 0.56 | 1.7 | 0.96 |
| *Control* | | T2–T1 | .037 | .027 | 6.18 | 53 | 17 | 4.39 | 1.98 | 2.1 | 3.67 |

*Note*: M = mean, SE = standard error, CI = confidence interval, ms = milliseconds, T1 = pretest, T2 = posttest, Gain = T2 minus T1 difference

groups (–2.52–5.39%). A similar pattern of results emerged for ABX RT scores. The main effect of *Time* was significant ($F(1,3680) = 61.34$, $p < .001$), as was the *Task Focus* × *Time* interaction ($F(1,3680) = 5.88$, $p < .015$), but no other main effect or interaction reached significance. The significant interaction was driven by the fact that the control group improved to a lesser extent (4.39%) than the meaning-focused groups (8.6–9.6%).

We next ran a similar linear mixed model on the accent rating scores (see Table 2 in Appendix C for model coefficients). The results showed a significant main effect of *Time* ($F(1,516) = 5.09$, $p = .024$), but none of the interactions involving *Viewing Mode* or *Task Focus* reached significance. Bonferroni-adjusted pairwise comparisons confirmed the trend we had identified for larger accent rating differences between pretest and posttest to occur under two of the four viewing conditions, when learners were exposed to the TV series with captions and were focusing on meaning ($t(516) = 1.57$, $p = .080$), and when they were exposed to it without captions, but were focusing on phonetic form ($t(516) = 1.58$, $p = .068$).

## DISCUSSION

The present study set out to investigate whether extensive exposure to multimodal input in the form of L2 captioned videos benefits L2 pronunciation skills. Previous research has claimed a beneficial effect of such exposure on second language listening comprehension and speech segmentation, reporting a perceptual advantage for a viewing mode with captions over a viewing mode without captions (Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015) or native language subtitles (Mitterer & McQueen, 2009). Based on such findings Mitterer and McQueen (2009) suggested that exposure to captioned video could benefit lexically guided perceptual learning and would consequently lead to the retuning of phonetic categories in L2 speech perception. The present study tested whether this was in fact the case through a set of L2 speech processing tasks and L2 phonological accuracy measures.

The main findings of the current study indicate that extensive viewing of TV series in the L2 enhances learners' speech processing skills. It enhanced not only the ability to segment speech, as already shown in previous research (Charles & Trenkic, 2015; Mitterer & McQueen, 2009), but also the ability to process speech faster and more efficiently, as suggested by the results of the *sentence verification task.* Thus, not only did speech segmentation skills improve but also benefits were visible at the level of sentence processing in a task that did not use materials related to the treatment. Such benefits in L2 speech processing were found for both the experimental groups exposed to captioned video as well as those exposed to uncaptioned video, with slightly larger gains for the uncaptioned viewing groups. This may be interpreted as an instance of transfer-appropriate processing (Lightbown, 2008), as the sentence verification task was performed auditorily only (without a visual orthographic representation), which is essentially the same type of exposure the participants exposed to uncaptioned videos received.

A similar interpretation is also possible for the results of the *Animacy Judgment* task, where learners assigned to captioned video groups obtained slightly larger gains in response latency. However, the small gains in speed of lexical access for some of the groups may be considered negligible overall, whereas more robust gains in L2 speech processing were observable at the sentence level. Unlike what previous research had

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

shown, participants exposed to both captioned and uncaptioned video improved their L2 speech processing skills. This might be explained by differences in the implementation of the treatment and the testing procedures in our study and those of previous research. In previous studies, testing was always administered immediately after each viewing session, and it was under these testing circumstances that the groups exposed to captioned video showed an advantage over those exposed to uncaptioned video or L1 subtitles. In contrast, our participants were tested once at pretest and once at posttest, after a period of 8 weeks of exposure, during which they had enough time to become familiar with the voices and speaking styles in the TV series. For instance, in Charles and Trenkic's (2015) study, the audiovisual materials were four unrelated documentaries each narrated by a different speaker. The clear advantage of captioned video over uncaptioned video for speech segmentation and listening comprehension found in their study may be due to the fact that participants were tested immediately after each viewing session and on unfamiliar materials. In contrast, in our study, extensive exposure to 16 excerpts from the same TV series is likely to have resulted in familiarity to speakers' voices, accent and speaking style, which may have mitigated group differences by helping participants in the uncaptioned viewing mode segment speech more efficiently. Additionally, the relatively high level of proficiency of our participants might also explain the lack of group differences between the two viewing modes.

Similarly, the viewing mode did not have large effects on phonological accuracy in perception, as assessed through the categorical discrimination of a difficult L2 sound contrast (ABX). In fact, gains in sensitivity to the contrasting L2 sounds we tested were inconsistent in that only the group exposed to captions in the meaning-focused condition improved to a larger extent than the control group did. Given the lack of a *Time* × *Viewing Mode* interaction our data cannot provide evidence of lexically guided perceptual learning having taken place to the extent of causing the retuning of phonetic categories in speech perception, as predicted by Mitterer and McQueen (2009) based on the interpretation of gains in their shadowing task. The consequences of lexically guided perceptual learning for the updating of phonetic categories and phono-lexical representation in L2 phonology remains an empirical question future research should address. In addition, this task focused exclusively on a single phonological dimension (accuracy and speed in the categorical discrimination of the /iː/-/ɪ/ contrast) and could not have possibly captured benefits for any other perceptual dimension beyond categorical perception. A more holistic measure of L2 perception might have proved more appropriate to capture potential treatment effects.

For phonological accuracy in production, a trend was observed in the analysis of L2 learners' accent ratings. Greater accent reduction occurred in the uncaptioned viewing mode in the pronunciation-focused condition and in the captioned viewing mode in the meaning-focused condition. These group differences suggest that benefits in pronunciation may occur under both viewing modes, but under different attention-directing conditions. Captioning appeared not to be effective in reducing learners' accent if combined with a focus on phonetic form, whereas uncaptioned video was effective at reducing accent if attention was being oriented toward phonetic form. This suggests that focusing on phonetic form while being exposed to captions might have been cognitively overloading (Mattys & Wiget, 2011), as the task of reading dynamic on-screen text in captioned video already requires high attentional demands and participants in the captioned viewing condition focusing on phonetic form had to deal with caption reading

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

while focusing on the pronunciation of words in the auditory input. Thus, it is possible that in a captioned video mode a focus on pronunciation did not allow learners to devote enough attentional resources to the processing of the auditory input, whereas in an uncaptioned video mode a focus on phonetic form was most effective in reducing accentedness. In the absence of dynamic text appearing on the screen, it would seem that learners were better able to focus on phonetic form and dedicate their attentional resources to the processing of the auditory input.

Several methodological issues need be taken into account when interpreting the results. Given the cognitive complexity of managing captions and aural input, benefits on L2 pronunciation may be moderated by cognitive individual differences in learners' ability to exploit multimodal input, as well as learners' motivation and interest to learn from this type of language exposure. Assessing individual differences in subtitle processing (Kruger & Steyn, 2014) is a necessary step to gain better understanding of how multimodal exposure benefits pronunciation gains. In addition, one of the limitations of the research design we implemented in this study is that only two out of the four experimental groups, the groups prompted with questions on meaning, were tested on comprehension. Thus, we could not ascertain to what extent including a focus on phonetic form through questions on pronunciation might have affected processing for meaning.

## CONCLUSION

The present study investigated the effect of exposure to L2 audiovisual materials on L2 pronunciation under different viewing modes and task focus conditions. The design of the experiment extended previous research by testing both L2 speech processing and L2 phonological accuracy in speech perception and production. The results revealed that L2 learners with a relatively advanced proficiency level improved their L2 speech processing skills by watching movies in the target language either with or without captions, providing empirical evidence that the viewing treatment served as a kind of training in L2 speech processing. In L2 production, a measure of accentedness revealed an interplay between the viewing modes and the task focus conditions to which L2 learners were exposed. Specifically, a focus on phonetic form resulted in pronunciation gains in the uncaptioned viewing mode, but not when videos were watched with captions, whereas in a viewing mode with captions pronunciation benefits were only visible when attention was directed to the plot. However, the asymmetry in gains might also be due to the relatively small amount of audiovisual input received (16 20-minute videos = 5.3 exposure hours), in line with the small overall amount of L2 input learners typically receive in a EFL context, compared to the much larger amounts of exposure learners might receive from watching captioned video in an ESL context.

Previous research has shown that audiovisual exposure can lead to incidental vocabulary learning, especially when watching captioned videos (e.g., Peters & Webb, 2018). For L2 pronunciation, our results suggest that incidental learning might occur only in the presence of captions. No evidence of incidental accent reduction for learners viewing uncaptioned videos was observed, which implies that, given the relatively small amount of input provided (5.3 hours), some kind of intentional focus on phonetic form while paying attention to the speech in the soundtrack is necessary for pronunciation to develop through exposure to audiovisual materials in an EFL context. Thus, in the context of the experimental time frame implemented in this study, benefits on L2 pronunciation

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

appeared to depend on the presence of an intentional focus on pronunciation, but it seems plausible that a longer or more intense exposure to audiovisual materials could lead to greater positive outcome of the treatment.

The use of audiovisual materials may serve as a means of enriching the limited language input in instructional settings for L2 pronunciation development. Future research should address the question of the effectiveness of viewing modes for improving L2 pronunciation further, especially comparing various L2 proficiency learner groups in both foreign and second language learning contexts.

## NOTES

[1]We use the terms "bimodal" and "multimodal" interchangeably. They both refer to exposure to language input presented through more than one modality simultaneously as it is the case of captioned videos, where the language presentation is auditory (speech) and visual (written text and image).

[2]Shapiro-Wilk normality tests (statistic > .9, $p > .05$) as well as visual inspection of Q-Q plots and histograms confirmed that the data was normally distributed.

[3]Nonparametric Kruskal-Wallis tests were used because both Shapiro-Wilk tests (all $ps < .05$) and visual inspection of Q-Q plots and histograms indicated lack of normality of the viewing habits measures.

## REFERENCES

Best, C., & Tyler, M. (2007). Non-native and second language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 15–34). John Benjamins.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PloS One*, *11*, e0158409.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within language subtitling. *Applied Psycholinguistics*, *23*, 509–533.

Bisson, M. J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 399–418.

Borrás, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, *78*, 61–75.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*, 1205–1224.

Charles, T. J., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bimodal input. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning: Principles, strategies and practical experiences* (pp. 173–198). Peter Lang.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, *70*, 604–618.

d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, *28*, 227–244.

Danan, M. (2004). Captioning and subtitling, undervalued language learning strategies. *Meta*, *49*, 67–77.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: an ESL teachers' pronunciation "toolbox." In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Iowa State University.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*, 741–773.

Flege, J. (2003). Methods for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–44). UniPress.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, *12*, 91–114.

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (Eds.) (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? [Special issue]. *Studies in Second Language Acquisition*, *39*.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, *97*, 720–744.

Kruger, J. L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, *7*. https://doi.org/10.16910/jemr.7.5.4.

Kruger, J.L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, *49*, 105–120.

Kruger, J. L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: An overview of eye tracking studies. *Refractory: A Journal of Entertainment Media*, *25*, 1–14.

Kruger, J. L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology*, *32*, 19–31.

Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 369–377.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z.-H. Han (Ed.), *Understanding second language process* (pp. 27–44). Multilingual Matters.

Markham, P. (2001). The influence of culture-specific background knowledge and captions on second language comprehension. *Journal of Educational Technology Systems*, *29*, 331–343.

Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). Routledge.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*, 145–160.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, *79*, 344–351.

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

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720–739.

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*, 381–399.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*, 578–596.

Muñoz, C. (2017). The role of age and proficiency in subtitle reading.An eye-tracking study. *System*, *67*, 77–86.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Nation, P. (2012). The BNC/COCA Word Family Lists. Unpublished manuscript. www.victoria.ac.nz/lals/about/staff/paul-nation.

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes*, *16*, 637–660.

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research.* Paper presented at the Second Language Research Forum, Toronto, Ontario, Canada.

Paivio, A. (1986). *Mental representations, a dual coding approach*. Oxford University Press.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577.

Peters, E., Heynen, E., & Puimege, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148.

Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 301–314). John Benjamins.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (2010). A psycholinguistic perspective on the acquisition of phonology. In C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallée (Eds.),

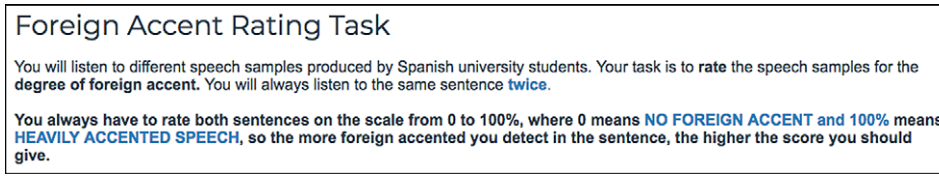
https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

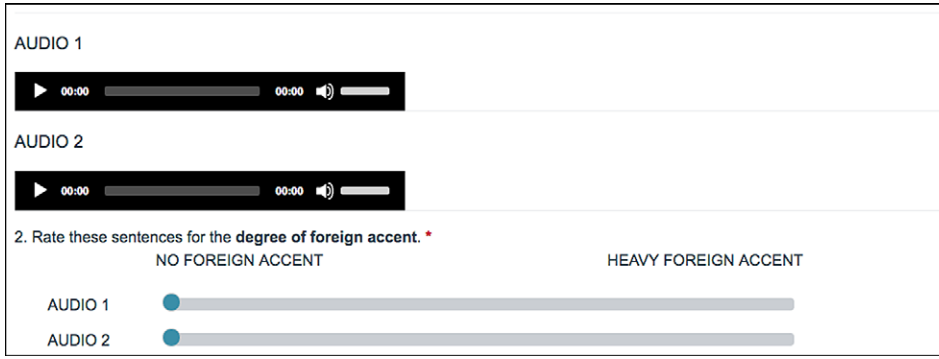
*Laboratory phonology 10: Variation, phonetic detail and phonological representation* (pp. 311–340). Mouton de Gruyter.
Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Unpublished PhD thesis). Victoria University, Wellington.
Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35*, 1–29.
Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics*, *36*, 377–409.
Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, *33*, 644–653.
Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*, 326–344.
Vanderplank, R. (1988). The value of teletext subtitles in language learning. *ELT Journal*, *42*, 272–281.
Vanderplank, R. (2010). Déjà vu?A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*, 1–37.
Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave Macmillan.
Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168).Routledge.
Webb, S., & Rodgers, M. (2009). Vocabulary demands of television programs. *Language Learning*, *59*, 335–366.
Wisniewska, N., & Mora, J. C. (2018). Pronunciation learning trough captioned videos. In *Proceedings of the 9th Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 204–215). Iowa State University.
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*, 497–528.

# APPENDICES

## APPENDIX A

## TASK INSTRUCTIONS




https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

## APPENDIX B

## EXAMPLES OF TASK FOCUS QUESTIONS

| Focus on Phonetic Form | Focus on Meaning |
|---|---|
| Is the word **scene** pronounced like **seen** or **sin**? Pick the right answer.<br>**a) seen**<br>b) sin<br>c) both | What made Luther think the suspect might be a taxi driver?<br>a) All the victims trusted him by default.<br>b) The suspect has a very good knowledge of the city.<br>c) He drives a white van. |
| The word **tough** rhymes with…<br>**a) stuff**<br>b) spot<br>c) hot | Why didn't Linda report her husband earlier?<br>a) Because she felt guilty and embarrassed<br>b) Because her friends laughed at her<br>c) Because her husband threatened he would kill himself |
| Which of the following words **DOES NOT** have the same vowel as in word M**U**RDER?<br>a) first<br>b) birth<br>c) hurt<br>**d) guard** | What kind of "justice" is Owen Lynch seeking?<br>a) He wants his dad Terry Lynch to be released from prison<br>b) All he wants is that his dad receives respect because he served in the army<br>c) He and his dad seek respect because he served in the army |
| The ending of the word **nature** is pronounced similar to…. Select 3 words that are pronounced similar.<br>**a) gesture**<br>**b) literature**<br>**c) picture**<br>d) injure<br>e) exposure<br>f) figure | What is the profile of the suspect according to Luther?<br>a) A soldier or an ex-soldier who has just came back from war and has depression<br>b) A soldier or an ex-soldier who is targeting police officers as his victims<br>c) A soldier or an ex-soldier who has a mental health problems |
| The sound in p**u**nish is the same as in…<br>a) v**a**nish<br>**b) m**o**ney**<br>c) p**u**rse | |

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

## APPENDIX C

## PARAMETER ESTIMATES OF LINEAR MIXED MODEL ANALYSES

TABLE 1. Results for the analysis of treatment effects

| | | *β* | *SE* | *t* | *95% CI* |
|---|---|---|---|---|---|
| *Shadowing* | *Intercept* | 72.231 | 17.07 | 4.23 | 36–107 |
| | *Time* | –4.200 | 2.22 | –1.89 | –8.56–0.16 |
| | *Viewing (Captions)* | 15.085 | 5.42 | 2.78 | 4.32–25.85 |
| | *Viewing (No Captions)* | 12.393 | 5.42 | 2.29 | 1.63–23.15 |
| | *Time × Viewing (Captions)* | –8.951 | 2.56 | –3.50 | –13.97–3.93 |
| | *Time × Viewing (No Captions)* | –8.881 | 2.56 | –3.47 | –13.90–3.86 |
| | *MaxScore* | –2.038 | 1.92 | –1.06 | –6.08–2.00 |
| *Animacy Judgment* | *Intercept* | 749.42 | 42.08 | 17.81 | 666–832 |
| | *Time* | 20.80 | 7.15 | 2.91 | 6.78–34.81 |
| | *Viewing (Captions)* | –2.19 | 34.97 | –0.06 | –71.69–67.32 |
| | *Viewing (No Captions)* | –13.58 | 34.85 | –0.39 | –82.85–55.69 |
| | *Time × Viewing (Captions)* | 2.41 | 8.32 | 0.29 | –13.90–18.72 |
| | *Time × Viewing (No Captions)* | –9.69 | 8.30 | –1.17 | –25.97–6.59 |
| *Sentence Verification* | *Intercept* | 2133.76 | 49.83 | 42.82 | 1795–2471 |
| | *Time* | 77.83 | 15.20 | 5.12 | 48.03–107.62 |
| | *Viewing (Captions)* | 31.28 | 45.41 | 0.69 | –58.93–121.48 |
| | *Viewing (No Captions)* | –29.87 | 45.25 | –0.66 | –119.75–60.02 |
| | *Time × Viewing (Captions)* | 23.31 | 17.65 | 1.32 | –11.29–57.91 |
| | *Time × Viewing (No Captions)* | 46.47 | 17.55 | 2.65 | 12.06–80.88 |

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press

TABLE 2. Results for the analysis of experimentalcondition effects (*Viewing* and *Task Focus*) on L2 phonology measures

| | | *β* | *SE* | *t* | *95% CI* |
|---|---|---|---|---|---|
| *ABX (accuracy)* | *Intercept* | 0.557 | 0.14 | 3.96 | 0.28–0.83 |
| | *Viewing (Captions)* | 0.031 | 0.18 | 0.17 | –0.33–0.39 |
| | *Viewing (No Captions)* | 0.094 | 0.18 | 0.52 | –0.26–0.45 |
| | *Task Focus* | 0.217 | 0.17 | 1.31 | –0.11–0.54 |
| | *Time* | –0.156 | 0.14 | –1.09 | –0.44–0.13 |
| | *Viewing (Captions) × Task Focus* | –0.054 | 0.24 | –0.23 | –0.52–0.41 |
| | *Viewing (Captions) × Time* | –0.084 | 0.19 | –0.45 | –0.45–0.28 |
| | *Viewing (No Captions) × Time* | 0.230 | 0.19 | 1.23 | –0.14–0.60 |
| | *Task Focus × Time* | –0.243 | 0.17 | –1.41 | –0.58–0.10 |
| | *Viewing (Captions) × Task Focus × Time* | 0.396 | 0.24 | 1.62 | –0.08–0.88 |
| *ABX (RT)* | *Intercept* | 7.096 | 0.05 | 146.29 | 7.00–7.19 |
| | *Viewing (Captions)* | –0.024 | 0.06 | –0.39 | –0.15–0.10 |
| | *Viewing (No Captions)* | –0.111 | 0.06 | –1.79 | –0.23–0.01 |
| | *Task Focus* | 0.12 | 0.06 | 2.16 | 0.01–0.23 |
| | *Time* | 0.043 | 0.02 | 2.28 | 0.01–0.08 |
| | *Viewing (Captions) × Task Focus* | –0.104 | 0.08 | –1.32 | –0.26–0.05 |
| | *Viewing (Captions) × Time* | 0.048 | 0.03 | 1.94 | 0.00–0.10 |
| | *Viewing (No Captions) × Time* | 0.039 | 0.02 | 1.61 | –0.01–0.09 |
| | *Task Focus × Time* | –0.043 | 0.02 | –2.03 | –0.09–0.00 |
| | *Viewing (Captions) × Task Focus × Time* | 0.012 | 0.03 | 0.38 | –0.05–0.07 |
| *Accent Ratings* | *Intercept* | 56.40 | 4.09 | 13.79 | 48–64 |
| | *Viewing (Captions)* | –4.88 | 5.22 | –0.93 | –15–5 |
| | *Viewing (No Captions)* | –2.07 | 5.22 | –0.40 | –12–8 |
| | *Task Focus* | –3.88 | 4.60 | –0.84 | –13–5 |
| | *Time* | 1.98 | 2.09 | 0.95 | –2–6 |
| | *Viewing (Captions) × Task Focus* | 4.74 | 6.55 | 0.72 | –8–18 |
| | *Viewing (Captions) × Time* | 0.78 | 2.67 | 0.29 | –4–6 |
| | *Viewing (No Captions) × Time* | –1.41 | 2.67 | –0.53 | –7–4 |
| | *Task Focus × Time* | 2.33 | 2.36 | 0.99 | –2–7 |
| | *Viewing (Captions) × Task Focus × Time* | –4.50 | 3.35 | –1.34 | –11–2 |

https://doi.org/10.1017/S0272263120000029 Published online by Cambridge University Press