UNIVERSITAT DE BARCELONA

Pronunciation learning through captioned videos:
Gains in L2 speech perception and production

Natalia Wisniewska

UNIVERSITAT DE BARCELONA

Pronunciation learning through captioned videos: Gains in L2 speech perception and production

Natalia Wisniewska

Supervised by Dr Joan Carles Mora

2020
Barcelona

Submitted in fulfillment of the requirements for the degree of Doctor in Linguistic, Literary and Cultural Studies

---

Abstract

This doctoral dissertation examines the benefits of extended exposure to multimodal input through captioned videos for second language pronunciation development. It investigates the effects of TV viewing on both L2 speech processing and perceptual and productive accuracy in learners' L2 phonology. It also analyses learners' eye-movements while watching captioned videos in order to relate the amount of on-screen text processing to pronunciation gains. Ninety Spanish/Catalan adult learners of English (EFL) were tested 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). The results revealed that exposure to authentic audiovisual materials in English can benefit the L2 pronunciation of post-intermediate/advanced EFL learners irrespective of viewing mode. Whereas previous studies had found larger benefits for captioned than uncaptioned viewing for speech processing and segmentation after short (e.g. single-session) exposures, the relatively long exposure treatment administered in the current study and inclusion of a form or a meaning-focused condition might have washed away potential advantages of one viewing mode over another. Viewing treatment benefits on L2 speech processing were larger on tasks assessing sentence-level than word- or segment-level gains. No significant benefits were found for phonological accuracy in perception. In production, the results revealed an interplay

between viewing mode and task focus, indicating that 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. Cognitive overload might explain why no benefits were obtained when attention was directed to pronunciation in a captioned viewing mode. Although large individual differences characterized L2 learners' caption reading behaviour, which was influenced by material-related as well as learner-related factors, the results indicated a relationship between amount of subtitle processing and foreign accent reduction. Viewing mode moderated foreign accent reduction, as incidental learning of pronunciation occurred only through exposure to captioned viewing without a focus on pronunciation, whereas in the absence of captions gains were driven by an intentional focus on pronunciation. Taken as a whole, this dissertation suggests that enriching the limited L2 input learners are exposed to in foreign language settings through the viewing of authentic audiovisual materials in target language may enhance L2 pronunciation development.

Resumen

Esta tesis doctoral examina los beneficios de la exposición prolongada a programas de televisión en inglés con subtítulos para mejorar la pronunciación. La tesis investiga los efectos de la exposición multimodal tanto en el procesamiento del habla como en la corrección fonético-fonológica a nivel perceptivo y productivo del inglés como lengua extranjera. Con este propósito, el estudio también analiza los movimientos oculares de los estudiantes mientras visualizan videos subtitulados y poder así relacionar la cantidad del texto procesado en los subtítulos con posibles mejoras en la percepción y producción fonético-fonológica. Noventa estudiantes universitarios, estudiantes de inglés como segunda lengua, fueron evaluados respecto de sus habilidades para el procesamiento del habla en inglés (segmentación, velocidad de acceso léxico y procesamiento del habla en frases) y corrección fonológica en la percepción (discriminación ABX) y producción (nivel perceptivo de acento extranjero) antes y después de un tratamiento de 8 semanas de exposición regular a materiales audiovisuales (películas de video). Los participantes fueron asignados aleatoriamente a cuatro condiciones experimentales que implicaban dos modos de visualización (con subtítulos o sin subtítulos) y dos condiciones de enfoque de la tarea de visualización de los vídeos (un enfoque en el que se induce una atención a la forma fonética de las palabras u otro en el que se induce una atención al significado). Los resultados indican que la exposición a materiales audiovisuales auténticos en inglés puede beneficiar el desarrollo de la pronunciación en los estudiantes de inglés de nivel relativamente avanzado. Los resultados muestran que los dos modos de visualización (con o sin subtítulos) pueden beneficiar el procesamiento del habla y que la duración de la

exposición puede mitigar la ventaja de la exposición subtitulada sobre la no subtitulada. Las mejoras después del tratamiento audiovisual fueron mucho más pronunciadas a nivel de frases que a nivel de palabras o segmentos. No se encontraron beneficios significativos en la corrección fonológica perceptiva. En producción, los resultados indicaron una relación entre el modo de visualización y el enfoque de la tarea, de tal manera que en la condición de atención a la forma fonética la pronunciación mejoró únicamente en ausencia de subtítulos, mientras que la condición de visualización con subtítulos condujo a la mejora en la pronunciación únicamente en la condición donde no se indujo la atención a la forma fonética. Aunque la cantidad de texto procesado de los subtítulos está sujeta a considerables diferencias individuales y se vio influenciada por factores relacionados tanto con el material audiovisual como con diferencias cognitivas individuales de los participantes, los resultados indicaron una relación entre la cantidad de texto procesado en los subtítulos y la reducción del acento extranjero condicionada por la modalidad de visualización. El aprendizaje incidental de la pronunciación a través del visionado de vídeos en este estudio tuvo lugar únicamente mediante una exposición subtitulada y sin atención a la forma fonológica, mientras que, la atención a la forma fonológica parece ser beneficiosa para la pronunciación en un visionado sin subtítulos. En su conjunto, esta tesis demuestra que la visualización de materiales audiovisuales auténticos en la lengua extranjera puede, además de facilitar un mayor contacto con la lengua inglesa en contextos de instrucción sometidos a un input lingüístico mínimo, mejorar el desarrollo de la pronunciación en una segunda lengua.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Dr. Joan Carles Mora, who has been the most supportive and encouraging supervisor a PhD candidate could have ever asked for. I am grateful not only for his excellent academic guidance and feedback, but even more so for opening the doors to numerous learning opportunities, which are the hallmarks of a successful and memorable PhD journey. My academic path would simply not be the same if it wasn't for his teaching, gràcies.

I would also like to thank Dr. Carme Muñoz and the GRAL Research Group for the opportunity given to collaborate and learn along their side and for enabling my further development as a researcher, especially by supporting my research stay at the Macquarie University in Australia. I also thank friends and colleagues from the English Linguistics Department at the University of Barcelona, especially Dr. Georgia Pujadas Jorba, for sharing their experience and some useful tips, which often made the navigation through the PhD journey much smoother.

I would like to extend my thanks to Dr. Jan-Louis Kruger, who accepted my visit at the Macquarie University and guided me through an excellent research stay, which allowed me to expand my knowledge on eye-tracking methodology and reading in multimodal contexts. It was indeed a perfect closure to the PhD experience, for which I am truly very grateful.

Big thanks to all the good friends, some of whom have been with me from the very beginning of the journey - Valentine and Ola - and some of whom I have met on the way in different parts of the world, especially Maurice, Debbie, Ian and Charlie, but

also others encountered at the final stage of writing in a backpackers' hostel, who would always throw a cheering and encouraging comment, thank you.

Special *dziękuję* (Polish "thank you") to my family, Marta, Wojtek, Babcia and my brother Igi, who have always been there for me, supported my ups and downs and always encouraged me to push forward, or sometimes even pushed for me, not only during the thesis writing, but ever since I first left Warsaw. If it wasn't for your constant support, moral and beyond, none of this crazy nomad life, academic progress nor personal growth would have happened. Thank you.

I would like to end with one more sincere thank you to my supervisor Joan Carles without whom the last five years in academia wouldn't have been as rewarding, thank you for all the support.

Table of contents

List of figures

List of tables

Chapter 1. Introduction

When learning a foreign language, one of the most challenging tasks is mastering its pronunciation. Any native listener without specialized linguistic knowledge can immediately label the speech of even highly proficient second language speakers as *foreign*, no matter how short the utterance (Flege, 1984). Despite the common belief that the ideal ultimate attainment stage in L2 pronunciation is to speak with a native-like English accent (Derwing, 2003; Scales, Wennerstrom, Richard &Wu, 2006; Timmis, 2002), pronunciation teaching approaches have changed towards prioritizing intelligibility and comprehensibility over nativelikeness (Murphy, 2014). Even though the native speaker model is not always treated as the epitome of language performance, classroom settings rarely offer optimal conditions for L2 pronunciation development, even if intelligibility and comprehensibility of speech (and not nativelikeness) are the ultimate goal. For many learners second language learning commonly takes place in a formal school setting as part of the compulsory curriculum where the pronunciation acquisition path too often seems to lag behind other linguistic dimensions such as vocabulary or grammar. Some of most important factors responsible for such state of affairs are: very limited amount of exposure (Muñoz, 2008; Piske, 2007), input quality (Foote, Holtby & Derwing, 2011) and the difficulty of integrating a pronunciation focus within a communicative teaching approach (Gurzynski-Weiss, Long & Solon, 2017; Mora & Levkina, 2017; Darcy, 2018; Darcy, Ewert & Lidster, 2012).

The idea of any sort of learning needs to be being constrained by school walls does no longer hold true. The way people access information has become limitless, and thus, has changed how and where we learn, with television and multimodal media exposure becoming an integral part of people's lives (Webb, 2015). Statistics on the Internet usage show that the digital population has reached 4.33 billion active users, with North America and Northern Europe both ranking first with a 95 percent Internet penetration rate among the population (Clement, 2019). The value of television for second language learning resides in its potential to provide large amounts of exposure to rich and authentic input, optimizing the conditions for improvement while offering an engaging and often self-selected content. Aided by easy access to technology, out-of-classroom sources of exposure have become a common and abundant supply of contact with the target language, especially when the target second language is English.

For the last decade, SLA researchers have been investigating the potential benefits of audiovisual sources of input for second language learning (Vanderplank, 2010, 2016) by examining its effects on a variety of language learning outcomes, especially listening comprehension (Kruger & Steyn, 2014; Markham, 2001; Montero Perez, Van Den Noortgate & Desmet, 2013; Vanderplank, 1988) and the acquisition of vocabulary (Bird & Williams, 2002; Bisson et al., 2014; d'Ydewalle & Van de Poel, 1999; Peters & Webb, 2018). A primary aim of this research has been to identify the most optimal viewing modes (Peters, Heynen & Puimege, 2016) by comparing the differential benefits of L1 subtitles and L2 subtitles (captions) for language learning and how viewing modes might interact with learners' individual differences in age and proficiency (Muñoz, 2017). This line of research is grounded in the notion of bimodal reinforcement from Paivio's (1986) Dual Coding Theory and Mayer's (2001) principles

of multimedia learning, which claim that the dual processing of auditory and visual information promotes learning by strengthening mental representations of perceived objects. Despite many consolidated findings, research on captioned videos has not yet examined its potential benefits for L2 pronunciation learning.

This doctoral thesis sets out to explore whether extended[1] exposure to multimodal input through captioned video can benefit L2 pronunciation development[2]. The general aim is to investigate the potential benefits of extended exposure to L2 captioned videos for L2 pronunciation by examining both L2 speech processing skills and perceptual and productive accuracy in learners' L2 phonology. The study involves a pre-/post-test design and it consists of an 8-week viewing treatment during which experimental participants (L1 Spanish/Catalan adult learners of English with advanced proficiency) were regularly exposed to audiovisual materials under two different viewing modes (viewing with or without L2 captions) and task focus conditions (inducing focus either on pronunciation or plot comprehension). The study also aims to investigate learners' eye-movements while watching captioned videos in order to trace changes in subtitle reading behaviour and relate individual differences in the amount of on-screen text processing to pronunciation gains.

The dissertation starts with a comprehensive literature review of the theoretical underpinnings of the relevant research areas concerning L2 pronunciation teaching and learning within and beyond classroom settings and the benefits of multimodal input for L2 pronunciation development (Chapter 2), followed by a brief rationale and the presentation of the research questions (Chapter 3). The chapter on methodology (Chapter 4) provides an overview of the study design, including a detailed description of the viewing treatment and the testing instruments with justification of their selection

as well as the procedures followed during the three-month treatment. The following chapter (Chapter 5) presents the results of the study and is divided into four subsections. Preceded by the viewing treatment overview and the analysis of treatment accuracy scores (5.1 and 5.1.1.), section 5.2 assesses the effect of the treatment on L2 speech processing skills under different viewing modes, followed by section 5.3, which assesses the effect of the treatment on L2 phonological accuracy as a function of task focus condition. Section 5.4 provides the analysis of learners' eye-gaze behaviour while watching L2 captioned videos and discusses pre-/post-treatment changes as well as examines the relationship between individual differences in on-screen text processing and pronunciation gains. Lastly, section 5.5 examines the role of individual differences in proficiency, attention control and phonological memory in L2 pronunciation learning through audiovisual exposure. Chapter 6 brings together the results of this dissertation by presenting the summary of the main findings and offers a thorough discussion. Chapter 7 acknowledges the limitations and offers suggestions for further research, which is followed by the final remarks and conclusions (Chapter 8).

Chapter 2. Literature review

This chapter presents the diverse theoretical and empirical sources of this dissertation. Starting off with the description of the challenges present in the classroom context for second language pronunciation development, it delves into the issue of the benefits of multimodal exposure, focusing specifically on audiovisual sources of language input and its relevance for second language pronunciation learning. This chapter offers an overview of the theoretical underpinnings underlying the benefits of multimodality, providing a comprehensive overview of the most relevant empirical studies in the area. It also discusses several factors that have been shown to play a mediating role in language learning through audiovisual input.

2.1 L2 pronunciation learning: classroom and beyond

It is a well-established fact that learning contexts (e.g. naturalistic or classroom settings, study abroad, short-term immersion in FL contexts) may affect L2 pronunciation learning, just as it may affect any other domains in language (e.g. grammar and vocabulary) (e.g., Collentine & Freed, 2004; Mora, 2008). In light of this dissertation, understanding how naturalist and classroom contexts impact the learning process is important both from a theoretical and applied perspective. From the theoretical standpoint, is it crucial for understanding the role that input plays in the acquisition of a second language, whereas the applied purpose is important for the relevance and efficiency of particular kinds of language instruction (Ellis, 1990). In simple terms, *naturalistic second language learning* can be characterized as learning through

immersion in the second language environment, whereas *foreign language learning* is characterized by learning in the classroom through formal language instruction. Throughout this dissertation, the terms SLA (Second Language Acquisition) or ESL (English as a Second Language) will be used to refer to both contexts of learning, and, when needed, additional differentiating terms will be introduced to identify the learning, which occurs exclusively within the classroom context (FL or EFL). SLA research has provided a lot of solid evidence for multiple factors having an impact on L2 speech acquisition. The most frequently discussed factors are L1 influence or the extent to which the L1 and the L2 differ phonologically (e.g. Flege, Bohn, & Jang, 1997), age of onset of L2 learning or first extensive exposure to the L2 (Munro, Flege, & MacKay, 1996), length of L2 exposure (Aoyama & Flege, 2011), and input quality and quantity (Flege, 2009), amongst others. Although many studies suggest that the later the language is learned, the lower the probability of high achievement in general proficiency (DeKeyser, 2000; Flege, Yeni-Komshian, & Liu, 1999) and, consequently, pronunciation, some individuals do not comply with this pattern (Birdsong, 1992; Ioup et al., 1994) and phonology is one of the language domains where large variability in learning outcomes can be observed (Bongaerts et al., 1997; Moyer, 1999, 2014). Moyer (2014), for instance, identifies three sets of factors that help explain exceptional outcomes in L2 pronunciation, namely, cognitive factors (language learning aptitude and talent), psycho-social factors (e.g. willingness to sound native, identification with the L2 culture, extroversion) as well as experiential factors. Age- and experience-related factors constitute the most widely investigated source of individual differences in L2 speech learning, Recent studies investigating the role of L2 experience in terms of in- and out-of-classroom exposure has found that input factors are more essential to the

development of L2 oral skills than starting age (Muñoz, 2014). For instance, variables related to L2 use such as the extent to which students avail themselves of out-of-classroom sources of L2 input have been found to affect L2 speech performance and development (Saito & Hanzawa, 2006), which emphasizes the importance of exposure to the L2 outside the classroom context.

Within classroom contexts predictors of successful phonological acquisition in immersion settings (e.g. an early start in L2 learning and frequency and amount of L2 use) do not appear to contribute much to explaining variance in phonological attainment among foreign language (FL) students (Cebrian, 2006; Fullana, 2006; Gallardo Del Puerto, García Lecumberri & Cenoz, 2006; Mora & Fullana, 2007), even after when FL students are immersed in the L2 environment during short stay abroad periods (Díaz-Campos, 2004; Mora, 2008, 2014; Muñoz & Llanes, 2014). This suggests that language learning in classroom contexts lacks the L2 use and exposure conditions for substantial development in oral skills and L2 phonological acquisition to take place (Muñoz, 2008). Providing opportunities for L2 pronunciation training and production practice could thus induce gains in speaking performance both outside the FL classroom (Saito & Hanzawa, 2016) as well as inside the FL classroom through pronunciation instruction (Burgess & Spencer, 2000; Darcy, Ewert, & Lidster, 2012; Sicola & Darcy, 2015).

For many learners second language learning commonly takes place in a formal school setting. Despite numerous aspects that can differentiate classroom environments (e.g. native vs. non-native teacher, nature of instruction, number of students), some are homogenized through educational policies such as the amount of hours of formal instructions per year. In Spain, similarly to other European countries, high school students receive on average 702 hours of formal instruction per academic year, from

which approximately 11% (77 hours) is typically dedicated to formal second language teaching (OECD, 2018). These numbers imply that, within formal school settings, students are exposed only to 2 to 4 hours of second language instruction weekly. The very limited amount of weekly exposure constitutes only the first challenge (e.g. Muñoz, 2008; Piske, 2007).

Within the classroom context, input is limited not only in terms of its quantity, but also quality, as most students are exposed to accented speech spoken by their peers as well as their teachers, who often lack specific pronunciation training (e.g. Foote, Holtby & Derwing, 2011). The results from an on-line survey of English pronunciation teaching practices comparing data from seven European countries (Henderson et al., 2012) revealed that Spanish teachers gave greater importance to pronunciation teaching than other language skills, but recognized the insufficient time and resources spent on it, advocating the need for specific training programmes. When asked about the causes of the lack of pronunciation focus in the classroom, teachers report implementation difficulties and the incompatibility of pronunciation teaching with curricular demands that do not prioritize oral skills. The results of this survey highlight that what makes pronunciation improvement particularly difficult as opposed to other language dimensions (e.g. vocabulary, grammar) is the fact that it does not often receive sufficient attention in the classroom (Thomson & Derwing, 2014). This is partly due to the lack of clear pedagogical guidelines and coherent goals (Darcy, Ewert & Lidster, 2012) and the shift in methodological focus from audiolingualism to communicative approaches to language teaching prioritizing attention to meaning and interaction (Levis, 2005).

Lastly, SLA research has provided substantial, empirical evidence showing that L2 pronunciation improvement is possible thanks to diverse types of intensive training, typically computer-mediated and taking place in a laboratory environment under the supervision of experts in the field of phonetics and phonology. For instance, High Variability Phonetic Training (HVPT) has proved to be effective in enhancing L2 learners' perceptive and productive abilities even after a short training period (e.g. Aliaga-García & Mora, 2009; Barriuso & Hayes-Harb, 2018; Rato et al., 2015; Thomson, 2018). Despite the proven effectiveness of lab training such as HVPT (*see* Thomson, 2012), the practicality of implementing such type of phonetic training within classroom settings is problematic, as it usually needs to be administered individually on a computer or other device and requires familiarity with certain experimental software and, more importantly, phonological expertise not all EFL teachers possess.

Pronunciation training in FL context is very challenging because of limited quantity of quality input and because it is hard to integrate a focus on pronunciation into current task-based communicative approaches to language teaching (Darcy, Ewert & Lidster, 2012Gurzynski-Weiss, Long & Solon, 2017; Mora & Levkina, 2017). Within classroom settings, where the communicative teaching framework is dominant, overdependence on decontextualized practice (Saito, 2012), lack of meaningful purpose for pronunciation feedback together with the scarcity of its occurrences (Saito, 2015) may hinder pronunciation development in a second language.

In contrast with classroom settings, learning in contexts where the target language is used as a default language of everyday communication (outside the classroom), usually referred to as naturalistic settings, is considered particularly beneficial as it provides access to diverse, authentic input ample in quantity (Flege,

2009; Lightbown & Spada, 2008). In such contexts, SLA research has provided extensive evidence of the following (see Saito & Hanzawa, 2016, for an overview):

- Extensive amounts of L2 input and interaction can lead both early and late L2 learners to achieve near-native L2 performance (Flege, 2009);

- The quality of L2 learners' speech is predicted by length of residence in an L2 speaking environment when the L2 is used as the main language of communication (but not the L1) (e.g. Flege, 2009; Flege & Liu, 2001);

- Frequency of L1 and L2 use strongly predicts the extent to which certain L2 learners benefit from their experience and how much the ultimate quality of their L2 performance improves after years of LOR (length of residence) (Flege, 2009);

- Even after short stay-abroad immersion programmes, school-instructed L2 learners show positive effects in the domain of L2 oral skills, especially for L2 speaking fluency (e.g., Freed, Segalowitz, & Dewey, 2004; Segalowitz &; Freed, 2004; Mora & Valls-Ferrer, 2012; *see* Mora, forthcoming for overview);

- Much learning is likely to happen within the first three to four months of immersion (e.g. study-abroad) in terms of fluency (Segalowitz & Freed, 2004), lexicogrammar accuracy (Mora & Valls-Ferrer, 2012).

The results from these and many other studies show that the learner's developing L2 system is enhanced with increased amount of relevant L2 experience through intensive exposure to the L2 (Saito, 2015). There is a general consensus among researchers that this statement holds true both within and beyond classroom settings. However, as classroom settings often constitute a 'minimal input' learning environment (Larson-Hall, 2008) due to the factors discussed in the previous paragraphs, SLA research has started to highlight the need for providing learners with additional sources of exposure outside classrooms (Muñoz, 2011; Derwing & Munro, 2015). Positive results of L2 pronunciation enhancement in EFL settings have been found when the amount of input is considerably large (e.g., Saito, 2015) or alternatively, enhanced through explicit pronunciation instruction (Gordon & Darcy, 2016; Kissling, 2013; Thomson & Derwing, 2014) or pronunciation feedback (Lyster, Saito & Sato, 2013).

2.2 TV exposure and second language learning

In the Information Age the opportunities for learning have expanded, largely as a result of digital culture. Research concerned with investigating how Digital Era impacts teaching and learning often reports on the shift from an authority-based learning towards a more autonomous information search (Mustapha & Kashefian-Naeeini, 2017) and learners' preference for "infotainment" or fusion between learning and entertainment. The term "digital natives" (Prensky, 2001) was coined to describe these new students, who enter the formal schooling system with radically different set of skills and needs as compared to those who acquired familiarity with digital technologies as adults. How this abundance of multimodal exposure to diverse sources of information

impacts today's students learning and processing is an empirical question researchers from different fields of study try to address. Just as any other acquisition of knowledge possible in the digital environment, learning a second language is not an exception to this rule.

With the increasing popularity of online streaming platforms, the total number of paying *Netflix* subscribers exceeded 148 million as of April 2019 and is constantly growing, and although obtaining exact figures on the number of subscribers by country is difficult due to companies' data privacy policies, a rough estimation suggests that on average people are exposed to over 8 hours of audiovisual input weekly, exclusively through online streaming platforms. Interestingly, although it is only a rough estimation, the hours of minimal online exposure to audiovisual materials exceeds by a factor of two the time dedicated to formal L2 instruction at school.

In light of the challenges an EFL classroom environment poses for creating optimal conditions for improving L2 pronunciation, 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, 1988; 2016). It is now well established that such materials offer large amounts of authentic language input and therefore can serve as a tool for learning.

Audiovisual exposure, such as when watching movies, can enhance second language learning by providing learners with an abundance of authentic input. As suggested by Webb and Rodgers (2009), extensive television viewing is an alternative way to increase the amount of L2 contact, as it has been shown to be an effective source of comprehensible input providing semantic support for the image (Rodgers, 2013). TV viewing is not only an abundant source of naturalistic input, but it can also contribute to

the facilitation of information processing. As suggested by Pujadas (2019, p. 9), "TV programmes - along with other audiovisual materials such as films, documentaries, or short videos - comply with Nation's (2007) five conditions for suitable input. Such suitability is explained in terms of the input being easily processed in large quantities, being comprehensible and presented in a format familiar to learners, being engaging and highly contextualized through image and dialogue.

For the last decades SLA research has been investigating the potential benefits of audiovisual sources of language input for second language learning (Vanderplank, 2010, 2016). 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, especially vocabulary (e.g., Baltova, 1999; Montero-Perez, Peters & Desmet, 2018; Webb, 2010) and listening comprehension (e.g., Kim, 2015; Markham, 2001; Montero-Perez, Peters & Desmet, 2014; Vanderplank, 1988). This research has also investigated the various conditions under which L2 learners watch audiovisual materials (Peters, Heynen & Puimège, 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 (e.g., Kruger & Steyn, 2014; Markham, 2001; Montero Perez et al., 2013; Vanderplank, 1988) and incidental (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) as well as explicit vocabulary learning (Pujadas, 2019).

Independently of the language domain being investigated in relation to the potential L2 benefits through TV exposure, there are several issues of relevance that are central in this discussion, such as explicit vs. implicit teaching, often implemented either through focus on form or through lack of explicit instructions. It is important to acknowledge that the term *incidental* learning can have various definitions (see Pujadas, 2019 for overview):

- learner's perspective: learning taking place "without intention, while doing something else" (Ortega, 2014: 94),
- research perspective: learning occurring without the expectation of being tested afterwards (Hulstijn, 2003),
- activity perspective: learning as a by-product of another activity.

Many studies investigating vocabulary acquisition in the context of TV exposure have adapted the third definition, where vocabulary learning occurs as by-product of another meaning-focused activity (i.e., watching videos for content comprehension). The results of these studies indicate that incidental vocabulary learning can take place through viewing short clips or full movies (e.g. Peters & Webb, 2018; Pujadas 2019) and TV series (e.g., Rodgers, 2013), acknowledging other factors influencing the benefits from the viewing activity such as the viewing mode (L2 captions, L1 subtitles), frequency of exposure to the audiovisual input or participants' proficiency level. The overall picture emerging for the domain of vocabulary acquisition in the context of TV exposure is that L2 captions are more beneficial for vocabulary acquisition than L1 subtitles (e.g., Danan, 2004; Vanderplank, 2010; Winke et al., 2010) for older/more proficient L2

learners and that learners with larger vocabulary knowledge benefit more than the learners with smaller vocabularies (e.g., Peter & Webb, 2018). Additionally, although the studies show mixed results, the frequency of occurrence has a positive effect on incidental word learning (e.g., Peters, 2016; see Pujadas, 2019 for overview). On the other hand, it has been shown that a way to optimize the effectiveness of vocabulary learning through TV exposure is by integrating the intentional or explicit approach with incidental learning (Schmitt, 2010). Research suggests that incidental and explicit learning approaches can be combined to accelerate the learning rate (e.g., Hulstijn, 2013; Nation, 2015). In the context of audiovisual exposure, this has been achieved by deliberately implementing a focus on vocabulary when the primary task is meaning-oriented. Such as for instance, in the context of TV viewing, by pre-directing attention to word-form through input enhancement or through a pre-task activity focusing on specific words, an approach that has been shown to yield gains in vocabulary (Pujadas, 2019).

For pronunciation, both *explicit* and *focus on form* often refer to the type of instructional approach, which involve intentional teaching of L2 phonetics relevant to segmental phonology features of sound units such as place and manner of articulation as well as suprasegmental phonology features such as stress, rhythm and intonation. Typically, in the FL classroom, phonetic instruction emphasizes the difference between learners' L1 and L2 phonological system in regards to phonemic inventories, articulation of analogous phones, grapheme-phoneme correspondence and phonological processes amongst other (Kissling, 2013). Thorough reviews of empirical studies on the benefits of instruction type (explicit vs. implicit) indicate that the results are in fact quite complex and sometimes even contradictory (Kissling, 2013; Piske at al. 2001). While

some studies report that pronunciation instruction has little to no effect on learners' pronunciation accuracy (e.g., Purcell & Suter, 1980; Suter, 1976), others suggest that explicit pronunciation instruction is essential in order to enhance EFL learners' oral abilities (e.g. Gordon & Darcy 2016; Saito, 2011; Thomson & Derwing, 2014). Additionally, while some studies conclude that instruction can help improve segmental production, but not comprehensibility (Derwing et al., 1997) the results from other studies suggest the opposite (Saito, 2011). The effectiveness of pronunciation instruction has been examined across many learners and contexts (e.g., different target language or proficiency levels in either FL, SL, lab or study abroad settings), pedagogical approaches (e.g., explicit, implicit, with or without feedback), target features (segmental, suprasegmental) by using different approaches to assess the outcomes of treatments of varying length. Some conclusions presented in a meta-analysis examining the effect of pronunciation instruction (PI) (Lee, Jang & Plonsky, 2014) and relevant for this dissertation are the following:

- Laboratory-based (as opposed to classroom-based) intervention may yield larger treatment effects due to increased experimental control; both, however, are effective;
- Although different linguistic foci (e.g. segmental vs. suprasegmental features) in conjunction with treatment features may influence the effects of PI, learners at different proficiencies can benefit from PI;
- The length of treatment may be related to its effectiveness, but longer treatments (i.e. longer than the median intervention of 4.25 h) generally produced larger effects;
- Including feedback in a program of PI can improve its effectiveness.

Research on the effects of TV exposure on L2 pronunciation learning is much scarcer than research on vocabulary acquisition or general listening comprehension. Thus, some issues already explored for the domains of vocabulary acquisition or listening comprehension, such as what viewing conditions (L2 captions, no captions or L1 subtitles) may be more beneficial for learning or whether explicit or implicit task focus is more effective, are yet to be explored for pronunciation. The following sections of the literature review aim at presenting the most relevant findings in regards to L2 pronunciation development through TV exposure, justifying how this dissertation seeks to fill some of these gaps in the literature.

2.3 Multimodality and L2 pronunciation learning

Contact with a language through audiovisual media, such as when watching a movie or a series, constitutes, what is called, multimodal exposure, which involves the presentation of language through two different modalities (auditory and visual) simultaneously. When watching subtitled movies, speech and sound effects are processed through the auditory channel, whereas on-screen-text (L1 subtitles or L2 captions) and moving images are processed through the visual channel. The main theoretical foundation of the benefits of multimodal exposure are explained by Mayer's (2001) Cognitive Theory of Multimedia Learning, which integrates previously formulated cognitive theories of learning such as Paivio's (1986) Dual Coding Theory and Chandler and Sweller's (1991) Cognitive Load Theory. The theoretical rationale is built on three main assumptions: (1) humans process information thorough two interacting systems: the verbal/auditory channel and the visual/pictorial channel (the

dual channel assumption); (2) humans' capacity for simultaneous processing of visual and auditory channel is limited (the limited capacity assumption); (3) meaningful learning requires a considerable amount of cognitive processing to integrate verbal and visual information into the existing knowledge (the active-processing assumption) (Mayer & Moreno, 2003).

Although multimedia learning theories are concerned with a universal approach to learning and therefore are not specific to language learning, but their theoretical underpinnings are applicable to SLA research. Multimodal input can facilitate L2 learners' processing by expanding the information sources (auditory, visual or both) they rely on and therefore support comprehension. Multimodality makes sensory information accessible in diverse semiotic codes and offers the opportunity to comprehend information through different channels (Guichon & McLornan, 2007). The reason why multimodal input can be beneficial for L2 pronunciation development is to be sought in the notion of bimodal reinforcement (Paivio, 1986; Mayer, 2014), which claims that the dual processing of auditory and visual information helps create and strengthen the mental representations of perceived objects, and therefore, promote learning. These theories suggest that the simultaneous activation of both systems results in better recall and greater depth of processing. When exposed to audiovisual materials such as a captioned movie with target language soundtrack, captions may help viewers with sound - symbol mapping in a way that, when the same word is presented through one mode only (e.g. just sound or just orthographic representation), might not. Thus, for L2 learners, multimodal exposure can facilitate correct mapping of the phonological representation of words to their orthographic form as well as facilitate detection of word boundaries, which improves listening comprehension thanks to the availability of the

on-screen-text. Thanks to visible words boundaries presented on the screen in captions and the simultaneously presented phonological representations of the words in the auditory input, learners have a change to strengthen the phonological representation or correct if, if stored incorrectly in the learner's mental lexicon.

In learning contexts where contact with target language is often limited, improving the understanding of oral discourse is one of the most challenging tasks L2 learners encounter (Graham, 2006; Kim, 2015). In such contexts, finding alternative ways to increase the amount of target language contact is crucial, especially because listening comprehension is considered to be at the core of L2 learning as the development of L2 listening skills has demonstrated a beneficial impact on the development of other linguistic skills (e.g. Dunkel 1991; Rost 2002).

Speakers of English as a foreign language at all levels of proficiency often have difficulties in understanding words in the continuous stream of speech due to the challenging task of making use of L2-specific word boundary cues to identify words (Cutler, Demuth, & McQueen, 2002; McQueen & Cutler, 1998; Norris, McQueen, & Cutler, 1995) and to a likely mismatch between the incoming auditory input and the phonological representations for L2 words in the learners´ L2 mental lexicon.

Speech segmentation describes the process of identifying the boundaries between the words in a continuous stream of speech (Charles & Trenkic, 2015). This process, particularly difficult for L2 learners, is fundamental for successful speech comprehension and for L2 pronunciation development. Research has shown that listeners use language-specific lexical segmentation strategies to identify word boundaries (Cutler & Butterfield, 1992; Cutler &Norris, 1988). For instance, native speakers of English rely heavily on stressed syllables in identifying the start of a new

word. However, the strategy used by the native speakers of English is not easily accessible for L2 learners since different languages rely on different segmentation strategies. It is extremely difficult to override the speech segmentation strategy dominant in learners' L1 when listening to L2 speech (Cutler, 2000). Thus, even after 9 to 10 years of English instruction and over a year of immersion in an English-speaking country, proficient L2 users may miss about 30% of the words they hear (Charles & Trenkic, 2015). Early research on multimodal input comparing traditional listening exercises (e.g., audio recorded conversation) and audiovisual materials (requiring simultaneous auditory and visual processing) found that the latter boosts foreign language processing, especially by enhancing listening skills (e.g., Brett, 1995; 1997; Meskill, 1996).

The rationale behind multimodal input being a driving force for L2 pronunciation development resides in learners' capacity for dual channel processing, which helps create and strengthen mental representations (i.e. phonological representation of words) when the same word is presented visually (text-on-screen) and auditorily (audio input). This type of multimodal exposure has the potential not only of boosting speech segmentation skills, but also of serving as a kind of perceptual training helping L2 learners minimize the mismatch between their existing and target L2 phonological representations. Although updating phonological representations in the course of language learning is challenging (Darcy, Daidone, Kojima, 2013; Darcy & Holliday, 2019), extensive multimodal exposure to naturalistic input, aided by the on-screen-text, may enhance segmentation skills and the mapping of phonological representations to the orthographic form of words. Phonological mapping can occur when watching movies with original language soundtrack and captions, as the

phonological target representation of a word is presented together with its orthographic representation. If the target phonological representation presented in the auditory input mismatches the phonological representation of the same word stored in learner's mental lexicon, the learner may notice the difference between these two when aided by the presence of the orthographic representation of the same word. Such auditory-visual mapping enabled through the visual input could be helpful in updating incorrect phonological representations in the learner's mental lexicon.

2.4 Captions and subtitles

There are several types of on-screen text, according to the language and format used. *Captions* (also referred as intralingual subtitles or same-language subtitles) refer to on-screen text presented in the same language as the original audio (e.g., English audio + English text), whereas *subtitles* (also referred as interlingual subtitling, standard or traditional subtitling) typically refer to on-screen text in the viewer's native language, which provides a translation of the foreign audio (e.g., English audio + Spanish text). For the sake of clarity, the terms *captions* (L2 audio + L2 text) and *subtitles* (L2 audio + L1 text) will be used throughout this dissertation. Although there are other types of on-screen texts that have been examined in SLA research such as reversed subtitling (L1 audio + L2 text) (e.g. Danan, 1992) or key-word captioning (L1 audio + L2 key-words) (e.g. Guillory, 1998), this dissertation focuses on full-captioning since this is the format most commonly available for viewers.

The results from the studies comparing language outcomes when viewers were exposed to captioning and subtitling versus no on-screen-text show that either on-

screen-text conditions lead to larger overall comprehension benefits than the no-text condition (Baltova, 1999; Garza, 1991; Gass et al, 2019; Markham, 2001; Markham & Peter, 2003; Montero-Perez et al, 2013, 2014; Neuman & Koskinen, 1992; Rogers & Webb, 2017; Winke, Gass & Syodorenko, 2010). However, whether the viewing with captions or subtitles is more beneficial in the context of multimodal exposure is not straightforward and is still a matter of debate. That is because studies show mixed results depending on which aspect of the language is being tested and what testing instruments are used as well as learners' proficiency level (*see* Pujadas & Muñoz, 2020 for an overview). On the one hand, there is a consensus regarding the superiority of subtitles in the viewer's native language regarding overall understanding of the presented content (Bianchi & Ciabattoni, 2008; Birulés-Muntané & Soto-Faraco, 2016; Markham, Peter & McCarthy, 2001; Markham & Peter, 2003), as reading in your native language, not surprisingly, makes comprehension effortless. On the other hand, the superiority of this viewing condition for general comprehension does not necessarily conform to what seems to be beneficial for L2 pronunciation development.

In terms of L2 pronunciation benefits, the debate over the superiority of captions or subtitles is much more transparent. Research shows that presenting aural input and on-screen-text in the same language (instead of native language subtitles) aids aural form recognition (Markham, 1999), helps establish form-meaning connection (Winke et al., 2010), enhances speech perception (Mitterer & McQueen, 2009) and segmentation (Charles & Trenkic, 2015), which are crucial skills for pronunciation development. There is substantial evidence of the value of captions for enhancing listening comprehension skills and for helping learners visualize the speech stream and identify word boundaries (Baltova, 1994; Bird &Williams, 2002; Charles & Trenkic, 2015;

Markham, 1999; Mitterer &McQueen, 2009; Sydorenko, 2010; Winke et al., 2010). Captions do not only aid disambiguating miscomprehended words from the auditory input, but also help the mapping between the phonological and orthographic representations of words (Garza, 1991), which in turn helps creating a more accurate memory trace, which may subsequently lead to better word identification from the aural input when the supporting text is not available (Bird &Williams, 2002; Garza, 1991). In the context of authentic video materials, the auditory input can be very demanding to process (Vanderplank, 2016a), as it may rely heavily on quick verbal interaction, full of unknown references, which need to be processed under time-restrictions posed by the presentation speed. If the ultimate goal is accuracy of comprehension, there is no doubt that L1 subtitles are more effective, especially for learners at lower levels of proficiency, since understanding your native language is effortless. However, if the goal is to develop better comprehension of L2 speech, which may subsequently lead to re-shaping the pre-established phonological representations, the only adequate viewing mode seems to be captioning, as the benefits of dual channel processing cannot be available to the learner otherwise.

2.5 L2 pronunciation benefits from exposure to captioned videos

Speakers of English as a foreign language at all levels of proficiency experience difficulties understanding words in the continuous stream of speech. Charles and Trenkic (2015) showed that L2 English university students failed to repeat back 30% of the spoken words they heard. This result is surprising, given that these students have spent more than one year in the United Kingdom studying in an immersion context. This

shadowing task performed in laboratory conditions consisted in repeating back short phrases, which participants had previously 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 learner's 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. Being exposed to the multimodal presentation mode with captions resulted in 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. This dissertation extends this research as it is designed to fill this gap in literature by testing whether the benefits in segmentation skills could extend to enhanced efficiency in L2 phonological processing and phonological accuracy.

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 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 had been exposed to, 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 exposure to multimodal input. Importantly, the advantage of captions over no captions or L1 subtitles the studies reviewed above report results from either a single or only a few viewing sessions and where always the result of an immediate post-test. To the best of

my knowledge the effects of extended exposure to captions have never been investigated in the context of L2 pronunciation improvement.

2.6 Individual differences in proficiency, attention and short-term memory

Investigating cognitive individual differences is important to inform scientific as well as pedagogical practices. It is crucial to include them in the design of this study in order not to misinterpret the findings, so that the pronunciation benefit found can be attributed to the viewing treatment conditions and not other confounding factors. Decades of extensive research on the role of aptitude and individual differences for second language learning has identified a variety of factors cognitive, experimental and psychological factors as potential predictors of ultimate attainment in L2 speech learning (*see* Mora, submitted). It is estimated that they may account for approximately 25% of the variance in L2 learning and to be related to oral production skills, but different aptitude components typically show differential predictive validity for different dimensions of language performance (Li, 2016)" (Mora, submitted). In the context of this study, testing participants' capacity of attention switching between different modalities (visual and auditory), ability to selectively attend to auditory signal as well as their memory for phonological details was consider to be essential.

It is important to acknowledge that individual differences play an important role in the acquisition of L2 phonology (Mora & Darcy, 2017), just as they play a role in other language domains. In the context of L2 pronunciation development, research on individual differences tries to identify the sources of variability associated with L2 learners' varying levels of ultimate attainment, investigating experience-related factors

(age of onset of L2 learning, amount and quality of L2 exposure and use), socio-psychological factors (e.g., motivation, personality, anxiety, learning strategies) and language learning aptitude and cognitive functions (e.g., working memory, attention, auditory processing skills) (*see* Mora, submitted*, for overview*).

Although empirical evidence supports the beneficial role of captions in aiding the decoding and segmentation of L2 speech by helping listeners map auditory input to linguistic form in running speech, L2 learners appear to vary greatly in their ability to do so. One of the factors that explain this variability is learners' proficiency level. Research has shown that a certain threshold in L2 competence needs to be reached for captioning to lead to L2 acquisition benefits (Muñoz, 2017; Neuman & Koskinen, 1992). This is because the presence of on-screen-text during viewing cannot compensate for unfamiliar vocabulary, expressions or grammatical structures. Most research (but see Guillory, 1989) has shown that for low level proficiency learners multimodal exposure is not fully aiding the learning process, because the input is not understood (Markham, 2001; 2003). Because of the interplay between proficiency level and gains in listening comprehension, Markham (2001) proposed a training sequence in order to scaffold listening comprehension development, especially if the L2 learning process usually takes place in a heavily reading-dependent classroom settings lacking sufficient aural input. The proposed sequence consists of firstly exposing students to audiovisual materials with L2 audio and native language subtitles, as initially better reading skills may aid comprehension that could be otherwise lost and then to replace subtitles with captions upon the second viewing of the same materials. The last step in the scaffolding sequence consists of presenting the audiovisual materials without the on-screen-text. The proposed sequence implies that the use of L1 subtitles supports the comprehension

of challenging materials, allowing students to initially use their stronger L1 reading skill (L1 subtitles), followed by the use of their weaker target language reading skills (captioned viewing mode), after which the students would be ready to rely on their target learning listening skills (uncaptioned mode). In a subsequent study, Markham and Peter (2003) found that captions were more helpful to advanced learners when the video materials were more abstract or complex.

Given that speech and text processing in the L2 is less efficient than it is in the L1 (Segalowitz, 2010), watching L2 captioned video is a particularly challenging activity from a cognitive load perspective (Mayer & Moreno, 2003), especially for low proficiency learners. It requires resolving competition and integration of different sources of L2 input without having control over the speed of the information flow: the auditory input (the soundtrack), the dynamic image (the scene) and the textual input (the captions). Under such circumstances, the L2 proficiency of learners as well as their use of cognitive resources (particularly attention) are likely to contribute substantially to language processing efficiency and, consequently, to how much they can benefit from the exposure to L2 captioned video for L2 pronunciation development. Although the impact of captions on cognitive load in multimodal input environments has yet to be determined (Kruger, Hefer & Matthew, 2013), it is already known that factors such as type of audiovisual presentation mode (e.g., Bisson et al., 2014), level of information redundancy between different information sources (Drew & Grimes, 1987; Grimes, 1991; Perego et al., 2010), speed of subtitle presentation (Fresno & Sepielak, 2020; Liao et al., *in press*; Szarkowska et al., 2016; Szarkowska & Gerber-Morón, 2018) or different visual display modes (Brown, Jones & Crabb, 2015) affect processing.

Mattys, Brooks & Cooke (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 sub-lexical tasks negatively affecting perceptual discrimination (Mattys & Wiget, 2011). Thus, as suggested earlier, 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.

Provided the learner surpasses the proficiency threshold, multimodal exposure without cognitive overload (Kruger et al., 2015) could serve as a form of perceptual training by reinforcing the link between the orthographic and auditory forms of words. Improving perceptual sensitivity to cross-language phonetic differences can help L2 learners overcome L1-based perception (Best & Tyler, 2007) and 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 can also have positive effects for higher-order levels of linguistic processing (e.g., lexical, semantic) and for listening comprehension skills. Therefore, serving as perceptual training, multimodal exposure might also lead to the updating of previously established phonological representations toward target phonological forms, by effectively improving L2 speech processing, segmental speech perception, and eventually, speech production.

Apart from proficiency level, other factors moderating perceptual learning that may have an impact on L2 pronunciation may be at play, such as cognitive individual differences. 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 pre-established phonological forms are likely to interfere with the processing of word forms in the auditory input, thus hindering pronunciation learning. According to Mayer (2001), exposure to multimodal materials can be beneficial to learners depending on the amount of information that can be held at a given time (i.e. working memory capacity). As summarized by Kam, Liu and Tseng (2020), it is plausible that L2 learners with better working memory capacity may benefit more from the addition of extra input in form of captions while watching videos, given that the constraint on working memory may force L2 learners to be selective in where they allocate their attention (Mayer, 2001). Due to the availability of multiple input sources, more efficient management of attentional resources may help learners selectively attend to auditory input or successfully store sounds in their phonological short-term memory.

Benefiting from multimodal exposure is a complex issue, as it may be affected by the interplay between various individual differences factors such as attention, phonological memory or the capacity to selectively attend to speech dimensions important for L2 pronunciation development. For instance, in the meta-analysis of 15 L2 captions studies, Montero Perez et al. (2013) concluded that "the relationship

between captioning effectiveness and proficiency level may be less problematic than [initially] outlined " (p.732), as some individual differences in cognitive profiles may play a more prominent role for benefiting listening comprehension in the context of multimodal input exposure. In a similar vein, provided that the learner has surpassed the proficiency threshold, it can also be difficult to determine what viewing mode (captioned on uncaptioned) may be more beneficial for L2 pronunciation development, taking into account the role cognitive individual differences in attention allocation or phonological memory may have.

2.7 Reading in multimodal contexts

The cognitive processes involved in eye-movement control when reading static text have been widely researched, thus now, we have come to a good understanding of the mechanisms at play (e.g., Reichle et al., 1998; Reichle et al., 1999, Reichle et al., 2003). However, the cognitive processes viewers engage in when processing dynamic text in captioned videos is substantially different and still not well understood. If presented on a difficulty spectrum, reading a novel (static reading on a static background) would occupy one end, while captions reading (fleeting text on dynamic background) would occupy the other end. In 1998, Rayner published an overview of 20 years of research on eye movements in reading and information processing in static text and the research in this field has been developing since. However, the level of insight regarding dynamic text reading and information processing in multimodal context is not quite the same, as the existing reading models (e.g. Reichle, Pollatsek, Fisher & Rayner, 1998; Reichle, 2020) have not examined reading from a multitasking perspective, and therefore do not

account for the presence of constant visual and/or auditory competition. Reading in multimodal contexts is more challenging because of the constant presence of competing input sources. The reader has to effectively manage the use of cognitive resources across different information sources without having control over the speed of its presentation (Kruger & Steyn, 2014). When a dynamic text (captions) is presented together with another visual stimulus (moving image), there are many factors that make reading cognitively demanding. One reason why reading in multimodal context can be challenging is that the high-acuity vision necessary to identify words is limited to the central 2 degrees of the visual angle (fovea), therefore other tasks requiring visual attention, for instance scene scanning, can be compromised. Another reason is that attention resources are limited and, in a multimodal context characterized by visual competition, the processing of different sources can result in a tradeoff. Given the limitation of the visual span, it is highly probable that reading interferes with any secondary visual tasks involving processing of complex visual information (for instance, paying attention to a visual detail and reading captions at the same time can be unfeasible). As it is the case of a subtitled video, time-restrictions posed by the disappearance of the fleeting text makes dynamic reading highly demanding, which adds to the complexity of multimodal processing. Under the circumstances described above, the L2 proficiency of learners as well as the use of their cognitive resources and speech processing skills are likely to contribute substantially to language processing efficiency and, consequently, to how much learners can benefit from exposure to L2 captioned video for L2 pronunciation development. On the one hand, reading in multimodal contexts can be supported by the presence of additional input sources (i.e. image and/or audio), which may contribute to the creation of a situation model

facilitating processing (*see* Liao et al., forthcoming). This means that in certain situations, even a superficial reading of the on-screen text could be sufficient for completion on an accurate situation model. On the other hand, the constant visual competition between caption reading and dynamic scene perception may also constitute a challenge, as one visual source can occupy learners' attentional resources causing an inevitable trade off. For instance, sudden onset of movement in the image such as a shot change (Krejtz, Szarkowska & Krejtz, 2013) could momentarily distract the viewer from caption reading, resulting in interrupted or incomplete reading. Likewise, an unfamiliar or low frequency word in the subtitles may cause the viewer to miss an important visual change as a result of momentary perceptual blindness (Schilling, Rayner & Chumbley, 1998).

In SLA research, many studies have implemented eye-tracking methodology to examine diverse aspects of second language processing (*see* Winke, Godfroid and Gass, 2013 *for an overview).* This is motivated by the fact that eye-gaze behaviour is possibly the richest online record of one's cognitive behavior when interacting with multimodal input and dynamic text reading. Eye-tracking provides quantitative evidence of visual attention allocation (*the eye-mind assumption*; Just & Carpenter, 1976; 1980) and it allows to record online eye-gaze movements, therefore numerous studies looked at subtitle processing implementing this methodology (Bisson, Van Heuven, Conklin & Tunney, 2012; d'Ydewalle & De Bruycker, 2007; d'Ydewalle & Gielen, 1992). The recorded eye-movements are classified into two distinct categories: *fixations* and *saccades*. Fixations are gaze points when the eyes stop scanning the scene, holding the central foveal vision in place so that the visual system can take in detailed information about what is being looked at (Land & Tatler, 2009), whereas saccades are rapid

movements of the fovea from one point of interest to the next. A large body of eye-tracking research supports the idea that information intake occurs mostly while we fixate our eyes on a specific point and not during the fast saccadic movement as then the image on the retina is of poor quality (Rayner, Smith, Malcolm, & Henderson, 2009). Visual perception is constructed upon alternating between these two types of eye-movement.

Despite growing interest in how competing visual and auditory input sources impacts language processing (Liao, Kruger & Doherty, 2020), the relationship between multimodal processing and language learning is not yet well understood. Determining the individual influence of each information source (text, audio, image) in order to understand the best combination for benefiting different linguistic aspects, proficiency levels or cognitive profiles is challenging. In the context of multimodal reading, the most common eye-tracking measures used in subtitling and captioning research are fixations count, gaze time or dwell time, proportion of skipped subtitles or, more recently, word skipping probability (*see* Doherty & Kruger, 2018 and Conklin, Pellicer Sanchez & Carrol, 2018 *for an overview*), but some researchers claim that these measures can only be interpreted in terms of attention allocation and do not provide in-depth detailed information regarding reading behavior and overall text processing.

The Reading Index of Dynamic Text (Kruger & Steyn, 2014) was developed to compensate for some shortcomings of the eye movement analysis in order to make more informed inferences about the visual processing of subtitles and the meaningful reading behaviour. It is based on the number of fixations on each word in every caption and its novelty resides in that the RIDT formula penalizes for certain behaviours that do not

ensure efficient processing such as skipping, perceptual jumps or re-fixating on the same word, which may hinder reading.

2.8 Phonology in reading

Baddeley's multi-component model of working memory (Baddeley & Hitch, 1974; Baddeley, 1986, 2000; *see Figure 1*.) is one of the most influential frameworks in explaining the functioning of cognitive processes. In this section, the role of phonology in reading will be illustrated utilizing this framework. This model consists of a central executive system that controls cognitive processes and coordinates three slave subsystems; the phonological loop, the visuo-spatial sketchpad and the episodic buffer. The phonological loop consists of a short-term store for verbal information that encodes phonological elements (sounds) and their serial temporal order in the form of auditory memory traces that decay rapidly; furthermore, it contains an articulatory rehearsal component, that prevents these auditory traces from disappearing through subvocal rehearsal - a silent articulation mechanism. The visuo-spatial sketchpad holds visual and spatial information in short-term memory. The episodic buffer integrates visuo-spatial and verbal information and links it to a long-term memory storage component.

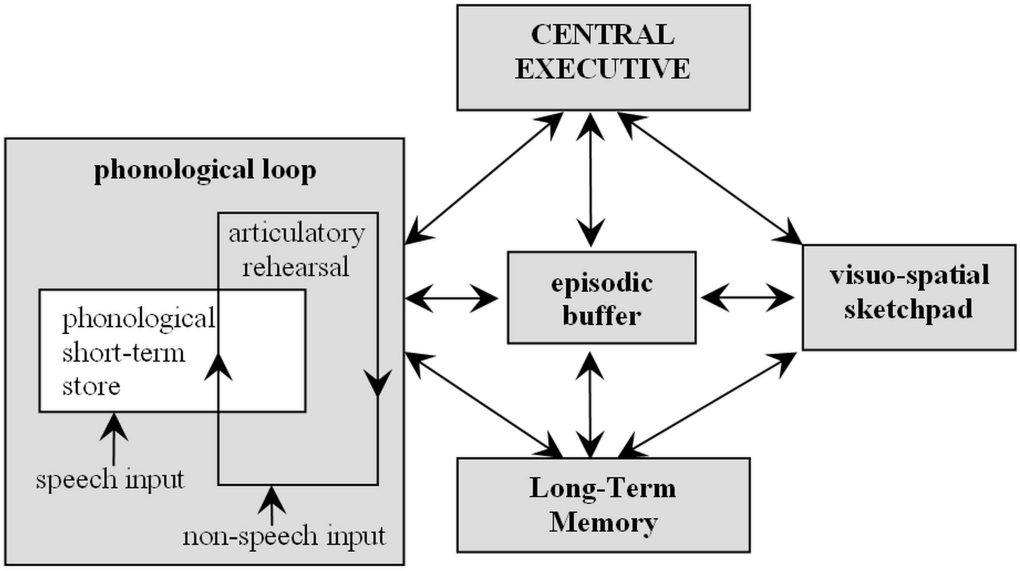


Figure 1 Baddeley's multi-component model of working memory (adapted from Baddeley, 1986; 2000; from Mora (2014: p.87)

When reading in their L1, readers momentarily store the recently read materials in the phonological loop, rather than in their visuo-spatial sketchpad. In other words, they do not see what they read, but hear it instead. This means that the visual trace is lost in favour of the phonological product (Walter, 2008, pp: 458). When reading in the second language, the phonological representation of what is being read may be unreliable or mismatching the target. When incorrect phonological representations are already established in the learner's mental lexicon, the similar orthographic form may reinforce wrong phoneme identification. This can be evident at times even if the higher order mental structure building is not affected. For instance, for Spanish learners the word *pirate* /ˈpaɪ.rət/ (Spanish *pirata* /piˈrata/) is often mispronounced as /ˈpirat/ due to its cognate status and orthographic resemblance with the L2 target. However, given the fact that phonological representations can be activated through orthographic input, hearing the target native-like production just as the learner's internal phonological

representation is activated through the visual orthographic input might help learner notice the gap (i.e. phonetic/phonological differences) between the activated phono-lexical representation and the native auditory input they have just heard. This could help updating the existing phono-lexical representations in the direction of the target forms just perceived.

When watching captioned videos, the presence of authentic auditory input can help bridge the mismatch between learner's inaccurate phonological representations and their target L2 forms. Although modifying existing phonological representations is a challenging task, extensive exposure to naturalistic input may enhance establishing correct symbol - sound mapping.

Chapter 3. Rationale

Substantial research evidence indicates that audiovisual exposure through TV series has a potential for aiding second language learning and for boosting perceptual learning leading to L2 pronunciation improvement. Due to their abundance and the ease of access, videos with original target language soundtrack may help bridge the gap between the lack of authentic classroom materials and the benefits of naturalistic, quality input, compensating for insufficient language exposure within classroom settings. Findings from previous studies suggest that viewing videos with captions and original soundtrack may foster L2 pronunciation improvement, given the facilitating role of captions for speech segmentation as well as the benefits of dual-channel processing. Nevertheless, none of the reviewed studies have directly tested the predicted beneficial role of multimodal exposure for L2 pronunciation benefits for L2 speech perception and production. Additionally, in none of the reviewed studies extended exposure to audiovisual materials was implemented, thus the superiority of captions was a product of single viewing sessions with an immediate post-test.

The studies reviewed above provide substantial empirical evidence for the beneficial effects of multimodal 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), suggesting that such exposure could benefit lexically guided perceptual learning and could consequently lead to the retuning of phonological form of L2 words in the

learners' mental lexicon or improvement of phonetic categories that make up the lexical form of L2 words. None of these studies tested whether this was indeed the case.

This dissertation aims at filling several gaps in the literature and sets out to investigate the effects of extended exposure to L2 captioned videos for L2 pronunciation. The primary aim is to extend previous research in this area by investigating the effects of multimodal exposure on L2 speech processing skills (beyond speech segmentation) and to test whether the benefits of perceptual learning indeed lead to L2 learners' retuning of phonological categories in perception and improved accuracy in production.

Moreover, previous studies attribute captions a beneficial role in enhancing perceptual learning without actually measuring the amount of on-screen-text processing. To my knowledge, this study is the first to investigate caption reading patterns in order to relate individual differences in the amount of on-screen text processing to pronunciation gains.

3.1 Research questions

This dissertation aimed at answering the following research questions:

1. Do learners' L2 speech processing skills and phonological accuracy benefit from exposure to L2 video clips?
2. Does viewing mode (i.e. captioned vs. uncaptioned) affect gains in L2 speech processing and phonological accuracy?
3. Does task focus (i.e. focus on phonetic form vs. focus on meaning) affect gains

in phonological accuracy under different viewing modes?

4. What is the relationship between the amount of on-screen-text processing and pronunciation gains?
5. Do individual differences in proficiency, attention control and phonological short-term memory explain gains in L2 speech processing and phonological accuracy as a function of viewing mode and task focus condition?

Chapter 4. Methodology

This chapter outlines the design of the study and describes the experimental setup adopted to achieve the objectives stated in the previous chapter. Section 4.1 discusses the general design of the study; section 4.2 details the demographics of the participants in the study; section 4.3 provides a detailed description of the viewing treatments; section 4.4 lists all the instruments used, justifies their selection and outlines the and operationalization of the constructs being measured in each task.

4.1 Research design

This study has a pre-/post-test design and it consists of an 8-week viewing treatment during which experimental participants were regularly exposed to audiovisual materials. The purpose of this design was to assess L2 pronunciation gains after controlled, regular exposure to multimodal input through videos presented either with or without captions either with or without explicit focus on pronunciation. The aim of the implemented experimental design was to gain better understanding of the role of captions and explicit focus on form for L2 pronunciation learning in the context of multimodal exposure. Participants watched 15 episodes of a British TV series *Luther* (BBC, 2010) under two different *viewing modes* (with or without L2 captions) and *task focus conditions* (inducing focus either on pronunciation or plot comprehension). This resulted in four experimental groups undergoing the viewing treatment with (1) captions and focus on phonetic form (Captions+FoPF), (2) no captions and focusing on phonetic form (NoCaptions+FoPF), (3) captions and focusing on meaning (Captions+FoM) and (4) no captions and focus on meaning (NoCaptions+FoM). A group of participants from the

same pool was recruited as a control group and did the pre-test and post-test assessment without getting exposure to the audiovisual materials. All participants performed a battery of tests (see Figure 2) one week before and one week after the treatment, which included L2 speech processing measures, L2 pronunciation accuracy measures and a set of global eye-gaze measures and the Reading Index of Dynamic Text (Kruger & Steyn, 2014). Gains in L2 pronunciation were assessed through the selection of L2 speech processing tasks (shadowing, animacy judgement, sentence verification) and L2 pronunciation accuracy (ABX discrimination and delayed sentence reading). Through eye-gaze measures (described in section 4.4.3), the amount of text processing and changes in caption reading behaviour before and after the viewing treatment were assessed. The pre-testing phase was preceded by a background questionnaire (Appendix A) and a battery of tests measuring individual differences in auditory selective attention, attention switching, phonological memory and included the elicited imitation task as a means of measuring L2 proficiency. The post-test contained an additional after treatment questionnaire (Appendix B) designed to measure participants' engagement with the viewing treatment as well as to assess the condition under which the viewing took place.

Week 1		Week 8
Pre-test **L2 speech processing** ● Shadowing ● Animacy judgement ● Sentence verification **L2 phonological accuracy** ● ABX discrimination ● Delayed sentence reading (foreign accent rating) **Eye-tracking measures** ● RIDT ● Global **Individual differences** ● Phonological memory ● Attention switching ● Auditory selective attention ● Proficiency (EIT) **Pre-test questionnaire**	**Viewing treatment** 2 video excerpts per week 4/5 treatment question per video (1) **Viewing mode** captions or no captions (2) **Task focus** phonetic form or meaning	Post-test **L2 speech processing** ● Shadowing ● Animacy judgement ● Sentence verification **L2 phonological accuracy** ● ABX discrimination ● Delayed sentence reading (foreign accent rating) **Eye-tracking measures** ● RIDT ● Global **Post-test questionnaire**

Figure 2 Experimental design

In order 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 efficiency in 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 difficult vowel contrast, and a foreign accent rating task provided a holistic assessment of accuracy in participants' L2 speech production. Eye-gaze measures were also collected before and after the viewing treatment in order to trace changes in subtitle reading behaviour and relate individual differences in the amount of on-screen text processing to pronunciation gains. Participants' proficiency, phonological memory and attention were tested in order to ensure that the potential benefit of the viewing treatment are not confounded with some of these cognitive individual differences and to disentangle their potential influence on the treatment gains. Phonological memory was tested as previous research has already provided evidence for its predictive role in L2 speech learning benefiting L2 oral fluency (O'Brien et al., 2007) or perceptual accuracy (Aliaga-Garcia, Mora & Cerviño-Povedano, 2010). Given the multimodal context in which the study is set and the attentional demands it poses, all participants' attention switching and auditory selective attention was also tested.

4.2 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 above or to a control group. In order to ensure that participant groups were comparable in terms of their exposure to L2 audiovisual materials, self-reports on frequency of viewing movies and series in English with original soundtrack and on viewing mode (with or without captions) (Table 1) were collected through the pre-test questionnaire (see Appendix A). 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).

In terms of viewing frequency, the majority of the participants (Captions + Form = 68%; NoCaptions + Form = 63%; Captions + Meaning = 78%; Control = 69%) self-qualified themselves as overall frequent viewers reporting watching movies either on a daily basis or 4-6 times per week with the exception of those assigned to the NoCaptions + Meaning group, in which only 41% reported watching movies with such frequency. When asked about the most frequently selected viewing mode, the majority opted for original language with either Spanish or Catalan captions (Captions + Form = 79%; NoCaptions + Form = 79%; NoCaptions + Form = 79%; NoCaptions + Meaning = 76%; Control = 84%) over English captions or no captions. Importantly however, the groups did not differ in terms of how frequently they watched movies in original language (English) with English captions ($X^2(16,85) = 25.58$, $p = 0.060$) or without them ($X^2(16,85) = 10.467$, $p = 0.841$). An elicited imitation task (Ortega et al., 2002) was used to obtain an estimate of inter-learner differences in L2 proficiency.

The elicited imitation task consisted of 30 auditorily presented sentences of increasing grammatical complexity participants were asked to repeat back (see section 4.4.4 for detailed task description and Appendix C for the list of sentences). The maximum score that could be obtained in the elicited imitation task is 120 (30 sentences x 4 points). The scores obtained 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 inter-learner variability. 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 ($F(4,85) = 0.14$; $p = 0.967$, $\eta^2 = .007$).

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

4.3 Viewing treatment

Participants watched the first season of the British series *Luther* (BBC One. 2010) divided into chronological excerpts of comparable length (approximately 20 minutes each) twice a week. The total viewing time of 15 sessions added up to approximately 5h of exposure. A survey was conducted to make sure that the participants had not seen the series selected prior to this experiment. They watched each session through an online platform *PlayPosit* (www.playposit.com). This editing platform was chosen as it allows users to create interactive online sessions by enriching the uploaded video content with interactive tools such as multiple choice or fill-in-the-blank questions. This platform allowed the researcher to embed questions into each viewing session, collect students' answers and provide immediate feedback after each answer. The sessions were made available to learners for watching online only for a limited period of time (3 days for each excerpt) to control for regularity of exposure (2 excerpts per week) as well as to keep control of learners' log times. The audiovisual materials were identical for all four experimental conditions and the learners watched the excerpts in the same chronological order. This type of viewing (narrow viewing) allows for natural development of the plot and for gradual accumulation of contextual knowledge, which facilitates comprehension (Webb, 2015). Participants received detailed instruction on how to use the platform and, after the treatment, were asked to report under what conditions the viewing had taken place. The precise record of participants' log times spent on the platform allowed the researcher to control that everyone followed the task procedure correctly (Appendix D). Lack of regularity in viewing or not completing all the viewing sessions resulted in

exclusion from the study. Additionally, participants were excluded if the registered time on more than three viewing sessions exceeded 45 minutes.

The viewing treatment had two distinctive conditions, namely, *viewing mode* and *task focus. Viewing mode* was operationalized between subjects by assigning participants to either a viewing condition *with captions* or *without captions*. Within each one of the viewing conditions, participants' attention was directed either to the phonetic form of words (*pronunciation focus*) or to the understanding of the plot (*meaning focus*). This was achieved by providing participants with two different types of questions (4-5 per video excerpt, 65 in total) that had to be answered while watching. The questions would unexpectedly appear on the left side of the screen (Figure *3*) causing the video to stop until the answer is provided. Questions on pronunciation served the purpose of directing learners' attention on the speech input focusing on the phonological form 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 E and F for a list of the treatment questions).

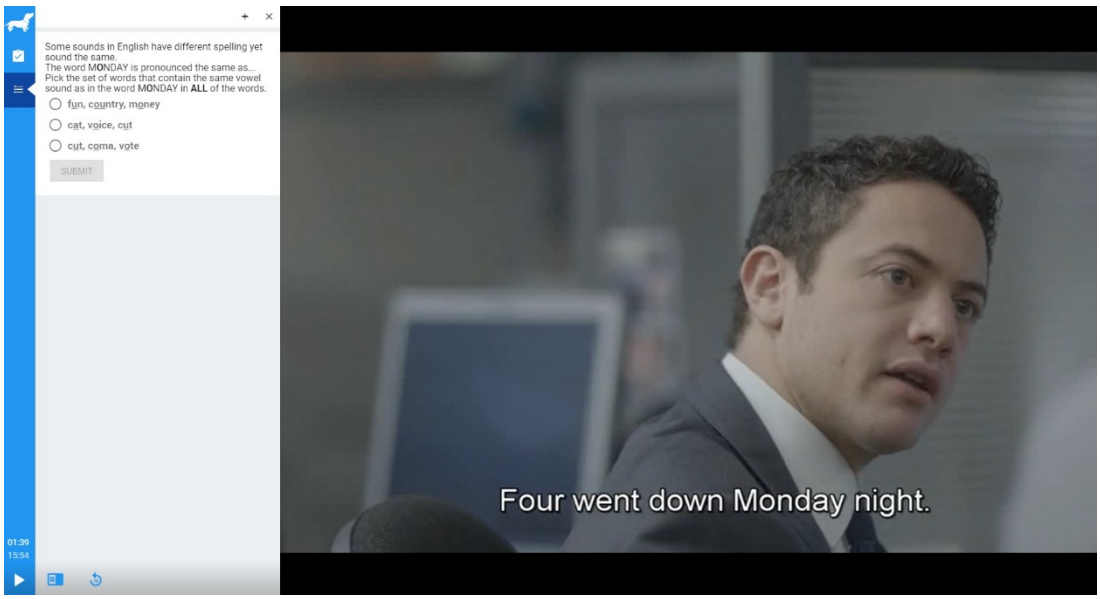


Figure 3 Treatment questions layout

The questions did not target a specific pronunciation feature, but rather focused on diverse potentially problematic phonological features for L1 Spanish/Catalan learners of English. Those targeted aspects such as /iː/-/ɪ/ and /ɑː/- /æ/ vowel contrasts, words with lack of orthographic transparency, regular past tense of the verbs (*-ed* ending) or *-ure* endings contained in the words such as *nature* or *gesture*, words with initial /s/ followed by a consonant, unstressed final schwa or words with vowel reduction. The answer to each question was always related to the auditory input the participants had just heard. Participants assigned to form-focused groups were expected to generally orient their attention to pronunciation while watching, as providing the correct answer to the questions on pronunciation was dependent on how much attention was paid to the auditory input.

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 a variety of formats was used, 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 this experimental design, watching video excerpts had an interactive component not present in normal video viewing. Nevertheless, the aim of the focus-on-meaning condition was to simulate the real-life situation of what happens when watching a movie or a series, i.e. to engage with the plot and focus exclusively on its comprehension as opposed to the other condition with an explicit pronunciation focus.

4.4 Testing

This section provides a detailed description of all the tests used in the study. It is divided into three subsections, each presenting the instruments used to assess the treatment outcomes in L2 speech processing (*4.4.1),* L2 phonology (*4.4.2*), eye movement behavior (*4.4.3*) and cognitive individual differences in attention and phonological memory (4.4.4).

4.4.1 L2 speech processing

Shadowing task

Following on Mitterer and McQueen (2009) and Charles and Trenkic (2015), a shadowing task was selected to measure speech segmentation skills. In this task participants were presented auditorily with short excerpts of speech and were ask to repeat them as accurately as they could. These speech samples were taken from the

audiovisual materials participants had been exposed to during the treatment (*Luther*) as well as from unfamiliar TV series participants had not been exposed to (*Sherlock*). In this way, the 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 listen to each sentence twice and were asked to repeat twice a total of 40 sentences (20 *old* items from *Luther* and 20 *new* items from *Sherlock)*. Each sentence was presented with a stimulus onset asynchrony of 3 times its duration. Old and new items were randomly mixed within these blocks. Speech samples were selected to vary in difficulty through word count (between 4 and 15 words in each sentence), but average word length was comparable across the two 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 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 per participant. Only the repetitions from the first trial were used to compute the final score. The maximum score that could be obtained on this task is 320, which equals to 162 words repeated from the series *Luther* and 158 words from the series *Sherlock* (see Appendix G).

Animacy judgement

To assess the 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 *non-living* (e.g. *chair*) item by pressing either the right or left shift key on the computer keyboard. Half of the items qualified as *living* and the other half as *non-living* and their presentation was fully randomized. 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)* (see Appendix H). The presentation of test items was preceded by eight practice trials providing feedback on accuracy and response time. The dependent measure for this task was the mean response latency (in milliseconds), which represents the processing time it took participants to access the lexical representation of the word appearing on the screen. The experiment was run in DmDX (Foster & Foster, 2003) and was preceded by a practice trial, which included 8 examples and provided two types of feedback: accuracy feedback ("Correct!" or "Wrong!") and reaction time feedback (e.g. 987 milliseconds or Too slow!) to prompt fast responses.

Sentence verification

This task was adapted from Munro and Derwing (1995) and measured speed of processing of L2 speech. Processing time was estimated by measuring how long it took listeners to assign true/false values to a number of short sentences (Appendix I). In this task participants were presented aurally with forty 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). To ensure that the participants listened to the whole sentence, the last word of each trial was always decisive for semantic disambiguation. 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. The 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.

4.4.2 L2 Phonology

ABX discrimination

To assess L2 perceptual learning in terms of phonological accuracy we administered a speeded ABX categorical discrimination task (Flege. 2003; Gottfried, 1984) adapted from Darcy, Mora & Daidone (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). Inter-stimulus 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 tri-syllabic non-words following the rules of English phonotactics and a CV.ˈCV.CV(C) structure (e.g. *fadittick*). The lax-tense /iː/-/ɪ/ vowel contrast was tested, 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 post-test. A response latency measure (RT) was also used 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). The dependent measures for this task were accuracy of correct responses and response latency (RT).

Delayed sentence production (foreign accent ratings)

This task was administered in order to measure learners' phonological accuracy in production, as assessed by native speakers' judgements of accentedness. 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 visible 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. The task was designed in a way that would allow for eliciting speech samples from memory, thus obtaining speech productions that would reflect learners' L2 phonology.

Given the number of participants in the study, three representative sentences containing complex phonological features (*see* Appendix J for a complete list of sentences) were selected for presentation to native speaker judges in a foreign accent rating task. These sentences were:

1. The path leads through a narrow street
2. She always thinks about her future
3. A gun was found at a crime scene.

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 pre-test and post-test). The sentences were presented in pre-test post-test pairs in randomized trials. In each trial, raters provided two ratings on a continuous foreign accent scale, one for each one of the two sentences in the pair. In half of the trials the pre-test sentence appeared first and in the other half the post-test 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. The two ends of the scale were labelled as 0 = "No Foreign Accent" and as 100 = "Heavy Foreign Accent". Due to the number of participants, there were 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 twelve native speakers rated ten percent of the speech samples in order to calculate an inter-rater reliability index. The Cronbach's alpha index for these ratings (360 items) resulted in a reliability coefficient of $\alpha = 0.862$, suggesting consistency of 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 average percentage score, the more target-like the L2 phonology of the participant was or the less foreign accented the speech was.

4.4.3 Eye tracking

The eye-movements were recorded to estimate how much text in the captions participants processed and to trace potential changes in eye-gaze behavior before and after the viewing treatment. Participants watched three video clips taken from the BBC series *Sherlock* (Gatiss et al., 2010-2017) presented with English captions. Table 2 shows the characteristics of each clip. Their eye movements were tracked and recorded on a Tobii T120 eye-tracker run on a 120Hz mode while they watched three short clips. The selection of this series for a pre-/post-test comparison is due to its genre specificity - *Sherlock*, just like the treatment series *Luther,* is a British crime drama with the lead character being a crime-solving detective. The viewing task was presented as a language comprehension task that required watching the clips carefully in order to answer a true/false comprehension question appearing on the screen at the end of each clip. The questions were designed to engage participants in the viewing activity and served the purpose of maintaining their attention while watching.

Table 2 Video clips characteristics

Clip	Duration	Number of captions	Number of words	Average number of words per caption	Average word length in characters
1	3 minutes 9 seconds	55	412	7.49	5.37
2	1 minute 29 seconds	30	220	7.33	5.21
3	1 minute 35 seconds	35	271	7.74	5.34
Total	6 minutes 13 seconds	120	903	7.52	5.30

The raw eye-gaze data with Velocity-Threshold Identification (I-VT) filter (Olsen, 2012) were used to calculate the Reading Index of Dynamic Text (Kruger & Stein, 2014) and *TobiiStudio* software was used to obtain global eye-tracking measures described below. To extract the essential information, all the captions in each video were marked individually as Area of Interests (AOI). This was done by drawing a box around each caption (Figure 4) leaving a border around it of approximately the length of two characters. The AOI for each caption was toggled on and off on the frame on which each caption appeared and disappeared.

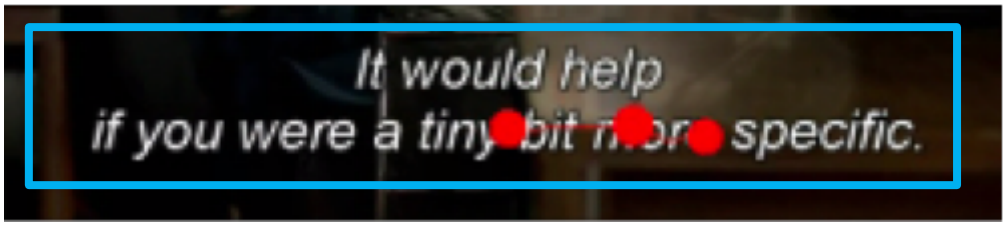


Figure 4 Example of an Area of Interest (AOI)

Global eye-tracking measures

Given the experimental design of the task and the technical properties of the eye-tracking equipment, all the measures being reported are global and based on the sentence-level analysis of the data. Some of the most widely used eye-tracking measures in caption and subtitling research (*see* Doherty and Kruger, 2018) are reported here to characterize subtitle reading behaviour, including:

- *fixation count:* the raw numerical count of fixation occurring in a particular AOI;
- *average fixation duration*: the length of a given fixation or a sequence of fixations reported in milliseconds;
- *dwell time*: a combination of the total duration of all fixations as well as saccades in a given AOI;
- *progressive saccades*: the raw numerical count of forward saccade occurring from left to right (here, in the direction of reading) in a particular AOI;
- *length of forward saccade*: the length of a given forward saccade reported in degrees of visual angle in a particular AOI;
- *regressions:* the raw numerical count of fixations occurring in the opposite direction to what is expected from linear reading (here, from right to left) in a particular AOI.

Reading Index of Dynamic Text (RIDT)

The RIDT (Kruger & Stein, 2014) assesses the amount of visual processing of dynamic text. It provides a reliable 0-1 index of how much text is processed in each caption, as it

was specifically designed to account for its dynamic nature. It is based on the number of unique fixations per standard word in any given caption and the average forward saccade length made by the viewer on each caption divided by the length of the standard word in the text as a whole (Figure 5). This measure is designed to estimate a unique visual information intake, thus it penalizes for specific re-fixations (when a saccade between two successive fixations is shorter than two characters) and regressions (only regressions longer than two characters are taken into account). The more text a viewer processes, the larger the index score, so that a viewer fixating on all the words in a given AOI in a consecutive manner would obtain a score approaching 1. This measure was used to explore the relation between the amount text processing and the pronunciation gains obtained from each task through a series of correlations.

$$\text{RIDT} = \frac{\text{number of unique fixations for } p \text{ in } s}{\text{number of standard words in } s} \text{ x } \frac{\text{average forward saccade length for } p \text{ in } s}{\text{standard word length for } v}$$

(s = subtitle; p = participant; v = video)

Figure 5 Reading Index for Dynamic Text (RIDT) formula (Kruger & Steyn, 2014)

From TobiiStudio software, the data was exported including information about fixation start and end, fixation duration and horizontal and vertical position. Following the data organization procedure (Kruger & Steyn, 2014), the extracted data was sorted by subject, then AOI, and then fixation start in order to calculate the following measures: number of fixations, number of re-fixations, number of regressions, saccade direction and length. Information regarding standard word count was calculated manually based on subtitle script. The length of the character in pixels was determined manually by

obtaining the caption frame dimension on the *x*-axis divided by the total amount of characters (including spaces) per each AOI.

4.4.4 Individual differences

This section presents the tasks used to assess individual differences in attention, phonological short-term memory and proficiency. As described in the literature review, in the context of multimodal exposure benefits it was considered essential to assess learners' cognitive differences in auditory selective attention (i.e. ability to attend to auditory cues), attention switching between three input modalities (audio, text, image) and phonological short-term memory and proficiency level. The subsections present each task separately.

L2 Proficiency

An elicited imitation task (Ortega et al., 2002) was used to obtain an estimate of inter-learner differences in L2 proficiency because it has been shown to discriminate well between different proficiency levels (Solon, Park, Henderson & Dehghan-Chaleshtori, 2019; Yan, Maeda, Lv, & Ginther, 2016). Scores from this task rely heavily on processing and automaticity, as learners are required to auditorily process and repeat back sentences of increasing grammatical complexity. This allowed the researcher 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 (see Appendix C). 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, Mackey & Plonsky, 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. The maximum score that could be obtained in the elicited imitation task is 120 (30 sentences x 4 points).

Attention switching

This novel task was based on the alternating runs paradigm (Monsell, 2003) and it was used to assess the ability of learners to switch attention between the auditory and written form of words. Participants' attention switching ability was tested to control for the potential contribution this cognitive ability might have on the treatment gains. In this task, each trial consisted of a square divided into 4 cells, one of which contained an avatar's face (male or female) and a female or male name written below it, which was presented on the computer screen simultaneously with an audio file of a recorded male or female voice saying the avatar's name (see Figure 6).

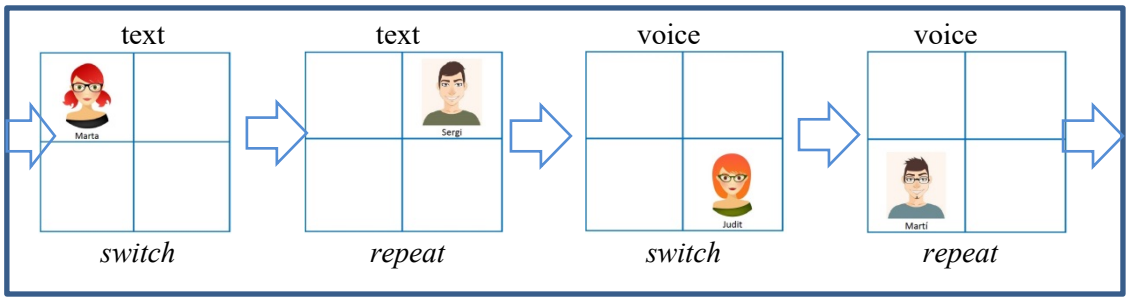


Figure 6 Example of a 4-trial run in the attention switching task

Such layout is attempting to reproduce the complex information processing demands of watching a captioned video (the image in the background, the caption at the bottom of the screen and the soundtrack). In the version of the task participants performed they were instructed to focus on the avatar's name (i.e. the text, T) whenever the visual cue (the avatar's face) appeared on any of the two top cells, and on the soundtrack (i.e. the voice, V) whenever the visual cue appeared on any of the two bottom cells and to decide whether the text or the audio identified a female or a male avatar. They were always asked to press the same left key for a "female" response (*Alt*) and the same right key for a "male" response (*AltGr*). It was therefore a forced-choice two-alternative decision task. Trials were presented in a clockwise fashion (**TTVVTTVVTTVV**...) creating sequences of switch (**TV**,**VT**) and repeat (**TT**,**VV**) trials in a predictable order. Participants were expected to obtain higher error rates and response latencies on switch than on repeat trials because switching between the text and the voice dimensions was expected to result in accuracy and response latency costs due to the difficulty participants may experience in re-allocating their attention on a different perceptual dimension, either the avatar's name (reading) or the speakers' voice (listening). There were 6 female and 6 male faces, and 6 female and 6 male names written below the faces spoken by 6 different male and female voices. The written names and the corresponding audiofiles were always gender-congruent as were the avatar's face and the voice saying the names, just as in watching captioned video the person on the screen speaks with his or her own voice and the spoken message coincides with what appears written in the caption. However, in order to avoid participants' correct identification be depend excessively on the basis of the image, 50 % of the trials in the reading condition were incongruent with the written name (i.e. a male name appeared below a female face. and

vice-versa). The same applied to the voice listening condition. A measure of attentional flexibility based on the response latency and accuracy switching costs (i.e. the difference in RT and accuracy between repeat and switch trials) was obtained for each individual participant.

Auditory selective attention

Participants' auditory selective attention was tested to control for the potential contribution this cognitive ability might have on the treatment gains, as those with better attention to auditory details could have a potential advantage on this type of treatment. Auditory selective attention task (Humes, Lee & Coughlin, 2006) has been found to predict L2 learners' performance on L2 sound discrimination task and phonetic training (e.g., Mora & Mora-Plaza, 2019; Moreira de Oliviera, 2020). Participants performed this task based on single-talker competition (Humes, Lee & Coughlin, 2006) in the L1 (Catalan) and in their L2 (English). The task consisted in listening to 64 trials of pairs of sentences (target vs. competitor) presented simultaneously. The two sentences in a pair were always different, one spoken by a male voice and one by a female voice (e.g. male: Ready CHARLIE go to BLUE SIX now; female: Ready TIGER go to RED EIGHT now). In every trial, a call signal (e.g. TIGER) appearing on the screen previous to the auditory presentation of the sentence cued the voice participants had to attend to for correctly identifying 1 of 4 colours and 1 of 8 digits visually presented on the screen (Figure 7). All the Catalan and English sentences were normalized for duration to 1700ms. Individual ASA scores were computed by adding up all correctly identified colours and digits (2 points for each trial) to a maximum score of 128 for each version of the task.

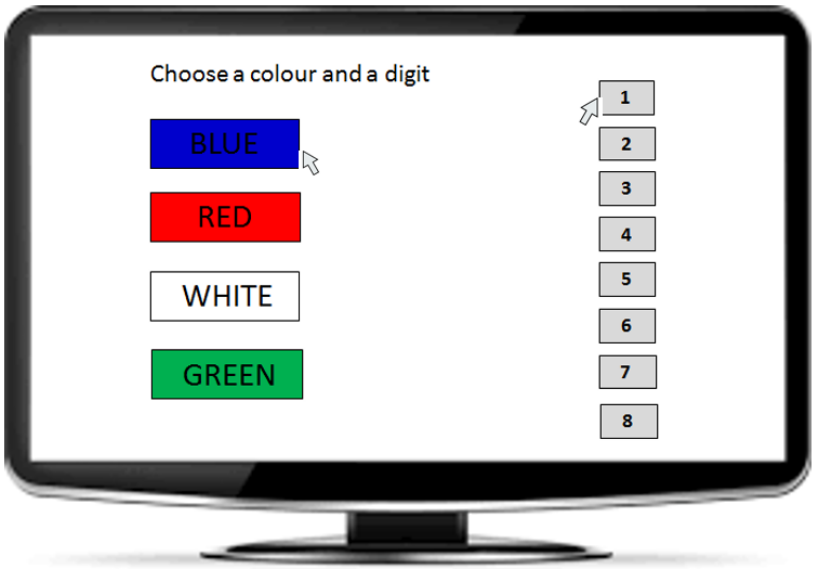


Figure 7 Auditory selective attention task display.

Phonological short-term memory

To assess PSTM, a serial nonword recognition task employing Danish nonwords developed by Cerviño-Povedano and Mora (2011) was used (see also Isaacs & Trofimovich, 2011). The task consisted in identifying whether the two sets of stimuli presented auditorily were identical or not. The use of nonwords in an unfamiliar language (Danish) forced participants to map phonetically unfamiliar sound sequences to their own phonology during nonword sequence recall. This more closely resembles what learners do when decoding L2 speech and still provides a valid phonological memory capacity measure unbiased by inter-subject differences in proficiency (French & O'Brien, 2008). Since the participants in this study were Spanish-Catalan bilinguals differing in how much they used Spanish or Catalan on a daily basis, the use of nonwords in an unfamiliar language was meant to neutralize the likely effect of inter-subject differences in degree of L1 dominance on the phonological memory measure.

Participants were presented with 24 trials for same-different discrimination, each consisting of a pair of sequences of Danish consonant-vowel-consonant (CVC) nonwords. Half of the pairs (12) consisted of two identical sequences of nonwords ('same pairs'); the other half (12) consisted of two sequences of nonwords that were identical except for a change in the serial position of one of the nonwords in the second sequence ('different pairs'). The sequences differed in item length (sequences of five, six and seven nonwords) with four same and four different pairs of sequences at each item length. For example, on any given trial participants would hear a pair of sequences of nonwords and they would need to decide whether the two sequences were the same; that is, whether the same items appeared in the same order (e.g., /tys, dam, rød, mild, fup/ vs /tys, dam, rød, mild, fup/) or in a different order (e.g., /tys, dam, rød, mild, fup/ vs /tys, dam, mild, rød, fup/). Each sequence was built in such a way that no vowel appeared more than once in the sequence and the nonwords varied maximally in their initial and final consonants. The 24 pairs of sequences were presented to participants in three blocks of five-, six-, and seven-item lengths. Participants were instructed to decide whether the nonwords were presented in the same or a different order by pressing one of two designated keys. DmDx was used to run the SNWR task and record subjects' responses. Simuli were presented via headphones at a rate of 750ms, sequences within a pair were separated by a 1500ms silence and sequence pairs were presented after a 500ms delay following the subject's response. The task was preceded by a short practice phase of 2 same order and 2 different-order sequence pairs. The number of correctly identified same/different nonword sequence pairs was used as a measure of phonological short-term memory (O'Brien et al. 2006. 2007; Aliaga-Garcia, Mora & Cerviño-Povedano, 2010). A weighted score was computed to reflect the greater

difficulty with increasing sequence length: Correct responses at sequence length 5 were assigned a score of 5, correct responses at sequence length 6 were assigned a score of 6, and correct responses at sequence length 7 were assigned a score of 7, for a maximum weighted score of 144.

This task provides a measure of participants' PSTM capacity by testing their ability to retain sequences of phonological units (speech sounds and their serial order) in their short-term memory; therefore, it is a task that taps into the phonological sound-processing required, for example, in the segmentation of strings of sounds into word-sized units in an L2.

4.4.5 Data Analysis

From the initial pool of one hundred and three candidates, a total of ninety subjects completed the 8-week viewing treatment. Lack of regularity in viewing or not completing all the viewing sessions resulted in exclusion from the study. For this reason, ten participants were excluded. Three participants were excluded due to technical problems resulting in partial data loss.

Prior to the descriptive and inferential data analysis, a screening procedure was run for all the reaction time (RT) measures (sentence verification task, animacy judgemenet, ABX). This data screening consisted in excluding RTs from inaccurate responses and RT values below or above three standard deviations from each subject's mean (3 SDs).

The eye-tracking data was also screened by excluding all recordings for which the recording accuracy was below 80%. Additionally, due to the nature of the study and the experimental design involving pre-/post-test comparison of the eye-movement

behaviour between the two testing times, it was crucial to maintain the recording accuracy homogeneous, i.e., that the difference in the recording accuracy does not exceed 10% between the testing times.

To assess the effects of the viewing treatment on pronunciation gains, a two-step mixed-effects model analysis was run for each task. For L2 speech processing, the first step consisted in running a linear mixed-effects model (LMM) with *Time* (T1, T2) and *Group* (Experimental, Control) and their interactions as fixed factors, including random intercepts for *Subject* and *Item*. For phonological accuracy, the data was fitted to a generalized linear mixed model (GLMM) with *Time* (T1, T2) and *Group* (Experimental, Control) and their interactions as fixed factors, including random intercepts for *Subject* and *Item*. Second step consisted in running a mixed-effects model (LMM for speech processing tasks and GLMM for phonological accuracy tasks) on experimental participants only with *Time* (T1, T2), *Viewing mode* (Captions, NoCaptions) and *Task focus* (FoPF - focus on phonetic form, FoM - focus on meaning) as fixed factors, including random intercepts for *Subject* and *Item*, to assess the effects of the viewing treatments.

The relationship between the amount of text processing and pronunciation improvement in speech processing and phonological accuracy was explored through a series of correlations between RIDT scores and pronunciation gains (difference between T2 score and T1 score).

Chapter 5. Results

This chapter presents the answers to the research questions that guided the design of the study. It consists of four subsections and each one of them lays out descriptive and inferential analyses carried out to provide detailed answers. *Section 5.1* presents the treatment summary and discusses the between-groups comparison regarding the condition under which the viewing treatment took place; *section 5.2* summarizes the viewing treatment outcomes in terms of accuracy scores; *section 5.3* assesses the effect of the treatment on L2 speech processing skills under different viewing modes, whereas *section 5.4* assesses the effect of the treatment on L2 phonological accuracy as a function of task focus condition; *section 5.5* provides analysis of learners' eye-gaze behaviour while watching L2 captioned videos and discusses pre-/post-treatment changes as well as examines the relation between individual differences in on-screen text processing and pronunciation gains; finally, *section 5.6* examines the role of individual differences in proficiency, attentional resources and phonological memory for benefitting L2 pronunciation development from audiovisual exposure. A brief summary of the results is presented at the end of each section.

5.1 Viewing treatment

To ensure the comparability of the treatment results, participants' self-reports on viewing conditions collected in form of a questionnaire were analyzed. The relationship between categorical variables addressed in the post-treatment questionnaire assessing the viewing treatment conditions, leaners' perceived learning and enjoyment (Appendix B) was explored statistically by chi-square tests with Cramér's V indexing the degree of

association between tested variables (see Appendix K). All the experimental groups turned out to be comparable in terms of the device used (X^2 (12) = 10.14, p = 0.606, V = 0.21), reporting having watched the series mostly on a computer screen and using headphones (X^2 (6) = 7.32, p = 0.29, V = 0.22). Similarly, they reported liking the series used in the treatment to a comparable degree (X^2 (9) = 13.866, p = 0.127, V = 0.252) and similar overall engagement with the plot (X^2 (3) = 2.795, p = 0.424, V = 0.194). Despite not having problems with answering treatment questions while watching (X^2(3) = 1.947, p = 0.583, V = 0.162), only those exposed to videos without captions while receiving plot comprehension questions reported that answering questions while watching was distracting (X^2(3) = 8.839, p = 0.032, V = 0.346) when compared to other experimental groups. No between-group differences were observed when participants were asked to assess the usefulness of the activity for pronunciation (X^2(12) = 10.122, p = 0.120, V = 0.265) as well as listening comprehension (X^2(12) = 2.701, p = 0.440, V = 0.191) agreeing on the fact that the viewing activity was helpful for both. Despite not having received any pronunciation focus, the usefulness ratings of participants assigned to meaning-focused groups did not differ from those receiving questions targeting pronunciation explicitly. This lack of differences may be explained by the fact that the voluntary participation in this experiment was presented as an optional activity granting extra credit for the phonetics class research participation component. Participants' answers regarding the role of captions in pronunciation learning differed significantly (X^2(3) = 18.165, p < 0.01, V = 0.495) as a function of the assigned viewing mode. Seventy percent of those exposed to captions claimed that they could focus on pronunciation due to the availability of the on-screen-text. Those viewing the series in the absence of captions claimed that their presence would distract them from focusing

on pronunciation ($X^2(3) = 11.512$, p = 0.009, $V = 0.394$) and that the reading of captions (if available) would tire them out ($X^2(9) = 14.112$, p = 0.003, $V = 0.437$).

5.1.1 Accuracy on responses to treatment questions

Table 3 shows the mean accuracy scores obtained by each experimental group when answering treatment questions on either plot comprehension or pronunciation under the two viewing conditions (see Appendix L for individual accuracy score on each viewing session). Figure 7 shows the progression of mean scores on each viewing session. Accuracy on treatment questions was above chance (0.33), ranging from 40.5 to 96.07 ($M = 74.77$, $SD = 10.68$) across participants, suggesting overall high level of performance and engagement with the activity.

Table 3 Mean accuracy scores on treatment questions

Group	Accuracy [%]	SD
Captions + FoPF	76.67	12.12
No captions + FoPF	77.47	9.15
Captions + FoM	79.67	10.13
No captions + FoM	67.42	8.41

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* x *Task Focus* interaction ($F(3,72) = 4.11$, $p = .046$,

η^2 = .054). The interaction arose because answering questions on plot comprehension in the absence of captions (No Captions + Focus on Meaning) yielded 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* = 76.67; *SD* = 12.12; *95% CI* = 71.69 - 83.76), No Captions + Focus on Phonetic Form (*M* = 77.47; *SD* = 9.15; *95% CI* = 72.80 - 81.63), Captions + Focus on Meaning (*M* = 79.67; *SD* = 10.13; *95% CI* = 72.37 - 82.14). Since treatment questions were designed to match the specific content of each video excerpt, the difficulty level was not identical on each session. Thus, a clear progression pattern was not expected and, as Figure 8 shows, it was not observed. The treatment questions were intended to direct learners' attention to either meaning or phonological form and the accuracy of responses was irrelevant as to the aim of the thesis.

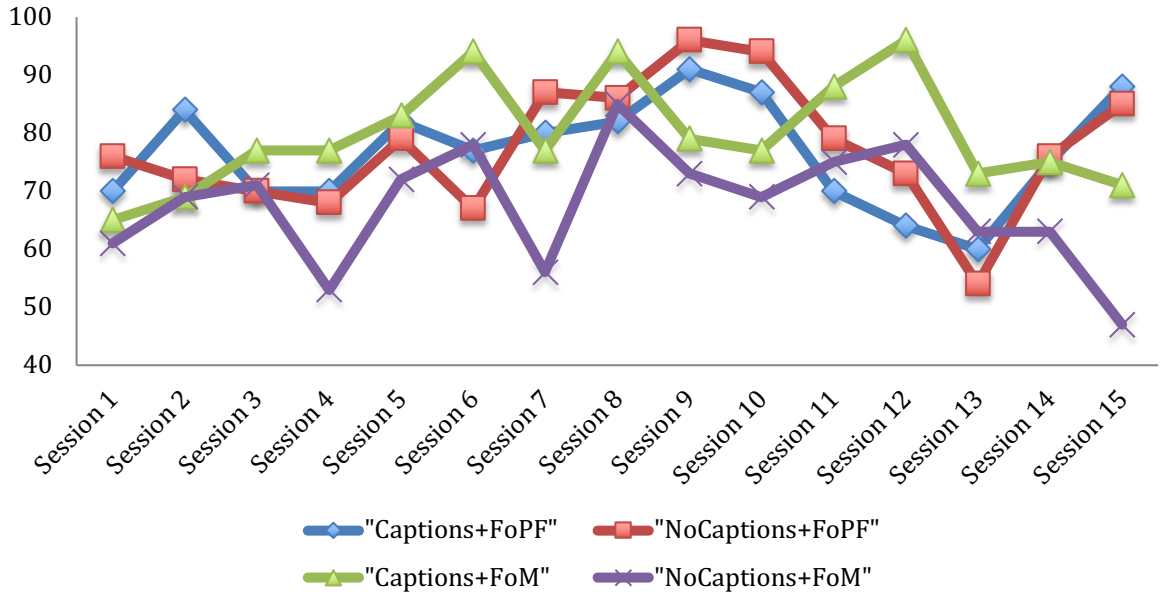


Figure 8 Progression of mean accuracy scores per viewing session

5.2 Treatment effects on L2 speech processing

Treatment effects on L2 speech processing skills were assessed through pre-/post-test changes in L2 learners' performance on the *Shadowing*, *Animacy Judgment* and *Sentence Verification* tasks. The following subsections present the analysis of each task. Firstly, a table showing descriptive statistics is presented, followed by a two-step inferential statistical analysis. For all L2 speech processing tasks, a two-step mixed-effects model analysis was run (see *Section 4.4.5* for details and Appendix M for parameter estimates for all the tests).

5.2.1 Shadowing

Overall, the descriptive statistics in Table 4 indicate positive gains between pre-test and post-test in L2 learners' segmentation skills. All experimental participants obtained larger gains in speech segmentation when compared to the control group. The results show that, after the viewing treatment, experimental participants were able to comprehend and segment between 10 and 15.4% more words when compared to the control participants, who could repeat back 4.2% more words.

Table 4 Pre-/post-treatment mean scores for shadowing task per group

		Shadowing (% words repeated)					
		Trained (Luther)			**Untrained (Sherlock)**		
Mean Scores	Test	**M**	**SE**	**95%CI**	**M**	**SE**	**95%CI**
Captions + FoPF	T1	55.99	1.935	52.18-59.80	61.95	1.807	58.40-65.50
(n = 18)	T2	70.35	1.716	66.97-73.72	70.54	1.667	67.26-73.81
NoCaptions + FoPF	T1	56.42	1.864	52.75-60.08	59.41	1.734	56.00-62.82
(n = 18)	T2	67.23	1.758	63.78-70.69	67.58	1.666	64.30-70.85
Captions + FoM	T1	59.04	1.904	55.30-62.79	60.71	1.691	57.38-64.03
(n = 19)	T2	71.05	1.725	67.66-74.44	71.77	1.462	68.89-74.64
NoCaptions + FoM	T1	53.37	1.970	49.49-57.24	57.29	1.810	53.73-60.84
(n = 16)	T2	68.84	1.771	65.36-72.30	70.02	1.650	66.78-73.26
Control	T1	51.42	2.493	46.51-56.33	53.16	2.356	48.52-57.80
(n = 12)	T2	55.62	2.456	50.78-60.46	60.42	2.284	55.92-64.92
Mean Gain Scores		**Gain**	**SE**	**% Gain**	**Gain**	**SE**	**% Gains**
Captions + FoPF	T2–T1	14.3	1.9	25.6	8.6	1.7	13.8
NoCaptions + FoPF	T2–T1	10.8	1.7	19.1	8.1	1.6	13.6
Captions + FoM	T2–T1	11.9	1.8	20.3	11.0	1.5	18.2
NoCaptions + FoM	T2–T1	15.5	1.9	28.9	12.8	1.7	22.3
Control	T2–T1	4.2	2.5	8.1	7.3	2.3	13.6

The shadowing task measuring speech segmentation included test items from the TV series participants had been exposed to (*Luther*) as well as test items from a series they had not been exposed to (*Sherlock*) to test for generalization of treatment gains to a new context. Table 5 shows mean scores obtained before and after treatment for each experimental condition and the control group. Although on average L2 learners were able to repeat slightly more words in test items at pre-test from the untrained series *Sherlock* (*M* = 58.9, *SE* = 1.8. 95% *CI* = 58.4 – 65.5) compared to the trained series *Luther* (*M* = 55.7, *SE* = 1.9, 95% *CI* = 52.2 – 59.8), gain sizes at post-test reached comparable scores for test items from both series (Sherlock: *M* = 68.6, *SE* = 1.7, 95% *CI* = 67.3 – 73.8 vs. Luther: *M* = 67.4, *SE* = 1.7, 95% *CI* = 66.0 – 73.7).

The first step of the analysis consisted in exploring the effects of treatment between the two testing times for experimental and control participants. A mixed effects model was run with *Time* (TI, T2), *Group* (Experimental, Controls) as fixed effects and *Subject* and *Item* as random factors and *MaxScore* as a covariate. The results showed a significant main effect of *Time* ($F(1, 6753) = 104.4$, $p < .001$) and the main effect of *Group* approached significance ($F(1, 84) = 3.2$, $p = .078$). Importantly, the model revealed that the *Time* x *Group* interaction was also significant ($F(1, 6753) = 12.0$, $p < .001$) (see Appendix M for parameter estimates). These results show that, although there was a consistent main effect of time and both experimental and control participants improved their performance between the two testing times, the Bonferroni-adjusted pairwise comparisons confirmed that those exposed to the viewing treatment (experimental group) obtained significantly higher scores than the control group at T2 ($p = .021$), but not at T1 ($p = .249$). In fact, the mean difference indicated that the

experimental participants improved twice as much as the control group (Experimental group's Mean Difference*:* 11.64; Control group's Mean difference*:* 5.75).

The next step consisted in testing for generalization effects. In a linear mixed-effects model, which was run on the experimental participants' data only, the following variables were included as fixed factors: *Series* (Luther, Sherlock), *Time* (T1, T2), *Viewing mode* (captions, no captions) and their interactions; *Subject* and *Item* constituted random intercepts. In this and all subsequent statistical analyses for this task, *MaxScore* (the maximum number of words a participant could repeat back on a given test trial) was included 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* ($F(1, 5911) = 216.8$, $p < .001$), but neither *Series* ($F(1, 5911) = 3.3$, $p = .078$) 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. Since there was no difference in participants' performance on the trained and untrained items, in subsequent analyses the items from both series (i.e. trained and untrained items) were used.

Finally, in order to test for treatment effects on speech segmentation skills, we ran a mixed-effects model with *Time*, *Viewing mode* and *Task focus* and their interactions as fixed effects (with *Subject* and *Item* as random intercepts and *MaxScore* a covariate). The main effects of *Viewing* mode ($F(1, 70) = 0.585$, $p = .447$) and *Task Focus* ($F(1, 70) = .009$, $p = .925$) did not reach significance, but the main effect of *Time* ($F(1, 5803) = 342.4$, $p < .000$) did. Additionally, the *Time x Viewing* interaction was not significant ($F(1, 5803) = .055$, $p = .815$), similarly to *Time x Task Focus* ($F(1,$

5803) = 3.44, p = .063) and the triple *Time x Task Focus x Viewing interaction* (F(1, 5803) = 3.26, p = .071), which also did not reach significance. 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 or while focusing on form vs. meaning. All experimental participants significantly outperformed the control group by obtaining higher gains on trained and untrained items. Both viewing modes appear to have affected speech segmentation skills positively. Similarly, the task focus did not differentiate between the gains on the Shadowing task.

5.2.2 Animacy judgement

Table 5 shows descriptive statistics for pre- /post-viewing treatment scores on the task measuring speed of lexical access. These results indicate overall reduction in response latencies after the viewing treatment. On this task, measuring accuracy serves the exclusive purpose of indicating participants' engaging with the task while making sure the answers they provide are not at chance level. The improvement in accuracy was not expected since all participants were familiar with the words used as stimuli, for which assigning the animate or inanimate value was not expected to pose any challenge per se (e.g., assigning inanimate value to the word "table" or an animate value to a word "cat"), thus on this task, a significant improvement in accuracy was not expected. Neither the experimental groups nor the control improved their accuracy between the two testing times, scoring almost at ceiling at pre-test and post-test (*Time:* F(1, 87) =3.97, p = .530; *Group*: F(1, 90) =3.34, p = .565). On this task, the reduction of response latencies indicates improvement in speech processing.

Table 5 Pre-/post-treatment mean RT and accuracy score for animacy judgement task

Mean Scores	Test	Response Latency [RT] M	SE	95%CI	Accuracy M	SE	95%CI
Captions + FoPF	T1	786.25	26.37	731.07-841.44	.92	.01	.90-.95
(n = 18)	T2	768.31	31.48	701.89-834.73	.93	.01	.92-.095
NoCaptions + FoPF	T1	782.32	30.09	689.34-815.31	.91	.01	.89-.94
(n = 18)	T2	767.05	31.06	701.81-832.30	.91	.01	.88-.93
Captions + FoM	T1	738.04	22.82	690.28-785.80	.90	.01	.87-.92
(n = 19)	T2	719.65	23.77	669.72-769.59	.92	.01	.90-.94
NoCaptions + FoM	T1	726.39	15.39	694.18-758.60	.89	.01	.87-.92
(n = 16)	T2	711.92	19.87	670.17-753.66	.91	.01	.89-.93
Control	T1	765.78	33.19	693.45-838.10	.92	.01	.90-.94
(n = 12)	T2	745.04	29.92	679.85-810.22	.92	.01	.89-.94
Mean Gain Scores		**Gain**	**SE**	**% Gain**	**Gain**	**SE**	**% Gain**
Captions + FoPF	T2–T1	21.8	16.1	2.7	.009	.03	.62
NoCaptions + FoPF	T2–T1	20.9	30.0	2.4	.004	.06	2.2
Captions + FoM	T2–T1	24.2	14.3	3.1	.002	04	.40
NoCaptions + FoM	T2–T1	14.1	10.4	2.0	.010	.05	2.9
Control	T2–T1	20.7	18.5	2.1	.001	.05	-.75

For the Animacy Judgement task, the same two-step analysis was run. First, a mixed effects model was run to assess the effects of fixed factors *Time* (T1, T2) and *Group* (Experimental, Control) and their interaction on response latencies. This analysis revealed no main effects of *Time (F*(1, 87) =1.77*, p* = .187) or *Group* (*F*(1, 90) = .109*, p* = .742), and the interaction between *Time x Group was also* non-significant *(F*(1, 87) =1.47*, p* = .702) suggesting that the experimental participants did not improve significantly between the two testing times when their performance was compared to the control group (see Appendix M for parameter estimates).

The second step of the analysis consisted in testing the effects of the viewing mode and task focus on the speed of lexical access after the exclusion of the control participants who did not undergo the viewing treatment. We ran a mixed-effects model with *Time*, *Viewing mode* and *Task focus* and their interactions as fixed effects (with *Subject* and *Item* as random intercepts. The results revealed no main effects of *Time (F*(1, 72) = 1.42*, p* = .238), *Viewing* (*F*(1, 76) = .109*, p* = .742) or *Task Focus (F*(1, 76) = 3.57*, p* = .062). Additionally, none of the interactions reached significance. Although the descriptive statistics may suggest that L2 learners exposed to captioned videos obtained larger gains (2.7 - 3.1%) than those exposed to uncaptioned videos (2 - 2.4%), the group difference did not reach significance at either T1 or T2. Overall, the improvement in speed of lexical access between the two testing times is negligible (see parameter estimates in Table 2, Appendix M).

5.2.3 Sentence verification

Table 6 shows descriptive statistics for mean scores obtained on this task. Overall, the descriptive statistics indicate positive gains between pre-test and post-test in L2

learners' speed of speech processing at a sentence level. 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.

Table 6. Descriptive statistics for sentence verification task

Mean Scores	Test	Response Latency [RT] M	SE	95%CI	Accuracy M	SE	95%CI
Captions + FoPF	T1	2268.67	13.24	2229.75-2281.74	0.83	.013	.81-.86
(n = 18)	T2	2159.72	12.46	2130.38-2179.29	0.88	.012	.85-.90
NoCaptions + FoPF	T1	2191.50	11.70	2173.82-2219.79	0.86	.013	.83-.88
(n = 18)	T2	2101.38	11.97	2077.20-2124.22	0.88	.012	.85-.90
Captions + FoM	T1	2211.88	12.50	2173.34-2222.43	0.82	.014	.79-.84
(n = 19)	T2	2137.53	12.30	2108.79-2157.10	0.84	.013	.81-.87
NoCaptions + FoM	T1	2222.74	12.05	2194.47-2241.80	0.82	.014	.79-.85
(n = 16)	T2	2072.79	11.19	2046.26-2090.20	0.86	.013	.83-.88
Control	T1	2188.89	16.10	2135.35-2198.66	0.81	.017	.78-.85
(n = 12)	T2	2120.72	14.30	2072.65-2128.87	0.85	.016	.82-.88
Mean Gain Scores		**Gain**	**SE**	**% Gains**	**Gain**	**SE**	**% Gains**
Captions + FoPF	T2–T1	-116.6	24.2	5.0	.04	.01	6.0
NoCaptions + FoPF	T2–T1	-101.1	16.9	4.6	.01	.01	2.6
Captions + FoM	T2–T1	-77.0	15.9	3.4	.03	.02	5.1
NoCaptions + FoM	T2–T1	-139.6	18.1	6.2	.04	.01	5.2
Control	T2–T1	-68.1	37.3	2.9	.04	.01	5.4

In the first step of the analysis, the effects of treatment between the two testing times for experimental and control participants were explored. The results revealed a significant main effect of *Time* ($F(1, 5814) = 134.6$, $p < .001$), non-significant effect of *Group* (i.e. experimental vs. control) ($F(1, 85) = .181$, $p = .672$) and a significant *Time* x *Group* interaction *Time* ($F(1, 5814) = 4.5$, $p = .033$). Bonferroni-adjusted pairwise comparison confirmed that, despite the fact that all participants (i.e. experimental and control) were able to assign correct semantic coherence value faster at post-test, the mean difference was greater for those who underwent the viewing treatment (*Mean* (Experimental group *Mean difference:* 112.96; Control group *Mean difference:* 77.83).

The second step of the analysis consisted in testing the effects of the viewing mode and task focus on the speed of sentence processing after the exclusion of the control participants who did not undergo the viewing treatment. The results showed a main effect of *Time* ($F(1, 4960) = 324.5$, $p <.001$), but neither *Viewing* ($F(1, 70) = 2.63$, $p =.109$), $p = .109$) nor *Task Focus* $F(1, 70) = .63$, $p = .430$) reached significance. The interaction between *Time* and *Viewing* was approaching significance ($F(1,4960) =3.47$, $p = .063$), whereas the triple interaction between *Time* x *Viewing* x *Focus* was significant ($p < .001$) (see Appendix M for parameter estimates). Although there were no differences at pre-test between captioned vs. uncaptioned condition, Bonferroni-adjusted pairwise comparison revealed that, after treatment, experimental participants who viewed the series without captions could assigned a semantic coherence value (thus, process sentences) twice as fast when compared to those who watched it with captions ($p = .053$). Although the difference in means between the captioned and uncaptioned groups focusing on form at pre-test was larger than the difference in means between captioned and uncaptioned group focusing on meaning (FoPF $p = .083$ vs.

FoM: p = .966), at post-test, these differences were comparable (FoPF: p = .212 vs. FoM: p = .134). These results suggest that although all participants significantly improved in speed of sentence processing from pre-test to post-test, those exposed to the TV series without captions obtained overall higher gains (5.4% uncaptioned vs. 4.2% captioned).

5.2.4 Summary

This section examined the effects of the viewing mode on L2 speech processing skills. The results revealed that, irrespective of the viewing mode, all the participants who underwent the viewing treatment could comprehend and repeat back on average 23.4% more of the words from the incoming auditory input. Irrespective of the viewing mode or task focus, L2 learners who viewed the series obtained significant speech segmentation gains (twice as high), when compared to leaners who did not watch the series. Moreover, the speech segmentation benefit obtained through the exposure to the TV series (Luther) could be generalized and seem to have helped participants to extend this skill to new materials from a previously unfamiliar TV series (Sherlock).

The results from the Shadowing task showed that, once again, irrespective of the viewing mode, experimental participants obtained significantly greater gains in the task measuring efficiency of sentence processing between testing times than the control group. Overall, these results suggest that after the extended viewing of the series, L2 learners could process sentences significantly faster, but those assigned to uncaptioned viewing mode obtained higher gains.

As for the efficiency of lexical access, the gains obtained between testing times were negligible overall, as none of the experimental groups outperformed the control. The gains obtained on the Sentence Verification task, which dealt with sentence-level processing as opposed to word-level processing, were much larger.

The benefits for L2 speech processing skills identified suggest that extended viewing of TV series in the target language, both with and without captions, might extend to L2 speech perception and production, leading to improvement in phonological accuracy. As suggested by previous research (Mitterer & McQueen, 2009), TV viewing may serve as a kind of perceptual training, which may lead to retuning of phonological categories and overall improvement of L2 speech. The following section examines whether this is true by assessing the effect of viewing mode and task focus conditions on L2 phonological accuracy in perception (L2 sound discrimination) and production (accentedness ratings).

5.3 Treatment effects on L2 phonological accuracy

Treatment effects on L2 phonological accuracy were assessed through an *ABX* categorical discrimination task and a delayed sentence production task (accent-rating). The following subsections present the analysis of each task. Firstly, a table showing descriptive statistics is presented followed by two-steps inferential statistical analysis. For both tasks, the data was fitted to a set of mixed-effects models (see *Section 4.4.5* for details and Appendix L for parameter estimates for all the test).

5.3.1 ABX

Table 7 shows that ABX discrimination accuracy and RT gains ranged considerably between experimental groups. The results of this task correspond to the average scores L2 learners on the test items only (/iː/-/ɪ/ contrast), as the control trials (/a/-/iː/ contrast) were excluded for the main analysis. 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 ($t(5993) = 4.39$, $p < .001$).

Table 7 Descriptive statistics for ABX task

			ABX Discrimination					
			Accuracy			RT (ms)		
Mean Scores			M	SE	95%CI	M	SE	95%CI
Experimental (n = 77)		T1	.649	.013	.62–.67	1244	25	1197–1293
		T2	.672	.013	.65–.70	1163	23	1119–1209
Captions	Form	T1	.660	.026	.61–.71	1271	50	1177–1372
(n = 38)	(n = 19)	T2	.679	.026	.63–.73	1197	47	1108–1293
	Meaning	T1	.586	.027	.53–.64	1291	51	1195–1394
	(n = 19)	T2	.643	.027	.59–.69	1178	46	1090–1272
NoCaptions	Form	T1	.668	.026	.62–.72	1267	51	1171–1370
(n = 39)	(n = 19)	T2	.704	.025	.65–.75	1218	49	1126–1317
	Meaning	T1	.674	.025	.62–.72	1173	45	1089–1264
	(n = 20)	T2	.657	.026	.60–.71	1080	42	1002–1165
Control (n = 13)		T1	.599	.033	.53–.66	1260	62	1144–1389
		T2	.636	.033	.57–.70	1207	59	1096–1329
Mean Gain Scores			Gain	SE	% Gain	Gain	SE	% Gain
Experimental		T2–T1	.023	.034	3.54	81	23	6.96
Captions	Form	T2–T1	.019	.014	2.88	74	9	6.18
	Meaning	T2–T1	.057	.027	9.73	113	19	9.59
NoCaptions	Form	T2–T1	.036	.028	5.39	49	20	4.02
	Meaning	T2–T1	–.017	.027	–2.52	93	18	8.61
Control		T2–T1	.037	.027	6.18	53	17	4.39

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

In the first place, the mixed-effects model analysis with *Time* (T1, T2), *Group* (Experimental, Control) and their interactions as fixed factors (with *Subject* and *Item* as random factors) was run on response latencies (reaction time) and accuracy separately. For accuracy, no main effects of *Time (*$F(1, 5788) = 2.8$, $p = .096$), *Group (*$F(1, 5788) = 2$, $t = 1.2$, $p = .157$) or their interaction ($F(1, 5788) = .118$, , $p = .731$) were found, which suggests lack of significant improvement between the two testing times.

As for response latencies on correct discrimination between the two vowel forming the contrasts, the results revealed significant main effects of *Time* ($F(1, 3710) = 29.1$, $p < .001$), but no significant effects or *Group* $F(1, 3710) = .233$, $p = .629$) or *Group* x *Time* interaction $F(1, 3710) = 1.36$, $p = .244$). Bonferroni-adjusted pairwise comparisons showed that, although all leaners were significantly faster at correctly discriminating between the target vowels at post-test, the mean difference between testing times was 33% faster in the experimental than the control group (*Mean Difference* Experimental = 80.9 vs. *Mean Difference* Control = 53.5). These results suggest that the experimental participants were providing correct responses faster than the controls.

To assess the effects of the treatment conditions on response accuracy, a mixed-effects model with *Time, Viewing, Task Focus* and their interactions was run. No main effects were found for *Time (*$F(1, 4920) = 3.1$, $p = .077$), *Task Focus (*$F(1, 4920) = 3.2$, $p = .072$), or *Viewing (*$F(1, 4920) = 2.5$, $p = .115$). Additionally, none of the interactions reached significance (*t*s >.7). Bonferroni-adjusted pairwise comparison revealed that none of the experimental groups improved their accuracy on ABX task significantly (*p*s >.2) with the exception of those learners who watched the series with captions while focusing on meaning ($p = .043$).

5.3.2 Accent rating

Table 8 shows mean accent rating scores obtained by each experimental group.

Table 8 Descriptive statistics for accent ratings

			Accent Rating (1–100)		
Mean Scores			M	SE	95%CI
Experimental (n = 77)		T1	53.9	1.6	51–57
		T2	52.2	1.6	49–55
Captions	Form	T1	53.0	3.3	46–60
(n = 38)	(n = 19)	T2	52.4	3.3	46–59
	Meaning	T1	54.3	3.2	48–61
	(n = 19)	T2	51.5	3.2	45–58
NoCaptions	Form	T1	53.3	3.3	47–60
(n = 39)	(n = 19)	T2	50.4	3.3	44–57
	Meaning	T1	54.9	3.2	49–61
	(n = 20)	T2	54.3	3.2	48–61
Control (n = 13)		T1	58.4	4.0	50–66
		T2	56.4	4.0	48–64
Mean Gain Scores			Gain	SE	% Gain
Experimental		T2–T1	1.70	0.8	3.21
Captions	Form	T2–T1	0.59	1.7	1.09
	Meaning	T2–T1	2.76	1.7	5.18
NoCaptions	Form	T2–T1	2.90	1.7	5.28
	Meaning	T2–T1	0.56	1.7	0.96
Control		T2–T1	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

A mixed-effects model was run on the accent rating scores (see Table 2 in Appendix M for model coefficients). The results showed that the main effect of *Time*

approached significance ($F(1, 434) = 3.74$, $p = .054$), but the interaction involving *Time* and *Group* did not ($F(1, 434) = .07$, $p = .780$). The Bonferroni-adjusted pairwise comparisons showed that only experimental participants who underwent the treatment significantly improved their accent rating scores ($p = .032$) from pre-test to post-test, whereas participants assigned to the control group did not ($p = .231$). Their accent ratings obtained at pre-test did not differ ($p = .381$).

The subsequent mixed-effects model run after the exclusion of control participants confirmed a significant main effect of *Time* ($F(1, 367) = 4.63$, $p = .032$), but neither *Viewing* ($F(1, 71) = .024$, $p = .878$) nor *Task Focus* ($F(1, 71) = .233$, $p = .631$) or their interactions reached significance. Bonferroni-adjusted pairwise comparisons revealed larger accent rating differences between pre-test and post-test to occur under two of the four viewing conditions, i.e., 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 ($t(516) = 1.58$, $p = .068$) (see parameter estimates in Table 2., Appendix L).

5.3.3 Summary

This section examined the effects of the viewing treatment on L2 phonological accuracy in perception and production. The results revealed that neither the viewing mode, nor task focus had large effects on phonological accuracy, as assessed through the categorical discrimination of a difficult L2 vowel contrast in the ABX task. Overall, the observed gains in the sensitivity to the tested contrast were small and only one experimental group (those exposed to captions while focusing on meaning) improved significantly and to a larger extent than the control group did.

For phonological accuracy in production, a trend was observed in the analysis of L2 learners' accent rating. Only two experimental groups outperformed the control group: greater accent reduction occurred in the uncaptioned viewing mode in the pronunciation-focused condition as well as in the captioned viewing mode in the meaning-focused condition. These between-group differences were only approaching significance ($p > 0.05$). The results seem to suggest that 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 or induced focus on pronunciation.

5.4 Eye-tracking

This section presents the analysis of eye-movement data. Eye-gaze data corresponding to the longest clip (Clip 1) were used in the computation of global measures and the Reading Index of Dynamic Text, as these data were estimated to be sufficient to provide an individual measure of reading behaviour.

Table 9 Descriptive statistics [mean (SD)] for global eye-tracking measure

Groups		N fixations	Total time on captions [s]	Fixation duration on captions [ms]	Total time on video [s]	Fixation duration on video [ms]	N fixations per caption
Cap+FoPF	T1	285.00 (130.56)	66.96 (31.55)	227.69 (31.52)	99.09 (41.99)	511.67 (160.89)	5.7
	T2	229.75 (120.58)	52.94 (29.72)	227.41 (25.20)	108.99 (37.01)	514.17 (115.01)	4.6
NoCap+FoPF	T1	292.06 (65.05)	64.67 (15.56)	223.57 (11.60)	100.53 (19.57)	461.87 (83.30)	5.84
	T2	208.87 (102.90)	47.79 (23.22)	227.12 (24.80)	109.93 (30.40)	505.63 (129.36)	4.18
Cap+FoM	T1	242.77 (92.12)	54.18 (21.09)	218.65 (16.23)	111.42 (19.31)	478.46 (88.02)	4.86
	T2	194.23 (91.84)	42.40 (20.69)	214.75 (20.67)	120.55 (16.92)	489.23 (81.79)	3.88
NoCap+FoM	T1	292.57 (130.80)	65.94 (30.41)	222.09 (31.59)	101.07 (35.68)	472.86 (102.01)	5.85
	T2	211.00 (151.66)	45.79 (32.55)	223.32 (28.49)	115.97 (33.97)	505.71 (142.33)	4.22
Ctrl	T1	290.56 (118.98)	61.81 (25.97)	216.31 (29.78)	94.46 (43.06)	426.67 (164.32)	5.81
	T2	224.78 (147.31)	49.94 (32.83)	229.41 (36.44)	113.01 (43.64)	475.56 (164.40)	4.5

Table 9 Descriptive statistics [mean (SD)] for global eye-tracking measure (continuation)

Groups		N forward saccades	Forward saccade length [pixels]	N regressions
Cap+FoPF	T1	119.25 (57.84)	95.22 (60.25)	56 (29.18)
	T2	89.67 (60.71)	79.13 (62.45)	37.42 (26.34)
NoCap+FoPF	T1	130.9 (30.9)	119.36 (31.84)	59.13 (21.32)
	T2	90.47 (46.11)	112.06 (52.42)	45.67 (25.38
Cap+FoM	T1	109.75 (41.69)	104.97 (66.3)	45.42 (26.93)
	T2	80.5 (38.66)	76.78 (62.83)	36.33 (24.96)
NoCap+FoM	T1	150.23 (59.05)	121.21 (30.89)	57.38 (29.67)
	T2	97.15 (74.78	82.39 (74.17)	41.85 (29.83)
Ctrl	T1	115.64 (53.96)	113.71 (50.22)	55.45 (24.73)
	T2	91.27 (76.3)	83.44 (56.36)	42.82 (31.52)

5.4.1 Global eye-tracking measures

The descriptive analysis of global eye-tracking measures (Table 9) shows that all participants had fewer fixations on captions (M = -66.86, SD = 13.82) when watching the video clips for the second time. The reduction in fixation count is marked especially for both experimental groups who underwent the uncaptioned viewing treatment as the fixation count decreased more (M = -82.38, SD = 0.81) as compared to those exposed to captions (M = -51.89. SD = 3.35) or the control group, who did not undergo the treatment (M = -65.78). A similar pattern is visible for the total dwell time spent on captions, as it is reduced for all participants (M = -14.94, SD = 3.19), but again the reduction of time spent on captions is especially marked for the two groups exposed to uncaptioned treatment (M = -18.51, SD = 1.63), who on average spent up to 20 seconds less looking at captions when compared with those who underwent captioned treatment (M = -12.9, SD = 1.12) or controls (M = -11.87, SD = 1.14).

The lack of easily noticeable consistent pattern in terms of mean fixation duration on captions was not observed. Although, the changes in mean fixation duration seem small enough to be neglected, this interpretation should be taken with caution as the analysis was performed exclusively on the sentence-level and does not take into account lexical properties of each word (e.g., frequency, length). Solid empirical evidence confirms the existence of systematic differences in the ease (or difficulty) of processing words in the text (see Rayner & Duffy, 1986 for overview). In reading, shorter words, more frequent words, and more predictable words receive shorter fixations when compared to longer, less frequent or less predictable words (see Rayner & Duffy, 1986; Rayner, Ashby, Pollatsek, & Reichle, 2004). These effects are robust and are also tied to lexical processing. The study by Liao et al. (in press) confirmed

word frequency effects in the reading of subtitles in terms of both gaze duration and total time although the effect size decreases with increasing subtitle speed. Thus, the lack of observed differences in fixation mean fixation duration between the two testing times might be a result of methodological shortcoming (no word-level data).

Results indicate that participant attended more to the image at the second viewing, which is confirmed by overall longer dwell time on the video.

The descriptive analysis suggests a consistent pattern in how the reading behaviour changed between testing times. For all participants, the observed change is characterized by the reduction in fixation count and the total amount of time spent on the captions as well as increased time spent looking on the image. Nevertheless, despite this consistent trend in the changes in reading behaviour between the two testing times, the individual data analysis shows highly variable inter-learner variation. To illustrate it, Figure 9 shows the range of individual variation in terms of fixation count on the captions. The higher the number, the larger the reduction in fixation count between testing times.

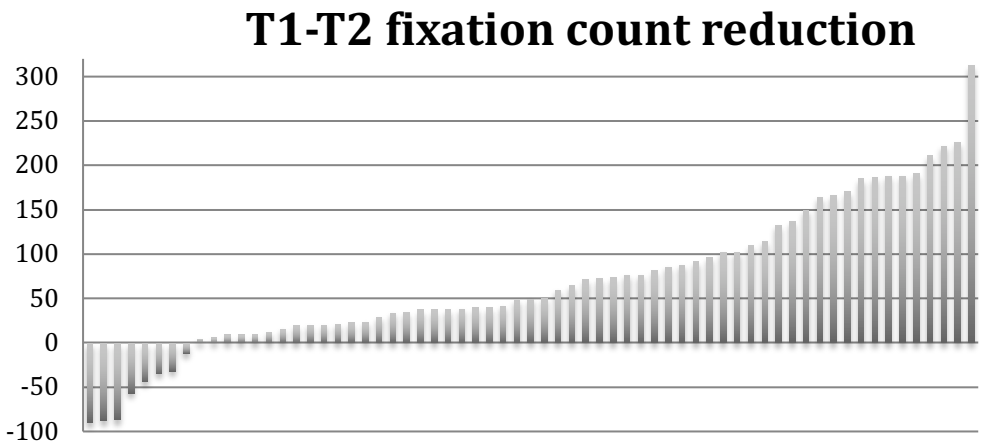


Figure 9 Fixation count reduction for each participant

The following section presents detailed analysis of individual dynamic text reading behaviour and how it relates to the pronunciation gains obtained from the treatment.

5.4.2 Reading Index of Dynamic Text (RIDT)

To assess how much text in each caption was read, an index score was calculated for each subtitle and for each participant and averaged across the video. The RIDT score correlated negatively with learners' proficiency (r = .-261, n = 63, p = .039), suggesting that the learners with higher proficiency level, read less captions than those with lower proficiency level. The analysis revealed a wide range of mean RIDT scores, ranging from 0.01 to 0.93 (within 0-1 index), confirming high inter-learner variability in the amount of processed text (Figure 10). Table 10 shows mean scores obtained by each experimental group. Overall, all participants had a lower index at the second viewing of the clip (i.e., after treatment), however the difference was the largest for two uncaptioned groups (NoCap +FoPF: M = -0.192; NoCap + FoM: M = -0.251) when compared to the captioned groups (Cap + FoPF: M = -0.145; Cap + FoM: M = -0.104). For participants assigned to the control group the difference between how much on-screen text was read at pre- and post-test was the smallest (M = -0.078). Descriptive data suggests that all participants skipped more text when viewing the clip for the second time and this difference was larger for those, who underwent uncaptioned treatment.

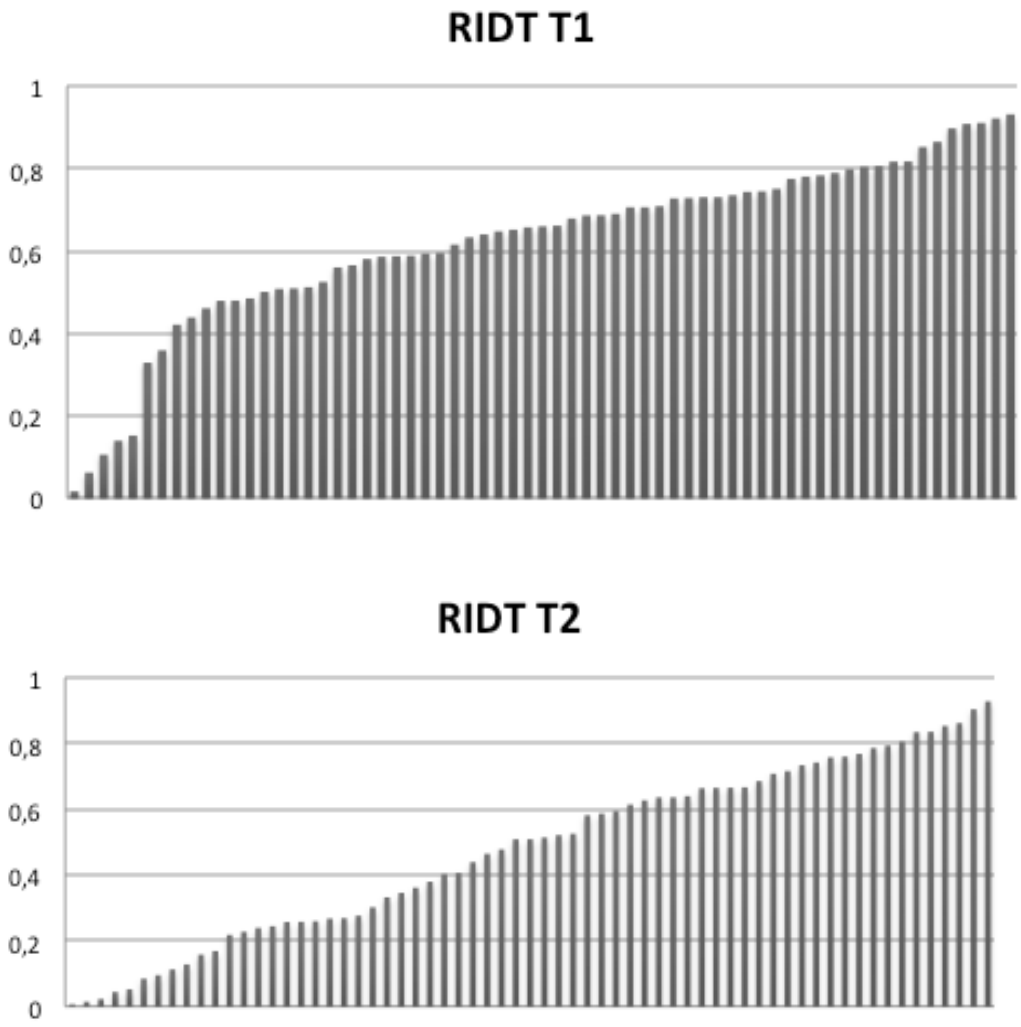


Figure 10 Mean RIDT score by subject

Table 10 Reading Index of Dynamic Text. Pre-/post-treatment mean scores per group

Group	RIDT T1 (*SD*)	RIDT T2 (*SD*)
Captions + FoPF	0.614 (0.285)	0.469 (0.272)
NoCaptions + FoPF	0.660 (0.126)	0.468 (0.222)
Captions + FoM	0.572 (0.209)	0.468 (0.221)
NoCaptions+ FoM	0.687 (0.219)	0.436 (0.276)
Control	0.594 (0.209)	0.516 (0.306)

To confirm this trend, an ANOVA with *Time* (T1, T2) as within-subject and *Viewing mode* (Captions, NoCaptions, Control) as between-subject factor and their interaction

was run to assess the effect of viewing treatment on subtitle reading. The dependent measure was the RIDT score The results yielded a main effect of *Time* ($F(1, 60) = 3$ 1.831, $p = .000$), which means that all experimental participants processed significantly less text in captions after being exposed to the treatment. *Time* x *Viewing* interaction *(*$F(2, 60) = 3.052$, $p = .055$) approached significance. The interaction was approaching significance because those participants exposed to the uncaptioned treatment obtained significantly lower reading indices ($t(54) = 3.850$, $p < .001$) when compared to those who underwent captioned viewing ($t(46) = 1.761$, $p = .085$) or those not exposed to the viewing treatment at all ($t(20) = 0.376$, $p = .711$). These results suggest that although all participants processed less text during the second viewing of the audiovisual materials, those exposed to the uncaptioned viewing condition tend to read less text than others.

The relationship between the amount of text processing and pronunciation improvement was explored through a series of correlations between RIDT scores and pronunciation gains (difference between T2 score and T1 score) on the tasks assessing L2 speech processing and pronunciation accuracy. Firstly, individual RIDT scores obtained by each participant at pre-test (T1) were correlated with gains on each task. The results show no relation between RIDT scores at T1 and gains in shadowing ($r =$ .113, n = 63, p = .380), sentence verification ($r = .008$, n = 63, $p = .949$), animacy judgment ($r =$.-200, n = 63, p = .118), ABX ($r =$.-097, n = 63, p = .450), or accent rating ($r =$.-025, n = 63, p = .844). Secondly, a similar series of correlations was run between the change in RIDT scores between testing times and the pronunciation gains. The results showed no relationship between the tested variables, which suggests that the greater reduction in text processing alone cannot predict pronunciation gains. Lastly, the final step of analysis consisted in looking at the relation between RIDT score at T1

and the pronunciation gains for each experimental group separately (Table 11). The RIDT score obtained at pre-test is treated here as baselines measure, reflecting individual differences in reading behaviour. For those assigned to the meaning-focused uncaptioned treatment, the results revealed a moderately strong correlation (r = .577, n = 13, p = .043) between how much text in the caption was processed and the gains in the animacy judgement task as well as a strong negative correlation between the RIDT scores and the gains from ABX task (r = -.812, n = 13, p = .001), suggesting that those who tend to read more, had greater gains in the speed of lexical access, but significantly lower gains in phonological accuracy when captions were not available.

Table 11 Correlations between T1 RIDT score and pronunciation gains

Group	Shadowing	Animacy judgment	Sentence Verification	ABX	Accent rating
Captions + FoPF *n=12*	.519 .084	-.373 .232	.132 .682	.016 .961	.001 .998
NoCaptions +FoPF *n=12*	.226 .417	.237 .396	-.279 .315	.001 .998	**.536*** **.039**
Captions + FoM *n=12*	.042 .897	.188 .558	-.229 .475	-.224 .483	**-.677*** **.016**
NoCaptions +FoM *n=13*	-.014 .963	**.577*** **.043**	.189 .537	**-.812**** **.001**	-.192 .530
Control *n=11*	.242 .501	-.045 .816	-.004 .990	.411 .209	.371 .262

Moreover, smaller accent reduction (here higher score equals to stronger foreign accent) for those assigned to pronunciation-focused uncaptioned viewing was significantly correlated with higher RIDT scores (r = .536, p = .039), whereas for the group assigned to the opposite experimental condition (meaning-focused captioned viewing), the reverse correlation (r = -.677, n = 12, p = .016) was found, where higher RIDT scores were significantly correlated with less accented speech. These results seem to suggest

that the participants who tend to read more text in captions, benefitted their pronunciation accuracy even when the explicit focus on pronunciation was absent, whereas the lack of captions did not result in accent reduction despite the availability of explicit pronunciation focus.

5.4.3 Summary

This section assessed the changes in captions reading between the two testing times and examined the relationship between individual reading behaviour and L2 pronunciation gains obtained from the viewing treatment. The descriptive analysis of the global eye-tracking measures revealed that all participants had fewer fixations on captions at the second viewing and attended more to the video image, which resulted in longer dwell time. This difference was the largest for those participants assigned to uncaptioned viewing condition. Statistical analysis with individual RIDT scores as the dependent measure yielded a *Time x Viewing* mode interaction approaching significance supporting the trend in the reading reduction observed in the descriptive analysis of the data. The correlation analysis between the individual reading behaviour and L2 pronunciation gains showed that those who read more and were assigned to the captioned viewing condition without explicit pronunciation focus received significantly better foreign accent ratings, whereas the speech of those who read more, but were assigned to uncaptioned viewing condition with explicit pronunciation focus received lower ratings. The results suggest an interesting interplay between individual reading behaviour and potential improvement in function of the viewing mode as well as explicit pronunciation focus.

5.5 Individual differences

This section presents the analysis of the tasks used to assess individual differences in attention control, phonological short-term memory and proficiency. The primary objective of this section is to examine the potential influence these individual differences might have had on the gains obtained from the viewing treatment. Firstly, the results of each task are presented descriptively. Secondly, correlation analyses between the scores from each task and speech processing and phonological accuracy gains are presented and discussed. Only correlations above .4 are presented and discussed. Table 12. presents descriptive statistics for participants assigned to the experimental groups only (control group is excluded).

Table 12 Descriptive statistics on tasks measuring individual differences

Measure	*Mean*	*SE*	*SD*	*Min*	*Max*	N
Attention switching *(Shift cost, milliseconds)*	178.19	13.64	99.32	-52.77	376.68	53
Auditory selective attention *(Acc, N* correct*)*	95.77	1.74	16.39	50	122	88
Phonological short-term memory *(Acc, N* correct*)*	92.20	2.13	19.97	24	144	88
Proficiency *(%)*	76.27	1.5	14.31	54	119	84

Note: Acc = Accuracy score

SE = Standard Error of the mean

SD = Standard Deviation

5.5.1 Attention switching

For each participant, two measures of attentional flexibility were calculated based on the response latency and accuracy switching costs (i.e. the difference in RT and accuracy between repeat and switch trials). As expected, learners' were more accurate on congruent trials (M = 95.93, SD = 4.76), i.e. when female or male picture corresponded to either female or mail name, as opposed to incongruent trials (M = 82.17, SD = 12.52), and faster on repeat trials (M = 974.69, SD = 213.69) as opposed to switch trials (M = 1138.98, SD = 245.41). These differences reported for accuracy on congruent vs. incongruent trials and reaction time on repeat vs, switch were significant ($p < .001$). On repeat trials, participants were overall faster when paying attention to text (M = 949.21, SD = 365.23) as opposed to when paying attention to voice (M = 983.32, SD = 350.96), however switching to text resulted in slightly higher switching cost (M = 1163.70, SD = 408.44) than switching to voice (M = 1100.49, SD = 378.43). Although there was no overall difference between the groups in terms of average switching cost latency ($F(4, 84) = 1.127$, $p = .350$), there was a large variability in terms of learners' performance on the repeat and switch trials.

The relation between attentional switching and speech processing and pronunciation accuracy gains was explored through a series of correlations. This analysis did not confirm any relationship between these variables with the exception of the gains from ABX task, where higher scores were weakly correlated with lower switching costs (r = -.215, p = 0.44, n = 88).

5.5.2 Auditory selective attention

This task assessed participants' ability to selectively attend to the specific cues presented auditorily. Mean accuracy scores obtained on the Catalan version of the task (M = 96.54, SD = 14.17) did not differ from those obtain on the English version (M = 95.77, SD = 16.39), which were also highly correlated (r = .737, p = .001 n = 84), suggesting the task reliably assessed the same cognitive skill. There were no between group differences in mean accuracy scores (F(4, 87) = 1.252, p = .296) obtained on this task. On average, participants identified correctly 74% of the targets, but there was noticeable inter-learner variability with scores ranging between 50 and 122 points.

The relation between auditory selective attention and pronunciation gains was explored through a series of correlations. The analysis revealed that the learners who were more successful in selectively attending to the right auditory cues were also those who had larger gains in speech segmentation (r = .258, p = .017, n = 86).

5.5.3 Phonological short-term memory

On the task assessing phonological short-term memory, participants correctly recognized identical sets of stimuli on 64% of the trials on average. Mean response accuracy decreased with the length of the stimuli sequence (5-nonwords: M = 5.46, SD = 1.67; 6-nonword: M = 5.15, SD = 1.39; 7-nonwords: M = 4.87, SD = 1.63), confirming that longer trials were more difficult, and therefore yielded more errors. To verify whether the groups were different in terms of phonological short-term memory, a one-way ANOVA was run with the weighted score as the dependent measure and group as the independent between-subject factor. Overall, there were no between groups differences observed in terms of accuracy scores when identifying identical trials (F(4, 86) = 0.370; p = 0.829) and the all experimental groups obtain similar scores on 5-

nonword sequence ($F(4, 86) = .410$; $p = .801$), 6-nonword sequence ($F(4, 86) = .697$; $p = .596$) and 7-nonword sequence ($F(4, 86) = .377$; $p = .824$). Similarly to the previous tasks, despite the lack of between group differences, there was large variability in terms of individual performance, with the weighed score ranging from the lowest 24 to the highest score of 144 points.

The relation between PSTM and pronunciation gains was explored through a series of correlations. No relation between the scores from this task and the gains from any task measuring pronunciation learning were found.

CHAPTER 6: Discussion

6.1 Introduction

This doctoral dissertation has explored the potential benefit of TV exposure for L2 pronunciation development. It has attempted to address gaps in the literature by (1) comparing the effects of two viewing modes (with and without captions) over an extended time of exposure; (2) has investigated whether the previously reported benefit of lexically guided perceptual learning (Mitterer and McQueen, 2009) could lead to the retuning of phonetic categories in L2 speech perception and production, (3) has explored the effect of form versus meaning focus while watching TV series and its interaction with two viewing modes, (4) has explored the relationship between subtitle processing and pronunciation gains and lastly, (5) has examined the relation between individual differences and L2 pronunciation gains from multimodal exposure.

This chapter provides a comprehensive overview of the main findings and offers the thorough interpretation of the results. It also highlights the pedagogical implications of the findings, acknowledges the limitations of the study and provides suggestions for future research. The main six findings from this dissertation are briefly summarized below:

1. Extended exposure to authentic audiovisual materials in English can benefit L2 pronunciation development of advanced EFL learners when the video content is presented either with or without captions.
2. The initial advantage of captioned over uncaptioned viewing, reported in previous studies, was not observed for L2 speech processing gains in this study.

Extended viewing seems beneficial for boosting these skills, irrespective of the viewing mode. The length of exposure could have mitigated the previously reported advantage of captions, as longer exposure allows learners to familiarize themselves with the accents and speaking style, boosting segmentation skills and diminishing the initial advantage of captions.

3. L2 pronunciation gains were much more prominent on tasks assessing sentence-level and not word- or segment-level outcomes.
4. For pronunciation accuracy, the results showed an interplay between the viewing mode and task focus condition, revealing that explicit focus on pronunciation maybe be crucial for boosting L2 speech production when viewers are exposed to uncaptioned video materials. However, the availability of captions may compensate for the lack of explicit focus, allowing for incidental pronunciation enhancement to some degree.
5. Caption reading behaviour of L2 learners is highly individualized, as indicated by a wide range of RIDT scores and it also seems to be altered through regular viewing treatment with specific characteristics.
6. Correlational analysis indicated a relation between the amount of subtitle processing and foreign accent reduction, which seems to be moderated by the viewing condition. The results indicated that learners who tend to read more text in subtitles are more likely to obtain better accent rating when exposed to captioned viewing, whereas accent reduction for those who read more is less likely if captions are not available.

6.2 Multimodal exposure effects on L2 speech processing

The main findings of this dissertation indicate that the extensive viewing of TV series in the second language enhances learners' speech processing skills. After the three-month exposure to the audiovisual materials, participants could not only comprehend and segment on average 13% more of the spoken input, as indicated by the results from shadowing task, but were also able to process speech faster and more efficiently, as suggested by the results of the sentence verification task. Gains in segmentation and speech processing at sentence-level extended to significantly improved performance on unfamiliar materials unrelated to the treatment, suggesting the generalization and applicability of these enhanced skills in other contexts.

Previous research on multimodal exposure (Charles & Trenkic, 2015; Mitterer & McQueen, 2009) attributed speech segmentation gains to the facilitating role of the on-screen text, reporting that it was through its availability that perceptual learning could happen. However, the findings from this dissertation showed that segmentation skills of all experimental participants benefitted from the TV exposure to almost an identical extent, regardless of whether the audiovisual materials were viewed with or without captions. This might be explained by differences in the implementation of the treatment and the testing procedures in this 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. Moreover, the participants were exposed only to a few sessions, ranging from 1 to 4. In contrast, in the current study design 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 viewing, 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, this dissertation showed that extensive exposure to 15 excerpts from the same TV series is likely to have resulted in progressive 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 over time. Additionally, the relatively high level of proficiency of the participants might also explain the lack of group differences between the two viewing modes. Thus, it appears that the advantage of captioned conditions is present when immediate comprehension of unfamiliar speech is required and when L2 learners do not have enough time to gain sufficient acquaintance with the speech.

This dissertation also showed that the benefits for L2 pronunciation development were visible not only in speech segmentation skills, but also in processing efficiency at sentence level in a task that did not use materials related to the treatment. The results from the sentence verification task showed that after the viewing treatment, all experimental participants were faster in assessing semantic coherence of aurally presented sentences. Such benefits were found for 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 in light of transfer-appropriate processing is also possible for the results of the Animacy Judgment task, where learners assigned to the captioned groups obtained slightly larger gains in response latency, possibly due to the visual presentation mode of this task. Importantly, the gains in speed of lexical access were rather small and for some of the groups may be even considered negligible overall.

Given all the evidence, for relatively advanced SLA leaners of English, extended exposure to original language audiovisual materials can lead to more robust L2 speech processing at the sentence level. The previously reported advantage of captions may be equalized in the context of prolonged exposure when the allocated exposure time allows for becoming familiar with the speech.

6.3 Multimodal exposure effects on phonological accuracy

Another goal of this dissertation was to test Mitterer and McQueen's (2009) prediction regarding the extent to which lexically guided perceptual learning could lead to the retuning of L2 learners' phonetic categories - a prediction based on the interpretation of gains in the shadowing task in their study. In order to investigate this issue, participants' phonological accuracy in perception and production was tested before and after treatment. The results showed that the viewing mode did not have large effects on phonological accuracy in perception, which was assessed through the categorical discrimination of a difficult L2 vowel contrast (ABX). In fact, gains in sensitivity to the contrasting L2 sounds tested were inconsistent, as 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* x *Viewing* interaction, the 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. It is important to mention that this task focused exclusively on a single phonological dimension (accuracy and speed in the categorical discrimination of the /iː/-/ɪ/ contrast), which is one of the most difficult contrasts for Spanish speakers and it could not have possibly captured benefits for other perceptual dimensions.

Accuracy gains in production were assessed through foreign accent ratings before and after the viewing treatment, where a trend was observed in the ratings of L2 learners' speech samples. Greater accent reduction occurred in the uncaptioned viewing mode in the pronunciation-focused condition and as 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 explicitly 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.

Participants exposed to audiovisual materials with captions, while receiving detailed questions on phonology, had to deal with captions reading while focusing on the pronunciation of words in the auditory input. Thus, it is possible that, with captions, a focus on pronunciation did not allow learners to devote enough attentional resources to the processing of the auditory input. In the context of multimodal exposure, optimizing the processing efficiency is crucial and the demand posed by the constant

multitasking might have been overburdening learners' working memory. On the other hand, in an uncaptioned viewing mode, 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. In the same line of interpretation, with the lack of captions, the explicit focus on pronunciation seems to enhance viewers' focus on the speech, which may result in beneficial pronunciation development.

To sum up, as for the role of the TV exposure for L2 pronunciation benefits in production, this data did could not confirm that the lexically guided perceptual learning could lead to updating of phonetic categories, either due to the specificity of the tested dimension or not sufficient time of exposure. Nevertheless, the task used to assess potential benefits of pronunciation accuracy in perception was focused exclusively on a single categorical dimension and could not capture any other improvement beyond segmental level in perception. A more holistic measure of L2 perception might have proved more appropriate to capture potential treatment effects. For production, it seems that the multimodal exposure with captions could serve as pronunciation training enhancing accuracy even without explicit pronunciation focus. However, when the captions are not available, an explicit pronunciation focus might be required to promote L2 pronunciation development.

6.4 Captions reading and L2 pronunciation improvement

This dissertation was the first to explore the link between pronunciation gains and multimodal exposure, relating the amount of subtitle processing to the viewing

treatment outcomes. Overall, the amount of text read before and after the viewing treatment varied significantly for all participants, who also read less text upon the second viewing. The difference in the amount of text processed between the two testing times was the largest for the experimental groups exposed to the uncaptioned viewing mode compared to those who watched the TV series with captions as well as the control group. This difference was indicated by the global eye-tracking measures through lower fixation count, shorter dwell time on captions and, most importantly, significantly lower RIDT scores. The two viewing modes affected how much text was read at the second viewing, as indexed by differentiated mean RIDT scores, which confirms that the results obtained cannot be explained merely by the fact that the participants saw the materials twice. Otherwise, no difference in the amount of processed text would have been observed as a function of the assigned viewing condition. Although the amount of processed text (presented in a multimodal context) is inevitably moderated by the nature of audiovisual materials and viewer-related factors, the results seems to suggest that, it can be also influenced by regular exposure leading to less dependency on captions. It is it possible that it is due to improved familiarization with speakers' voices and accents along the viewing exposure.

Overall, the analysis of subtitle reading patterns at an individual level indicated that the amount of on-screen-text processed varies substantially, as suggested by a large range of RIDT scores. Although the initial exploration of the data through correlations did not show any relation between RIDT scores (measured before treatment) and pronunciation gains, when the specificity of the viewing treatment was accounted for (i.e., when the between-groups comparison included both the viewing mode and the task focus condition as opposed to the comparison between the viewing mode only), an interesting interplay was revealed. This interplay indicated a relation between the

amount of subtitle processing and foreign accent reduction, which seems to be moderated by the assigned viewing condition - those who tend to read more text in captions were more likely to obtain better accent ratings when exposed to captioned viewing, whereas accent reduction for those who tend to read more is less likely if captions are not available. When watching TV series in English, L2 learners who rely more on captions may benefit their L2 pronunciation more when the text is available even without explicit pronunciation focus. On the other hand, if learners tend to process more text in the captions, and therefore opt for relaying on their presence, they may benefit their pronunciation to a smaller degree.

The final point to be raised in regard to the correlational analysis concerns the strongest negative correlation, which suggests that learners who read more text in captions and had been assigned to uncaptioned viewing with no explicit pronunciation focus obtained the smallest gains in the task assessing their sensitivity to a phonetic contrast (ABX). Interestingly, for learners assigned to captioned viewing, but still without the focus on pronunciation, the strength of this negative correlation diminished, and for those who watched the materials with captions and with the added focus on phonetic form the relation between reading and the gains became gradually stronger. Nevertheless, the strength of these correlations did not reach significance (with one exception) resulting in lack of robustness of this finding, therefore the trend, although worth mentioning, should be taken with cautious. Future research should either expand on the total exposure time or opt for more robust testing of perceptual gains in order to be able to shed more light on multimodal exposure validity for segmental accuracy and beyond.

6.5 Role of individual differences in the context of multimodal exposure

Measuring L2 learners' cognitive individual differences in attention control, phonological short-term memory and proficiency was crucial for the design of this study. Its inclusion allowed for the closely examination of the role they may have in the context of multimodal exposure on the gains obtained from the viewing treatment.

Overall, the results did not reveal any clear pattern that would indicate an interaction between the cognitive skills tested in this study and pronunciation gains in the context of multimodal exposure. The correlational analysis showed that learners who were more successful in attending to the right auditory cues (auditory selective attention task) obtained larger gains in speech segmentation. Similarly, those who are better at switching their attention between different modalities obtained higher scores in the task measuring sensitivity to phonetic contrast (ABX). Nevertheless, the strength of these correlations was weak (< 0.4). These findings suggest that the benefits obtained from the extended, regular exposure to TV series could occur irrespective of individual differences in auditory selective attention, learners' attention switching capacity or proficiency. Although the results from this dissertation showed that the difference in proficiency did not significantly affect the gains obtained from the viewing treatment, it is important to keep in mind that such experimental design could not inspect the issue of proficiency in details. This is due to the fact that all the learners had advance level of proficiency. Provided wider range in proficiency level, its effects on the viewing treatment gains in L2 pronunciation might differ significantly from what the effects observed for the proficient learners participating in this study were.

Chapter 7. Limitations and future directions

This dissertation provided some evidence that the use of audiovisual materials may serve as a tool for enriching the limited language input in instructional settings for L2 pronunciation development. However, it has some limitations that need to be acknowledged.

Firstly, due to the extensive and continuous involvement required to complete the viewing treatment, the final number of participants in each experimental condition was relatively small, especially after exclusion of those who did not complete all sessions or whose eye-tracking data did not pass the recording accuracy threshold. Secondly, obtaining an eye-tracking record for every viewing session implemented in this study was not feasible because it would significantly compromise not only the length, but also the integrity of the treatment. Therefore, maintaining rigorous control of every viewing session was not possible within this study design. However, strict revision of log times (how much time each participant spent on the Playposit platform used for the implementation of the treatment) allowed the researcher to discriminate between those who did not comply with the participation policy. Although, possibly more faithful to the real environment of multimodal exposure on daily bases, the lack of strict control of the conditions under which the viewing took place is a common shortcoming of studies in which treatment is not performed in a lab setting.

There are some other limitations relative to the implementation of the treatment. Plot comprehension was tested only for those participants who were assigned to the meaning-focused groups; those answering pronunciation-focused questions did not receive plot comprehension questions, as this would have doubled the task demands by introducing another confound, compromising the between-groups comparability.

However, the post-treatment assessment of engagement with the viewing activity did not reveal any differences between groups, as all experimental participants claimed that they liked the series and felt engaged with the activity to the same degree.

This dissertation was the first to relate L2 pronunciation gains to how much text in captions is processed. The resolution of the eye-tracker prevented reliable word-level data analysis, allowing for reporting of global eye-gaze measures only. Future research could extend the design by relating pronunciation gains to local (word-level) eye-tracking measures. Additionally, continuous tracking of potential changes in reading patterns on a word-level basis could further improve the understanding of how multimodal exposure benefits L2 pronunciation development and highlight the necessity of word occurrence frequency and saliency.

Another limitation concerns the selection of the instruments through which gains in pronunciation were assessed. As the viewing treatment was extensive and required prolonged engagement from the students participating in this study, the battery of tests, although complete, could not be extended any further. Deciding on the final battery of tests was a challenging task, since previous literature on L2 pronunciation assessment in the context of captioned video was scarce. This was especially true for phonological accuracy tests. Thus, the test selection was not informed by previous research, which made it difficult to expand the testing of suprasegmental features. The test selected to assess pronunciation gains in perception was the ABX task. As mentioned in section 6.3, this task could have been too specific and therefore could not capture any benefits beyond the sensitivity to one phonetic contrast. Testing more phonetic features or selecting a test that measures phonological accuracy on a word or sentence level could have been more appropriate and more informative.

Lastly, although participation in this study was voluntary, it was offered to students undertaking undergraduate studies in fields related to languages or linguistics. This means that all the participants were undertaking a course in phonetics/phonology during the duration of the viewing treatment. Although the focus on pronunciation was made explicit exclusively to those who were assigned to pronunciation-focused groups, the rest of the participants (those assigned to the meaning-focused groups) were also given the opportunity to participate in this experiment as a part of their phonetics/phonology course. Thus, regardless of the task focus condition assigned to, all participants were to some extend aware that this research is related to second language pronunciation.

Chapter 8. Final remarks

The present dissertation investigated the effects of exposure to L2 audiovisual materials on L2 pronunciation under different viewing modes and task focus conditions. The design of this study extended previous research by testing both L2 speech processing and L2 phonological accuracy in speech perception and production as well as by relating learners' pronunciation gains obtained through the viewing treatment to how much of the on-screen text in captions was processed. The results revealed that L2 learners with advanced proficiency level improved their 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. Thus, on the bases of results from this dissertation is seems that viewing audiovisual materials in the target language may benefit second language processing skills.

For L2 production, the results indicated an interplay between the viewing mode and the task focus condition. A focus on phonetic form resulted in more pronounced gains in the uncaptioned viewing mode, but not when the videos were watched with captions. When watching the series with captions pronunciation gains were visible when learners' attention was directed to the plot. Previous research has shown that audiovisual exposure in target language can lead to incidental vocabulary learning, especially when watching captioned videos (e.g. Peters &Webb, 2018). The results from this dissertation shed some light on the issue of incidental learning for L2 pronunciation in production, suggesting that it may occur only in the presence of captions. No evidence of incidental accent reduction for learners viewing uncaptioned videos was observed. This could imply that some kind of intentional focus on phonetic form while

paying attention to the speech in the soundtrack might be necessary for pronunciation to develop through exposure to audiovisual materials. It is essential to consider the relatively small amount of input (15 x 20-minute videos = 5 hours of exposure) implemented in this study when interpreting the results. In the context of the experimental time frame, benefits on L2 pronunciation may appear to depend on the presence of an intentional focus on pronunciation. However, it is plausible that a longer or more intense exposure to audiovisual materials could lead to greater positive outcomes of the treatments.

NOTES

[1] In the context this dissertation, extended exposure refers to the regular viewing of audivisual materials over a three-month period, involving a delayed post-test occurring one week after the viewing treatment has finished.

[2]Throughout this dissertation the term "multimodal" and "bimodal" input are used 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 language presentation is auditory (speech) and visual (written text and image).

Appendix A. Pre-test questionnaire

Qüestionari: Aprenentatge de l'anglès

Benvolguts/des estudiants,

Us estaríem molt agraïts si ens poguéssiu ajudar tot contestant les següents preguntes sobre l'aprenentatge del l'anglès. El Grup de Recerca GRAL de la Universitat de Barcelona porta a terme aquest estudi per tal d'entendre millor el rol d'exposició i contacte amb l'anglès fora de l'aula. El qüestionari té una durada de 5 minuts. Tota la informació proporcionada es tractarà de forma CONFIDENCIAL i només s'utilitzarà amb finalitats de recerca.

Moltes gràcies per la vostra participació!

- **Quina és la frase que millor descriu els teus els teus hàbits de veure sèries o pel·lícules en versió original o amb subtítols en anglès?**

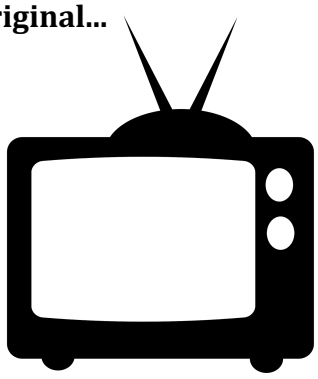
1.) Veig moltes sèries o pel·lícules en versió original (o amb subtítols en anglès) - si m'enganxo, no paro.
2.) Segueixo algunes sèries en versió original (o amb subtítols en anglès).
3.) De tant en tant veig algun capítol, però mai de forma regular.
4.) Mai veig sèries en versió original o amb subtítols en anglès.

- **Veus pel·lícules i sèries en anglès en versió original?**

	Mai	Entre 1-3 cops / mes	Entre 1-3 cops / setmana	Entre 4-6 cops / setmana	Cada dia
Amb subtítols en català/castellà					
Amb subtítols en anglès					
Sense subtítols					

- **T'agrada veure pel·lícules i sèries en anglès en versió original...**

	Gens	No gaire	Un mica	Bastant	Molt
Amb subtítols en anglès					
Sense subtítols					



❖ **Has vist alguna d'aquestes sèries? Si l'has vist...**

- en versió original **sense** subtítols escriu **VO**
- en versió original **amb** subtítol en anglès escriu **VOS**
- **doblada** o **amb** subtítol en català/castellà escriu **D**

13 Reasons Why		*American Gods*		*American Horror Story*		*Ballers*		*Big Bang Theory*	
Big Little Lies		*Black Mirror*		*Breaking Bad*		*Dexter*		*Doctor Who*	
Family Guy		*Friends*		*Game of Thrones*		*Homeland*		*House of Cards*	
How I Met Your Mother		*Lost*		*Luther*		*Mad Men*		*Modern Family*	
Mr. Robot		*Narcos*		*Orange is the New Black*		*Outlander*		*Peaky Blinders*	
Prison Break		*Sense8*		*Sherlock*		*Silicon Valley*		*Six Feet Under*	
Sons of Anarchy		*Stranger Things*		*The Crown*		*The Handmaid´s Tale*		*The IT Crowd*	
The Mentalist		*The Simpsons*		*The Sopranos*		*The Walking Dead*		*The Wire*	
True Detective		*Twin Peaks*		*Vampire Diaries*		*Vikings*		*Westworld*	

❖ **Quines altres sèries veus habitualment o has vist en versió original en anglès?**

.....

.....

.....

❖ **Per què les mires en anglès?**

.....

.....

.....

Appendix B. Post-test questionnaire

Questionnaire

Group 1 and 3

*Required

1. Email address *

_____

2. Name and Surname *

_____

3. What group did you belong to? *

Mark only one oval.

- ◯ 1
- ◯ 3

How did you watch the videos?

4. 1. On what device where you watching the videos? *

Mark only one oval.

- ◯ Always on a mobile phone
- ◯ Always on a computer
- ◯ Always on a tablet
- ◯ On different devices, but predominantly computer
- ◯ On different devices, but predominantly mobile
- ◯ Other: _____

5. 2. Did you use headphones while watching the videos? *

Mark only one oval.

- ◯ I always used them
- ◯ I sometimes used them
- ◯ I never used them

6. 3. Did you watch the sessions regularly? *

Mark only one oval.

- ◯ I watched the sessions regularly and have seen ALL all of them on time
- ◯ I watched the sessions regularly and have seen MOST of them on time
- ◯ I wasn´t regular with my watching, but I have seen ALL the sessions
- ◯ I missed some sessions

What did you learn?

7. 4. In general, do you think the exercise of watching the series was helpful for your English skills? *

Mark only one oval per row.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

Tell us if you think you have improved on the following aspects:

8. 5.

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
learning new vocabulary					
pronunciation					
listening comprehension					
grammar					
spelling					

9. 6. Did you watch the series WITH or WITHOUT subtitles

Mark only one oval.

- With
- Without

10. 7. If you could choose, would you rather watch the series WITH or WITHOUT subtitles? *

Mark only one oval.

- With
- Without
- No difference

11. 8. Choose the statements you consider true for you *

Tick all that apply.

- [ ] I needed subtitles to understand the dialog
- [ ] I didn´t need subtitles to understand the dialog
- [ ] Subtitles helped me to focus on pronunciation
- [ ] I found subtitles distracting my focus on pronunciation
- [ ] I wasn´t focusing on pronunciation
- [ ] I got tired of reading subtitles

Did you like the series?

12. 9. In general, did you like the series Luther?

Mark only one oval per row.

	Really dislike	Dislike	Neither like nor dislike	Like	Really like
Your answer	◯	◯	◯	◯	◯

13. 10. Pick the sentences, which best describe your attitude while watching: *

Tick all that apply.

- [ ] I was engaged with the plot
- [ ] I wasn´t engaged with the plot
- [ ] I was paying attention to improve my listening comprehension skills
- [ ] I was paying attention to improve my pronunciation skills
- [ ] I had no problems answering the questions while watching
- [ ] I couldn´t pay attention to answer the questions while watching
- [ ] I found the questions engaging
- [ ] I found the questions distracting

Other: ☐ _____

14. 11. Did you find this activity useful for your pronunciation?

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

15. 12. Did you find this activity useful for practicing listening comprehension?

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

16. Any comments about the series or the activity?

Thank you for your participation!

This is the end of the Research Task!

Questionnaire

Group 2 and 4

*Required

1. Email address *

_____

2. Name and Surname *

_____

3. What group did you belong to? *

Mark only one oval.

- ◯ 2
- ◯ 4

How did you watch the videos?

4. 1. On what device where you watching the videos? *

Mark only one oval.

- ◯ Always on a mobile phone
- ◯ Always on a computer
- ◯ Always on a tablet
- ◯ On different devices, but predominantly computer
- ◯ On different devices, but predominantly mobile
- ◯ Other: _____

5. 2. Did you use headphones while watching the videos? *

Mark only one oval.

- ◯ I always used them
- ◯ I sometimes used them
- ◯ I never used them

6. 3. Did you watch the sessions regularly? *

Mark only one oval.

- ◯ I watched the sessions regularly and have seen ALL all of them on time
- ◯ I watched the sessions regularly and have seen MOST of them on time
- ◯ I wasn´t regular with my watching, but I have seen ALL the sessions
- ◯ I missed some sessions

What did you learn?

7. 4. In general, do you think the exercise of watching the series was helpful for your English skills? *

Mark only one oval per row.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

Tell us if you think you have improved on the following aspects:

8. 5.

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
learning new vocabulary	◯	◯	◯	◯	◯
pronunciation	◯	◯	◯	◯	◯
listening comprehension	◯	◯	◯	◯	◯
grammar	◯	◯	◯	◯	◯
spelling	◯	◯	◯	◯	◯

9. 6. Did you watch the series WITH or WITHOUT subtitles

Mark only one oval.

◯ With
◯ Without

10. 7. If you could choose, would you rather watch the series WITH or WITHOUT subtitles? *

Mark only one oval.

◯ With
◯ Without
◯ No difference

11. 8. Choose the statements you consider true for you *

Tick all that apply.

- [ ] I needed subtitles to understand the dialog better
- [ ] I didn´t need subtitles to understand the dialog
- [ ] I think subtitles would helpe me to focus on pronunciation
- [ ] I think subtitles would distract me from focusing on pronunciation
- [ ] I wasn´t focusing on pronunciation
- [ ] I think I would get tired of reading subtitles

Did you like the series?

12. 9. In general, did you like the series Luther?

Mark only one oval per row.

	Really dislike	Dislike	Neither like nor dislike	Like	Really like
Your answer	◯	◯	◯	◯	◯

13. 10. Pick the sentences, which best describe your attitude while watching: *

Tick all that apply.

- [ ] I was engaged with the plot
- [ ] I wasn´t engaged with the plot
- [ ] I was paying attention to improve my listening comprehension skills
- [ ] I was paying attention to improve my pronunciation skills
- [ ] I had no problems answering the questions while watching
- [ ] I couldn´t pay attention to answer the questions while watching
- [ ] I found the questions engaging
- [ ] I found the questions distracting

Other: ☐ _____

14. 11. Did you find this activity useful for your pronunciation?

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

15. 12. Did you find this activity useful for practicing listening comprehension?

Mark only one oval per row.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
Your answer	◯	◯	◯	◯	◯

16. Any comments about the series or the activity?

Thank you for your participation!
This is the end of the Research Task!

Appendix C. Elicited imitation task stimuli

1. I have to get a haircut
2. The red book is on the table
3. The streets in this city are white
4. He takes a shower every morning
5. What did you say you were doing today?
6. I doubt that he knows how to drive that well
7. After dinner I had a long peaceful nap
8. It is possible that it will rain tomorrow
9. I enjoy movies which have a happy ending
10. The houses are very nice but too expensive
11. The little boy whose kitten died yesterday is said
12. That restaurant is supposed to have very good food
13. I want a nice big house in which my animals can live
14. You really enjoy listening to country music don't you
15. She's just finished paining the inside of her apartment
16. Cross the street at the light and then just continuous straight ahead
17. The person I am dating has a wonderful sense of humor
18. She only orders meat dishes and never eats vegetables
19. I wish the price of town houses would become affordable
20. I hope it will get warmer sooner this year that it did last year
21. A good friend of mine always takes care of my neighbor's three children
22. The black cat that you fed yesterday was the one chased by the dog
23. Before he can go outside, he has to finish cleaning his room
24. The most fun I have ever had was when we went to the opera
25. The terrible thief whom the police caught was very tall and thin
26. Would you be so kind as to hand me the book which is on the table
27. The number of people who smoke cigars is increasing every year
28. I don't know if the 11:30 train has left the station yet
29. The exam wasn't nearly as difficult as you told me it would be
30. There are a lot of people who don't eat anything at all in the morning

Appendix D. Participants' log times and accuracy on each viewing session [min]

Sessions 1-5

100%	17m	✓	86%	26m	✓	75%	20m	✓	100%	21m	✓	86%	25m	✓					
75%	21m	✓	71%	20m	✓	75%	21m	✓	100%	20m	✓	86%	21m	✓					
50%	34m	✓	57%	39m	✓	0%	21m	✓	50%	21m	✓	86%	31m	✓					
75%	19m	✓	71%	21m	✓	75%	23m	✓	67%	19m	✓	29%	22m	✓					
75%	20m	✓	71%	22m	✓	75%	19m	✓	33%	21m	✓	29%	22m	✓					
100%	17m	✓	71%	21m	✓	100%	20m	✓	100%	22m	✓	71%	22m	✓					
75%	20m	✓	57%	21m	✓	50%	19m	✓	100%	24m	✓	57%	23m	✓					
75%	18m	✓	57%	19m	✓	100%	29m	✓	100%	20m	✓	100%	21m	✓					
100%	20m	✓	86%	28m	✓	75%	19m	✓	100%	21m	✓	100%	26m	✓					
100%	19m	✓	86%	22m	✓	100%	22m	✓	83%	19m	✓	86%	21m	✓					
	-		14%	78m	◔	75%	38m	✓	50%	23m	✓	71%	57m	✓					
75%	20m	✓	71%	21m	✓	100%	21m	✓		-			-						

100%	26m	✓	71%	21m	✓	50%	21m	✓	100%	25m	✓	43%	25m	✓
75%	19m	✓	71%	22m	✓	100%	20m	✓	100%	21m	✓	86%	22m	✓
	-			-		75%	19m	✓	67%	21m	✓	71%	23m	✓
100%	20m	✓	86%	22m	✓	25%	19m	✓	83%	21m	✓	57%	21m	✓
75%	18m	✓	57%	21m	✓	100%	19m	✓	100%	23m	✓	71%	21m	✓
75%	19m	✓	71%	27m	✓	75%	22m	✓	67%	24m	✓	71%	24m	✓
	-		14%	78m	◔	75%	38m	✓	50%	23m	✓	71%	57m	✓
75%	20m	✓	71%	21m	✓	100%	21m	✓		-			-	
	-			-			-		33%	20m	✓	57%	21m	✓
100%	22m	✓	71%	22m	✓	75%	20m	✓	50%	22m	✓	57%	31m	✓
100%	18m	✓	71%	21m	✓	100%	19m	✓	100%	20m	✓	57%	22m	✓

Sessions 6-10

100%	19m	✓	100%	20m	✓	57%	18m	✓	100%	19m	✓	91%	21m	✓
100%	33m	✓	100%	18m	✓	57%	32m	✓	57%	28m	✓	64%	18m	✓
100%	20m	✓	100%	51m	✓	100%	23m	✓	100%	23m	✓	100%	23m	✓
83%	19m	✓	100%	19m	✓	100%	20m	✓	100%	20m	✓	100%	20m	✓
	-			-			-			-			-	
100%	25m	✓	100%	19m	✓	100%	19m	✓	43%	21m	✓	55%	21m	✓
	-			-			-			-			-	
100%	25m	✓	100%	19m	✓	100%	19m	✓	43%	21m	✓	55%	21m	✓
	-			-			-			-			-	
100%	38m	✓	100%	20m	✓	100%	20m	✓	100%	20m	✓	55%	29m	✓
100%	19m	✓	100%	18m	✓	86%	19m	✓	71%	19m	✓	91%	21m	✓

33%	22m	✓	50%	32m	✓	71%	40m	✓	71%	24m	✓	73%	44m	✓
100%	22m	✓	100%	20m	✓	100%	18m	✓	86%	18m	✓	64%	19m	✓
67%	44m	✓	100%	36m	✓	86%	58m	✓	86%	20m	✓	91%	22m	✓
33%	18m	✓	50%	18m	✓	43%	20m	✓	57%	19m	✓	27%	23m	✓
100%	47m	✓	75%	21m	✓	57%	19m	✓	57%	21m	✓	45%	32m	✓
100%	19m	✓	75%	18m	✓	86%	18m	✓	71%	18m	✓	73%	19m	✓
100%	17m	✓	100%	18m	✓	100%	34m	✓	86%	21m	✓	100%	23m	✓
83%	25m	✓	100%	20m	✓	100%	19m	✓	100%	19m	✓	91%	24m	✓
	-			-			-			-			-	
100%	18m	✓	100%	18m	✓	100%	19m	✓	100%	19m	✓	100%	21m	✓
100%	20m	✓	100%	18m	✓	100%	17m	✓	71%	18m	✓	64%	24m	✓
83%	21m	✓	75%	25m	✓	86%	20m	✓	71%	24m	✓	91%	23m	✓

Session 11-15

75%	20m	✓	100%	21m	✓	80%	26m	✓	80%	24m	✓	100%	24m	✓
100%	57m	✓	80%	19m	✓	80%	20m	✓	80%	32m	✓	80%	19m	✓
	-			-			-			-			-	
100%	21m	✓	60%	20m	✓	100%	20m	✓	100%	18m	✓	80%	32m	✓
100%	21m	✓	100%	19m	✓	80%	19m	✓	60%	17m	✓	100%	17m	✓
50%	21m	✓	80%	22m	✓	20%	20m	✓	60%	19m	✓	20%	19m	✓
	-			-			-			-			-	
100%	27m	✓	80%	19m	✓	60%	20m	✓	60%	18m	✓	60%	18m	✓
	-			-			-			-			-	
75%	51m	✓	80%	23m	✓	80%	26m	✓	100%	18m	✓	60%	22m	✓
50%	20m	✓	80%	19m	✓	40%	20m	✓	100%	18m	✓	60%	18m	✓

100%	22m	✓	60%	24m	✓	40%	21m	✓	40%	21m	✓	60%	22m	✓
100%	19m	✓	100%	18m	✓	80%	19m	✓	100%	18m	✓	60%	21m	✓
50%	99+m	✓	60%	82m	✓	40%	19m	✓	60%	39m	✓	40%	20m	✓
100%	19m	✓		-		80%	24m	✓	40%	17m	✓	80%	21m	✓
50%	26m	✓	20%	59m	✓	20%	42m	✓	20%	28m	✓	40%	26m	✓
100%	19m	✓	80%	18m	✓	60%	19m	✓	60%	18m	✓	60%	18m	✓
100%	19m	✓	40%	18m	✓	60%	22m	✓	40%	24m	✓	80%	18m	✓
100%	28m	✓	100%	17m	✓	20%	18m	✓	60%	17m	✓	80%	17m	✓
100%	21m	✓	100%	20m	✓	80%	20m	✓	100%	20m	✓	100%	21m	✓
100%	23m	✓	100%	22m	✓	100%	22m	✓	60%	20m	✓	100%	21m	✓
	-			-			-			-			-	
100%	27m	✓	80%	19m	✓	60%	20m	✓	60%	18m	✓	60%	18m	✓

Appendix E. List of Focus on Phonetic Form treatment questions and answers

		Questions
Session 1	1	Is the word ***scene*** pronounced like ***seen*** or ***sin***?
	2	Is the word ***gun*** pronounced similar to...
	3	Which pronunciation of the past tense (*-ed* ending) is different than the rest? Pick the odd one out.
	4	The word ***crime*** rhymes with…Write one word that rhymes below
	5	Which word is pronounced differently than the rest? Pick the odd one out.
Session 2	1	Focus on the way the word *wrong* is pronounced. What do these **two sets of words** have in common when it comes to pronunciation? **Write short answer below.**
	2	The pronunciation of the word evil is more similar to…
	3	Which word is pronounced differently than the rest? Pick the odd one out.
	4	The word ***gesture*** is pronounced alike…Pick the correct set of words.
Session 3	1	Is the word ***perversity*** accented the same way as the word ***unambiguous***?
	2	Are these two verbs (TOUCH**ED,** FIR**ED**) in the past tense pronounced the same or different?
	3	The first vowel in the word oven is pronounced the same as in..
	4	The highlighted part of these two words are pronounced in the same: pr**ou**d and c**ou**ntry
Session 4	1	Some sounds in English have different spelling yet sound the same.
	2	The word *M**o**nday* is pronounced the same as...
	3	Which word in this sentence is pronounced so that it seems a letter is omitted?
	4	The beginning of which words sound alike the beginning of the word ***statement***? Pick the set in which ALL of the beginnings sound the same as the word statement
	5	Write one past form of any verb where the -ed ending is pronounced the same as the

		word received
Session 5	1	The word ***tough*** rhymes with…Pick the word that rhymes.
	2	Which of the following words DOES NOT have the same vowel as in word MURDER? Pick the odd one out.
	3	How is the word ***accepted*** pronounced?
	4	Focus on the words: skull splat straight, the beginning of these words is pronounced the same as in... Pick the correct answer.
Session 6	1	Does the word ***actually*** have more syllables than the words ***practically*** and ***naturally? Write short answer below***
	2	The word nature is pronounced similar to... Select 3 words that are pronounced similar.
	3	Pick 3 words that share the same vowel sound that you hear in word ***secret***
	4	The sound in **punish** is the same…Pick the right answer.
Session 7	1	Pick **4** words that have the same vowel sound like in the word **blood**
	2	The past tense (**-ed** ending) of the verb **kick** was pronounced with what sound?
	3	In which word the vowel is pronounced differently to **beating** /ˈ**bi**:**.tɪŋ/** ?
	4	The pronunciation of the word **feel** and **fill** is the same or different?
Session 8	1	The word **pleasure** doesn't rhyme with...
	2	The word ***fan*** is pronounced the same as ***fun***?
	3	What letter of the word lamb is spelled but NOT pronounced? Write the letter and hit the spacebar 3 times to submit the answer.
	4	The word supermarket is pronounced with /u/. Do you hear the same sound in suspend?
	5	Pick **3** words that contain **different** vowel sound that the one in the word **feel**

Session 9	1	The letter **u** in words **j<u>u</u>dgment** or **j<u>u</u>dge** are pronounced like...
	2	Is the **/t/** sound in the word **complaint** pronounced or is it silent?
	3	Which pronunciation of the word **standard** is correct? **Explain why.**
	4	How is the highlighted part of the word **ass<u>au</u>lt** pronounced?
Session 10	1	Is the word ***serial*** pronounced the same as ***cereal***?
	2	How is the word *th<u>ough</u>t* pronounced? Write one word where the spelling of *-ough* is pronounced differently?
	3	Which word could be the "hidden one" embedded in the word fantasy? Pick the correct hidden word.
	4	Choose 3 words that contain the same vowel as in the word **sl<u>ee</u>ping. Pick only 3 words otherwise your answer won't be scored.**
Session 11	1	The beginning of the word ***<u>e</u>ngage*** is pronounced with...Pick the right vowel.
	2	The letter a in cab sounds the same as in far.
	3	Do the words ***tr<u>ou</u>ble*** and ***p<u>u</u>blic*** contain the same vowel or are they pronounced differently?
	4	The past tense of the verb ***passed*** is pronounced with which sound?
	5	The word busy starts off with the same sound as the word.... Pick one word that contains the same vowel.
Session 12	1	Which word is hidden inside the word ***stealing***?
	2	The vowel in *m<u>a</u>ny* is pronounced as... Pick the right sound.
	3	Pick 2 verbs for which the past tense (*-ed)* is pronounced the same way as *pack<u>ed</u>* Pick only two verbs!
	4	Is ***<u>serious</u>*** pronounced the same way as ***<u>cereal</u>***?

Session 13	1	The word ***bluffing*** and ***cutting*** share the same vowel sound.
	2	Is the ***d*** in **diamonds** pronounced or is it silent?
	3	Pick 2 combinations of sounds that **ARE NOT** present in the word ***actually***?
	4	The sound in *tongue* is the same as in... Pick the word with the matching vowel sound.
	5	Is the letter o in the word police pronounced?
Session 14	1	Pick 2 words that contain different vowel then the one in *k**i**dnap*
	2	How was the word *neither* pronounced?
	3	*Torture* and *nature* share the same final sound. Which word is the odd one out because it´s pronounced differently?
	4	The word **mistake** and **mean** share the same vowel.
	5	The words **love** and **word** share the same vowel?
Session 15	1	The beginning of the word ***fountain*** is alike...
	2	Both words appropriate and action share the initial vowel sound. Is it true or false?
	3	How does the past tense (*-ed)* of the verb **want** sound?
	4	The word *tr**ou**ble* shares the same vowel with the word... Pick a word with the same vowel

		A	B	C	D	E	F
Session 1	1	sin	seen				
	2	fun	mud	ran			
	3	closed	used	helped	happened		
	4	*Write short answer*					
	5	rough	enough	though	Tough		
Session 2	1	**W**RONG, **W**RAP, S**W**ORD, **W**ROTE, **W**RIST	LIS**T**EN, CAS**T**LE, FAS**T**EN, WHIS**T**LE, HUS**T**LE				
	2	**e**vening and **e**vent	**e**volution and **e**vent				
	3	caught	though	ought	thought		
	4	nature, literature, picture	injure, exposure, figure				
Session 3	1	Yes, the stress falls on the same syllable	No, the stress falls on a different syllable				
	2	Yes, the -ed ending is pronounced as /t/ in both cases	No, -ed in touched is pronounced /t/ and -ed in fired is pronounced as /d/	Yes, the -ed ending is pronounced as /d/ in both cases			
	3	out	over	none of the above			
	4	True - they sound the same	False - these are different sounds				
Session 4	1	**fun, country, money**	**cat, voice, cut**	**cut, coma, vote**			
	2	discover, what done, learn, kind, man, actually					
	3	establish, structure, estress	stamp, estate, est	star, special, stroop			
	4	*Write short answer*					
Session 5	1	stuff	spot	knot	hot		
	2	first	birth	hurt	guard		
	3	/**ək**ˈsep.tɪd/	/**əs**ˈ□ep.tɪd/				
	4	/ˈes.□ə.moʊ	/ep.ɪk/	/ˈstræ□.əl/			

Se□sion 6	1	*Write short answer*					
	2	literature	picture	injure	exposure	figure	
	3	kn**ee**, s**i**t, s**ea**t, st**i**ll, f**ee**t, h*i*s, unt**i**l					
	4	as in v**a**nish	as in m**o**ney	as in p**u**rse			
Session 7	1	M**o**nday, b**u**t, b**u**tter, b**o**wl, sp**oo**ky, m**a**tch, b**a**ck, f**u**n					
	2	/d/	/**i**d/				
	3	reaching	feeling	siting			
	4	Same - different spelling, the same pronunciation	Different - both spelling and pronunciation differe				
Session 8	1	measure	treasure	gesture			
	2	True	False				
	3	Yes - these two are the same sounds	No - these two are different vowels				
	4	*Write short answer*					
	5	field, fizzy, filling, eat, gene, fist					
Session 9	1	a	u	e			
	2	Yes, the /t/ is pronounced	No, the /t/ is not pronounced				
	3	/ˈs□æn.□əd/	/ˈəstæn.dəd/				
	4	Alik□ □□/ soun	A□ike /o/ sound	Alike /e/ sound			
Session 10	1	Yes - they have different spelling, but the same pronunciation *Write short a□sw□r*	No - both the spelling and pronunciation□□re differ□□t				
	3	/fæn/	/fʌn/				
	4	sipping	slip	□easing	w□eping	sneezing	kidding
Session 11	1	**/i/** like in experiment	**/e/** like in evidence	**/ə/** like in amazing			

	2	True	False				
	3	Yes, both are pronou□c□d□w□th □a	A□ No, they contain □if□erent □□wel so□nds				
	4	/t/	/d/	/id/			
	5	**bi**□□□r		□e□ut	bird		
Session 12	1	steel	still	style			
	2	/ʌ/ like in f**u**n	/ae/ like in b**a**g				
	3	washed	helped	wanted	cleaned		
	4	Yes, the highlighted pa□ts□□r□ □ron□un□ed t□e sa□e	No, the □i□hlighte□ part□ are pronounced differently			lo□□d	
Session 13	1	True	False				
	2	Yes, it is pronounced	No, it is silent				
	3	/tʃu/	/tu/	/æk/	/ʌk/		
	4	m**u**ch	dr**o**p	b**a**ck	bl**o**ck		
	5	Yes, it's prono□nc	N□, □t□s □ot p□onounc□d				
Session 14	1	m**i**nute	sl**i**ppery	sc**e**nery	d**i**a□□t		
	2	/ˈn□ɪ.ðər/	/ˈni:.ðər/				
	3	structure	gesture	literature	culture	pleasure	furniture
	4	True	False				
	5	Yes, it´s the same vowel	No, they don´t share th□ v□□el				
S□ssion 15	1	**FA**mi□y	**FO**rce				
	2	True	False				
	3	/t/	/d/	□id			
	4	b**u**b□l	A. sl**a**ng	A. n**ou**n	A. r**oa**d		

Appendix F. List of Focus on Meaning treatment questions and answers

Session 1	1	According to Luther, what's suspicious about this crime scene?
	2	What can be deduced after seeing the fight scene?
	3	What did the man mean by describing Luther as "nitroglycerine"?
	4	Why does Luther think it was Alice Morgan who killed her parents?
	5	Why is Luther letting Alice Morgan free?
Session 2	1	What is Luther planning on doing?
	2	What was the purpose of Luther's visit?
	3	Why does Zoe doesn't want to be with Luther anymore?
	4	Why is Alice threatening Zoe?
Session 3	1	What does Ian mean when he says "brief John"?
	2	What does Terry Lynch want?
	3	What does Ian find suspicious?
	4	What is Luther trying to do?
Session 4	1	What is the profile of the suspect according to Luther?
	2	What is Alice Morgan "investigating" about Luther?
	3	According to Luther, what's suspicious about this crime scene?
	4	What kind of "justice" is Owen Lynch seeking?
	5	What does Alice want?

Session 5	1	Where did Alice hide the gun?
	2	Can Luther use the gun as the evidence against Alice?
	3	What's the purpose of the conversation between Mark North and Rose?
	4	What does Luther mean when he asks Ian if he is worried about "being on the devil's side without knowing"?
Session 6	1	Choose **all** statements, which are **TRUE**
	2	What does Rose order Luther to do?
	3	According to the police, who is responsible for the death of all the victims? Give a short answer.
	4	What does Alice mean when she says that Zoe felt she "**had lost Luther to the deads**"?
Session 7	1	What does Luther refer to when he asks if DS Ripley is "**superstitious about these sort of things**"?
	2	Why Lucian the police didn't catch Burgess in the past?
	3	Why is Luther suggesting to DS Ripley to look at unsolved cases?
	4	Who is the man being interrogated by Luther?
Session 8	1	What task did Luther give to Benny?
	2	Why Lucian Burgess uses the term "**sacrificial lamb**"?
	3	What is most likely about to happen?
	4	Why is Luther asking advice to Alice Morgan?
Session 9	1	Why John Luther chooses not to tell anyone they found the body?
	2	Did Justin Ripley break the promise he gave to Luther?
	3	What does "**waiting on warrants**" mean?
	4	Did Mark North do what Alice told him to?

Session 10

1 What does Luther mean when he says that "**the man is escalating fast**"?
2 Why would the police want to interrogate Henry Madsen?
3 According to DS Ripley, is described behavior of the suspect unusual?
4 What are the predictions regarding the murderer's profile?

Session 11

1 What made Luther think the suspect might be a taxi driver?
2 Why did the girl get into the cab?
3 Is Luther´s initial prediction about the profile of the suspect sustained?
4 What is Martin implying when he says ¨**the timing chafes my brain**¨?

Session 12

1 What did the comment about Luther's "merciless approach" refer to?
2 What does Mark know?
3 Why didn't Linda report her husband earlier?
4 Why Luther insists that Linda reveals her lover's name?

Session 13

1 What does James Carrodus do?
2 What made Luther suspicious about that man?
3 What's Luther plan?
4 What can be deduced from the conversation between Bill and Ian Reed?
5 Did the plan work out?

Session 14

1 What does Sugarman mean when he says Evie is ¨**an asset of very limited benefit**¨?
2 Pick a statement that is ***true*** considering the conversation between Luther and Evangeline.
3 Is Ian Reed telling the truth?
4 What valuable information did Ian Reed **NOT** share with Luther?
5 Why Ian Reed kills Bill?

Session 15

1 Did Patrick do the fake passports for the kidnappers?
2 Why did Ian kill Daniel?
3 Is Ian telling the truth that he did not know about the kidnapping?
4 Why did Zoe ask Ian if something was wrong?

		A	B	C	D
Session 1	1	The shooter knew the layout of the place	There was no gun	The shooter killed the dog	
	2	Luther is going through some difficulties in his marriage	Luther and his ex-wife Zoe are already divorced	Zoe is lying to her lover, because she doesn't want to leave Luther	
	3	Luther's past actions could have problematic consequences for him and the people he works with	Luther's behavior is unpredicted therefore he cannot be trusted	Luther is not good at what he does as a police investigator	
	4	Because she hated them	Because she doesn't show the feeling of "survivor guilt"	Because she is proud of it and doesn't want to alibi herself	
	5	He cannot prove she committed the crime	The opinions about Alice's guilt are divided therefore her case needs further investigation	Luther is still not sure how Alice is related to this crime case	
Session 2	1	He thinks the only way to catch Alice Morgan is to wait until she makes a mistake.	He doesn't have a plan yet, but he admires Alice's brilliance	He wants to incriminate Alice by using falsifies evidence	
	2	He wanted to inform Alice that he knew she kept the gun	He wanted to interrogate her and find out where she is hiding the gun	He wanted to threaten Alice	
	3	She thinks Luther is boring and she fell in love with another man	She suffers because Luther has been spending all the energy on chasing criminals	She thinks Luther doesn't care about her anymore	
	4	She is looking for her next victim	She wants to threaten Luther using Zoe as a tool	Power Play with Luther excites Alice and she is on a killing spree	
Session 3	1	He means that Justin should get hold of John and tell him what they found	He means that Justin should take the briefcase to John and tell him what they found	He wants revenge for being incarcerated	
	2	He wants to spend less time in prison	He wants his son to be caught by the police		
	3	He thinks it's strange that Terry didn't memorize his son's number	He thinks it's strange that tactical unit forces got involved in the case	He thinks it's strange that they found the cell phone	
	4	To strip Terry off the respect by making some compromising photos public	To threaten Terry by prolonging his sentence	To give money in the envelope in exchange for Terry's son whereabouts	
Session 4	1	A soldier or an ex-soldier who has just	A soldier or an ex-soldier who is	A soldier or an ex-soldier who has	

		came back from war and has depression	targeting police officers as his victims	mental health problems	
	2	She wants to know if what happened to Henry Madsen was an accident	She is investigating how to incriminate Luther for Henry Madsen's murder	She wants to know more about Luther's ex-wife Zoe to threaten him	
	3	That the suspect didn't kill the victim, although he is a good shooter	That the suspect attacks during the day not night	That no one had seen him escaping the crime scene	
	4	He wants his dad, Terry Lynch, to be out of prison	All he wants is to receive respect because he served in the army	All he wants is that his dad receives respect because he served in the army	
	5	She wants to know why Henry Madsen fell down	She wants to threaten Luther and put his position at work in danger	She wants to blackmail Luther and attack his ex-wife Zoe	
Session 5	1	Inside the dog's digestive track	Inside Alice room	It is still unclear	
	2	No, the gun was burnt therefore its forensic value is not valid	Yes, the gun is the only evidence that can incriminate Alice	No, Alice have committed a "perfect crime" and left no evidence of it	
	3	Mark is filing a complaint against Luther	Mark came to testify that Luther did not attack him	Mark is filing a complaint about a possible harm Luther can cause due to jealousy	
	4	Luther's comment reveals suicidal thoughts caused by the breakup with Zoe	Luther's reflecting whether his love for Zoe makes him a good or bad man	Luther admits his morality as a policeman is faulty	
Session 6	1	Zoe is angry that Luther is so devoted to his job	Zoe doesn't blame Luther for being so devoted to his job	Zoe doesn't want to be with Luther anymore	Zoe still hopes her relationship with Luther has future
	2	Rose orders not to proceed to Owen Lynch’s location	Rose orders to proceed to Owen Lynch’s location	Rose orders to proceed and sends back-up police support	
	3	Write a short answer			
	4	That Luther’s job had absorbed him completely and would never leave him in peace	That, due to his job, Luther has been acting like a dead man	That Luther’s job requires seeing a lot of crime	
Session 7	1	A. He refers to satanism and cult-related matters	He refers to theory that it is the family member who almost always commits the crime	He refers to believing in god	
	2	A. Because an undercover police operation didn't go as planned	Because the evidence against him were stolen	Because he was brutally beaten and ended up in the hospital	
	3	A. Luther thinks Lucian Burgess uses his	Luther thinks Lucian Burgess has an	Luther thinks Lucian Burgess has	

		old victims' blood to write on the walls	admirer who might also be involve in the crimes	committed more crimes that they think
	4	A. He is an ex-police officer who worked on Burgess' case in the past	He used to be Burgess friend, who reported him to the police	He is an ex-police officer who specialized in murder cases related to cults and satanism
Session 8	1	To identify the girls who appear in the video	To identify location where the beating is taking place	To spy on the girls who appear in the video
	2	He uses this phrase to emphasize the innocence of the victims	He pretends to be a victim, who police wants to unjustly incriminate	He thinks he is a target for the police, because of the controversial business he runs
	3	A. The man is going to interrogate Luther and suspend his investigation	The man is going to present evidences that Luther beat up Mark North	The man is going to take over the investigation and Luther will loose his job
	4	He tries to understand what compels Lucian Burgess to kill	He looks for evidences against Alice to incriminate her	He wants to know if these two crime cases are connected
Session 9	1	Because the body belong to another woman, meaning there are more victims than initially expected	Because Luther intends to catch the killer himself and bring justice on his own	Because this is the only way he can catch the killer and incriminate him with valid evidences
	2	No, because he did not compromise Luther's operation nor gave out his whereabouts	Yes, because he thinks that Luther's decisions are against the law	Yes, because he doesn't want Luther to catch the killer on his own
	3	That Luther's investigation is about to be suspended	That Luther is about to receive a document authorizing to make an arrest	That Lucian Burgess' name will be made public so that everyone knows he is a murderer
	4		Yes, he withdrew the complaint	No, he did not listen to Alice
Session 10	1	That the man kills his victim more rapidly each time	That the man uses more violence each time he kills	That the man has a well-designed plan, which makes it difficult to catch him
	2	A. Because the bodies of many victims were found while he was in coma	Because the police wants to know if Luther carried the investigation according to the protocol	Because Henry Madsen's crime could be link to Alice Morgan's murder case
	3	Yes, it's unusual because of the rapid escalation	No, because his victims are completely random	No, the intervals between the killings are described as very typical
	4	That he is married and that his wife is	That he has psychiatric problems and	That he is married and that he knows

		about to leave him	that he is a loner	David Bowie
Session 11	1	All the victims trusted him by default.	The suspect has a very good knowledge of the city.	He drives a white van.
	2	A. Because she wanted a ride and he offered it for free	Because Graham tricked her into getting in	Because she was lost and Graham knew the area
	3	Yes, as he still thinks the suspect is a taxi driver.	No, as he thinks the suspect cannot be a taxi driver	Partially, as he thinks the suspect might have been a taxi driver in the past
	4	That he finds the sudden death of Henry Madsen suspicious	That he is directly accusing Luther of Henry Madsen´s death	That he is surprised Henry Madsen survived the accident
Session 12	1	Luther's interrogation strategy being cruel	Luther seeking the harshest punishment for Graham	Luther treating Linda as if she was another suspect
	2	That Zoe slept with Luther	That Zoe is planning to leave him	That Luther has been threatening Zoe
	3	Because she felt guilty and embarrassed	Because her friends laughed at her	Because her husband threatened he would kill himself
	4	Because the lover might be Graham's next victim	Because Luther wanted to check if Linda is telling the truth	Because the lover has to be treated as potential suspect
Session 13	1	He is an art dealer and he stole some diamonds	He is a businessman doing some illegal business with Russians	He is a really wealthy art collector
	2	The fact that he did not give the kidnappers what they asked for	The fact that he is a dealer, who does illegal business	The fact that he asked about Ian Reed
	3	Luther wants to buy some time to open up lines of negotiation with the kidnappers	Luther wants to trick the kidnappers and give them the diamonds from the evidence safe	Luther wants to buy time by giving the borrowed diamonds and get more diamonds if necessary
	4	That Ian is being involved in the money laundering and that he knew about the robbery	That Ian is a good friend of Bill's and he is trying to help him not to get in serious trouble	That Ian is a money launder and that he knew about the kidnapping
	5	No, because James didn't deliver the diamonds	Yes, it worked out as planned	Yes, James wasn't supposed to deliver the diamonds
Session 14	1	That Luther cannot hurt her nor use her in any way against him	That Evie will deny that Sugarman is involved in the kidnapping	That Luther cannot demonstrate with evidence that Evie is involved
	2	Luther presented to Evangeline a manipulated version of Daniel´s audio	Luther feels pity for Evangeline, because of what Daniel said about her	Luther used the original recording of Daniel´s voice to get information about

		recording		the kidnappers whereabouts	
	3	Yes	No		
	4	That it is Bill, his friend, who is responsible for the death of Tom and Jessica	That it was his plan to make Tom Mayer try to rescue Jessica	That it is Bill, who has the diamonds	
	5	Because Ian doesn´t want the police to discover that he was involved	Because Ian got carried away and decided to bring justice by himself	Because Ian knows that the Bill is not telling the truth	
Session 15	1	Yes, because he was forced to	No, because he didn´t want to help Sugarman	Yes, because he wanted to make money	No, because he had already problems with the police
	2	Because wanted to silence Sugarman	Because, he wanted revenge for his friend´s death	Because Ian lost his nerves as Sugarman killed his friend	
	3	Yes	No		
	4	Because he asked for an alcoholic drink	Because she is surprised by his visit	Because she realized he looked worried and troubled	

Appendix G. Shadowing task stimuli list

Item	Series	Sentence	N words
1.	Luther	Are we clear on that	5
2.	Luther	That would be illegal	4
3.	Luther	What you are doing is wrong	6
4.	Luther	What do you mean by that	6
5.	Luther	None of it seems right to me	7
6.	Luther	I wish I could tell you I had	8
7.	Luther	She lives in a flat near campus	7
8.	Luther	I don't see how this is relevant	7
9.	Luther	Do you know where he was Friday night	8
10.	Luther	So how are things with you and your husband	9
11.	Luther	Why did your wife turn her face from you John	10
12.	Luther	Three victims in five weeks spread across London	8
13.	Luther	These are very special circumstances	5
14.	Luther	What if you only catch people who make mistakes	9
15.	Luther	Now I know a lot of people in this prison	10
16.	Luther	I have been wondering, why do you think he does it?	11
17.	Luther	Are you superstitious about that sort of thing	8
18.	Luther	It's just things have got really complicated	8
19.	Luther	In the end I just gave up trying to make him proud	12
20.	Luther	Cause the thing about you Terry, is that you are a hard bastard, aren't you	15
21.	Sherlock	That's not a word I'd use	8
22.	Sherlock	What exactly am I supposed to be doing here	9
23.	Sherlock	He could look at you and tell you your whole life story	12
24.	Sherlock	You asked me to come I am assuming it's important	11
25.	Sherlock	It's password-protected	4
26.	Sherlock	Wait, it wasn't a date	5
27.	Sherlock	I hired you to do a job	7
28.	Sherlock	So you're working here tonight	6
29.	Sherlock	Cap of tea would be lovely, thank you	8
30.	Sherlock	Afraid they don't see it like that	7
31.	Sherlock	This is what you wanted isn't it	7
32.	Sherlock	The police don't consult amateurs	5
33.	Sherlock	Do you know where I could get a cab	9
34.	Sherlock	You risk your life to prove you're clever	9
35.	Sherlock	We don't know a thing about each other	8
36.	Sherlock	What kind of result do you care about?	8
37.	Sherlock	Can I maybe decide that for myself?	7
38.	Sherlock	What you couldn't be bothered to get up	8
39.	Sherlock	We're still no closer to finding him	8
40.	Sherlock	So you're doing well, you've been abroad a lot	11
	Total		**320**

Appendix H. Animacy judgment task stimuli list

Language	Word	FREQcount	CDcount	FREQlow	CDlow	SUBTLwf	Lg10WF	SUBTLcd	Lg10CD	Length (letter)
English	accountant	315	191	303	185	6,18	2,4997	2,28	2,2833	10
English	ant	273	111	200	95	5,35	2,4378	1,32	2,0492	3
English	armchair	21	19	21	19	0,41	1,3424	0,23	1,301	8
English	assistant	1643	977	1414	861	32,22	3,2159	11,65	2,9903	9
English	baker	698	264	83	63	13,69	2,8445	3,15	2,4232	5
English	balcony	373	267	369	264	7,31	2,5729	3,18	2,4281	7
English	basket	672	428	643	421	13,18	2,828	5,1	2,6325	6
English	bear	2928	1500	2386	1317	57,41	3,4667	17,88	3,1764	4
English	bee	528	248	407	202	10,35	2,7235	2,96	2,3962	3
English	belt	1242	838	1180	810	24,35	3,0945	9,99	2,9238	4
English	bicycle	337	239	316	228	6,61	2,5289	2,85	2,3802	7
English	bird	2318	1196	2007	1095	45,45	3,3653	14,26	3,0781	4
English	boat	4885	1457	4750	1426	95,78	3,689	17,37	3,1638	4
English	book	9026	3244	8226	3075	176,98	3,9555	38,67	3,5112	4
English	bottle	2588	1561	2510	1537	50,75	3,4131	18,61	3,1937	6
English	building	5078	2411	4690	2305	99,57	3,7058	28,74	3,3824	8
English	bull	1403	665	917	497	27,51	3,1474	7,93	2,8235	4
English	butcher	434	291	359	263	8,51	2,6385	3,47	2,4654	7
English	butterfly	281	186	197	144	5,51	2,4502	2,22	2,2718	9
English	candle	409	291	393	279	8,02	2,6128	3,47	2,4654	6
English	carrot	195	123	156	104	3,82	2,2923	1,47	2,0934	6
English	ceiling	426	350	421	346	8,35	2,6304	4,17	2,5453	7
English	cheese	1991	1130	1731	1025	39,04	3,2993	13,47	3,0535	6
English	chicken	3148	1570	2555	1394	61,73	3,4982	18,72	3,1962	7

English	computer	3011	1224	2717	1170	59,04	3,4789	14,59	3,0881	8
English	cupboard	127	99	122	97	2,49	2,1072	1,18	2	8
English	dancer	831	499	769	481	16,29	2,9201	5,95	2,699	6
English	daughter	8739	2993	8576	2957	171,35	3,9415	35,68	3,4763	8
English	desk	2239	1404	2204	1391	43,9	3,3502	16,74	3,1477	4
English	dolphin	141	72	116	62	2,76	2,1523	0,86	1,8633	7
English	door	14895	5286	14339	5228	292,06	4,1731	63,02	3,7232	4
English	duck	1263	703	938	562	24,76	3,1017	8,38	2,8476	4
English	employee	590	429	553	415	11,57	2,7716	5,11	2,6335	8
English	farmer	604	319	389	262	11,84	2,7818	3,8	2,5051	6
English	father	28279	5446	21688	5139	554,49	4,4515	64,93	3,7362	6
English	fish	4258	1708	3618	1562	83,49	3,6293	20,36	3,2327	4
English	folder	83	61	83	61	1,63	1,9243	0,73	1,7924	6
English	fox	1102	409	326	197	21,61	3,0426	4,88	2,6128	3
English	fridge	502	392	499	390	9,84	2,7016	4,67	2,5944	6
English	friend	21384	6178	20982	6147	419,29	4,3301	73,65	3,7909	6
English	frog	603	271	504	245	11,82	2,781	3,23	2,4346	4
English	gardener	214	141	208	135	4,2	2,3324	1,68	2,1523	8
English	girl	28413	6143	27239	6094	557,12	4,4535	73,24	3,7885	4
English	glasses	1689	1008	1626	982	33,12	3,2279	12,02	3,0039	7
English	goat	537	338	480	315	10,53	2,7308	4,03	2,5302	4
English	hairdresser	126	84	123	82	2,47	2,1038	1	1,9294	11
English	hunter	936	385	430	273	18,35	2,9717	4,59	2,5866	6
English	keyboard	92	66	88	63	1,8	1,9685	0,79	1,8261	8
English	king	6592	1688	3319	993	129,25	3,8191	20,12	3,2276	4
English	kitchen	2974	1815	2807	1750	58,31	3,4735	21,64	3,2591	7
English	knife	2387	1255	2304	1215	46,8	3,378	14,96	3,099	5
English	lady	11071	4129	8981	3792	217,08	4,0442	49,23	3,616	4
English	lawyer	4055	1686	3959	1669	79,51	3,6081	20,1	3,2271	6
English	librarian	147	75	101	68	2,88	2,1703	0,89	1,8808	9
English	mountain	1805	885	1408	735	35,39	3,2567	10,55	2,9474	8
English	mouse	975	400	750	323	19,12	2,9894	4,77	2,6031	5

English	nanny	531	183	306	156	10,41	2,7259	2,18	2,2648	5
English	napkin	184	137	170	130	3,61	2,2672	1,63	2,1399	6
English	newspaper	1208	752	1152	729	23,69	3,0824	8,97	2,8768	9
English	orange	1138	684	866	542	22,31	3,0565	8,15	2,8357	6
English	pencil	503	391	469	373	9,86	2,7024	4,66	2,5933	6
English	picture	7061	3178	6868	3119	138,45	3,8489	37,89	3,5023	7
English	pillow	581	453	553	443	11,39	2,7649	5,4	2,6571	6
English	plane	4872	1691	4782	1671	95,53	3,6878	20,16	3,2284	5
English	plumber	229	150	217	146	4,49	2,3617	1,79	2,179	7
English	priest	1336	522	1194	477	26,2	3,1261	6,22	2,7185	6
English	road	5709	2766	5043	2567	111,94	3,7566	32,98	3,442	4
English	scarf	239	165	235	163	4,69	2,3802	1,97	2,2201	5
English	shark	764	258	595	232	14,98	2,8837	3,08	2,4133	5
English	sheep	685	392	614	372	13,43	2,8363	4,67	2,5944	5
English	shirt	2365	1400	2330	1383	46,37	3,374	16,69	3,1464	5
English	sister	9207	3046	7971	2911	180,53	3,9642	36,31	3,4839	6
English	skates	215	112	203	108	4,22	2,3345	1,34	2,0531	6
English	spider	515	231	346	186	10,1	2,7126	2,75	2,3655	6
English	stamp	302	222	275	206	5,92	2,4814	2,65	2,3483	5
English	star	4149	1821	3118	1543	81,35	3,618	21,71	3,2605	4
English	stone	2072	938	1158	737	40,63	3,3166	11,18	2,9727	5
English	street	7557	3411	5281	2861	148,18	3,8784	40,67	3,533	6
English	suitcase	683	371	660	364	13,39	2,8351	4,42	2,5705	8
English	surgeon	838	500	780	472	16,43	2,9238	5,96	2,6998	7
English	thief	1238	704	1143	670	24,27	3,0931	8,39	2,8482	5
English	trousers	263	185	249	180	5,16	2,4216	2,21	2,2695	8
English	wallet	1163	689	1143	677	22,8	3,066	8,21	2,8388	6
English	whale	574	231	459	209	11,25	2,7597	2,75	2,3655	5
English	window	4386	2442	4287	2410	86	3,6422	29,11	3,3879	6
English	wine	3078	1518	2905	1470	60,35	3,4884	18,1	3,1816	4
English	woman	22166	5925	20418	5744	434,63	4,3457	70,64	3,7728	5
English	worm	516	271	430	260	10,12	2,7135	3,23	2,4346	4

Appendix I. Sentence verification task stimuli list

Item	Sentence
1.	Gasoline is an excellent drink
2.	Elephants are big animals
3.	The Queen of England lives in Washington
4.	Spaghetti grows on tall trees
5.	Hot and cold are opposites
6.	The sun always sets in the north
7.	The inside of an egg is blue
8.	August is a winter month
9.	It always snows in July
10.	March has thirty-eight days
11.	Most people wear hats on their feet
12.	The stars come out in the day
13.	Exercise is good for your health
14.	Japan is a wealthy country
15.	Wednesday is the first day of the week
16.	All men can have babies
17.	All dogs have fifteen legs
18.	Shakespeare wrote many fine plays
19.	Most teenagers like rock and roll
20.	Some people love to eat chocolate
21.	Some people keep dogs as pets
22.	Young children can be very noisy
23.	Some roses have a beautiful smell
24.	Hungry cats like to chase mice
25.	People eat through their noses
26.	Ten dollars is less than ten cents
27.	You can start a fire with a match
28.	Red and green are colours
29.	Many houses are made of bricks
30.	There are many cities on the moon
31.	Italy is a country in Europe
32.	Many people drink coffee for breakfast
33.	You can buy beer at church
34.	The American flag has stars and stripes
35.	Gold is a valuable metal
36.	Milk comes from yellow chickens
37.	Most swimsuits have long sleeves
38.	A monkey is a kind of bird
39.	Ships travel on the water
40.	You can buy a burger at McDonalds

Appendix J. Delayed sentence production task stimuli list

Item	Feature	Sentence
1.	/æ/	Pack your bags
2.	/æ/	My ankle hurts a lot
3.	/æ/	Your cat is always happy
4.	/æ/	He is not a good match for you
5.	/æ/	Turn on the fan, it´s hot
6.	/ʌ/	I am visiting my uncle
7.	/ʌ/	That´s a nasty cut
8.	/ʌ/	He isn´t enjoying it much
9.	/ʌ/	Don´t spoil the fun
10.	/ʌ/	Alcohol is a drug
11.	/iː/	Ill patient seeks help
12.	/iː/	My feet are cold
13.	/iː/	Only a sheep and a cow
14.	/iː/	It was the cheap one
15.	/iː/	There is a seat at the front
16.	/ɪ/	They sit at the same table
17.	/ɪ/	I'll have the chips please
18.	/ɪ/	It keeps you fit I think
19.	/ɪ/	He went by ship to India
20.	/ɪ/	I feel sick today
21.	/-tʃə(r)/	What a nice gesture
22.	/-tʃə(r)/	She always thinks about her future
23.	/-tʃə(r)/	Lying is against his nature
24.	/-tʃə(r)/	This photo fails to capture her beauty
25.	/-ɪkli,-əli/	He is typically wrong
26.	/-ɪkli,-əli/	It´s basically all we need
27.	/-ɪkli,-əli/	He is practically a genius
28.	/-ɪkli,-əli/	He is actually saying the truth
29.	/d/ or /t/	She remembered to do the shopping
30.	/d/ or /t/	They talked and laughed all night long
31.	/d/ or /t/	Sara worked until late
32.	/d/ or /t/	I listened to the special guest
33.	initial /s/	The train station is near
34.	initial /s/	Try to speak spontaneously
35.	initial /s/	I thought all spiders are dangerous
36.	ortography	Karate is a tough sport
37.	ortography	Though they are smart, I don´t like them
38.	ortography	The path leads through a narrow street
39.	other	A gun was found at the **crime scene**
40.	other	I asked her about the money

Appendix K. Results of chi-square tests for post-test questionnaire

		Groups								Chi-square and V Cramera statistics
		Cap +FoFP		NoCap+FoPF		Cap +FoM		NoCap+FoM		
		n	%	n	%	n	%	n	%	
Device use	always on a computer	15	88,2%	16	84,2%	14	73,7%	15	78,9%	Chi2(12) =10,142; p = 0,604; V = 0,214
	always on a mobile phone	0	0,0%	0	0,0%	1	5,3%	0	0,0%	
	always on tablet	0	0,0%	1	5,3%	1	5,3%	0	0,0%	
	different devices, but predominantly computer	1	5,9%	2	10,5%	3	15,8%	4	21,1%	
	different devices, but predominantly mobile	1	5,9%	0	0,0%	0	0,0%	0	0,0%	
Headphone use	always used headphones	12	70,6%	15	78,9%	9	47,4%	13	68,4%	Chi2(6) = 7,323; p = 0,292; V = 0,222
	sometimes used headphones	5	29,4%	4	21,1%	8	42,1%	4	21,1%	
	never used headphones	0	0,0%	0	0,0%	2	10,5%	2	10,5%	
Viewing regularity	I watched the sessions regularly and have seen all of them on time	10	58,8%	10	52,6%	9	47,4%	7	36,8%	Chi2(6) = 6,037; p = 0,419; V = 0,202
	I watched the sessions regularly and have seen MOST of them on time	6	35,3%	9	47,4%	8	42,1%	8	42,1%	
	I wasn´t regular with my watching, but I have seen ALL the sessions	1	5,9%	0	0,0%	2	10,5%	4	21,1%	
	I missed some sessions	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
Treatment helpfulness	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	Chi2(9) = 9,296; p = 0,410; V = 0,205
	Disagree	0	0,0%	0	0,0%	0	0,0%	1	5,3%	
	Neither agree nor disagree	0	0,0%	1	5,3%	0	0,0%	1	5,3%	
	Agree	9	52,9%	12	63,2%	10	52,6%	14	73,7%	
	Strongly agree	8	47,1%	6	31,6%	9	47,4%	3	15,8%	
Treatment helpfulness new vocabulary	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	1	5,3%	Chi2(12) = 20,357; p = 0,061; V = 0,303
	Disagree	0	0,0%	2	10,5%	0	0,0%	1	5,3%	
	Neither agree nor disagree	2	11,8%	5	26,3%	0	0,0%	7	36,8%	
	Agree	12	70,6%	10	52,6%	12	63,2%	8	42,1%	
	Strongly agree	3	17,6%	2	10,5%	7	36,8%	2	10,5%	
Treatment helpfulness	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	1	5,3%	Chi2(9) = 14,315; p = 0,112; V = 0,256

pronunciation	Disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
	Neither agree nor disagree	1	5,9%	0	0,0%	3	16,7%	5	26,3%	
	Agree	6	35,3%	10	52,6%	4	22,2%	8	42,1%	
	Strongly agree	10	58,8%	9	47,4%	11	61,1%	5	26,3%	
Treatment helpfulness listening comprehension	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	1	5,3%	Chi2(9) = 5,648; p = 0,775; V = 0,159
	Disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
	Neither agree nor disagree	1	5,9%	0	0,0%	1	5,3%	0	0,0%	
	Agree	8	47,1%	9	47,4%	9	47,4%	7	36,8%	
	Strongly agree	8	47,1%	10	52,6%	9	47,4%	11	57,9%	
Treatment helpfulness grammar	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	1	5,3%	Chi2(12) = 14,154; p = 0,291; V = 0,252
	Disagree	0	0,0%	1	5,3%	1	5,3%	1	5,3%	
	Neither agree nor disagree	3	17,6%	9	47,4%	11	57,9%	11	57,9%	
	Agree	13	76,5%	8	42,1%	6	31,6%	5	26,3%	
	Strongly agree	1	5,9%	1	5,3%	1	5,3%	1	5,3%	
Treatment helpfulness spelling	Strongly disagree	0	0,0%	0	0,0%	2	10,5%	2	10,5%	Chi2(12) = 13,029; p = 0,367; V = 0,244
	Disagree	0	0,0%	1	5,3%	1	5,3%	2	10,5%	
	Neither agree nor disagree	2	12,5%	8	42,1%	6	31,6%	7	36,8%	
	Agree	12	75,0%	7	36,8%	7	36,8%	6	31,6%	
	Strongly agree	2	12,5%	3	15,8%	3	15,8%	2	10,5%	
Viewing condition	with subtitles	17	100,0%	0	0,0%	19	100,0%	0	0,0%	Chi2(3) = 74,000; p < 0,01; V = 1,000
	without subtitles	0	0,0%	19	100,0%	0	0,0%	19	100,0%	
Preference for viewing condition	with subtitles	9	52,9%	11	57,9%	10	52,6%	7	36,8%	Chi2(6) = 3,942; p = 0,685; V = 0,163
	without subtitles	4	23,5%	6	31,6%	5	26,3%	9	47,4%	
	no difference	4	23,5%	2	10,5%	4	21,1%	3	15,8%	
I need subtitles to understand the dialog	0	14	82,4%	8	42,1%	12	63,2%	12	63,2%	Chi2(3) = 6,612; p = 0,102; V = 0,290
	1	3	17,6%	11	57,9%	7	36,8%	7	36,8%	
No need for subtitles to understand the dialog (1/3) & Did not need sub to understand dialogue (2/4)	0	13	76,5%	15	78,9%	14	73,7%	11	57,9%	Chi2(3) = 2,500; p = 0,475; V = 0,184
	1	4	23,5%	4	21,1%	5	26,3%	8	42,1%	
Subtitles helped me to focus on pronunciation (1/3) & I think subs would help (2/4)	0	4	23,5%	15	78,9%	7	36,8%	15	78,9%	Chi2(3) = 18,165; p < 0,01; V = 0,495
	1	13	76,5%	4	21,1%	12	63,2%	4	21,1%	
Subtitles distracted my focus on pronunciation (1/3) I think sub would distract me (2/4)	0	14	82,4%	8	42,1%	15	78,9%	8	42,1%	Chi2(3) = 11,512; p = 0,009; V = 0,394
	1	3	17,6%	11	57,9%	4	21,1%	11	57,9%	

Was not focusing on pronunciation (1/3) & I was not focusing on pronunciation (2/4)	0	17	100,0%	19	100,0%	18	94,7%	16	84,2%	Chi2(3) = 6,065; p = 0,109; V = 0,286
	1	0	0,0%	0	0,0%	1	5,3%	3	15,8%	
I got tired of reading subtitles (1/3) & I would get tired of reading sub (2/4)	0	17	100,0%	15	78,9%	19	100,0%	12	63,2%	Chi2(9) = 14,112; p = 0,003; V = 0,437
	1	0	0,0%	4	21,1%	0	0,0%	7	36,8%	
Degree of liking the series	Really dislike	0	0,0%	0	0,0%	0	0,0%	0	0,0%	Chi2(9) = 13,866; p = 0,127; V = 0,252
	Dislike	1	6,3%	0	0,0%	0	0,0%	0	0,0%	
	Neither liked nor dislike	2	12,5%	3	15,8%	2	10,5%	3	15,8%	
	Dislike	7	43,8%	11	57,9%	3	15,8%	9	47,4%	
	Really like	6	37,5%	5	26,3%	14	73,7%	7	36,8%	
Engaged with plot	0	2	11,8%	5	26,3%	2	10,5%	5	26,3%	Chi2(3) = 2,795; p = 0,424; V = 0,194
	1	15	88,2%	14	73,7%	17	89,5%	14	73,7%	
Not engaged with plot	0	15	88,2%	16	84,2%	19	100,0%	16	84,2%	Chi2(3) = 3,296; p = 0,348; V = 0,211
	1	2	11,8%	3	15,8%	0	0,0%	3	15,8%	
Paying attention to improve listening comprehension (1/3) & Paid attention to improve listening comprehension (2/4)	0	5	29,4%	8	42,1%	6	31,6%	3	15,8%	Chi2(3) = 3,192; p = 0,363; V = 0,208
	1	12	70,6%	11	57,9%	13	68,4%	16	84,2%	
Paying attention to improve pronunciation (1/3) & Paid attention to improve pronunciation (2/4)	0	8	47,1%	7	36,8%	12	63,2%	12	63,2%	Chi2(3) = 3,801 p = 0,284; V = 0,227
	1	9	52,9%	12	63,2%	7	36,8%	7	36,8%	
No problem with answering questions while watching	0	7	41,2%	11	57,9%	11	57,9%	11	57,9%	Chi2(3) = 1,474 p = 0,688; V = 0,141
	1	10	58,8%	8	42,1%	8	42,1%	8	42,1%	
Problem with answering questions while watching (1/3) & Could not pay attention to answering questions (2/4)	0	17	100,0%	18	94,7%	18	94,7%	17	89,5%	Chi2(3) = 1,947 p = 0,583; V = 0,162
	1	0	0,0%	1	5,3%	1	5,3%	2	10,5%	
Engaged with questions	0	8	47,1%	10	52,6%	7	36,8%	11	57,9%	Chi2(3) = 1,848 p = 0,604; V = 0,158
	1	9	52,9%	9	47,4%	12	63,2%	8	42,1%	
Found questions distracting	0	17	100,0%	18	94,7%	19	100,0%	15	78,9%	Chi2(3) = 8,839 p = 0,032; V = 0,346
	1	0	0,0%	1	5,3%	0	0,0%	4	21,1%	
Activity usefulness	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	Chi2(12) = 10,122 p = 0,120; V = 0,265

for pronunciation	Disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
	Neither agree nor disagree	1	6,3%	0	0,0%	2	11,1%	5	26,3%	
	Agree	7	43,8%	11	57,9%	9	50,0%	11	57,9%	
	Strongly agree	8	50,0%	8	42,1%	7	38,9%	3	15,8%	
Activity usefulness for listening comprehension	Strongly disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	Chi2(12) = 2,701 p = 0,440; V = 0,191
	Disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
	Neither agree nor disagree	0	0,0%	0	0,0%	0	0,0%	0	0,0%	
	Agree	6	35,3%	11	57,9%	7	36,8%	7	36,8%	
	Strongly agree	11	64,7%	8	42,1%	12	63,2%	12	63,2%	

Appendix L. Individual mean accuracy score on treatment questions

		Viewing treatment session																
Participant	**Group**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**Mean [%]**	**SD**
1	Group 1	86	100	75	86	100	73	71	71	75	50	60	40	40	60	100	72,47	19,97
2	Group 1	86	100	75	71	75	64	86	100	100	100	60	100	80	100	100	86,47	14,72
3	Group 1	86	50	0	57	50	91	86	86	100	67	40	60	40	60	50	61,53	25,79
4	Group 1	29	67	75	71	75	27	57	43	50	33	80	40	80	70	100	59,80	21,87
5	Group 1	29	33	75	71	75	45	57	57	75	100	40	20	20	20	50	51,13	24,57
6	Group 1	71	100	100	71	100	73	71	86	100	100	60	60	60	80	100	82,13	16,60
7	Group 1	43	100	50	71	100	100	86	100	100	100	100	80	80	100	75	85,67	19,14
8	Group 1	86	100	100	71	75	91	100	100	100	83	80	80	80	80	100	88,40	10,74
9	Group 1	57	83	25	86	100	100	100	100	100	100	80	100	100	60	100	86,07	22,44
10	Group 1	71	100	100	57	75	64	71	100	100	100	100	60	80	100	100	85,20	17,25
11	Group 1	71	67	75	71	75	91	71	86	75	83	20	60	20	80	50	66,33	21,30
12	Group 1	57	100	50	57	75	91	100	57	100	100	80	40	60	40	100	73,80	23,39
13	Group 1	100	100	100	57	75	64	57	57	100	100	80	60	20	100	100	78,00	24,80
14	Group 1	100	100	75	86	100	100	100	100	100	100	100	100	80	100	100	96,07	8,40
15	Group 1	86	83	100	86	100	100	100	100	100	83	100	60	100	100	100	93,20	11,58
16	Group 1	57	33	100	71	75	55	43	100	100	100	60	60	60	80	100	72,93	22,82
17	Group 1	57	50	75	71	100	55	100	100	100	100	60	100	80	80	75	80,20	18,87
18	Group 1	75	100	50	71	100	57	86	65	100	50	40	60	40	100	75	71,27	22,00
19	Group 1	57	100	100	71	100	100	100	86	71	91	60	100	40	80	50	80,40	20,95
20	Group 2	86	100	75	71	75	70	100	57	100	100	60	80	80	60	75	79,27	15,15
21	Group 2	86	83	75	0	50	40	100	100	100	88	100	100	40	80	75	74,47	29,52

22	Group 2	57	67	100	86	75	60	57	57	100	100	60	40	40	60	100	70,60	21,53
23	Group 2	71	33	50	57	75	80	71	57	75	100	60	60	60	60	65	64,93	15,12
24	Group 2	100	33	100	71	50	70	86	100	100	88	100	80	80	80	75	80,87	19,65
25	Group 2	86	100	50	71	100	80	100	100	100	100	100	80	40	100	100	87,13	19,74
26	Group 2	57	67	50	57	75	60	71	100	100	100	100	80	80	100	100	79,80	18,96
27	Group 2	50	67	57	86	100	70	86	100	100	88	100	80	60	100	25	77,93	22,67
28	Group 2	86	50	50	86	50	50	100	71	75	88	80	100	0	60	50	66,40	26,14
29	Group 2	100	100	75	71	100	100	100	100	100	100	100	80	60	80	100	91,07	13,81
30	Group 2	86	83	100	71	75	100	100	86	100	88	80	80	40	80	100	84,60	15,89
31	Group 2	71	100	75	86	100	60	86	100	100	88	60	60	40	100	100	81,73	19,63
32	Group 2	43	50	25	71	75	40	57	71	100	88	40	60	80	40	100	62,67	23,25
33	Group 2	71	50	75	71	75	80	71	86	100	100	60	100	20	100	100	77,27	22,55
34	Group 2	71	67	100	71	100	60	100	100	100	100	60	40	60	60	100	79,27	21,25
35	Group 2	86	100	75	71	75	90	100	100	75	100	80	100	80	80	100	87,47	11,51
36	Group 2	86	100	100	71	75	70	100	100	100	63	60	60	60	80	75	80,00	16,38
37	Group 2	71	100	75	86	100	90	100	86	100	88	100	60	60	80	75	84,73	14,22
38	Group 2	57	100	25	43	50	50	43	57	100	88	100	0	0	40	100	56,87	34,49
39	Group 3	60	75	100	100	75	88	50	100	100	75	100	100	100	80	50	83,53	18,81
40	Group 3	60	100	50	20	100	100	100	75	100	75	100	100	80	100	25	79,00	28,23
41	Group 3	75	40	75	80	100	88	75	100	50	75	75	100	80	60	100	78,20	18,21
42	Group 3	75	80	75	60	75	100	75	100	100	75	50	100	60	60	75	77,33	16,35
43	Group 3	100	80	75	100	100	63	75	100	100	75	100	75	40	60	50	79,53	20,15
44	Group 3	50	60	100	60	75	100	100	100	100	75	100	100	80	80	50	82,00	19,62
45	Group 3	75	40	75	100	100	100	100	75	100	75	100	75	20	80	100	81,00	24,07
46	Group 3	75	40	100	80	50	100	100	100	25	75	75	100	80	60	100	77,33	24,34
47	Group 3	25	100	50	80	100	100	25	100	75	75	100	100	60	60	100	76,67	27,30
48	Group 3	50	60	100	80	75	100	25	100	75	75	75	100	100	80	75	78,00	21,36
49	Group 3	50	80	75	100	75	100	100	75	100	100	100	100	100	80	75	87,33	15,57
50	Group 3	100	80	50	60	100	88	100	100	75	75	100	100	80	100	50	83,87	18,69

51	Group 3	100	80	75	80	100	88	75	75	100	75	100	100	60	100	75	85,53	13,44
52	Group 3	75	40	75	80	75	63	75	75	50	50	75	100	60	65	70	68,53	14,62
53	Group 3	75	80	75	80	50	56	75	100	100	100	100	75	60	60	75	77,40	16,69
54	Group 3	75	80	100	100	100	88	50	75	100	75	100	100	80	80	75	85,20	14,78
55	Group 3	100	60	75	80	50	38	50	75	75	75	75	100	80	80	50	70,87	18,00
56	Group 3	25	40	50	60	25	50	25	25	75	50	75	25	60	20	0	40,33	21,83
57	Group 3	75	60	50	100	50	88	100	100	75	50	100	100	40	80	50	74,53	22,78
58	Group 4	80	75	75	80	50	100	50	100	100	50	75	50	60	60	0	67,00	26,10
59	Group 4	60	50	100	40	50	44	50	100	25	75	25	75	40	80	0	54,27	27,98
60	Group 4	20	50	100	60	60	67	25	100	75	75	75	75	20	40	50	59,47	25,59
61	Group 4	40	75	75	20	100	78	75	75	50	75	75	100	80	40	75	68,87	22,07
62	Group 4	60	50	50	60	100	100	25	75	100	75	100	75	60	80	50	70,67	22,82
63	Group 4	80	50	50	60	75	78	25	100	100	75	100	75	40	60	100	71,20	23,52
64	Group 4	40	75	75	80	75	78	100	100	50	75	100	100	100	80	75	80,20	18,18
65	Group 4	80	75	75	80	100	67	50	100	100	75	100	75	40	80	50	76,47	19,05
66	Group 4	80	100	75	60	75	100	75	100	75	75	100	100	60	60	75	80,67	15,45
67	Group 4	80	50	75	20	25	44	25	50	50	50	25	75	80	20	50	47,93	21,87
68	Group 4	80	50	50	20	50	44	100	100	75	25	75	100	60	60	50	62,60	25,40
69	Group 4	60	100	100	40	100	100	50	100	50	100	75	25	60	60	100	74,67	26,76
70	Group 4	100	75	75	80	75	89	25	75	100	75	100	75	60	60	25	72,60	23,08
71	Group 4	60	100	25	80	75	78	75	100	75	75	50	75	80	80	75	73,53	18,23
72	Group 4	60	100	100	40	50	78	75	100	50	75	75	100	60	80	0	69,53	27,56
73	Group 4	60	75	75	40	75	78	50	75	100	75	50	50	60	60	0	61,53	22,86
74	Group 4	80	50	75	0	75	78	100	75	100	50	50	75	60	40	50	63,87	25,35
75	Group 4	20	75	25	60	75	100	75	50	100	75	100	100	70	80	50	70,33	25,67
76	Group 4	80	50	75	80	75	78	75	100	50	50	100	100	80	80	25	73,20	21,23
77	Group 4	50	60	75	40	0	75	25	25	75	50	50	80	60	70	50,00	52,33	22,59

Appendix M. Parameter estimates of mixed-effects models for L2 speech processing tasks

Table 1. Parameter estimates for the fixed-effects model with *Time* (T1, T2) and *Group* (Experimental, Control) as fixed effects.

		β	*SE*	*t*	*Sig.*	*95% CI*
Shadowing	*Intercept*	76.55	22.58	3.39	.001	32.0-120.9
	Time [T1,T1]	-5.72	1.57	-3.63	.000	-8.8-[-2.6]
	Group [Exp, Ctrl]	11.64	4.94	2.35	.021	1.1-21.5
	Time x Group	-5.89	1.70	-3.47	.001	-9.2-[-2.6]
	MaxScore	-2.32	1.29	-1.79	.081	-4.9-.30
Animacy Judgement	*Intercept*	745.03	31.18	23.90	.000	683.3-806.8
	Time [T1,T1]	20.74	22,37	.93	.357	-23.7-65.2
	Group [Exp, Ctrl]	-5.73	33.69	-.17	.865	-72.5-61.0
	Time x Group	-9.30	24.22	-.384	.702	-57.4-38.8
Sentence Verification	*Intercept*	2133.77	49.83	42.89	.000	2035.3-2232.2
	Time [T1,T1]	77.83	15.20	5.12	.000	48.0-107.6
	Group [Exp, Ctrl]	.251	42.66	.006	.995	- 84.5-84.9
	Time x Group	35.13	16.44	2.16	.033	2.9-67.4

Table 2. Parameter estimates for mixed-effects model with *Time* (T1, T2), *Viewing* (Captions, no captions), and *Focus* (FoF, FoM) as fixed effects

		β	*SE*	*t*	*Sig.*	*95% CI*
Shadowing	*Intercept*	87.58	11.39	7.68	.000	64.6-110.5
	Time	-14.10	1.27	-11.05	.000	-16.6-[-11.6]
	Viewing [Captions, NoCptions]	1.97	4.98	.396	.693	-8.0-12.0
	Focus [FoPF, FoM]	-2.02	4.98	-.406	686	-11.9-7.9
	Time x Viewing	2.56	1.78	1.443	.149	-.92-6.1
	Time x Focus	4.61	1.78	2.59	.010	1.1-8.1
	Viewing x Focus	1.05	7.05	.150	.881	-32.0-15.1
	Time x Viewing x Focus	-4.54	2.25	-1.80	.070	-9.5-.4
	Max_score	-2.62	1.30	-1.74	.089	-4.9-.4
Animacy Judgement	*Intercept*	712.19	25.30	28.14	.000	662.0-762.4
	Time	14.19	19.02	.75	.458	-23.7-52.1
	Viewing [Captions, NoCptions]	3.30	35.80	.09	.927	-67.7-74.3
	Focus [FoPF, FoM]	52.12	35.80	1.46	.148	-18.9-123.1
	Time x Viewing	8.36	26.90	.31	.757	-45.3-62
	Time x Focus	-26.20	26.90	-.97	.334	-79.8-27.4
	Viewing x Focus	-2.11	50.80	-.04	,967	-102.8-98.6
	Time x Viewing x Focus	24.39	38.30	.64	.526	-51.9-100.7
Sentence Verification	*Intercept*	2087.93	47.60	43.86	.000	1993.9-2182.0
	Time	149.48	12.50	11.96	.000	124.9-173.9

Viewing [Captions, NoCptions]	67.09	44.24	1.52	.134	-21.0-155.2
Focus [FoPF, FoM]	31.78	44.20	.719	.474	-56.2-119.8
Time x Viewing	-69.00	17.76	-3.88	.000	-103.8-34.2
Time x Focus	-49.57	17.55	-2.82	.005	-84.0-15.2
Viewing x Focus	-10.66	62.96	-.17	.000	-136.1-114.7
Time x Viewing x Focus	91.29	25.09	3.63	.000	42.1-14.5

Appendix L. Parameter estimates of mixed-effects models for L2 phonological accuracy tasks

Table 1. Parameter estimates of pre-/post-treatment time effect for experimental and control groups

		Coefficient	*Std. Error*	*t*	*Sig.*	*95% Confidence Interval*	
						Lower	Upper
ABX [accuracy]	*Intercept*	.558	.1422	3.921	.000	.279	.836
	Group [Exp, Ctrl]	.160	.1542	1.037	.300	-.142	.462
	Time [T1,T1]	-.156	.1435	-1.090	.276	-.438	.125
	Time x Group	.053	.1558	.343	.731	-.252	.359
ABX [RT]	*Intercept*	7.096	.0492	144.137	.000	6.999	7.192
	Group [Exp, Ctrl]	-.037	.0531	-.700	.484	-.141	.067
	Time [T1,T1]	.043	.0190	2.282	.023	.006	.081
	Time x Group	.024	.0205	1.166	.244	-.016	.064
Accent Ratings	*Intercept*	55.27	4.19	13.19	.000	46.85	63.69
	Group [Exp, Ctrl]	2.30	1.92	1.20	.23	-1.47	6.07
	Time [T1,T1]	-3.10	4.18	-.74	.46	-11.39	5.19
	Time x Group	-.58	2.08	-.28	.78	-4.66	3.50

Table 2. Parameter estimates for the analysis of experimental condition effects (*Viewing* and *Task Focus*) on L2 phonological accuracy

		Coefficient	*Std. Error*	*t*	*Sig.*	*95% Confidence Interval*	
						Lower	Upper
ABX [accuracy]	*Intercept*	.652	.1114	5.855	.001	.434	.871
	Viewing [Captions, NoCaptions]	-.065	.1591	-.407	.684	-.377	.247
	Focus [FoPF, FoM]	.215	.1598	1.342	.180	-.099	.528
	Time	.071	.1201	.588	.556	-.165	.306
	Viewing x Focus	-.049	.2277	-.213	.831	-.495	.398
	Time x Viewing	-.312	.1696	-1.837	.066	-.644	.021
	Focus x Time	-.240	.1722	-1.393	.164	-.577	.098
	Time x Viewing x Focus	.390	.2436	1.599	.110	-.088	.867
Accent Ratings	*Intercept*	54.32	3.56	15.25	.001	47.02	61.63
	Time	.58	1.57	.37	.71	-2.52	3.67
	Viewing [Captions, NoCptions]	-2.81	4.42	-.63	.53	-11.61	6.00
	Focus [FoPF, FoM]	-3.88	4.43	-.88	.38	-12.68	4.93
	Time x Viewing	2.19	2.23	.98	.33	-2.19	6.57
	Time x Focus	2.33	2.24	1.04	.30	-2.07	6.72
	Viewing x Focus	4.73	6.30	.75	.46	-7.81	17.27
	Time x Viewing x Focus	-4.50	3.18	-1.42	.16	-10.75	1.75

References

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. Perception & Psychophysics, *59*(7), 1059-1068.

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M.A.Watkins, A.S Rauber, & B.O Batista (Eds.) Recent research in second language phonetics/phonology: Perception and production (pp. 2-31). Newcastle upon Tyne, England: Cambridge Scholars.

Aoyama, K. & Flege, J.E. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. *Journal of the Phonetic Society of Japan, 15(3), 5-13*.

Askildson, L. R. (2011). Theory and pedagogy of reading while listening: Phonological recoding for L2 reading development. Journal of Linguistics and Language Teaching, *2*(2), 267-285.

Baddeley, A. (1992). Working memory. Science, *255*(5044), 556-559.

Baddeley, A. (2010). Working memory. Current biology, *20*(4), R136-R140.

Baddeley, A. D. (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). Psychonomic Bulletin & Review, *7*(3), 544-549.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. The Quarterly Journal of Experimental Psychology, *33*(4), 439-454.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. Psychological review, *105*(1), 158.

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. Canadian modern language Review, *50*(3), 507-531.

Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. Canadian Modern Language Review, *56*(1), 31-48.

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*(4), 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(2), 399-418.

Brett, P. (1995). Multimedia for listening comprehension: The design of a multimedia-based resource for developing listening skills. System, *23*(1), 77-85.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, *28*(2), 191-215.

Carlet, A. & Souza, H. Kivistö de. (2018). Improving L2 Pronunciation Inside and Outside the Classroom: Perception, Production and Autonomous Learning of L2 Vowels. *Ilha do Desterro*, *71*(3), 99-123.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and instruction, *8*(4), 293-332.

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.

Conklin, K., Pellicer Sanchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of memory and language*, *31*(2), 218-236.

Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. Psychological science, *13*(3), 258-262.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance*, *14*(1), 113.

Barriuso, A & Hayes-Harb, C. (2018). *High Variability Phonetic Training as a Bridge From Research to Practice*. The CATESOL Journal, 30, 177-194.

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.

Bianchi, F. & T. Ciabattoni (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In Baldry, A., Pavesi, M. & Torsello, C. Taylor (eds.), *From didactas to ecolingua*. Trieste: Edizioni Università di Trieste, 69–80.

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*(4), 509-533.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 706-755.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PloS one*, *11*(6).

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*(2), 399-418.

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*(2), 399-418.

Bongaerts, Theo & Summeren , Chantal & Planken, Brigitte & Erik Schils ,. (1997). Age and Ultimate Attainment in the Pronunciation of a Foreign Language. Studies in Second Language Acquisition. 19, 447-465.

Brett, P. (1995). Multimedia for listening comprehension: The design of a multimedia-based resource for developing listening skills. System, *23*(1), 77-85.

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. System, *25*(1), 39-53.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and cognitive processes*, *27*(7-8), 1205-1224.

Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., & Jay, C. (2015). Dynamic subtitles: the user experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (pp. 103-112).

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*(4), 604-618.

Clement, J. Global digital population as of July 2019 (in millions). Statista. July 18, 2019. Accessed on July 23, 2019. https://www.statista.com/statistics/617136/digital-population-worldwide/

Collentine, J., & Freed, B. (2004). LEARNING CONTEXT AND ITS EFFECTS ON SECOND LANGUAGE ACQUISITION: Introduction. *Studies in Second Language Acquisition, 26*(2), 153-171.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological review*, *100*(4), 589.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language learning*, 42(4), 497-527.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal des traducteurs/Meta: Translators' Journal*, *49*(1), 67-77.

Darcy, I. (2018) Powerful and Effective Pronunciation Instruction: How Can We Achieve It? *CATESOL Journal*, *30*, 13-45

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*, 372-420.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference.

Darcy, I. & Holliday, J. (2019). Teaching an old word new tricks: phonological updates in the L2 lexicon. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 10-26). Ames, IA: Iowa State University.

Derwing, T. (2003). What do ESL students say about their accents?. *Canadian Modern Language Review*, *59*(4), 547-567.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(4), 499-533.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in second language acquisition*, *26*(2), 249-273.

Doherty, S., & Kruger, J. L. (2018). The development of eye tracking in empirical research on subtitling and captioning. *Seeing into screens: Eye tracking and the moving image*, 46-64.

Drew, D. G., & Grimes, T. (1987). Audiovisual redundancy and TV news recall. *Communication Research*, *14*(4), 452-461.

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL quarterly*, *25*(3), 431-457.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European psychologist*, *12*(3), 196-205.

d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, *28*(3), 227-244.

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In *Eye movements and visual cognition* (pp. 415-427). Springer, New York, NY.

Flege, J. E. (2009). Give input a chance. *Input matters in SLA*, 175-190.

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*(3), 692-707.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, *41*(1), 78-104.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada journal*, 1-22.

Freed, B., Segalowitz, N., & Dewey, D. (2004). Context of learning and second language fluency in french: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs. *Studies in Second Language Acquisition, 26*(2), 275-301.

Fullana, N. 2006. The development of English (FL) perception and production skills: Starting age and exposure effects. In Muñoz, C. (ed), Age and the rate of foreign language learning. Clevedon, UK: Multilingual Matters, 41-64.

Gallardo Del Puerto, F., García Lecumberri, M. L., & Cenoz, J. (2006) Age and native language influence on the perception of English vowels. In B.O. Baptista & M. A. Watkins (ed.) *English with a Latin Beat: Studies in Portuguese/Spanish Interphonology* (pp. 57-69). Amsterdam: John Benjamins.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, *24*(3), 239-258.

Gass, Susan & Winke, Paula & Isbell, Daniel R. & Ahn, Jieun (2019). How captions help people learn languages. A working-memory, eye-tracking study. Language Learning & Technology *23*(2), 84-104.

Gatiss, M., Moffat, S., Vertue, B., Eaton, R., Jones, &Vertue, S. (2010-2017). *Sherlock,* BBC One.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56-92.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*(2), 165-182.

Grimes, T. (1991). Mild auditory-visual dissonance in television news may exceed viewer attentional capacity. *Human Communication Research*, *18*(2), 268-298.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. System, 36(1), 85-93.

Guillory H.G. (1998), The effects of keyword captions to authentic French video on learner comprehension, in: “Calico Journal”, 15 (1/3), 89-108.

Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis?. *Studies in Second Language Acquisition*,*39*(2), 213-224.

Henderson, A., Curnick, L., Frost, D., Kautzsch, A., Kirkova-Naskova, A., Levey, D., ... & Waniek-Klimczak, E. (2015). The English pronunciation teaching in Europe survey: Factors inside and outside the classroom. In *Investigating English pronunciation* (pp. 260-291).

Hulstijn, J.H. (2013). "Incidental learning is second language acquisition". In Chapelle C.A. (Ed.). *The encyclopedia of applied linguistics (*Vol.5, pp. 2632-40) Chichester: Wiley-Blackwell.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in second language acquisition*, *16*(1), 73-98.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation*, *8*(2), 139-143.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kam, E. F., Liu, Y. T., & Tseng, W. T. (2020). Effects of modality preference and working memory capacity on captioned videos in enhancing L2 listening outcomes. *ReCALL*, *32*(2), 213-230.

Kim, H. S. (2015). Using Authentic Videos to Improve EFL Students' Listening Comprehension. *International Journal of Contents*, *11*(4).

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners?. *The modern language journal*, *97*(3), 720-744.

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*,*49*(1), 105-120.

Kruger, J. L., Hefer, E., & Matthew, G. (2013). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In *Proceedings of the 2013 Conference on Eye Tracking South Africa* (pp. 62-66).

Land, M., & Tatler, (2009). *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second language research*, *24*(1), 35-63.

Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language learning*, *57*(1), 87-118.

Lee ,J., Jang, J., Plonsky, L. (2015) The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis, *Applied Linguistics*, *36(*3), 345–366.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly*, *39*(3), 369-377.

Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *Journal of Specialised Translation*, *33*, 70-98.

Liao, S., Yu, L., Reichel, E.D., Kruger, J.L. (In press) Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading.*

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. *Understanding second language process*, 27-44.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, *123*(2), 107.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language teaching*,*46*(1), 1-40.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals 32*(3), 321-328.

Markham, P. (2001). The influence of culture-specific background knowledge and captions on second language comprehension. *Journal of Educational Technology Systems*,*29*(4), 331-343.

Markham, P., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, *31*(3), 331-341.

Markham, P., Peter, L.A., & McCarthy, T.J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals 34*(5), 439-445.

Marsden, E., Mackey, A., & Plonsky, L. D. (2016). The IRIS repository: Advancing research practice and methodology. In *Advancing Methodology and Practice: The IRIS.*

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145-160.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive psychology*, *59*(3), 203-243.

Mayer, R. (2001). Multimedia Learning. Cambridge: Cambridge University Press.

Mayer, R. (20014). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

McQueen, J.M., & Cutler, A. (1998) Spotting (different kinds of) words in (different kinds of) context. *Proceedings of the fifth international conference on spoken language processing*, Sydney, Vol. 6, pp. 2791-2794.

Meskill, C. (1996). Listening skills development through multimedia. *Journal of Educational Multimedia and Hypermedia*, *5*(2), 179-201.

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, *79*(1), 344-351.

Montero Perez, Maribel & Peters, Elke & Clarebout, Geraldine & Desmet, Piet (2014). Effects of Captioning on Video Comprehension and Incidental Vocabulary Learning. Language Learning & Technology 18.1, 118-141.

Montero Perez, Maribel & Van Den Noortgate, Wim & Desmet, Piet (2013). Captioned video for L2 listening and vocabulary learning. A meta-analysis.System 41(3), 720-739.

Mora, J. C. (submitted). Aptitude and individual differences. In Tracey M. Derwing, Murray J. Munro & Ron I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking*. Routledge.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. *A portrait of the young in the new multilingual Spain*, 241-263.

Mora, J. C. (2014). Inter-subject variation in L2 speech perception and cognitive abilities. In Casesnoves, R., Forcadell, M. & Gavaldà, N. (ed.) *Ens queda la paraula. Estudis de lingüística aplicada en honor de M. Teresa Turell*. Barcelona: Institut Universitari de Lingüística Aplicada. 83-101.

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*(2), 381-399.

Mora, J. C. & Mora-Plaza, I. (2019) Contributions of cognitive attention control to L2 speech learning. In Nyvad, A. M., Hejná, M., Højen, A., Jespersen, A. B., &

Sørensen, M. H. (eds.) *A Sound Approach to Language Matters - In Honor of Ocke-Schwen Bohn. Dept. of English, School of Communication & Culture, Aarhus University, Denmark.* 477-499.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *Tesol Quarterly*, *46*(4), 610-641.

Moreira de Oliveira, D. (2020) Auditory selective attention and performance in high variability phonetic training: The perception of Portuguese stops by Chinese L2 learners. PhD Thesis, Universidade do Minho, Portugal.

Moyer, A. (1999). ULTIMATE ATTAINMENT IN L2 PHONOLOGY: The Critical Factors of Age, Motivation, and Instruction. *Studies in Second Language Acquisition, 21*(1), 81-108.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418-440.

Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(03), 313-334.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied linguistics*, *29*(4), 578-596.

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

Muñoz, C., & Llanes, Á. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal*, *98*(1), 432-449.

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, *42*, 258-269.

Mustapha, R., & Kashefian-Naeeini, S. (2017). Moving Teaching and Learning into the Digital Era. *Journal of English Language & Translation Studies*, *5*(3), 27-36.

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, *1*(1), 2-13.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as "comprehensible input": Effects of incidental word learning from context for effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*(1), 95-106.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209.

OECD (2019), *Education at a Glance 2019*: *OECD Indicators*, OECD Publishing, Paris.

Ortega, L., (2014). *Understanding second language acquisition*. Routledge.

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

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media psychology*, *13*(3), 243-272.

Perez, M. M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720-739.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*,*40*(3), 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.

Peters, Elke (2019). The Effect of Imagery and On-Screen Text on Foreign Language Vocabulary Learning from Audiovisual Input. TESOL Quarterly 53(4), 1008-1032.

Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. *Language experience in second language speech learning. In honor of James Emil Flege*, 301-314.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, *29*(2), 191-215.

Pollatsek, A., Bolozky, S., Well, A. D., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and language*, *14*(1), 174-180.

Prensky, M. (2001). Digital natives, digital immigrants. *On the horizon*, *9*(5).

Pujadas Jorba, Geòrgia (2019). Language Learning through Extensive TV Viewing. A study with adolescent EFL learners. Doctoral dissertation: Universitat de Barcelona.

Pujadas, G., & Muñoz, C. (2020). Examining adolescent efl learners'tv viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, 1-25.

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language learning*,*30*(2), 271-287.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (2010). A psycholinguistic perspective on the acquisition of phonology. *Laboratory phonology*, *10*(3), 311-340.

Rato A., Rauber A.S., Kluge D.C., dos Santos G.R. (2015) Designing Speech Perception Tasks with TP. In: Mompean J.A., Fouz-González J. (eds) Investigating English Pronunciation. Palgrave Macmillan, London.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological science*, *20*(1), 6-10.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720-732.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, *14*(3), 191-201.

Reichle, E. D. (2020). *Computational models of reading: A handbook*. Oxford: Oxford University Press. Manuscript in press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading.*Psychological review*, *105*(1), 125.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*,*26*(4), 445.

Rodgers, Michael P. H. & Webb, Stuart (2017). The Effects of Captions on EFL Learners' Comprehension of English-Language Television Programs. CALICO Journal 34.1, 20-38.

Rodgers, R. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes and captions* (Unpublished doctoral dissertation). Victoria University of Wellington.

Rost, M. (2002). Listening tasks and language acquisition. In *JALT 2002 at Shizuoka Conference Proceedings. Retrieved September* (Vol. 25, p. 2012).

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology.*Language awareness*, *20*(1), 45-59.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly*, *46*(4), 842-854.

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563-595.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences.*Applied Psycholinguistics*, *37*(4), 813-840.

Scales, J., Wennerstrom, A., Richard, D., & Wu, S. H. (2006). Language learners' perceptions of accent. *Tesol Quarterly*,*40*(4), 715-738.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270-1281.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. London: Routledge.

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in At Home and Study Abroad Contexts. *Studies in Second Language Acquisition, 26*(2), 173-199.

Solon, M., Park, H. I., Henderson, C., & Dehghan-Chaleshtori, M. (2019). Revisiting the Spanish elicited imitation task: a tool for assessing advanced language learners?. *Studies in Second Language Acquisition*, *41*(5), 1027-1053.

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated?. *TESOL quarterly*, *42*(2), 181-207.

Suárez, M. D. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference?. *The Language Learning Journal*, *47*(4), 497-517.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language learning*, *26*(2), 233-253.

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language learning & technology*, *14*(2), 50-73.

Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PloS one*, *13*(6).

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, Ł., & Kruger, J. L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures*, *17*(2), 183-204.

Thomson, R. I. (2012). ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong. In *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 224-233).

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, 4(2), 208-231.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344.

Thomson. R. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. Language Learning 62.4. 1231–1258.

Timmis, I. (2002). Native-speaker norms and International English: a classroom view. *ELT journal*, *56*(3), 240-249.

Van Orden, G. C., Johnston, J. C., & Hale, L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 371.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, *40*(3), 191-210.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, *42*(4), 272-281.

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language teaching*, *43*(1), 1-37.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching.* London: Palgrave Macmillan.

Vanderplank, R. (2016a). 'Effects of'and 'effects with'captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners?.*Language Teaching*, *49*(2), 235-250.

Vanderplank, Robert (2019). 'Gist watching can only take you so far': attitudes, strategies and changes in behaviour in watching films with captions. Language Learning Journal *47*(4), 407-423.

Walter, C. (2008). Phonology in second language reading: Not an optional extra. *TESOL quarterly*, *42*(3), 455-474.

Webb, S. (2015). Extensive viewing: Language learning through watching television. D. Nunan, J.C. Richards (Eds*.),* Language learning beyond the classroom, Routledge, New York/London (2015), pp. 159-168.

Webb, S., & Rodgers, M. P. (2009). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-366.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eye-movement recordings in second language research. *Studies in Second Language Acquisition*, *35*(2), 205-212.

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65-86.

Wisniewska, N., & Mora, J. C. (2018). Pronunciation learning trough captioned videos. In *Proceedings of the 9 Annual Pronunciation In Second Language Learning And Teaching Conference* (ISSN 2380-9566) (p. 204).

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*(4), 497-528.