The role of captioned video in developing speech segmentation for learners of English as a second language

Tendai Juma Charles

PhD

University of York

Education

April 2017

Abstract

Over three decade's worth of research into the effects of captioned video on second language learning suggests that it may improve L2 listening comprehension (Vanderplank, 2010). However, there are several limitations to the experimental designs of some early studies in this field (Vanderplank, 2013), thus there is a need for more robust research to be conducted. Mitterer and McQueen (2009) designed an innovative experiment to investigate the effects of captioned video on L2 speech perception. The aim of this research is to build upon their work, primarily by investigating whether repeated exposure to captioned video, delivered by several speakers from a broadly similar variety of English, could (a) lead to improved *speech segmentation* when listening to speakers that the learner has never heard before, and (b) improved L2 listening *comprehension*. For example, whether watching a number of documentaries with subtitles could lead to enhanced listening comprehension when watching subsequent subtitles-free documentaries delivered by different presenters.

The main contribution of this study is that it is the first to specifically investigate the ability of participants to segment the speech of (a) previously encountered utterances; (b) different utterances by the same speaker; and (c) different utterances by different speakers of a similar accent. A pre-test / intervention / post-test experimental design was performed multiple times on international university students in the UK. Participants who watched captioned video during the treatment phase, consistently outperformed control groups, which suggests that L2 learners of English can improve their L2 listening skills (and more specifically, their L2 speech segmentation ability) by simply watching same-language subtitled TV programmes on a regular basis.

Table of Contents

List of tables

List of figures

Acknowledgements

I would like to thank Dr. Danijela Trenkic, my Supervisor and mentor, for her highly supportive supervision during my time as her postgraduate student. As the first member of my family to work and study within the field of Higher Education, I was in need of guidance. Dr. Danijela advised me to regularly attend academic conferences, she helped me to publish my first academic paper, and she prepared me for a career in academia. The constant support and encouragement I received from her whilst facing the hurdles of postgraduate research is priceless and I am forever grateful.

I would also like to dedicate this work to my beloved children and thank them for their patience whilst I spent countless hours away from home working on this research.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Some of the findings presented in "Chapter 3: Study 1" were published jointly with my Supervisor here: Charles, T. J. & Trenkic, D. (2015) *Subtitles and language learning: Principles, strategies and practical experiences.* Gambier, Y., Caimi, A. & Mariotti , C. (eds.). Bern: Peter Lang, p. 173-198.

1.1 The challenge of listening in a second language

Although listening in a native language (L1) appears to be an effortless task, it is actually a highly complex mental process (Vandergrift, 2011), which takes up 40% - 50% of a person's communication time (Gilman and Moody, 1984; Burley-Allen, 1995; Feyten, 1991). Listening requires the use of linguistic knowledge for 'bottom-up' processing, which enables listeners to identify what words are actually present within a given speech signal (Flowerdew and Miller, 2005). At the same time, listeners use contextual knowledge of the topic at hand for 'top-down' processing, in order to interpret the intended meaning of what the speaker is saying (Ur, 1984). It is this combination of bottom-up and top-down cognitive processes that makes listening comprehension a rather complex task, because listeners must identify words in a speech stream as well as apply schematic knowledge to those words in an attempt to accurately understand the speaker's intended message (Buck, 2001).

This thesis focuses on the fundamental problem with listening, which, for listeners, is the fact that speech comes as a continuous stream of sound without visible boundaries between words. When a person reads an orthographic sentence such as '*British weather is so awful*', she can visually see gaps between the words. These gaps (or spaces) indicate to the reader where one word starts and ends. However, if that person were to listen to an oral version of the same sentence, it would be spoken as '*britishweatherissoawful*', without any breaks between words. How then, does the listener know '*soawful*' are the two separate words '*so*' and '*awful*'? Indeed, how does she know "*awful*" is one word, as opposed to '*awe*' and '*full*'? This is known as the problem of 'speech segmentation', which this thesis defines as the cognitive process of identifying word boundaries in a continuous stream of real-time, authentic speech.

Listeners use lexical (Norris, 1994) and phonetic (Vroomen et al., 1998) cues in the speech stream to help them segment these words, however, cues vary from one language to another, depending upon its rhythmic structure (McQueen et al., 2001). The rhythmic structures of languages can be syllable-timed, mora-timed, or stress-timed. In a syllable-timed language like French, the temporal duration of every spoken syllable is approximately equal (Roach, 1982), whereas, in a mora-timed language (like Japanese), it is the duration of each mora that is equal, not the spoken

syllables. However, in a stress-timed language like English, it is the duration between two stressed syllables that is roughly equal (Cutler, 1990).

This is significant because research has indicated that due to the different rhythmic structures of languages, L1 speakers instinctively develop language-specific segmentation strategies (Cutler, 2000). For instance, in a stress-timed language such as English, spoken syllables are either strong or weak, with strong syllables containing full vowels and weak syllables containing reduced vowels (Cutler, 1990). Take the two-syllable word '*British*' as an example, when articulated by L1 speakers, it has the stress pattern 'Oo', which means it sounds like '*BRITish*'. Similarly, '*weather*' follows the same stress pattern, so is pronounced as '*WEAther*', with the vowel of the weak syllable sounding like a schwa. Therefore, considering that 90% of spoken content words in English are stress-initial (Cutler and Carter, 1987), when pronounced with the correct stress pattern, the sentence '*britishweatherissoawful*' is actually spoken as '*BRITishWEAtherissoAWful*'. Consequently, L1 speakers of English use the Metrical Segmentation Strategy (Cutler and Norris, 1988) when listening to spoken discourse, whereby they segment speech whenever they hear a strong syllable, because they instinctively assume this is likely to be the start of a new word.

By using the Metrical Segmentation Strategy, it is easy for L1 speakers of English to segment a real-time continuous stream of spoken English discourse (Flowerdew and Miller, 2005). However, because this language-specific segmentation strategy is an integral part of an L1 English speaker's listening process, she applies it to any language she listens to, even if it is not the most efficient strategy to use for segmenting speech in that language (Cutler, 2000). For example, it was noted earlier that French is a syllable-timed language, thus L1 speakers of French segment spoken French discourse by using a syllabically based segmentation strategy (Cutler and Norris, 1988). Yet, when L1 speakers of English listen to French, they *do not* use a syllabically based segmentation strategy, rather they apply the Metrical Segmentation Strategy when segmenting French speech, which actually slows their ability to process what they are listening to. Cutler (2000, p. 1) aptly refers to this phenomenon as "listening to a second language through the ears of a first".

These cross-linguistic differences between languages are what make second language (L2) speech segmentation such an arduous task (Flowerdew and Miller, 2005). Because "it is harder [*for L2 listeners*] to determine which bits of the acoustic blur that hits your ears are the beginnings and ends of words… which bits *are* words, and what words they might be..." (Brown 1996, p. 2). Indeed, L2 language learners frequently cite their inability to accurately segment speech as a fundamental hindrance to their L2 listening comprehension (Goh, 2000).

1.2 Bimodal input theory

Even though L2 listening, and speech segmentation in particular are difficult tasks, through adequate experience, exposure and training, it is possible for L2 learners to improve their abilities (Cutler, 2000). One proposal, dating back to Price's (1983) initial study, claims that exposure to 'bimodal input' can be a form of training to aid the L2 listening process. 'Bimodal input' can be defined as the simultaneous presentation of aural and matching orthographic input. For example, when a person watches a TV show with same-language subtitles, they listen to aural input while reading an orthographic representation of the words they hear. This proposal has triggered a large amount of research on its benefits for literacy development, vocabulary acquisition, general language training, and the improvement of L2 listening comprehension (*see* Vanderplank, 2010 for a comprehensive review).

Focusing on L2 listening comprehension specifically, a review of early studies in this body of research reveals a commonly used methodological approach, wherein participants are divided into a control group and a bimodal group, who each watch a TV show for a short period of time. The control group watches the TV show without subtitles, whilst the bimodal group watches the same TV show with same-language subtitles. Following the viewing, participants sit some form of listening comprehension test, and the groups' results are then compared. For the majority of studies that adopt this type of methodological approach, researchers frequently report that participants in the bimodal condition outperform participants in the control condition on the listening comprehension tests. Hence, proponents of bimodal input theory argue that these consistent findings provide sufficient support for the claim that bimodal input can aid L2 listening comprehension.

One could argue that such a claim is not warranted though, because although these studies show that participants' comprehension scores were better in the bimodal condition, they do not prove that bimodal input actually improved *listening* comprehension. On the contrary, it is possible to assume that their higher comprehension scores were merely a result of *reading* the subtitles, especially since the so-called 'listening comprehension' tests used in these early experiments never required participants to listen to anything.

More recent investigations into the potential benefits of bimodal input on L2 listening comprehension have improved on such methodological limitations. For example, Mitterer and McQueen (2009) were the first researchers in this field to use a test that actually requires participants to listen while being tested. The aim of their study was to investigate whether watching subtitled TV programmes could help L2 learners of English adapt to unfamiliar regional accents. 121 Dutch people who were "fluent in English" (ibid, p. 2) participated in this experiment. These participants watched 25-minutes of a TV show that was either spoken in a heavy Australian accent or a strong Scottish accent. They were divided into groups and either watched the programme with L1 subtitles, or L2 subtitles (bimodal condition), or without subtitles.

After the viewing, participants performed a shadowing task, wherein they heard 160 short, pause-bound utterances, and then repeated whatever words they were able to segment. The utterances were excerpts from the TV shows, half of which were previously presented to participants during the viewing; whilst half of them were from scenes the participants had not watched. Crucially, participants in the bimodal condition scored significantly higher on both types of excerpts in this test than the other groups, suggesting that they were indeed able to adapt to these regional accents. Such a test is a valid approach to testing the bottom-up processing of listening (Buck, 2001), and due to the robust design of the experiment, coupled with their findings, Vanderplank (2013) says "For me, this excellent study is almost as big a breakthrough as Karen Price's article" (p. 7).

Mitterer and McQueen's (2009) study was highly informative, however, three questions arise, which could warrant further research. First of all, recall that L1 speakers of English use the Metrical Segmentation Strategy to segment speech; well,

this is actually the same for L1 speakers of Dutch, because Dutch is also a stress-timed language. So the first question here is, would participants whose L1 has a completely different phonological structure to English, also perform so well on their test? Secondly, the participants in Mitterer and McQueen's (2009) study clearly had a very high level of English. Would a replicated experiment on lower level L2 learners yield similar results? Finally, their experiment was conducted on one day, and they found significant results after one single exposure to bimodal input. How would participants benefit from repeated exposure to bimodal input?

1.3 The present study

Early studies took the approach of dividing participants into a bimodal or non-bimodal group, making them watch a TV programme in those conditions, and then testing listening comprehension with a written summary or multiple-choice question (MCQ) test. It could be argued that the listening tests they employed lacked test construct validity because they required participants to read and write, but did not require them to actually listen to anything. Mitterer and McQueen (2009) designed a better experiment because participants were expected to listen to short excerpts and segment them. The findings of their research are undoubtedly insightful, however, their study was conducted on (a) very advanced level L2 learners, (b) on learners whose L1 shares the same rhythmic structure as the L2, and (c) the duration of their study was very short.

The present study attempts to contribute new knowledge to the field by addressing these issues. Using Mitterer and McQueen's (2009) research as a basis, the aim of this study is to investigate the effects of bimodal input on (a) *intermediate* level users of English, (b) learners whose L1 does *not* have a stress-timed rhythmic structure, and (c) to see whether *repeated* exposure to bimodal input can improve L2 speech segmentation. Furthermore, whereas Mitterer and McQueen (2009) found that learners in the bimodal condition were able to adapt to the accents of specific speakers, this study aims to discover whether bimodal input can lead to *generalisation of learning*, whereby the learners are able to better comprehend the speech of new speakers from a broadly similar accent.

1.4 Organisation of this thesis

Having introduced the overarching aims of this study in Chapter 1, Chapter 2 will discuss bimodal theory, review several studies that look at the effects of bimodal input on L2 listening, and then state the aims of this present study. Chapter 3 explains how L2 listening was measured in this study, describes the main features of the experiments that make up research, and presents the methods used for statistical analyses. Chapter 4 then presents the first study, which is comprised of three pilot experiments that investigated the general L2 speech segmentation ability of international university students. After that, Chapter 5 describes the second study, which focused on the effects of bimodal input on L2 speech segmentation. The third and final study, which explores the effects of bimodal input on both L2 segmentation and general listening comprehension, is then presented in Chapter 6. Finally, Chapter 7 summarises findings from the three studies, discusses possible pedagogic implications, and proposes suggestions for future research.

CHAPTER 2: LITERATURE REVIEW

This literature review discusses evidence that bimodal input may help language learners improve their L2 listening comprehension in English. The chapter begins by discussing two conceptual models that provide a theoretical basis for understanding why bimodal input may aid the L2 listening comprehension process. Research into bimodal input is then critically reviewed and a number of limitations in the literature are identified. The chapter then looks at speech segmentation, highlighting cross-linguistic differences between languages, and proposes that bimodal input can help L2 speech segmentation. Next, recent research in the field is presented, and crucial gaps in the literature are identified. Finally, the aims and research questions of this thesis are stated.

2.1 Bimodal input

In 1983, a groundbreaking article was published by MATESOL entitled *Closed-captioned TV: An untapped resource*, authored by Prof. Karen Price of Harvard University. As captioned video was growing in popularity during that era, Price set out to investigate whether exposure to captioned video would significantly improve second language listening comprehension. Price recruited 500 university students from a variety of L1 backgrounds and divided them into two groups, one group watched captioned videos in the English language, while the other group watched the same video without captions. Half of the participants in each group had one viewing, whilst the other half had two viewings. All participants then sat a listening comprehension test. Despite coming from different L1 backgrounds, and with varying levels of education, Price found that participants who watched captioned videos significantly outperformed participants who were not exposed to captions, even after a single viewing.

Price's (1983) research was pioneering because it was the first study that investigated this. It led to over thirty years worth of research into the potential effects of captioned video viewing on L2 language learning. This section looks at the 'theory' of bimodal input: claims about its effects on L2 reading and vocabulary acquisition, and research on its effects on L2 listening comprehension, going on to discuss methodological concerns before specifically reviewing literature on its effects on speech segmentation.

2.1.1 Bimodal input theory

To-date, no conceptual model has been developed to explicitly propose why watching captioned video may be beneficial for language learners, or, cognitively speaking, how it improves their abilities. This raises the question – why would viewing captioned video have such a significant impact on L2 comprehension?

Paivio's (1986) Dual Coding Theory may offer a conceptual explanation and is sometimes cited in the literature as potential theoretical support (Danan, 1992; Harji et al. 2010). Essentially, Dual Coding Theory (*see* Figure 1) proposes that (a) the human mind is composed of two subsystems, one specifically for processing visual stimuli (imagery) and the other for processing verbal stimuli (words); and (b) humans remember information and learn it better when they receive it as a combination of imagery and words (rather than only one or the other) because the activation of both these subsystems leads to enhanced long-term memory storage and recall (Paivio and Csapo, 1973).

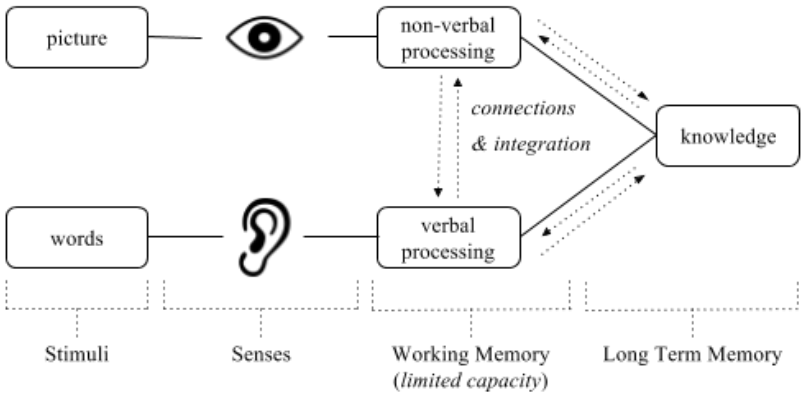

Figure 1. Simplified model of Paivio's (1986) Dual Coding Theory (adapted from Koumoundouros, 2015)

A number of studies have provided empirical support for Dual Coding Theory (*see* Clark and Paivio, 1991; Paivio and Begg, 1981; Paivio, 1986). Most of these experiments have focused on presenting participants with word-word pairs, picture-picture pairs, and picture-word pairs to activate both cognitive subsystems (Paivio, 2013). Research on Dual Coding Theory has been applied to cognitive phenomena such as problem-solving, mnemonics, concept learning, language learning and

bilingualism (ibid). However, it should be noted that literature on Dual Coding Theory focuses on the dual presentation of an image (non-verbal input) with text (verbal input). As far as Dual Coding Theory is concerned, aural and orthographic input are both representations of words, which would be handled by the one cognitive subsystem that handles verbal processing, and a video (with or without captions) introduces a third modality of imagery, which is processed by the subsystem that handles non-verbal processing.

That understanding differs slightly from the body of research into captioned video, which follows Price (1983) because her research focuses on the simultaneous presentation of spoken words (aural input) with accompanying text (orthographic input). The researchers in this field acknowledge that images are present on screen, but usually do not stress their impact on second language learning in this context. This may be because often, when watching captioned video, the spoken and written words do not necessarily match the image. It is important to note this because the term 'bimodal input' in this thesis (and the literature) refers to the combination of aural and matching orthographic input, regardless of the images displayed on screen. Having said that, there is no doubt that orthographic input is a type of visual stimuli that is received via the eyes, therefore, bimodal input may find theoretical support in Dual Coding Theory because, as Figure 1 indicates, there is a connection between the non-verbal and verbal subsystems.

A more recent theory that provides a conceptual model for understanding what happens when L2 learners watch captioned video is the Cognitive Theory of Multimedia Learning (Mayer and Moreno, 2003). This theory is built upon Paivio's (1986) Dual Coding Theory and Baddeley's (1998) theory of working memory (discussed in Chapter 3). Looking at Figure 2, there are two rows that represent two information-processing channels – namely, the auditory/verbal channel at the top and the visual/pictorial channel at the bottom. There are also five columns that signify modes of knowledge representation, which are:

1. Physical representations (e.g. words and/or pictures that are presented to the L2 learner);
2. Sensory representations (in the ears and/or eyes of the L2 learner);

3. Shallow working memory representations (e.g. sounds and/or images attended to by the L2 learner);
4. Deep working memory representations (e.g. verbal and pictorial models constructed by the L2 learner); and finally
5. Long-term memory representations (e.g. the learner's relevant prior knowledge).

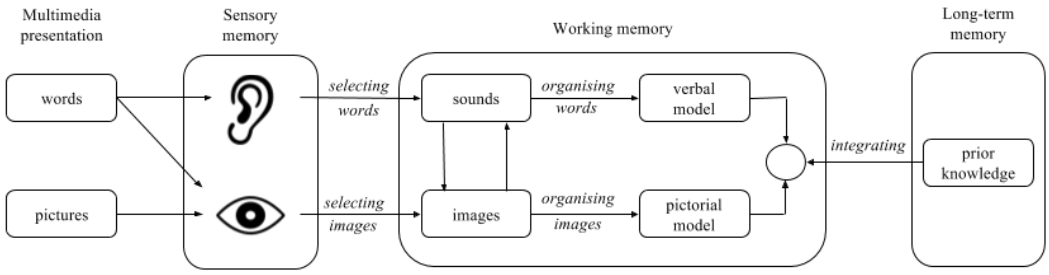

Figure 2. Cognitive Theory of Multimedia Learning (adapted from Mayer and Moreno, 2003)

The arrows in this model symbolise cognitive processing. Significantly, the arrow from 'words' to the eyes represents orthographic input reaching the eyes; whilst the arrow from 'words' to the ears represents aural input reaching the ears; and the arrow from 'pictures' to the eyes represents images reaching the eyes. The arrow labeled 'selecting words' signifies the L2 learner's attention to some of the auditory input received from the ears, and the arrow labeled 'selecting images' refers to the L2 learner's paying attention to some of the visual input received by the eyes. The arrow labeled 'organising words' is indicative of the L2 learner's construction of a meaningful representation of the incoming verbal input, and the arrow labeled 'organising images' signifies the L2 learner's construction of a meaningful representation of the incoming visual input. Lastly, the arrow labeled 'integrating' denotes a merging of the verbal model with the pictorial model and relevant prior knowledge (Mayer and Moreno, 2003).

Empirical research has used this model as a theoretical framework and found that learners are better at comprehending input when it is presented both visually, as an image, and verbally, as spoken or written input (Mayer and Moreno, 2003). However, this body of research also makes two interesting claims: (a) that captioned video

actually impairs overall comprehension of the input, due to a 'redundancy effect', which happens because the learner gets overloaded with too much of the same input, i.e. the combination of subtitles, and oral narration (Mayer et al. 2001; Moreno and Mayer, 2002); and (b) that if learners are exposed to text and matching speech without any form of animation or video input, they learn better (Moreno and Mayer, 2002) because the brain's non-verbal subsystem is not overloaded by imagery and subtitles (Mayer and Moreno, 2003).

From these two theories, it can be proposed that when watching captioned video there are two sources of verbal input coming from different modalities; however different verbal modalities actually interact with one another when activated, because hearing a word automatically activates knowledge of how it is spelled (cf. Seidenberg and Tanenhaus, 1979). Therefore it is possible that the simultaneous presentation of spoken and written forms of a word (i.e. bimodal input) may leave more memory traces than simply presenting one modality and result in enhanced learning. There have been a plethora of studies in the past thirty years supporting this claim, and an online bibliography of over 150 publications is maintained by Burger (2013). Many of the peer-reviewed publications on this list have been reviewed by Vanderplank (2010; 2013) and more recently by Perez et al. (2013) in a meta-analysis. As rightly noted by Vanderplank (2016) many of these studies are similar, in that they follow the same fundamental experimental design but have been carried out in different contexts. The following section briefly describes a selection of these studies, which show how bimodal input may affect L2 reading and vocabulary acquisition.

2.1.2 Bimodal input and L2 reading / vocabulary acquisition

There are a number of studies dating back to the 80s that claim watching captioned videos can have positive effects on L2 (and even L1) *reading* comprehension (*see* Goldman and Goldman, 1988; Bean and Wilson, 1989; Neuman and Koskinen, 1990; Markham and Peter 2003; Kothati et al. 2002; Kothari et al. 2004; Chen, 2012; Baharani and Ghafournia, 2015). Taking a recent publication as an example, Baharani and Ghafournia (2015) investigated the effects of multimodal texts on the English reading comprehension of Iranian secondary school students. They defined a multimodal text as "a combination of spoken and written languages, still or moving images, which can be presented on paper or electronic screen" (p. 161). Eighty female

Iranian students participated in this study, who were classified as intermediate level users of English according to an Oxford Placement Test. Participants were randomly assigned to one of four treatment groups; (a) a control group, which read a printed text normally, (b) a multimodal printed group, which read a printed text with accompanying pictures, (c) a multimodal non-printed group, which watched captioned videos, and (d) a multimodal printed and non-printed, which read a printed text with accompanying pictures and watched captioned videos. The duration of this experiment was just over two months, and reading comprehension of the texts was measured with a twenty-point multiple-choice question test. Comparing performance on pre- and post-tests revealed that participants in the multimodal printed group outperformed other participants; however, participants who watched captioned videos outperformed the control group. Ultimately, the authors suggested that watching captioned videos is beneficial for reading comprehension, L2 learner motivation, and even target language pronunciation.

Similarly, a number of studies dating back to the 90's claim that watching captioned videos can have positive effects on L2 *vocabulary* acquisition (*see* Baltova, 1999; Hui, 2005; Yuksel and Tanriverdi, 2009; Zarei, 2009; Harji et al. 2010; Sydorenko, 2010; Zarei and Saddeghi, 2011; Karakas and Sariçoban, 2012; Pak, 2012; Ahm, 2013; Hu and Huang, 2013; Rodgers, 2013; Gorjian, 2014; Hoogendyk et al. 2014; Naghizadeh and Darabi, 2015; Rostam et al. 2015; Ebrahimi and Bazaee, 2016; Peters et al. 2016). Again, to take a recent example, Peters et al. (2016) investigated whether L2 learners of English would benefit more from L1 subtitles or L2 subtitles for vocabulary learning. For the sake of brevity, only the first of two experiments they conducted is reported here. Essentially, 28 Belgian secondary school students participated in this study, which spanned two weeks. Participants were assigned to either the 'L1 subtitles' or the 'L2 subtitles' groups, and in week one, they sat a vocabulary size test, a vocabulary form recognition test, and a meaning recall test. In week two, participants watched 13 minutes of a documentary about eating insects. Thirty-nine target lexical items were selected from this 13-minuted segment of the documentary and were used in the posttest. Ultimately, participants in the 'L2 subtitles' group outperformed the 'L1 subtitles' group on the vocabulary form recognition test, but scores were not significantly different for the meaning recall test. The authors concluded by stating that "[t]he findings of the two studies presented

here show that EFL learners can indeed learn new words when watching a TV program with L1 subtitles or captions in class" (Peters et al. 2016, p. 146).

Crucially, the findings of Baharani and Ghafournia (2015) and Peters et al. (2016) are consistent with earlier studies. Research consistently demonstrates the beneficial effects of bimodal input for L2 reading and vocabulary acquisition. However, because the scope of this thesis is L2 listening comprehension specifically, the following section reviews studies focusing on this area.

2.1.3 Bimodal input and L2 listening comprehension

There are a number of studies dating back to the 80s which claim that watching captioned videos can have positive effects on L2 *listening* comprehension (*see* Markham, 1989; Huang and Eskey, 1999; Yoshino et al. 2000; Chai and Erlam, 2008; Tsai, 2010; Hayati and Mohmedi, 2011; Latifi et al. 2011; Etemadi, 2012; Sharif and Ebrahimian, 2013; Barekat and Farrokhian, 2014; Zareian, et al. 2015; Bahrani and Tondar, 2016; Saed et al. 2016; Rodgers and Webb, 2017). Three publications will be presented here as illustrative examples of how studies in this field are usually conducted. To begin with, Markham (1989) investigated the role of captioned video on second language listening comprehension. Participants in Markham's (1989) experiment were 76 university-level students studying English as a second language. In groups, these participants watched a TV programme without subtitles, then another TV programme with subtitles, or vice-versa. Following each viewing, participants underwent multiple-choice reading comprehension tasks. The results indicated that participants demonstrated better performance on the test after being exposed to the TV programme with subtitles than after they had watched the TV programme without subtitles. This finding led Markham to support the claim that watching captioned videos improved students' L2 listening comprehension.

Some years later, Markham et al. (2001) investigated the effects of bimodal input on L2 listening comprehension. They conducted an experiment on 169 American university students, who were intermediate-level learners of Spanish as an L2. Participants watched a 7-minute DVD about Apollo 13, either with L2 subtitles or with L1 subtitles; and a control group watched the video without subtitles. Immediately after this viewing, participants were given 10 minutes to write a

summary in their L1 (i.e. English) about the video, after which they completed a 10-item multiple-choice comprehension test. The results of these tests indicated that participants exposed to L2 subtitles outperformed the other two groups on both the written summary and the multiple-choice test; suggesting (again) that watching L2 subtitles leads to better L2 listening comprehension.

Precisely 10 years later, Hayati and Mohmedi (2011) investigated the effects of L2 subtitles on listening comprehension. The conducted an experiment on 90 undergraduate Iranian students, who were intermediate-level learners of English as an L2. The study was performed over a 6-week period where in each session, participants watched a 5-minute (approx.) DVD about natural disasters either with L2 subtitles, or with L1 subtitles, or without subtitles (the control group). Immediately after each viewing, participants completed a 10-item multiple-choice comprehension test, with every question containing language that was actually used in the DVD. There was also a final comprehension test, and participants were asked to write a review of the entire procedure. The results of these tests indicated that participants exposed to L2 subtitles outperformed the other two groups on the multiple-choice test; suggesting that watching captioned videos with L2 subtitles leads to better L2 listening comprehension. Feedback from participants regarding the experiment also implied that they found L2 subtitles to be useful.

From these three sample studies, three main points can be drawn, as they are consistent with most studies of the effects of captioned videos on L2 listening comprehension. Firstly, the experimental design – quite simply, participants watch a TV programme and then sit a listening comprehension test to determine how well they understood what they viewed. Secondly, the materials – participants are usually required to watch a portion of a TV programme such as a documentary, sitcom or film. Thirdly, and most importantly, are the treatment conditions. Most studies have a 'control' group who watch the TV programme normally (i.e. with sound but not subtitles); then there is an L1 subtitles group (i.e. a group who watch a TV programme with English being spoken but subtitles onscreen in the viewers' native tongue); and finally an L2 subtitles group (i.e. a group who watch a TV programme spoken in English whilst watching English subtitles. To differentiate between L1 and L2 subtitles, the literature uses a variety of terms such as '*inter*lingual subtitles',

'*intra*lingual subtitles', 'same-language subtitles', 'closed captions', and 'teletext'. For the sake of simplicity, this thesis will refer to the presentation of L2 sound and L2 text as 'bimodal input'.

Critically, the three studies presented above consistently found that bimodal input can lead to improved L2 listening comprehension. Several methodological issues, however, make it difficult to accept this claim at face value. In fact, this is true not only for these three studies, but also for most other research prior to 2010 because "many had weak research designs and/or a poor selection of materials" (Vanderplank 2013, p. 240). The following section discusses some of these methodological problems and suggests how they can be addressed.

2.1.4 Addressing methodological concerns in the literature

Looking at the body of research related to the effects of captioned video on L2 listening comprehension, it is clear that there is a high level of reliability, because time and time again, participants exposed to bimodal input (L2 sound and L2 text) outperform participants in other treatment conditions on comprehension tests. However, there are two fundamental methodological issues, which may cast some doubt on the validity of these studies; namely, (a) test construct validity, and (b) experimental design.

2.1.4.1 Test construct validity

Test construct validity is a major factor to consider when designing listening tests, and is concerned with the extent to which the test actually tests what it claims to be testing (Buck, 2001). For example, imagine a gap-fill task on a listening test that requires the test-taker to listen to a monologue, spot the word '*accommodation*', and write it on the exam paper. A student with poor writing skills sits this test, accurately hears the target word, but writes it as '*acomodation*', so receives a score of zero due to the misspelling. A test of this nature lacks construct validity because it is clear the student heard the word correctly, however, due to her weak writing skills is penalised.

Of the three publications presented in the previous section, it can be seen that Markham (1989) used multiple-choice reading comprehension tasks to test participants' understanding of the TV programmes they watched. Thus, the question

here is to what extent does a multiple-choice *reading* comprehension test actually test participants' L2 *listening* comprehension? According to Rost (2002) tests of listening comprehension that require students to read are as much reading tests as they are listening tests. This is quite simply because participants must possess the ability to read, understand and comprehend the actual test questions in order to answer them. Furthermore, many comprehension tests (regardless of being in written or spoken form) are mostly memory tests (Brown, 1990), because they test not so much what participants heard, but rather, (a) what they can actually remember from what they heard, or (b) what they can effectively reconstruct from what they have heard based upon their background knowledge of the topic at hand.

Markham et al.'s (2001) study was designed slightly differently and improved upon Markham's (1989) experiment. However, questions regarding test construct validity still arise. For example, referring to the data analysis used for the written summary, Markham et al. (2001, p. 442) said "the researchers counted the number of written idea units, elaborations, and distortions generated by the participants". Such a description sounds like a rubric more suitable for a writing test than for a listening test. In order for students to write a summary with a reasonable number of 'idea units', they needed to have advanced writing skills. They also needed to be good at writing quickly, as they were only given ten minutes to complete this task.

In Hayati and Mohmedi's (2011) experiment, the DVD viewings were very short and realistically, participants can guess the answers on a multiple-choice test, consequently, it is not possible to make empirical claims based upon MCQ tests alone. Indeed, the authors themselves identified this as a potential limitation to their study, stating: "Assessment of the learning outcome was measured only with multiple-choice tests. The problem lies with the need to devise alternative assessment techniques that tap various aspects of listening comprehension" (p. 191).

This issue of test construct validity is prevalent throughout the literature, as is evident from the thorough meta-analysis conducted by Perez et al. (2013), who identified two main types of assessment used in the literature; namely, receptive listening comprehension tasks and productive listening comprehension tasks. Studies that employed a receptive listening assessment type would use written multiple choice

questions, aural multiple-choice questions, and true/false tests. Studies that employed a productive listening assessment type would use recall protocol tests, open-ended questions, and fill-in-the-gaps tasks. Ultimately, the findings of Perez et al.'s (2013) meta-analysis suggest that it is better to use receptive listening assessment tasks when testing L2 listening comprehension in bimodal research experiments.

2.1.4.2 Experimental design

Regarding the experimental design of these studies, several questions can also be raised. Recall that the main claim of Markham's (1989) study is that watching captioned video leads to better *listening* comprehension. Yet, it is possible that the improved performance of participants in the bimodal group was due to their good reading ability in the L2, and not because they were actively engaged in the act of listening. In fact, a group of hearing-impaired students would also perform better on their comprehension test after watching a programme with subtitles than without. Now, it is possible that watching the captioned video may have improved participants' general comprehension of the material, however, claiming that *reading* subtitles improved their *listening* may not be justifiable. Crucially, the design of studies like this cannot warrant any conclusions regarding participants' *listening* skills.

As for the Markham et al. (2001) study, a potential problems lies in the actual choice of treatment materials presented to participants. Although it was commendable of the researchers to use authentic language materials, it is reasonable to assume that the sample of educated, American university students recruited would have been well-versed in (or at least aware of) the story of Apollo 13. Therefore, it is possible that their background knowledge of the topic may have influenced their ability to construct and write a summary about it, without needing to comprehend what was presented to them during the experiment.

Finally, reviewing the experimental design of Hayati and Mohmedi's (2011) experiment, they (like many other researchers) randomly assigned participants to one of three treatment conditions; namely, a group that watched the video with L2 subtitles, or a group that watched it with L1 subtitles, or one that was given no subtitles, as visualized in Table 1. Hayati and Mohmedi (2011) found that the L2

subtitles group outperformed the L1 subtitles group, which outperformed the no subtitles group. However, it could be possible to conclude from this that reading the subtitles alone aided their performance on the comprehension tests. If a fourth group was included in their study, a group who were exposed to orthographic input (i.e. they read subtitles) but not aural input (i.e. there was no oral narration), then more meaningful conclusions could be drawn about the effectiveness of bimodal input.

Table 1. The three treatment groups in Hayati and Mohmedi's (2011) study

Group	*Orthographic Input*	*Aural Input*
'L2 subtitles'	✓	✓
'L1 subtitles'	✓	✓
'No subtitles'	X	✓

2.1.4.3 Individual differences between learners

Another issue that may be a cause for concern is most studies in the literature do not consider individual differences between participants. For example, Markham (1989), Markham et al. (2001), and Hayati and Mohmedi (2011) classified participants as 'intermediate-level' users of English, and then proceeded with their experiments. Yet, none of them did a pre-test to obtain an accurate measurement of participants' listening comprehension ability. Similarly, since viewing captioned video requires that participants engage in reading, why did they not measure participants' reading proficiency prior to the experiment? It is possible to assume that students who are faster readers or have better reading comprehension skills may benefit more from subtitles than those who are poor readers. This is a point that was also raised by Vanderplank (2013), who having reviewed Winke et al.'s (2010) publication, said:

> "There are clearly a number of key variables regarding language proficiency which need to be taken into account: it isn't so much a matter of reading comprehension and listening comprehension but the ability to follow (i.e. decode) what is being said that is the problem for L2 viewing. When you add captions as a supposed support – which assumes the ability to read and understand at reasonable speed – you introduce a reading speed variable" (Vanderplank 2013, p. 242).

Additionally, referring back to Dual Coding Theory and the Cognitive Theory of Multimedia Learning, both models discuss the role of working memory and its ability to process two modalities of input. Why then was working memory capacity not measured in some sort of pre-test, to determine whether it accounted for individual differences in performance on the comprehension tasks? Moreover, differences in vocabulary size may also affect how much a person benefits from reading subtitles whilst watching a TV programme, as students with larger vocabularies may outperform those with less vocabulary knowledge. Most studies do not consider these variables when exploring the effects of bimodal input on L2 listening comprehension.

2.1.4.4 How should the effects of bimodal input be investigated?

These criticism regarding test construct validity and experimental design are by no means intended to be an attack on previous research, because again, there is a high level of reliability in the findings of those studies. The purpose of this commentary is simply to identify where improvements can be made in order to move this field of research forward continually. Rost (2002) argues that many listening tests used by researchers, teachers and examiners tend to lack construct validity because they test general comprehension and language proficiency rather than actual listening ability itself. According to Buck (2001), valid listening tests clearly focus on either (a) 'the sound system' (i.e. the perceptual processing of speech), (b) the 'understanding of literal meanings', or (c) 'tasks beyond literal meanings' (i.e. the utilisation of what has been heard). Therefore, in order to accurately test the effects of captioned video on L2 listening, researchers may consider beginning by clearly stating which of these three capacities they are investigating, and then selecting a measurement tool suitable for testing that particular faculty. Fortunately, there are a couple of studies in the literature that achieved this perfectly, and these will be discussed in section 2.2.2.

2.2 Bimodal input and L2 speech segmentation

Thus far, this literature review has made the following claims: (a) that Dual Coding Theory and the Cognitive Theory of Multimedia Learning offer a conceptual framework for understanding why watching captioned videos may lead to improved L2 comprehension; (b) there is a body of research that shows how watching captioned videos is beneficial for L2 reading and L2 vocabulary acquisition; (c) there are a plethora of studies that provide evidence in support of the idea that watching

captioned videos can lead to better L2 listening comprehension; however (d) improvements need to be made to the overall experimental design of these studies. Following Buck's (2001) principles, this section begins by "defining the construct" (p. 94) by establishing which part of the listening process will be focused on in this thesis.

2.2.1 The problem with L2 speech segmentation

2.2.1.1 Understanding the L1 listening process

To begin with, let us take a step back, away from the literature on captioned video, and reflect on the actual listening process. When a person utters a sentence, a sound wave is produced that travels through the air to the outer ear of the listener, it is funneled through the ear canal until it vibrates the eardrum, which then signals the cochlea to transmit an electrical impulse via the auditory nerve to the auditory cortex located in the brain. This is the first phase of the listening process and describes how the listener is able to *hear* the sound that the speaker produced (Greenberg & Ainsworth, 2006). Comprehending the sentence is a much more complex task, which entails what the literature refers to as 'bottom-up' and 'top-down' processing (Flowerdew and Miller, 2005). In bottom-up processing, speech segmentation occurs, whereby the spoken utterance is decoded into small units of sound. These units are then used to identify individual words within the utterance, and these words are then analysed at the syntactic level to understand the speaker's intended message (Ur, 1984).

Conversely, in top-down processing, the listener uses her knowledge of the context, situation and other relevant cues to determine what the speaker intends. For example, when a teacher starts a 9.00 am class by saying '*good morning students*', often L1 listeners can merely hear the phoneme /g/ or /'gʊd/ and predict that the phrase "good morning" will follow based upon their prior experiences of lessons at this time. Consequently, it should be noted that bottom-up and top-down processing may occur simultaneously or in any convenient order (Buck, 2001). For example, Vanderplank (2014) describes how L1 listeners comprehend news reports, explaining that they often use top-down processing because they tend to have background knowledge about most topics. However, when they are exposed to unfamiliar topics or unfamiliar

accents, they are likely to depend on bottom-up processing in order to follow every word that is spoken.

There are a number of theoretical models that attempt to explain the cognitive process of listening comprehension, such as McClelland and Rumelhart's (1986) Parallel Distributed Processing model of cognition, Kintsch's (1998) Construction-Integration model, and Anderson's (2005) three-phased model of the cognitive listening process. None of these theories have been proven to be factual, but based upon the general literature about the L1 listening process, a simplified diagram can still be constructed as in Figure 3 (alternatively, *see* Cutler and Clifton 1999, for a more detailed explanation of L1 listening comprehension).

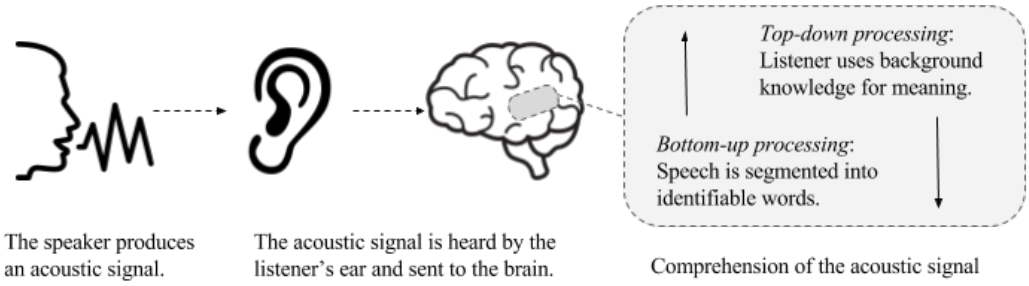

Figure 3. A simplified view of the L1 listening process

2.2.1.2 Difficulties with L2 listening

Although listening in a native language (L1) appears to be an effortless task, it is actually a highly complex mental process (Vandergrift, 2011), and possibly the most difficult L2 language skill for a person to learn (Vandergrift, 2004). There are many factors at play which can hinder the L2 listener's ability to comprehend an acoustic message; some of the bottom-up processing factors include: the pace at which the speaker delivers the message, her pitch, her accent, any background noise, the lexical items present in the message, the syntactical complexity of the message, and the learner's attention (Buck, 2001). Authors such as Wilson (2008) have also discussed several top-down processing issues, such as the listener's awareness of the topic of the acoustic message, its context, and sociocultural knowledge.

Using a sample of forty Chinese EFL students, whose English language ability ranged from beginners to advanced, Goh (2000) identified ten consistent listening comprehension problems these participants were facing, and categorised them as difficulties with perceptual processing, parsing the input stream and utilization issues (*see* Table 2). Remarkably, her results revealed that 22 out of 40 participants said they were unable to recognise words they already knew when listening to an authentic text in real-time. This problem was experienced by both beginners and advanced users of English.

Table 2. Problems related to different phases of listening comprehension (Goh, 2000, p. 59)

Perception	*Parsing*	*Utilisation*
Do not recognise words they know	Quickly forget what is heard	Understand words but not the intended message
Neglect the next part when thinking about meaning	Unable to form a mental representation from words heard	Confused about the key ideas in the message
Cannot chunk streams of speech	Do not understand subsequent parts of input because of earlier problems	
Miss the beginning of texts		
Concentrate too hard or unable to concentrate		

This is particularly interesting because it is something L2 learners of any language complain about; they say that it is difficult to listen to a spoken sentence and identify all lexical items in the message, even if they already have prior knowledge of those words. Indeed, in a study of British students who were learning French, Graham (2006) found that participants experienced this exact problem, they were unable to identify individual words in a stream of real-time speech. Fundamentally, the actual problem here is that speech is a continuous stream of sound, and unlike printed text, is not divided into clearly recognisable words (Quene & Koster, 1998). Research on the physical speech signal itself reveals that cessations of speech energy can appear both within individual words and between them (Anderson, 2005). This suggests that "[a]lthough well-defined gaps between words seem to exist in speech, these gaps are

often an illusion... It is our familiarity with our own language that leads to the illusion of word boundaries" (p. 58).

This point is illuminated by the research of Warner (2005) who recorded undergraduate students at the University of Arizona who were native speakers of English, during real conversations with their friends or family. She recorded one of these students asking the question '*Why, what weekend were you guys gonna be there?*' and presented a spectrogram of that utterance (*see* Figure 4). Spectrograms are visual representations of acoustic signals, which are used by phoneticians to analyse speech; they display frequency (on the vertical axis) against time (on the horizontal axis) with greater energy represented by darker shading (Cutler and Clifton, 1999). To the untrained eye, it appears as though there are four chunks of sound, as opposed to nine distinct chunks representing each individual word. Furthermore, when a word such as '*weekend*' is extracted from this acoustic signal to be examined on its own, it appears to the untrained eye as two separate chunks of sound (*see* Figure 5).

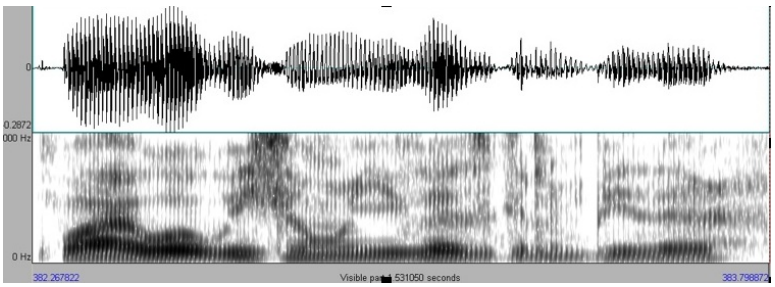

Figure 4. Spectrogram of the utterance "Why, what weekend were you guys gonna be there?" (Warner, 2005).

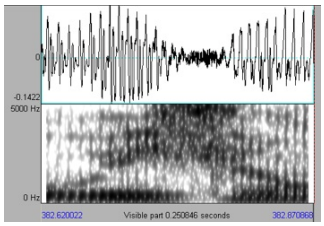

Figure 5. Spectrogram of the utterance "weekend" (Warner, 2005).

In order to accurately segment a speech stream (such as Figure 5) into identifiable words, the listener uses acoustic cues. Cues can include things such as a period of silence that precedes a word (Umeda, 1975), the aspiration of an initial stop (Nakatani

& Dukes, 1977), or intonation of speech (Dilley et al., 2010). An in-depth review of these cues is beyond the scope of this thesis (*see* Mattys et al. 2005 for more detail). What is relevant here is that due to cross-linguistic differences between languages, L1 segmentation cues differ from one language to another.

2.2.1.3 L1 vs. L2 speech segmentation

One of the most important cross-linguistic differences between languages is that they differ in metrical rhythmic structure, which can be syllable-timed, mora-timed or stress-timed (McQueen et al., 2001). In syllable-timed languages such as French and Spanish, the duration of every syllable is roughly equal (Mehler et al., 1981). Take the word '*palace*' for example, in French there is a clear and distinctive syllable boundary between '*pa-*' and '*-lace*' (Cutler et al., 1986). In English, however, the /l/ belongs to both the first and second syllables, so is pronounced as '*pal-*' and '*-lace*'. Thus L1 listeners of syllable-timed languages use syllables as a segmentation unit, which can be used to determine boundaries between words (ibid).

Whereas French and Spanish are syllable-timed languages, the metrical rhythmic structure of Japanese is based on the mora, a sub-syllabic unit. A syllable can be composed of (a) an onset, (b) a nucleus, and (c) a coda. Using the word '*can*' as an example, /c/ is the onset, which can be one or more consonants at the beginning of a word. /a/ is the nucleus, which in most languages is one or more vowels, and finally, /n/ is the coda, which may be one or more consonants located at the end of a word. This is important because the mora differs in that it can be a nucleus (e.g. /a/), or an onset plus nucleus (e.g. /ca/), or simply the coda (e.g. /n/). Cutler et al. (1993) explain this concept using the Japanese phrase '*shinshin-to*' (meaning: in silence they fall), which is composed of three syllables (*shin*, *shin*, *to*) but five morae (*shi*, *n*, *shi*, *n*, *to*). Significantly, L1 listeners of mora-timed languages use morae as units for speech segmentation, which is very different to L1 listeners of syllable-timed languages.

English is a stress-timed language, which makes use of strong and weak syllables. The Metrical Segmentation Strategy proposed by Cutler and Norris (1988) suggests that L1 listeners of English segment words at the onset of a strong syllable that is stressed. They conducted a series of experiments that supported their hypothesis, finding, for example, that it was easier for L1 listeners to segment the target word

'*mint*' from '*mintesh*' than '*mintaye*' because the former is composed of a strong and weak syllable, whereas the latter is composed of two strong syllables. When a word such as '*mintesh*' is composed of a strong and weak syllable, it is easier to detect the target word because it does not trigger segmentation. A word such as '*mintaye*' with two strong syllables, however, is automatically segmented by the L1 listener as '*min-taye*', slowing down detection of the target word '*min*t' because it is formed from letters on either side of the segmentation point. The Metrical Segmentation Strategy has been widely accepted by subsequent researchers on L1 speech segmentation (Vroomen, and De Gelder, 1995), particularly because 90% of spoken content words in the English language are stress-initial (Cutler and Carter, 1987). Dutch shows a similar pattern (van Heuven & Hagman, 1988 cited in Lin, 2013) and research suggests that L1 listeners of Dutch also use the Metrical Segmentation Strategy (Vroomen et al., 1996).

An awareness of these cross-linguistic differences between languages is important for appreciating the complexity of speech segmentation in a second language, because it is these differences that make L2 speech segmentation such an demanding task (Flowerdew and Miller, 2005). This is because L1 speakers of a language inherently use their native segmentation strategy when trying to segment speech in an L2, a phenomenon described by Cutler (2000) as "listening to a second language through the ears of a first". Essentially, this means when an L1 speaker of French is learning the English language, and listening to a native speaker talk, she will automatically (and incorrectly) attempt to segment speech using syllabic cues. This raises a question – which is that since research into the effects of captioned videos claims to enhance L2 listening comprehension in general, what affect does it have on L2 *speech segmentation* in particular?

This section has presented an overview of the L1 listening process, identified speech segmentation as the primary challenge of L2 listening comprehension, and looked at how speech segmentation differs from one language to another due to cross-linguistic differences. The next section discusses how bimodal input may improve L2 speech segmentation.

2.2.2 How does bimodal input help with L2 speech segmentation?

In section 2.1.4 it was noted that although many earlier studies about the effects of watching captioned videos on L2 listening comprehension suffered from drawbacks in terms of experimental design, there were a couple of studies that were actually designed very well. One such study was conducted by Mitterer and McQueen (2009), who investigated the effects of watching captioned video on speech perception during L2 listening, and more specifically, regional accents. They conducted an experiment with 121 Dutch university students, who were described as advanced-level learners of English as an L2. These participants watched 25-minutes of either an Australian sitcom (*Kath and Kim*) or a Scottish film (*Trainspotting*), either with L2 (English) subtitles (bimodal input), or with L1 (Dutch) subtitles, or without subtitles. Immediately after this viewing, participants were tested by means of a shadowing task. This task required them to listen to short utterances extracted from the DVDs (80 from each), and for each utterance, the participants had to repeat whatever words they heard. This task quite simply tested participants' ability to identify words in a continuous speech signal, i.e. lexical segmentation. Additionally, during the shadowing task, participants were presented with 'old items' (utterances they had been exposed to during the viewing), and 'new items' (utterances extracted from the same DVDs but which they had not been exposed to). All of these utterances were spoken by the lead character in the film and/or sitcom.

The results of this experiment revealed that participants exposed to bimodal input outperformed the other groups on 'old items' (previously heard utterances). Mitterer and McQueen (2009) claimed that this finding suggests bimodal input leads to better L2 listening at the speech perception phase. Furthermore, these participants also performed better on the 'new items' (previously unheard utterances from a familiar speaker), suggesting that their learning generalised to new input. Again, the authors postulated that this suggests "listeners were able to retune their perceptual categories to characteristics of the exposure speakers, leading to long-term changes in speech perception" (Mitterer and McQueen 2009, p.4).

This study by Mitterer and McQueen (2009) is highly innovative, and no other experiment within the literature has been conducted in this manner. Vanderplank (2013) describes it as "almost as big a breakthrough as Karen Price's article" (p. 7), a

major claim, considering that Price's (1983) research pioneered the way for all subsequent studies discussed in this literature review. Crucially, there are two major contributions from this study: (a), that it developed a task appropriate for specifically measuring lexical segmentation at the initial perceptual processing stage of listening comprehension, and (b), the claim that bimodal input can cause an L2 listener to *retune* "their perceptual categories to characteristics of the exposure speakers" (Mitterer and McQueen 2009, p.4). This is a groundbreaking finding, that effectively offers a solution to the problem of applying an L1 segmentation strategy to the L2 speech stream, because bimodal input has the potential to help an L2 listener retune (or adapt) their perceptual processing faculties.

2.2.3 Bimodal input and the generalisation of learning

2.2.3.1 Listening to previously encountered utterances

'Old items' in Mitterer and McQueen's (2009) study referred to utterances that participants had been first exposed to during the DVD viewing, and then again on the listening test. Since participants under bimodal input conditions were better at segmenting these utterances during the listening test, it suggests that reading a word whilst listening to it spoken helps with later L2 speech segmentation. Bird and Williams had in fact made a similar claim in 2002, in another well-designed and robust study.

Bird and Williams (2002) investigated the effects of bimodal input on the implicit and explicit learning of spoken words and non-words. In their experiments, vocabulary was presented to advanced-level L2 learners of English under three treatment conditions: (a) text with sound (bimodal input), (b) text-only (no sound), and (c) sound-only (no text). The authors looked at the effects of these treatments on word learning, measured by improvements in how long it took the learners to remember the words, as well as recognition memory. The results revealed that vocabulary presented in the form of bimodal input led to better recognition memory for spoken words and non-words when compared to the other two presentation conditions. Bird and Williams (2002) concluded from their experiments that bimodal input may have a significant effect on both long-term implicit and explicit learning of spoken word forms. Essentially, their research provides support for the notion that bimodal input

can improve L2 speech segmentation when listening to previously encountered utterances.

2.2.3.2 Listening to previously unheard utterances

'New items' in Mitterer and McQueen's (2009) study referred to utterances that participants had not been previously exposed to but which were spoken by a speaker they had already listened to. Since participants under bimodal input conditions were better at segmenting these utterances during the listening test, it suggests that reading a word whilst listening to it helps with later L2 speech segmentation of completely new aural input.

Although there is no other study that explicitly investigates this concept of 'new items', there is evidence of the generalisation of learning in other studies too. For example, a recent study by Bahrani and Tondar (2016) investigated the effects of bimodal input on L2 listening comprehension. Forty six intermediate-level L2 learners of English participated in their study and were assigned to either the bimodal group or the control group (no subtitles). The researchers adopted a pre-test / treatment / post-test experimental design, which spanned six weeks, and the duration of treatment was 6 hours per week. At the pre-test phase, participants sat the listening section of an IELTS test, then for six weeks they watched documentaries in one of the two treatment conditions, and finally at the post-test stage, they sat another IELTS listening test. Intriguingly, the bimodal group significantly outperformed the control group on the post-test. This is significant because there was absolutely no relationship between the documentaries and the IELTS tests. The content of the IELTS would not even be regarded as 'new items' according to Mitterer and McQueen's (2009) definition, but would have to be considered as '*brand new*' because these are utterances that had never been heard before by participants, which were spoken by unfamiliar speakers. This suggests that six weeks of watching captioned videos was a form of training that led to the retuning of perceptual processing within these participants.

2.3 Recent developments in the field

With over thirty years of research consistently finding that bimodal input has positive effects on L2 listening comprehension, some recent studies have begun to look at different ways bimodal input can be used. In a recent plenary speech, Vanderplank (2013) differentiated between research on the 'effects *of*' and 'effects *with*' captioned videos on L2 language learners. In this speech, he presented an example of how researchers at the University of Brighton are using subtitles with interactive TV technologies so that ordinary language learners at home can exploit the benefits of bimodal input. He also spoke about research conducted on autonomous, independent language learners, who watched streamed movies online with subtitles, to enhance their L2 language learning. He concluded by speaking about an initiative at Oxford University on Captions and Lifelong Language Learning, which is aimed at supporting adult language learners to benefit from watching captioned videos. Similarly, a recent publication by Culbertson et al. (2017) has adopted a completely new take on bimodal input research, through the development of an online crowdsourced language learning system, which employs subtitles that users can interact with.

However, some researchers are still investigating the effects *of* captioned videos on L2 listening comprehension and the results have not always been completely positive, suggesting there is still a need for further research. For example, Rodgers and Webb (2017) identified a research gap in the literature, which was that most studies have previously only exposed participants to very short viewings of the treatment material (see for example the research by Markham, 1989, described previously). Three hundred and seventy two (372) Japanese university students participated in their study, by watching ten 42-minute episodes of an American TV programme in either a bimodal group condition or a control group (no subtitles) condition. Out of the ten viewings, the bimodal group only significantly outperformed the control group on three episodes. Similarly, in a recent study, van der Zee et al. (2017) investigated the effects of bimodal input presentation for people using Massive Open Online Courses (known as MOOCs) for studying online. One hundred and twenty five (125) participants watched lectures online either with or without subtitles and then sat multiple-choice question comprehensions. The results of this study did not reveal any statistically significant effect of bimodal input presentation.

2.4 Aims of this research

This section summarises what has been discussed in this literature review, identifies several research gaps in the literature, and states the research questions for this study.

2.4.1 Summary of key points in the literature

Having reviewed the literature, there are several main points to note before moving onto the next section. These are: (a) *Significant authors and studies*: Research in this field was pioneered by Price (1983), who contributed nothing further to the field after this date. In the last 34 years, Vanderplank (1988; 1990; 1992; 2010; 2013; 2014; 2016) has been the leading figure in research on the effects of watching captioned videos on L2 language acquisition. Montero Perez et al. (2013) conducted a thorough meta-analysis of research in this field, specifically for L2 listening and vocabulary learning. (b) *Theoretical underpinning*: Dual Coding Theory (Paivio, 1986) and the Cognitive Theory of Multimedia Learning (Mayer and Moreno, 2003) both suggest that language learners understand new input most easily when it comes as both visual and aural input rather than in only one modality. (c) *Reliability of findings*: Most studies about the effects of watching captioned video on L2 listening comprehension have followed a similar experimental design. Participants are assigned to one of three groups; namely, L2 subtitles, L1 subtitles, and no subtitles. They watch a TV programme and are tested on their listening comprehension. Usually, in experiments designed this way, the results indicate that participants who watch the TV programmes with L2 subtitles understand the material better than other groups.

(d) *Methodological limitations*: Unfortunately, many studies lack test construct validity because they claim to test L2 listening but use a tool that measures other skills such as reading or memory. (e) *L2 speech segmentation*: Due to cross-linguistic differences between languages, many learners of a second language subconsciously use L1 segmentation cues when listening to speakers of the target L2. (f) *Generalisation of learning*: Several studies have shown that bimodal input can lead to long-term improvements in L2 perceptual processing (speech segmentation).

2.4.2 Research gaps to explore

There are a number of research gaps in the literature.

2.4.2.1 The L2 proficiency of participants

Mitterer and McQueen (2009) were the only researchers to specifically investigate the effects of bimodal input on L2 speech segmentation. However, their study focused on advanced level Dutch learners of English. Like English, Dutch speakers use the Metrical Segmentation Strategy (Cutler 2000) when segmenting continuous speech, which presumably makes it easier for Dutch learners to listen to L2 English. Many overseas students in the UK, however, appear to have difficulties following university lectures delivered by well-educated native speakers (Zeng, 1997), and even understanding carefully scripted TV programmes, such as the news or documentaries. One question, therefore, is whether L2 English speakers from different (and more distant than Dutch) L1 backgrounds, and with lower proficiency in English, could also benefit from bimodal input to improve their listening.

2.4.2.2 Duration of the study

Mitterer and McQueen (2009) claim that bimodal input leads to long-term improvements in L2 perceptual processing, however, their experiment was only conducted over one day. Surely, in order to make valid claims about long-term improvements in learning, adaptation and perceptual processing, measurements should be taken on more than one occasion. Hence, there is a gap in the research and a need for experiments that take repeated measurements on multiple occasions.

2.4.2.3 Generalisation of learning

Mitterer and McQueen's (2009) study differentiated between 'old' items and 'new' items, showing how bimodal input can help L2 listeners to adapt to a particular speaker's voice. However, no other study has investigated these types of utterances to confirm the reliability of their findings. Another interesting question, not addressed by Mitterer and McQueen's (2009) study, is whether exposure to bimodal input, delivered by a number of speakers from a broadly similar variety of English, could lead to improved lexical segmentation when listening to speakers that the learner has never heard before – for example whether watching a number of BBC documentaries with subtitles leads to superior listening comprehension when watching subsequent

subtitle-free documentaries delivered by different presenters. Thus, there is a need for research on what could be described as 'brand new items', meaning utterances participants have not heard previously and from a speaker they are unfamiliar with.

2.3.2.4 L2 speech segmentation and comprehension

Since most research has considered the effects of watching captioned videos on L2 listening *comprehension*, whilst Mitterer and McQueen (2009) are the only researchers to investigate its effects on L2 *segmentation*, there is a need for a study that looks at both segmentation and comprehension in order to determine whether enhanced segmentation leads to enhanced comprehension.

2.3.2.5 Measuring individual differences

Most studies in the literature do not make enough of an effort to determine participants' proficiency levels on various skills. Therefore, there is a need for a study that carefully measures participants' reading proficiency, vocabulary size, and working memory span, in order to determine which second language learners benefit the most from watching captioned videos in developing their speech segmentation abilities.

2.4.3 Research questions

Having identified the gaps in the literature above, five central research questions were posed for this thesis project. These were:

(a) "RQ1: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?"

(b) "RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?"

(c) "RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?"

(d) RQ4: Do reading proficiency, vocabulary size and working memory play a role in the usefulness of subtitles for L2 speech segmentation?

(e) RQ5: Does repeated watching of subtitled programmes lead to long-term improvement in L2 listening comprehension?

3.1 How L2 Listening Segmentation & Comprehension was Tested in this Study

As is evident from the literature review and stated research question, second language listening *segmentation* was the main focus of this study and designing a valid test to measure this was crucial. The shadow task developed by Mitterer and McQueen (2009) was replicated as a foundation for testing speech segmentation in this study because it was used to test participants' ability to segment previously heard input, and new input by a familiar speaker. Additionally, of the numerous studies conducted in the field of bimodal input, Vanderplank (2013) praised their experimental design for its empirical soundness. The following sections describe the speech segmentation test and listening comprehension tests used in this study, as well as several skills tests that were used in the final experiment.

3.1.1 Speech Segmentation Test

In order to test speech segmentation a shadow task was developed, which replicated the design of Mitterer and McQueen's (2009) measurement tool, but with slight modifications. Recall that their test was composed of 80 utterances, 40 'old items' and 40 'new items' taken from DVDs. They do not report how these were extracted but they provide enough information for their materials to be duplicated. In this study, to actually extract the utterances from DVDs, iFunia Studio's 'Media-Converter' software was used to convert the DVDs into MP3 audio files, then NCH Software's 'WavePad Sound Editor' app was used to edit these master files, and generate an audio file for each individual utterance. Then, for each excerpt, a <beep> sound was added 750ms before it, which replicated the 'warning tone' used by Mitterer and McQueen (2009). A silent period of 3.5 times the excerpt's duration was then added to the sound file. For example the excerpt "Its made of ordinary stuff" (an utterance extracted from the lecture) is 1.5 seconds in duration, so 5.25 seconds of silence was added after it. This provided enough time for participants to respond. Finally, the individual sound files were joined together in a new audio file, which started with a 5-second countdown. Ultimately, the sound file was arranged as follows:

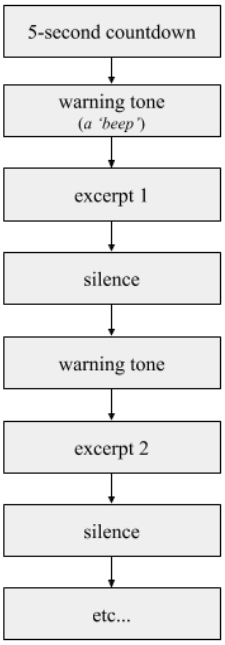

Figure 6. Design of shadowing task used in the experiments

It should be noted at this point that Experiments 1, 2, 3 and 4 all use different DVDs, so rather than overwhelming the reader with too much information here, a separate chapter has been written for each experiment. These chapters look more closely at the materials used, the procedure followed, scoring methods, and the overall design of each experiment respectively. The traditional format of a thesis is to have a methodology chapter followed by a results chapter, however, for theses such as this one, which is composed of multiple experiments with different research aims, there is a trend for chapters to presented accordingly.

In terms of justifying the actual speech segmentation test used, there are alternative approaches towards testing speech segmentation. Psycholinguists have several methods for testing speech segmentation in laboratories (see Mehler et al. 1981; Saffran et al. 1996), and for over twenty years, many researchers have used the 'word-spotting task' designed by Anne Cutler (Cutler and Shanley, 2010). The main disadvantage of adopting these methods, is that they are very artificial by nature, requiring participants to identify words between nonsense utterances. Conversely, the main advantage of Mitterer and McQueen's (2009) shadowing task, is that

participants are exposed to authentic speech in real-time, which is good because findings from such a test can be generalised to real-world situations.

3.1.2 Testing L2 Listening Comprehension

Reviewing literature about the effects of watching captioned video on L2 listening comprehension, it is evident that most studies require participants to watch a captioned video, and then take a multiple-choice question (MCQ) quiz that tests their understanding of what they watched (*see* Başaran and Köse 2013; Hayati and Mohmedi 2011; Taylor 2005). In some studies, the researchers themselves designed the comprehension tests, which is good in the sense MCQs are quick and easy to score. However, a disadvantage of only using MCQs is participants can guess the answers, and receive points that are not deserved. In other studies, the listening section of well-known language tests was used, such as the TOEFL (*see* Ghasemboland and Nafissi 2012; Huang and Eskey 1999). Using a well-established language test such as TOEFL has the advantage of (a) incorporating multiple question types (such as gap-fill, matching), and (b) these tests have been designed, piloted, and peer-reviewed by specialists with expertise in exam writing.

To test listening comprehension in this study, the 'Cambridge English: Advanced (CAE)' exam (C.E.L.A., n.d.) was selected. CAE exams are divided into four sections, which test reading, writing, listening and speaking. The listening section is composed of 30 questions, that includes a combination of multiple-choice questions, matching tasks, and sentence completion questions. Each question is worth one point, which means the highest score someone can obtain is 30. Audio recordings come from monologues (such as lectures) and interactive speakers (such as interviews), and the duration of the exam is approximately 40-minutes.

The CAE exam was selected as a listening comprehension test because it is designed for L2 users of CEFR B2 to C1 proficiency, which is equivalent to the level of most international students in the UK (as mandated by British university entrance requirements). The CAE was also selected because it is delivered in a different format to the IELTS exam, which many international students were familiar with, a factor that may influence performance. Also, from a pragmatic perspective, 40-minutes was deemed as a reasonable amount of time for the weekly tests, considering students

would sit it after watching 30-minutes of a documentary and doing the shadowing task. Moreover, past papers were readily available online, with model answers that could be downloaded and used for scoring. Accordingly, four CAE past papers were used (*see* Appendix E) and delivered in a counterbalanced approach across weeks and between participants.

3.1.3 Testing Vocabulary

In Experiment 5, participants were tested on their vocabulary size, reading proficiency and working memory capacity, as these were identified as factors that may influence or affect a person's performance on the listening segmentation and comprehension tasks. There are many tools available for testing these skills, and a review of these tools was performed using the IRIS database (Marsden et al., 2016) and other Internet resources before selecting the methods described in this section.

To test vocabulary size, Nation's (2010) online 'Vocabulary Size Test' was used <http://my.vocabularysize.com/>. This test measures an individual's written receptive vocabulary size (Nation and Beglar, 2007), making it is useful for determining how many words a person knows well enough to read in the English language. The test is composed of 140 lexical items, with each item representing 100 word families. A 'word family' is the base form of a word plus its inflected forms (such as third person *-s* or *-ing*) and derived forms from affixes such as *-able*, *-ism* and so forth (Hirsh and Nation 1992). Word families are suitable units for measuring receptive vocabulary because intermediate- to advanced- L2 learners of English are readily able to ascertain the meanings of affixed members of a word family. For example, L2 speakers who know the word 'produce' are likely to know what is meant by the words 'producing' and 'producer' (Beglar, 2009).

Well-educated native speakers of English generally know around 20,000 word families (Nation, 2006), but the 'Vocabulary Size Test' only samples from the most frequent 14,000 word families in the English language, such that if a person answers every question correctly on the test, it is assumed she knows all of these word families. Research suggests that undergraduate L2 speakers of English who sit this test usually score in the 5,000 to 6,000 word family range, whilst doctoral-level L2 speakers tend to score around 9,000 (Nation and Beglar, 2007). As will be discussed

in Chapter 6, the participants in this experiment scored approximately 7,500 word families, which means they are able to comprehend the large majority of spoken English discourse without assistance (Hu and Nation, 2000), and they should be able to comprehend standard subtitles, considering knowledge of 8,000 word families is sufficient for the purpose of reading an English novel without help (Nation, 2006).

To take the test, the participant first opens the website, sees a homepage and clicks on the 'Start the test now' icon (*see* Figure 7), then selects her native language from a menu. Once a language has been selected, simple test instructions are displayed (*see* Figure 8).

Figure 7. The vocabularysize.com homepage page

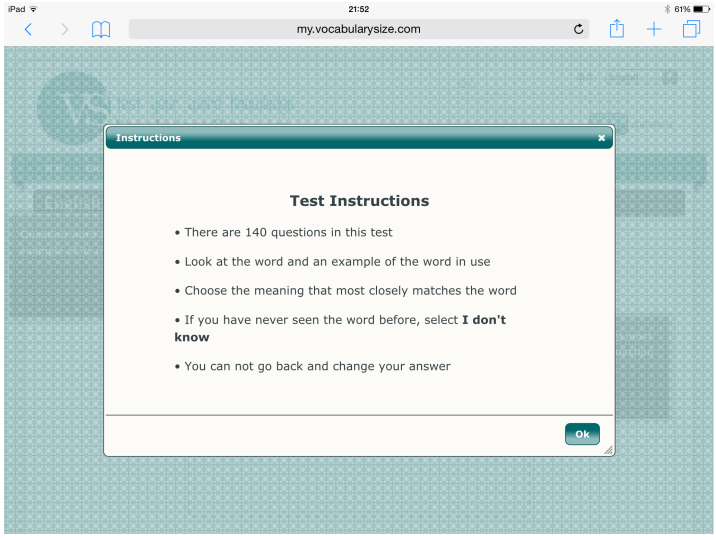

Figure 8. The vocabularysize.com test instructions

After reading the test instructions, an example test-item is presented, with labels clearly explaining each component of the test (*see* Figure 9). When the participant answers the practice question, they are allowed to proceed to the actual test. During the test, a lexical item is presented as a stand-alone word in English, and within an example sentence. The participant chooses the correct translation of this word from a multiple-choice list (*see* Figure 10). The Participant cannot navigate backwards through the test, and her score is presented at the end (*see* Figure 11).

Figure 9. The vocabularysize.com explanation of test questions

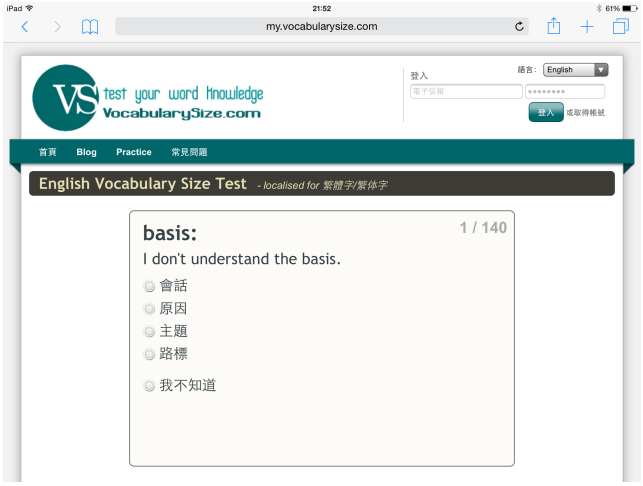

Figure 10. A vocabularysize.com test item

Figure 11. The vocabularysize.com test results page

3.1.4 Testing Reading Proficiency

To test reading rate and comprehension, the York Adult Assessment-Revised (YAA-R) (Warmington et al., 2013) was used. YAA-R is an assessment battery consisting of tests of writing, spelling, reading and phonological skills, that aims to assess the abilities of pre-university students and identify whether these students have special educational needs (specifically dyslexia). The reading test component of YAA-R was selected for this experiment because the passage was a non-fictional text designed specifically for university-level students, and research suggests it is a reliable measure of students' abilities (ibid). Furthermore, from a pragmatic perspective, there was a concern that overwhelming participants with lengthy tests may have lead to fatigue, and YAA-R was a shorter alternative to the CAE exam's reading comprehension section.

Participants read aloud a 492-word passage about The History of Chocolate (*see* Appendix D), whilst being timed and recorded. Reading rate was then calculated as $\frac{NumberofWords}{ReadingTime}\ x\ 60$, for example, Participant 1 spent 290 seconds reading the passage, so her reading rate was calculated as $\frac{492}{290}\ x\ 60\ = 102$ words per minute (WPM). After reading the passage, participants were asked to answer 15 comprehension questions. Each question was worth one point, so the total points possible was 15.

For example, this is a sample sentence from the passage: "*Chocolate is now enjoyed all over the world but until the late sixteenth century it was only found in Central and South America*". This is the first of the fifteen comprehension questions: "*Prior to the 16th Century, where was chocolate found?*". The answer key clearly states that test-takers must answer with both 'Central and South America' in order to score 1 point.

Warmington et al. (2013) administered this test on 126 native speakers of English and found that their average reading rate was 164 WPM (M = 164.15; SD = 20.39), and reading comprehension scores were around 10 (M = 9.74; SD = 2.30). As Chapter 6 will demonstrate, the average reading rate for participants in this study was 97 WPM, while their reading comprehension score was approximately 9 out of 15. These scores indicate that although the L2 students in this experiment read subtitles slower than L1 speakers, their level of comprehension may be similar.

3.1.5 Testing Working Memory

Building upon theories of short-term memory, it has been suggested that working memory is a cognitive system that can store a limited amount of information for a limited period of time, whilst simultaneously processing incoming stimuli. Figure 12 represents a model of working memory proposed by Baddeley and Hitch (1974, cited in Baddeley, 2003), which is composed of three components. The main component is the Central executive, which processes demanding cognitive tasks such as mental arithmetic, and allocates incoming data to one of two relevant sub-components. The Visuospatial sketchpad is one of these sub-components and processes information in a visual form, whereas the other subcomponent, the Phonological loop, processes aural and orthographic input.

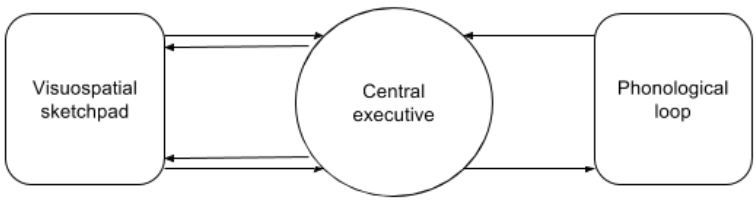

Figure 12. The three-component model of working memory (adapted from Baddeley 2003)

There are three commonly used approaches to the measurement of working memory capacity, known as the counting span task, the operation span task, and the reading span task (Wilhelm et al., 2013). In a counting span task, participants are required to count a series of shapes that are presented to them, and remember these count totals for later recall (Case et al. 1982; Engle et al., 1999). In an operation span task, participants are required to solve a series of mathematical equations whilst remembering words that are presented to them for later recall (Turner and Engle, 1989). In a reading span task, participants read sentences and verify the accuracy of these sentences, whilst remembering specific words that are presented to them for later recall (Daneman and Carpenter, 1980). Although these three task types present participants with different stimuli, their underlying structure and implementation are very similar. Essentially, they are designed in a way that forces the working memory system to store data (e.g. words) while engaging in a task that demands the manipulation or processing of other data (e.g. determining the accuracy of a sentence) (Conway et al., 2005).

To test working memory capacity, it seemed appropriate to select a span task that requires participants to listen to sentences, as they would in the actual shadowing task during the experiments. In Baddeley and Hitch's (1974) model of working memory, this would require engagement of the phonological loop. The IRIS database (Marsden et al., 2016) is an online repository of materials, stimuli, and instruments that can be downloaded without charge for the purpose of academic research on L2 language learning. The IRIS database was used to find a span task that met the needs of this study, and after reviewing several instruments, a paper-based version of the 'Phonological working memory span test' (Winke, 2013) was selected.

This test (*see* Appendix D) was adapted from previously validated working memory tasks (Mackey et al., 2010; Winke et al., 2003), and is composed of 48 unique sentences. Half the sentences are grammatically accurate and half are semantically plausible, which meant there were four types of sentence: (a) grammatical and plausible, (b) grammatical and implausible, (c) ungrammatical and plausible, and (e) ungrammatical and implausible.

Participants heard sentences played from an audio file in sets of three, four, or five. For each sentence, they had to decide whether it was plausible, whilst simultaneously deciding whether the sentence was grammatical. After giving the participant a few seconds to decide, the audio file would then say 'turn to the next page'. On the following page, participants were asked 'What was the last word of each sentence that you just heard?', and they were given spaces to write their answers. For example, participants heard the following three sentences spoken:

a. Working at the feet is a great job for Tom, but standing all day is beginning to wear on his **bank**.
b. My older sister's furniture were really small, but she has a lot of apartment and no space for **storage**.
c. She never goes camping because she has serious issues about have to use non-modern, antiseptic **bathrooms**.

Whilst listening to those sentences, they would see the options below and indicate whether the sentences were plausible and grammatical:

	Sense/Nonsense		Grammatical/Ungrammatical	
1)	S	N	G	U
2)	S	N	G	U
3)	S	N	G	U

When instructed to, the participants would then turn to the next page and answer the following question:

What as the last word of each sentence that you just heard?
Write them on the lines below.

1. _____
2. _____
3. _____

Winke (2013) designed this test according to principles proposed by Conway et al. (2005) in their 'user guide' on the administration of working memory span tasks, which includes guidelines for various approaches to scoring. In this experiment, the Partial-Credit Unit method was used for scoring, which represents "the mean

proportion of elements within an item that were recalled correctly" (Conway et al, 2005, p. 775). Essentially, a maximum of 2 points were possible, with half a point awarded for a correct 'grammatical' question, half a point awarded for a 'plausible' question, and one full point awarded for accurately recalling the sentence's final word. With 48 sentences presented in the test, the maximum score a participant could achieve was 96. In this study (as Chapter 6 showed), the mean score achieved by participants of 0.65 meant on average they were able to accurately recall up to 5 words, whilst processing grammatical accuracy and plausibility of sentences presented to them.

3.2 Main Characteristics of the Experiments

There were five experiments that made up this study, two of the pilots were conducted on one day in order to create a suitable shadow task test. The remaining three experiments were constructed using a pre-test / treatment / post-test design. In this type of experimental design, measurements were taken before treatment at the pre-test phase, after treatment at the immediate post-test phase, and then again after treatment at the delayed post-test stage. Additionally, participants were randomly assigned to a treatment group (Campbell and Stanley, 1966). In this study, this was actualised by recruiting students to participate in the experiment, assigning them to a group, then inviting them to attend the pre-test in 'week 1'. They sat the pre-test individually, in a specific room at a specified time and day, then the following week returned for the first treatment. During the first treatment, they watched a DVD under particular conditions (which are explained in more detail in subsequent chapters), then sat the immediate post-test. This was repeated in week 3, and finally, they returned in week 4 to be tested without receiving any treatment.

Experiments are composed of independent and dependent variables. Independent variables can be manipulated in an attempt to observe these effects on a dependent variable (Cohen et al., 2013). For example, in this study, participants were divided into groups, with each group receiving a different treatment. These treatments were manipulated and so 'group' is an independent variable. A shadow task was used as an instrument to measure participants' segmentation ability, their scores were the outcome of the treatment they received and so 'score' is the dependent variable.

Table 3. Independent and dependent variables across experiments 1 – 5

Experiment	*Independent variables*	*Dependent variables*
1	Genre (type of TV programme)	Shadow task score (measured from 0 - 100)
2	Group (bimodal, no subtitles, no sound) Time (week 1…4)	Shadow task score (measured from 0 - 100)
3	Documentary (type of documentary)	Shadow task score (measured from 0 - 100)
4	Group (bimodal, no subtitles, no sound) Time (week 1…4)	Shadow task score (measured from 0 - 100)
5	Group (bimodal, no subtitles, no sound) Time (week 1…4) Reading rate (measure in seconds) Reading comprehension (measured from 0 – 15) Vocabulary size (measure by word families) Working memory (measure from 0 to 100)	Shadow task score (measured from 0 - 100) Listening comprehension score (measured from 0 - 100)

3.2.1 Validity and Reliability

Two key concerns for any experimental research study are those of validity and reliability.

3.2.1.1 Validity

Validity refers to the ability of the test to accurately test what it claims to be testing, and the ability to justify that 'x' causes 'y' (Mackey and Gass, 2005). For example, in this study, does the shadow task actually test segmentation? And is it actually bimodal input that contributes to improving segmentation? These were major concerns when designing the experiments in this study. There are a multitude of threats to validity, factors that may have a negative affect on the validity of the experiment, which include history effects, maturation, testing effects, and instrumentation (Cohen et al., 2013). These factors may affect participants' or groups' scores on variables due to events that change the condition of a study. For example, after sitting the shadow task

in week 1 and week 2, it may become easier for participants to take the test in week 3 because they have been trained on it several times. 'Study 1' was conducted to ensure this validity was met: it consisted of three pilot experiments that were used to test the measurement tool and treatment tools. Weaknesses were identified and addressed when conducting later experiments.

3.2.1.2 Reliability

Reliability refers to the ability of a measurement tool to produce consistent results if repeated on multiple occasions (Mackey and Gass, 2005). For example, if a participant in this study scored 65% on the shadowing task, it was assumed she should be able to score approximately 65% again if she took the test again under the same conditions. Threats to reliability are factors that may cause inconsistent results using the measurement tool, and include environmental changes, researcher error, and participant changes (Lund and Lund, 2013). To minimise the threat of environmental and participant changes, all participants sat the test in the same room, under the same conditions, and when appropriate, at the same time as their previous test. For example, if participant 4 sat the pre-test at 11am on a Monday, she returned to the same room for the next test on the following Monday at the same time. To minimise the threat of researcher error, the precise testing procedure was written on a piece of paper and referred to whilst testing each participant. The shadow task in this study appears to be reliable because it was used in Experiment 2, Experiment 4, and Experiment 5 and on each occasion, participants' results were similar.

3.2.2 Ethical Considerations

All research projects within the social sciences must consider a number of fundamental ethical principles (Campbell and Stanley, 1966). In this study, five ethical factors were considered, namely (a) informed consent, (b) anonymity / confidentiality, (c) minimizing the risk of harm, and (d) the right to withdraw (Cohen et al., 2013). To begin with, each participant was given a written consent form, which was read aloud to them whilst they read it. The consent form stated the purpose of the study, the procedures involved, explained that it was voluntary, that they could withdraw at any time, and that a nominal payment would be made in gratitude for their time. The design of the consent form was shortened and amended slightly from Study 1 to Study 3, as is evident in Appendix A.

Concerning anonymity and confidentiality, the participants were informed that all of their responses would be anonymised and that their names would not appear in any publication. Each participant was assigned a number, and data was analysed based on these numbers. Minimizing harm to participants was a concern because they may have experienced stress during the tests or fatigue due to the tests' duration. They were informed that there were no expectations, that the tests would not affect their academic studies / grades in any way, and that they had the right to withdraw or discontinue the experiment at any time.

3.2.3 Participants

Sampling is a crucial factor to consider in experimental design as it can impact the quality of results and findings (Cohen et al., 2013). The target population for this study was international undergraduate or postgraduate students studying for a degree in the UK who spoke English as a second/foreign language at a level that satisfied British university entrance requirements. In order to select a sample of participants from a population, one can choose to adopt probability sampling techniques or non-probability sampling techniques. Probability sampling techniques use clearly defined methods of random selection in order to select participants. Non-probability sampling techniques on the other hand, may require the researcher to make subjective judgments in order to select participants (Lund and Lund, 2013). This study adopted a non-probability sampling technique known as convenience sampling to recruit participants.

Convenience sampling was used because it was a quick and cost effect way of getting students to take part in this research. Students were approached in the university's library, the research aims were explained, and they were offered a small payment of £10 if they attended all sessions of the experiment. Although this was easy to do, a limitation of this method is that it may not accurately represent the target population because there may be over-representation or under-representation of specific groups, which could lead to sampling bias (Lund and Lund, 2013). There were three types of participant who took part in this study, and each of them is described in this section.

Native speakers of English were recruited to provide a baseline measure for understanding the performance of non-native speakers. They were not used officially

as a ‘control group’ because the control group in this study was the ‘no subtitles’ group who watched treatment materials under normal conditions (i.e. whilst listening but without reading). Rather, the native speakers were used to test the complexity of the shadowing task, because if L1 speakers found it difficult, then surely L2 speakers would also find it challenging. For Experiments 1 to 3, a different native speaker was invited to sit the shadowing task for each test. The 3 native speakers were British, female, university students, and during general discussions with them after completing a task, they described it as very easy. For Experiment 4, it was decided that more native speakers should be recruited to test the shadow task. The aim was to recruit twelve native speakers, but in the end it was only possible to recruit nine. These nine native speakers were a mixture of students and teachers, some British and others American.

The second type of participants in this study were mixed-L1 international students. For Experiments 1 and 3, the students who agreed to participate in this research came from Europe, Asia and the Middle East. They were all around the same age, had similar IELTS scores and had spent similar amounts of time studying English. However, considering that L2 listening can be influenced by L1 linguistic constructs, the variance in participants’ scores may be due to L1 differences. Therefore, it was decided that participants should be selected from the same L1. Chinese students were selected to participate in Experiments 4 and 5 simply because there are many international students in the UK from China. Although using a sample of students from a single L1 is not representative of the entire UK international student population, at least the group of participants used was very homogenous across a variety of variables. In total, 112 international students and 12 native speakers participated in this study.

3.3 Statistical Analysis

A combination of both descriptive and inferential statistics was used to analyse data collected from participants in this study. This section describes what those methods were, explains how the results should be interpreted, and justifies why these methods were selected.

3.3.1 Descriptive Statistics

The term 'descriptive statistics' refers to forms of analysis that describe data in a simple and meaningful way (Lund and Lund, 2013). Five types of descriptive statistic were used frequently in this study, namely (a) measures of central tendency, (b) measures of spread, (c) graphical description, (d) tabulated description, and (c) statistical commentary. Each of these is explained below.

A measure of central tendency is a single value that identifies the central location within a set of data, and commonly used measures of central tendency are the mean, median and mode (Field, 2013). The mean (*M*) is always used as a measure of central tendency in this study, as it is the average value for any given data set. For example, in Experiment 5, participants sit a reading comprehension test, where the maximum score possible is 15. The mean score was calculated as 8.84, which means most participants achieved a score close to this figure. Now, it is common practice to only report actual findings in the results chapter of a thesis; however, for the sake of readers who do not have a strong background in statistics, it was decided that a few examples would be included in this chapter for the benefit of clarifying what the results actually mean.

A measure of spread is a single value that indicates how spread-out scores within a given data set are (Field, 2013). Returning to the example above, not every participant scored 8.84, some scored higher and others lower. Commonly used methods to summarise how spread-out these scores are include range, absolute deviation and standard deviation (Lund and Lund, 2013). The standard deviation (*SD*) is always used as a measure of spread in this study, and is coupled with the mean. The closer the standard deviation is to zero (0) the closer all values in the data set are to the mean. So for the reading comprehension test, the standard deviation was 1.89, meaning most students achieved a score close to 8.84.

A graphical description of data is usually some sort of graph or chart that is used to visualise the data set. Commonly used methods to visualise a data set are pie charts, line graphs, bar graphs, and X-Y plots (Field, 2013). Bar charts are always used in this study to show changes in scores over time and to compare scores between groups. These charts also include error bars. Error bars can be used to represent the standard

error, the standard deviation or the 95% confidence interval of the mean. Error bars in this study always show the 95% confidence interval, which is "an interval constructed such that in 95% of samples the true value of the population mean will fall within its limits" (Field, 2013, p. 47). These error bars are visually informative because if they do not overlap, it means there is a statistically significant difference between mean scores. For example, Figure 13. below is taken from Experiment 2. Looking at the error bars from week 2 to week 3 for the 'no subtitles' group, it is clear to see there is an overlap for the bars; but looking at the 'bimodal' group, there is no overlap, indicating there is a significant difference from week 2 to week 3 on their scores.

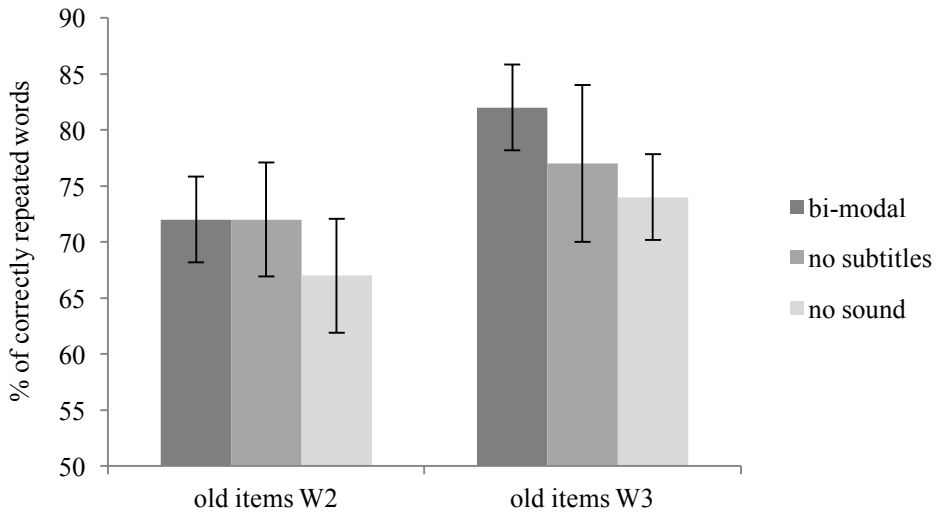

Figure 13. Error bars example (data from Experiment 2)

The two other types of descriptive statistics present in this study are well-known and self-explanatory. Tabulated descriptions are basically tables formatted according to APA standards, which make it easy for the reader to see the data summarised; statistical commentary is simply a written description of the data set.

3.3.2 Inferential Statistics

This research study targeted a very specific population: international students in the UK who speak English as a second or foreign language. There are thousands of such students living throughout the country and it was impossible for all of them to participate in this research project. Consequently a sample of students was recruited as representative of the population. In order to generalise the results of this study to

the entire population (rather than those who partook in the experiments) it was necessary to use inferential statistics. Four types of inferential statistic were used in this study, namely (a) Analysis of Variance, (b) Pearson's correlation, (c) regression analysis, and (d) t-tests. Each of these is explained below.

3.3.2.1 Analysis of Variance (ANOVA)

The mixed ANOVA compares mean differences between groups that are divided on two independent variables, in order to determine if there is an interaction between them and the dependent variable. These independent variables are referred to as a 'between-subjects factor' and a 'within-subjects factor', and the aim of a mixed ANOVA is to ascertain whether there is a significant effect or interaction in either of them (Field, 2013). Taking Experiment 4 as an example, the dependent variable is "score" (i.e. the number of words participants are able to accurately segment and repeat on the shadow task), whereas the within-subjects factor is "time" (i.e. week 1 – week 4), and the between-subjects factor is "group" (i.e. the treatment conditions that participants were placed into). Essentially, the experiment lasts for a period of four weeks, and speech segmentation is measured on four occasions, which represents the four levels of the within-subjects factor. The mixed ANOVA reveals whether there was an interaction between the between-subjects factor and within-subjects factor on the dependent variable.

In order to perform a mixed ANOVA, four key assumptions need to be met (Lund and Lund, 2013). Firstly, there should be no 'outliers' in the data set. An outlier is any extreme value that does not follow the participants' usual behavior patterns. For example, if most participants scored approximately 58% on a the shadow task, but one of them scored 82%, this could negatively affect the findings, because it may alter the overall mean score, distort differences between groups, and may cause problems when generalising this sample's results to the population. Therefore, before a mixed ANOVA is performed, any outliers in the data must first be identified; and in this study, they were identified through the use of boxplots. Using SPSS, boxplots such as the one in Figure 14. below, use small circles to show data points that are 'far' from the median value (indicated by the horizontal line in the bars); and extreme values that are very far from the median are represented by a star symbol.

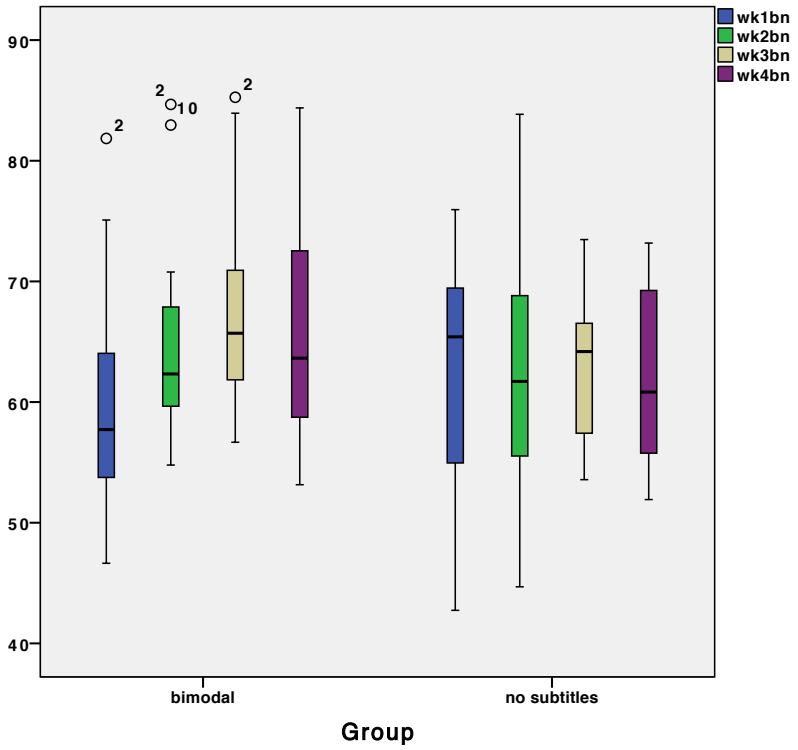

Figure 14. Boxplot example (data from Experiment 5)

When outliers have been identified, they can be completely removed from the data, or kept as they are, or their values can be normalised. Any outliers in this study's data were normalised by capping them at the group's mean plus two standard deviations (Field, 2013).

The second assumption is that the dependent variable should be normally distributed, which means that all data points should be evenly distributed around the mean, rather than skewed in one direction. Normality of distribution can be determined using numerical methods such as skewness/kurtosis values or the Shapiro-Wilk test; or by graphical methods such as histograms or Nomal Q-Q plots (Lund and Lund, 2013). In this study, the Shapiro-Wilk test of normality was used, which tests the null hypothesis that data is distributed equally to a normal distribution. If the result of this test is greater than 0.05 (i.e. $p > 0.05$), then the data is normally distributed. If the result is not normally distributed, then the data can either be transformed, or the mixed ANOVA can be proceeded with anyway because it is a relatively robust statistical analysis (which is what was done in this study).

The third assumption is that of homogeneity of variances, which (continuing with the Experiment 5 example) means there should be equal variance between categories of "group" (the between-subjects factor), at each category of "time" (the within-subjects factor), for "score" (the dependent variable). Unequal variances may be problematic because they can affect the Type 1 error rate (Field, 2013). Levene's test of equality of error variances was therefore used, which tests the null hypothesis that variances within the dataset are equal. If the result of this test is greater than 0.05 (i.e. $p > 0.05$), then there is homogeneity of variance in the dataset (Lund and Lund, 2013). This assumption was always met in this study.

The fourth assumption that needs to be met is the assumption of sphericity, which means the variance of differences between groups should be equal. This is a particularly critical assumption to be meet because if it is violated, results may be invalid (Field, 2013). Mauchly's Test of Sphericity was used in this study to determine whether this assumption was met. Essentially, this tests the null hypothesis that variances of differences between the categories of the within-subjects factor (i.e. "time") are equal. If the result of this test is greater than 0.05 (i.e. $p > 0.05$), then the assumption of sphericity has been satisfied (Lund and Lund, 2013). This assumption was always met in this study.

After fulfilling the above-mentioned assumptions, and running the mixed ANOVA through SPSS, the next step was to determine whether a statistically significant interaction existed between the between-subjects and within-subjects factors. Taking an example from Experiment 4, a mixed ANOVA with 'group' as the between-subject factor and 'time' as the within-subject factor revealed a main effect of 'group' and the result read as "$F(2,45) = 12.68$, $p < 0.01$", which has the following meaning:

Table 4. How to understand the result of a mixed ANOVA (adapted from Lund and Lund, 2013)

Part of result	*Meaning*
F	Shows an F-test (i.e. comparing to an F-distribution)
2 in (2,45)	Specifies degrees of freedom for the interaction
45 in (2,45)	Indicates degrees of freedom for the error
12.68	The obtained F-value
$p < 0.01$	Indicates the probability of obtaining the observed F-value, given the null hypothesis is true.

The result above represents a 'by-subject' analysis (also known as an 'F_1'), and to increase the robustness of this finding, a 'by-item' analysis was also conducted (known as an 'F_2'), which yielded the following result: ($F_1(2,45) = 12.68$, $p < 0.01$, $F_2(2,158) = 147.52$, $p < 0.01$). In both cases the main effect of 'group' was statistically significant. However, the mixed ANOVA does not report how or where the differences between groups were (Field, 2013), thus pairwise comparisons or planned contrasts were calculated in order to determine how the groups differ.

Besides the mixed ANOVA, this study also included a one-way repeated-measures ANOVA in Experiment 1 to determine whether there was a statistically significant difference between participants' scores on the shadowing task that used utterances from different genres. Type of 'programme' was the independent variable (within-subject factor) and 'score' was the dependent variable. A repeated-measures ANOVA was used because the same group of participants was being measured on the dependent variable more than twice under different conditions (Field, 2013), which in this case were the types of programme. Additionally, a basic one-way ANOVA was used to determine whether there were significant differences between the number of words used in excerpts / utterances for the shadowing task.

3.3.2.2 Pearson's Correlation

The Pearson correlation analysis reveals whether a linear relationship exists between two variables, and specifically the strength and direction of that relationship (Lund and Lund, 2013). In Experiment 5, a Pearson correlation was performed to determine whether a relationship exists between participants' L2 speech segmentation ability

and their L2 listening comprehension. This was done by comparing their weekly scores on the shadow task with their weekly scores on the comprehension tests.

Like the ANOVA analyses, the Pearson correlation can only be performed successfully if certain assumptions are met (Lund and Lund, 2013). Firstly, both the variables being tested must be continuous variables, and they should be paired. Secondly, there should be a linear relationship between these variables. To determine whether linearity exists, a scatterplot should be generated and inspected. For example, Figure 15 below plots participants' scores on the shadow task against the comprehension test in week 4 of Experiment 5, and it can be seen that there was a positive relationship.

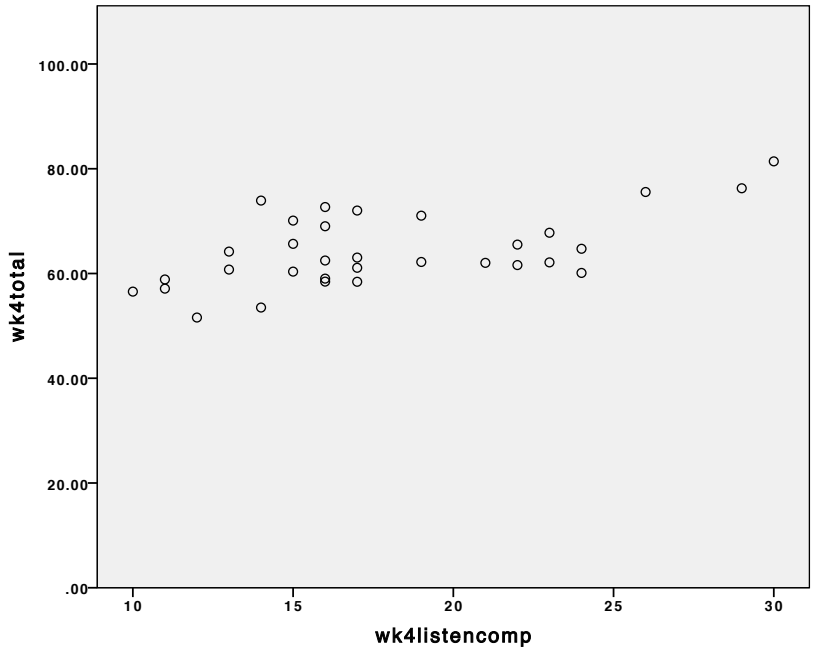

Figure 15. Scatterplot example (data from Experiment 5)

Additionally, there should be no significant outliers in the data set and it should be distributed normally. Once these assumptions are met, the analysis can be run, the results of which produce a correlation coefficient value (*r*) ranging from +1 to -1 (Field, 2013). A result greater than (r >) 0.5 indicates a strong positive correlation, < 0.3 to < 0.5 shows a moderate correlation, < 0.1 to < 0.3 indicates a small correlation, a zero (0) value represents no association, whilst -1 shows a negative relationship

(Lund and Lund, 2013). For example, at the pre-test phase of Experiment 5, the Pearson's correlation produced this result: $r(30) = 0.38$, $p < 0.05$. This reveals a moderate correlation between shadowing task scores and comprehension scores.

3.3.2.3 Hierarchical Multiple Regression Analysis

According to Lund and Lund (2013) a hierarchical multiple regression analysis is useful for (a) predicting a dependent variable based upon multiple independent variables, and (b) determining the relative contribution of these independent variables to the total variance of the model (the overall fit). In Experiment 5, a hierarchical multiple regression analysis was performed to determine how much of the variance in shadowing task scores (the dependent variable) amongst participants in the bimodal group could be explained by 'working memory', 'vocabulary size', 'reading rate', and 'reading comprehension' (the independent variables). A standard multiple regression analysis can also do this, but a hierarchical multiple regression analysis has the added benefit of determining which of these independent variables had a statistically significant affect on the dependent variable (Lund and Lund, 2013).

Similar to the previous inferential analyses discussed, there are certain assumptions that must be met before the multiple regression analysis can be run (Lund and Lund, 2013). Firstly, there should be one dependent variable and two or more independent variables. Secondly, there should be no significant outliers and data should be normally distributed. Thirdly, there should be a linear relationship between the dependent variable and each of the independent variables. After the analysis has been run, SPSS generates an R^2 value, which represents the variation in the dependent variable explained by the independent variables. For example, in Experiment 5, the hierarchical multiple regression analysis produced a R^2 of 0.49, with 'reading comprehension' and 'working memory' as the independent variables, meaning they explained 49% of the variability of the dependent variable (i.e. the shadow task score).

This R^2 value alone is not sufficient because it is based upon the sample of participants and regarded as a positively-biased estimate of variation, meaning it is larger than it should be if and when generalized to a larger population (Lund and Lund, 2013). Consequently, an 'adjusted R^2' is also generated, as this corrects the bias

and presents a value that can be generalized to the population. Continuing with the above example, the hierarchical regression analysis produced an adjusted R^2 of 0.42, suggesting that 'reading comprehension' and 'working memory' can explain 42% of the variance on shadowing task scores for international students who are exposed to bimodal input. In addition to producing these results, the hierarchical multiple regression analysis also runs an ANOVA to determine the model's statistical significance. Ultimately, the result read as ($F(2, 13) = 6.46$, $p < 0.05$, R^2 0.49, *adj.* R^2 0.422), which revealed a statically significant variance due to the independent variables.

SPSS also produces results for each independent variable with a coefficient value. Following up on the previous example, the coefficient (β) value was 0.94, which indicates that participants with higher 'reading comprehension' ability were scoring higher on the shadowing task. This value was produced along with a t-test and 95% confidence interval, such that the final result read as $\beta = 0.94$, $t(13) = 3.49$, $p < 0.05$, revealing this was statistically significant.

3.3.2.4 Welch's T-Test

An independent-samples t-test is useful for comparing two groups of participants on a particular variable and determining whether a significant difference exists between them (Field, 2013). In Experiment 5, the 'bimodal' and 'no subtitles' groups were compared on variables such as 'age' and 'number of months living in the UK'. The assumption of normality was met and there were no outliers in the data, however the assumption of homogeneity of variances was violated. Consequently, Welch's t-test was used as this accommodates unequal variances and delivers a valid test result (Lund and Lund, 2013).

3.4 An Overview of the studies conducted

Three separate yet connected studies were conducted throughout this research project. An overview of these studies is presented below and the following three chapters present the findings of these studies.

3.4.1 Study 1

Study 1 was entitled 'international students and L2 speech segmentation' because the primary aim of this study was to determine whether speech segmentation is challenging for international students in the UK. Mitterer and McQueen's (2009) shadow task was replicated, modified and used to test a sample of students at a British university. This study was was based on five research questions. Firstly, 'how difficult is L2 speech segmentation for international students'? This was investigated with an initial pilot experiment that used the modified shadow task. The result of that pilot led to the construction of a second pilot experiment, which aimed to investigate whether 'repeated watching of subtitled programmes leads to long-term improvement in L2 segmentation abilities (a) when listening to previously encountered utterances, (b) when listening to different utterances by the same speaker, and (c) when listening to different utterances by different speakers of a broadly similar accent'. Finally, a third pilot experiment was conducted to address the question: 'how difficult is L2 speech segmentation for international students when listening to different speakers of a broadly similar accent'? The results of these pilot experiments warranted further investigations, and consequently a second study was performed.

3.4.2 Study 2

Study 2 was entitled 'effects of bimodal input on L2 speech segmentation' because it looked at whether exposure to captioned programmes could aid international students, as a result of the findings of Study 1, which established that there are weaknesses in the segmentation ability of international students in the UK. Parts of Mitterer and McQueen's (2009) research was replicated, in that there were similar groups and variables used. However, this study went a step further by using different types of participants, adding an additional group, introducing a new variable, and extending the duration of the experiment to four weeks, with the aim of answering the following research questions: 'does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities (a) when listening to previously

encountered utterances, (b) when listening to different utterances by the same speaker, and (c) when listening to different utterances by different speakers of a broadly similar accent'?

3.4.3 Study 3

Study 1 showed that international students in the UK have limited L2 speech segmentation ability, and Study 2 demonstrated that bimodal input can be used to improve their segmentation skills. Study 3 was entitled 'effects of bimodal input on L2 listening comprehension' because its focus was general listening comprehension. This was deemed to be the appropriate next step in the research because segmentation ability in and of itself is not useful if it does not contribute towards understanding. This study replicates the experiment used in Study 2, but builds upon it in two-significant ways. First, a listening comprehension test is added as a new component, and a data analysis performed to determine whether a relationship between segmentation ability and listening comprehension exists. Secondly, the experiment is run over five weeks instead of four; during the additional week, participants are tested on a number of skills with the objective of answering the research question 'Does reading proficiency, vocabulary size and working memory play a role in the usefulness of subtitles for L2 speech segmentation and comprehension'?

CHAPTER 4: STUDY 1 - INTERNATIONAL STUDENTS AND L2 SPEECH SEGMENTATION

4.1 Experiment 1

According to Mitterer and McQueen's (2009) study watching captioned videos helps L2 listeners to better segment speech and comprehend heavily accented variations of English, which in turn can lead to long-term improvements in L2 listening. They tested L2 speech segmentation ability by using a shadowing task, wherein participants would hear a short, pause-bound utterance, and then verbally repeat the words they heard. Having reviewed other research into the effects of watching captioned videos on L2 listening, this approach appeared to be the most empirically valid, and consequently was used as the basis for experiments in this thesis.

As a first step towards building upon Mitterer and McQueen's (2009) experiment, their shadowing task was replicated using slightly different materials. A test was conducted to discover whether it would be suitable for use with a sample of students who were at an intermediate proficiency level and from very different L1 backgrounds. This was achieved through the creation of Experiment 1, which essentially aimed to establish the difficulty of L2 speech segmentation for intermediate-level, international university students in the UK. This section states the initial research questions, describes the students who participated in the experiment, explains how it was designed, implemented and scored, and then closes with a discussion of the results.

4.1.1 Research questions

The aim of Experiment 1 was to establish the difficulty of L2 speech segmentation for intermediate-level, international university students in the UK. As an initial, exploratory study, the primary research question at this stage was simply "how difficult is L2 speech segmentation for international students?". In an attempt to answer this question, Mitterer and McQueen's (2009) shadowing task was replicated, but with slight alterations and different materials.

4.1.2 Participants

One native speaker of English participated in the experiment as a control, providing a baseline for understanding the L2 data. She was a 23-year old British female who was studying towards a PhD in Education at the time. Additionally, a convenience sample of ten international students at a British University, participated in this experiment

(*see* Table 5). They spoke a variety of L1s (4 Chinese, 2 Greek, 1 Arabic, 1 Kurdish, 1 Spanish, 1 Latvian), were all studying at the same university and had a mean IELTS listening test score of 6.5 (*SD* = 0.56), excluding one Greek participant who had never sat the IELTS examination.

Table 5. Summary of the participants' demographic information in Experiment 1

N	***Age***	*Gender*	*Degree*	*No. of months living in UK*	*No. of years studied English during life*
10	*M* = 23.40; *SD* = 3.44	7 females 3 males	6 Postgrads 4 Undergrads	*M* = 14.70; *SD* = 13.23	*M* = 8.80; *SD* = 4.92

4.1.3 Design and Materials

In order to make a representative sample, six types of programme were used to create the test materials, which were a documentary, a film, an academic lecture, a news report, a sitcom, and a stand-up comedy (*see* Table 6). Ten audio excerpts were made from each of these programmes, which equated to sixty excerpts in total. These excerpts were verbal utterances bounded by pauses. For example, "whats the tipping point" is an excerpt that was taken from the stand-up comedy programme, and "from Broadwater farm" is an excerpt that was extracted from the news report. Utterances (*see* Appendix C.1 for the complete list) were selected that (a) had no background music, (b) had no background noises, (c) were spoken clearly in British English, and (d) were of CEFR upper B2-level complexity according to the 'Text Inspector Profile' online tool (Cambridge University Press, 2012). The overall average number of words in an excerpt was 7 (*M* = 6.67; *SD* = 2.17), and there were no statistically significant differences in number of words per excerpt between programmes, as determined by a one-way ANOVA $F(1,4) = 0.071$, $p > 0.05$.

Table 6. Summary of the materials used in the listening test for Experiment 1

Genre	*Title*	*Speaker*
Documentary	'Meerkats'	David Attenborough
Film	'The Duchess'	Keira Knightly
Lecture	'The Birth of the Early Modern Period: From Galileo to Descartes'	Dr. Peter Millican
News	2011 Tottenham London Riots	Jane Hill
Sitcom	'Blackadder'	Rowan Atkinson
Stand-up	'Russell Brand in New York'	Russell Brand

4.1.4 Procedure and Scoring

Participants were met individually, and informed about the purpose of the study, and nature of the listening test. They were given time to ask questions for clarification, and if they agreed to participate in the research, were asked to sign a consent form (*see* Appendix A.1). They then completed a language background questionnaire (*see* Appendix B.1), which (a) asked about demographic information, (b) asked them to estimate on a scale from 0% to 100% how accurately they thought they would repeat excerpts from different programmes, and (c) whether or not they had previously viewed the TV shows that were to be presented to them during the experiment.

Participants then took the listening test individually in an empty room, in a quiet part of the university, to reduce possible interruptions or background noise. Each participant sat on a chair facing a blank wall, wearing over-ear noise-reducing headphones, which were connected to a laptop. The laptop contained the audio file of the utterances, and this was played using iTunes. The participants' responses were recorded using WavePad, whilst the experimenter sat at the back of the room during the experiment.

The sixty utterances were grouped in blocks of ten by programme, and presented to participants in the same order. This means that participants heard ten utterances from the documentary, then ten utterances from the film and so on. Participants were asked to repeat the words they heard, and prior to starting the test, the experimenter demonstrated how to do this. At the end of the experiment, participants were paid a nominal fee for their time.

4.1.4.1 Scoring

As previously stated, during the listening test, participants were asked to repeat the words they heard in a series of utterances. For example, when hearing the utterance "this group of meerkats", participants were required to repeat these four words. Following the approach used by Mitterer and McQueen (2009), both function and content words were counted. If they accurately repeated all four words, they would score 100%, if they accurately repeated three out of four words, they would score 75% for that utterance, and so on. Similarly, in an utterance such as "I think the obvious point is this", if they accurately repeated all seven words, they would score 100%, if they accurately repeated six out of seven words, they would score 86% for that utterance, and so on. The mean scores for each utterance were then used for descriptive and inferential data analyses.

4.1.5 Results

4.1.5.1 Questionnaire Results

The questionnaire (distributed before the listening test) asked participants to estimate how accurately they would be able to repeat what they heard from the six types of TV programme. The results of this question are presented in Table 7, and suggest that participants expected to accurately repeat approximately 70% of excerpts in the listening test (M = 70.75; SD = 13.10).

Table 7. Participants' estimates of their ability to repeat words from the programmes

Genre	*N*	*M*	*SD*
Documentary	10	64.5	20.28
Film	10	82.5	9.20
Lecture	10	80.5	9.85
News	10	73.5	17.33
Sitcom	10	70.0	14.39
Stand-up comedy	10	54.5	22.97

4.1.5.2 Segmentation Test Results

The native speaker who acted as a control accurately repeated 100% of the utterances, whilst the international participants accurately repeated 71.30% (*SD* = 10.04) of the utterances they heard during this experiment. Table 8 shows participants' scores on the shadowing task by programme.

Table 8. Participants' actual scores on the shadowing task

Genre	*N*	*M*	*SD*
Documentary	10	72.22	10.64
Film	10	59.83	13.89
Lecture	10	71.87	12.26
News	10	74.23	10.91
Sitcom	10	70.60	14.37
Stand-up comedy	10	77.97	11.82

A one-way repeated-measures ANOVA was performed with 'programme' as the independent variable and 'score' as the dependent variable. A repeated-measures ANOVA was used because the same group of participants was being measured on the dependent variable more than twice under different conditions, which in this case were the types of programme. Segmentation scores were normally distributed for all programmes as assessed by Shapiro-Wilk's test ($p > .05$) and Mauchly's test of sphericity was not statistically significant, indicating that the assumption of sphericity was met, $\chi^2(14) = 12.207$, $p = 0.613$.

The ANOVA revealed a statistically significant main effect of programme ($F_1(5, 45) = 7.53$, $p < 0.05$, $F_2(9,531) = 12.13$, $p < 0.05$) which suggests that there was a significant difference on participants' performance on the shadowing task, depending upon the type of programme they watched. Furthermore, pairwise comparisons suggested that participants' ability to repeat words from the film (*M* = 59.83; *SD* = 13.89) was lower than their ability to repeat words from the stand-up (*M* = 77.97; *SD* = 11.82). There were no other differences between programmes.

4.1.6 Discussion

The primary aim of this initial experiment was to establish how difficult L2 speech segmentation is for international students at UK universities. Using a modified version of Mitterer and McQueen's (2009) shadowing task, the listening test used in this experiment simply required participants to repeat the words they heard in a series of sixty short utterances bounded by pauses. The results revealed that as a group, participants were only able to accurately repeat 71.30% (*SD* = 10.04) of the utterances they heard. The results indicate that this group of international students in the UK was unable to accurately repeat approximately 30% of the words they heard, as spoken by British L1 speakers of English. Potentially, their performance could be even worse when faced with a real-world context, as opposed to sitting in a lab under controlled conditions that minimised the effects of background noise and other distractions.

This finding is consistent with the self-reported listening difficulties that international students have stated in other research studies. For example, Liu (2013) conducted surveys with Chinese students in the UK to investigate the challenges they face as international students. Some of the participants in her study stated that they were only able to understand 20% - 50% of what they heard during academic lectures. It appears this self-reported figure is consistent with some of the predictions participants made for Experiment 1 and their actual results. For example one participant only managed to segment 56.21% of the words he heard from the lecture genre.

Moreover, from the questionnaire results, it can be seen that participants in Experiment 1 accurately predicted their performance. They predicted that they would only be able to repeat approximately 70% of the words they heard, suggesting that they themselves are well-aware of their own difficulties with L2 listening. A further interesting finding was that participants predicted their highest performance would be on the film utterances and their lowest would be on the stand-up comedy utterances. Presumably this prediction was made because out of the six genres available, they had more successful experiences watching movies than other types of programme, and minimal success when watching stand-up comedy due to the predominance of culturally biased jokes, which are difficult to comprehend for L2 learners of every language (Bell, 2007).

In actuality, however, participants' highest scores were on the stand-up comedy excerpts, whereas their lowest scores were on the film excerpts. This particular result further illuminates the fact that the listening task used in this experiment only tests speech segmentation, and not higher-level comprehension. Participants were able to hear, segment, and repeat words spoken by Russell Brand (the stand-up comedian) more successfully than words spoken by any other speaker. However, it is impossible to determine how much of his speech they actually comprehended.

In summary, conducting this initial pilot study answered three preliminary questions. Firstly, it demonstrated that the shadowing task used by Mitterer and McQueen (2009) with advanced speakers of English in Holland, could be a suitable approach for testing the L2 speech segmentation ability of international students in the UK, which is a very different population. It was a successful trial run because the participants were able to follow test instructions accurately, hear the excerpts clearly, and repeat words within the given time frame without being rushed.

Secondly, although this experiment was not specifically designed to test differences between genres, the results indicate that participants perform differently when presented with different types of programme. Consequently, it was decided that in future experiments it would be better to keep this variable constant, by only using TV programmes from one genre. Thirdly, the task revealed that despite studying English in their native countries and then living in the UK for more than a year, these participants faced clear problems with understanding standard British English, because they were unable to segment approximately 30% of what they heard.

Having established this, the next question was 'can these students be helped?', and 'if yes, then how?'. Given that, as stated in the literature review (*see* Chapter 2), there is a body of research that suggests watching captioned videos can improve L2 listening, the next question was 'how could an intervention study be designed to investigate this?'. Experiment 2, discussed in the next section of this chapter, attempts to address this question.

4.2 Experiment 2

In Experiment 1, Mitterer and McQueen's (2009) shadowing task was successfully replicated and piloted on 10 international students at a British university. The results revealed that these participants were unable to repeat approximately 30% of the utterances they heard. Having reviewed the literature, it was evident that watching captioned videos had the potential to help such students improve their listening, however, previous research studies were often conducted on a single day. For Experiment 2, the objective was to expose participants to captioned videos multiple times, on multiple days, to see whether multiple exposures could lead to continued improvement in performance and L2 speech segmentation ability. This section presents Experiment 2's research questions, describes its participants, explains the design of the experiment, and how it was scored, then closes with a discussion of the results.

4.2.1 Research question

The primary aim of Experiment 2 was to see whether *repeated exposure* to bimodal input could help international students, such as those in Experiment 1, to improve their overall performance on the listening test, assuming that participants in Experiment 1 were truly representative of their population. Again, Mitterer and McQueen's (2009) research was referred to for guidance in experimental design, and it was noted that in their study, they also investigated participants' ability to segment what they described as 'old' and 'new' items.

For Experiment 2, the first research question was to see whether presenting captioned videos could help listeners to segment speech that they had heard previously. To investigate this, they were tested on some utterances that they had already heard, and these were what Mitterer and McQueen (2009) referred to as 'old items'. The second question was to determine whether watching captioned videos could also help listeners to segment speech that they had not previously heard but that was spoken by a familiar speaker. To investigate this, they were tested on some utterances spoken by someone familiar, but with utterances they had not heard, and these were called 'new items' (ibid).

Most importantly, the third question aimed to investigate whether the effects of watching captioned video could lead to generalised learning, such that if an L2 listener is in a real-world situation, faced with an unfamiliar speaker (of a broadly similar accent) who speaks utterances not previously encountered, she would then be able to segment speech effectively. To test this, materials were used from various documentaries, with different speakers and unfamiliar utterances, which were labelled in this experiment as 'unrelated items'. These three utterance types can be visualised in Table 9 below:

Table 9. The three types of variables investigated in Experiment 2

	Familiar speaker	*Unfamiliar speaker*
Utterance previously heard	'old item'	
Utterance not heard	'new item'	'unrelated item'

4.2.2 Participants

Again, a native speaker of English (different from the one used in Experiment 1) participated in the experiment as a control, providing a baseline for understanding the L2 data. This participant was a 20-year old British female who was studying an undergraduate degree at the time. Additionally, a sample of twelve international students at a British University participated in this experiment (*see* Table 10). They were all Chinese females studying on an MA course in Applied Linguistics course.

Table 10. Summary of the participants' demographic information in Experiment 2, divided by the 3 treatment groups.

Group	***N***	*Age*		*IELTS Listening Score*		*No. of months Living in UK*		*No. of years Studied English during life*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bimodal	4	23.50	1.00	6.63	0.48	6.25	1.26	13.00	1.83
No subtitles	4	23.50	1.00	7.38	0.75	9.00	6.06	12.00	2.31
No sound	4	23.75	0.50	7.00	0.58	8.25	0.96	12.00	0.82

4.2.3 Design and Materials

Three treatment groups were created for this experiment, namely a 'bimodal' group, a 'no subtitles' group and a 'no sound' group. The 'bimodal' group would watch a

given DVD with subtitles turned on and sound playing. The 'no subtitles' group would watch the same DVD but with subtitles turned off whilst sound was playing. In previous research, this has always been considered a 'control' group because these are the normal conditions in which people watch DVDs.

Finally, the 'no sound' group would watch the same DVD with subtitles turned on but sound set to mute. The 'no sound' group was introduced to control for whether a potential advantage in the bimodal condition stems from the subtitles alone (i.e. to rule out the possibility that participants are not processing the spoken input in the bimodal condition) or from the combination of subtitles and sound. The three conditions can be visualised in Table 11.

Table 11. The 3 treatment groups used in Experiment 2

Group	*Orthographic Input*	*Aural Input*
'Bimodal'	✓	✓
'No Subtitles'	X	✓
'No Sound'	✓	X

A pre-test – immediate – post-test design was employed over a four-week period. Participants completed a shadowing task in week 1 as a pre-test, to determine their pre-existing ability to segment L2 speech. In weeks 2 and 3 they watched 30-minutes of two different documentaries, in their assigned condition. The treatment was spread across two weeks in order to explore the cumulative effect on listening of watching programmes with subtitles on listening. In each week, an immediate post-test was conducted. In week 4, participants sat a delayed post-test.

4.2.3.1 Treatment Materials

Experiment 1 used different types of TV programme, but for this experiment only one type of programme was used, that being documentaries. Documentaries were used because the type of monologue present in them is similar to the type of academic lecture that students attend at university, and they come with high-quality, ready-made subtitles that can easily be turned on or off on a TV or computer. Five documentaries were used in this experiment (*see* Table 12).

Table 12. Summary of the materials used in the listening test for Experiment 2

Documentary	*Title*	*Speaker*
1	'Far Eastern Odyssey'	Rick Stein
2	'Blue Planet'	David Attenborough
3	'Wonders of the Universe'	Brian Cox
4	'Meerkats'	Simon King
5	'How We Built Britain'	David Dimbleby

Documentaries 1 and 3 were used as the treatment materials. In week 2, participants watched the first 30-minutes of Documentary 1 before sitting the listening test, and in week 3, they watched the first 30-minutes of Documentary 3 before sitting the listening test. They did not watch any DVDs during week 1 or week 4, as visualised in Table 13 below.

Table 13. The DVD materials that participants watched in Experiment 2

	Week 1	*Week 2*	*Week 3*	*Week 4*
Documentary	N/A	Documentary 1	Documentary 3	N/A

4.2.3.2 Listening Test Materials

The pre-test was composed of 20 utterances (*see* Appendix C.2 for the complete list) from the 5 documentaries, totaling 100 utterances. The immediate post-tests consisted of 120 utterances each. Of those, forty were 'old items', i.e. utterances that the participants were exposed to during the 30-minute viewings in weeks 2 and 3. Forty items were 'new', i.e. utterances from the same documentary, but from the part to which participants were not exposed. Again, these items were introduced to check whether learning gets generalised to new utterances produced by the same speaker. Finally, there were forty 'unrelated items' i.e. utterances from another documentary which the participants did not watch (utterances from Documentary 2 in week 2 and Documentary 4 in week 3), again, this condition was employed to explore whether learning generalises to different speakers of a broadly similar accent (i.e. standard British English).

The delayed post-test consisted of 160 utterances. Of those, 40 were 'old items' (20 from Documentary 1 and 20 from Documentary 3), 40 were 'new items' (20 from

Documentary 1 and 20 from Documentary 3), 40 were 'unrelated items' (20 from Documentary 2 and 20 from Documentary 4) and there were 40 'final unrelated items' – these were items from Documentary 5, which participants had not previously encountered (*see* Table 14 for summary of experimental design).

Table 14. Design of Experiment 2

	Pre-test (week 1)	*Immediate post-test (week 2)*	*Immediate post-test (week 3)*	*Delayed post-test (week 4)*
Excerpt type:		40 'old items' 40 'new items' 40 'unrelated items'	40 'old items' 40 'new items' 40 'unrelated items'	40 'old items' 40 'new items' 40 'unrelated items' 40 'final unrelated items'
Source:	All 5	Documentaries 1 & 2	Documentaries 3 & 4	All 5
Total:	100 excerpts	120 excerpts	120 excerpts	160 excerpts

Utterances were selected in accordance with the same criteria that were used in Experiment 1, which means there was no background music or noise, they were spoken clearly in British English, and were of CEFR upper B2-level complexity according to the 'Text Inspector Profile' online tool. The overall average number of words in an excerpt was 5 ($M = 5.27$; $SD = 1.87$), and there were no statistically significant differences in number of words per excerpt between programmes, as determined by a one-way ANOVA $F(1,2) = 0.14$, $p > 0.05$. Essentially, the listening test itself was created in the same way as it was for Experiment 1.

4.2.4 Procedure and Scoring

Participants attended 4 sessions across 4 weeks. They were trained and tested individually. They all provided informed consent and were paid a nominal fee for their time. The procedure for the shadowing task was identical to the one described for Experiment 1. During the training phase, participants individually watched the video on a laptop computer, with headphones in the two sound conditions, whilst the researcher sat behind them in the same room. The listening test was scored in the same way as it was for Experiment 1.

One participant in the 'no sound' group performed as an outlier, scoring over 2 standard deviations above her group's mean scores. This participant had an IELTS listening score of 8.5, which was distinctively higher than any of the other participants (the other participants' IELTS listening scores were $M = 6.86$; $SD = 0.45$). Rather than excluding her from the analysis, her scores were normalized, capping them at her group's mean plus two standard deviations (Field, 2013).

4.2.5 Results

4.2.5.1 Overall

The native speaker who acted as a control accurately repeated 100% of the utterances. As for the twelve international students, their overall scores on the four tests (pre-test, first immediate post-test, second immediate post-test, delayed post-test) are summarised in Figure 16. The chart includes error bars showing 95% confidence intervals, which is useful for visually inferring statistically significant differences between mean scores (Field, 2013). If error bars overlap, there are no significant differences between scores; however, when bars do not overlap, this indicates a significant difference between scores (*see* Chapter 3).

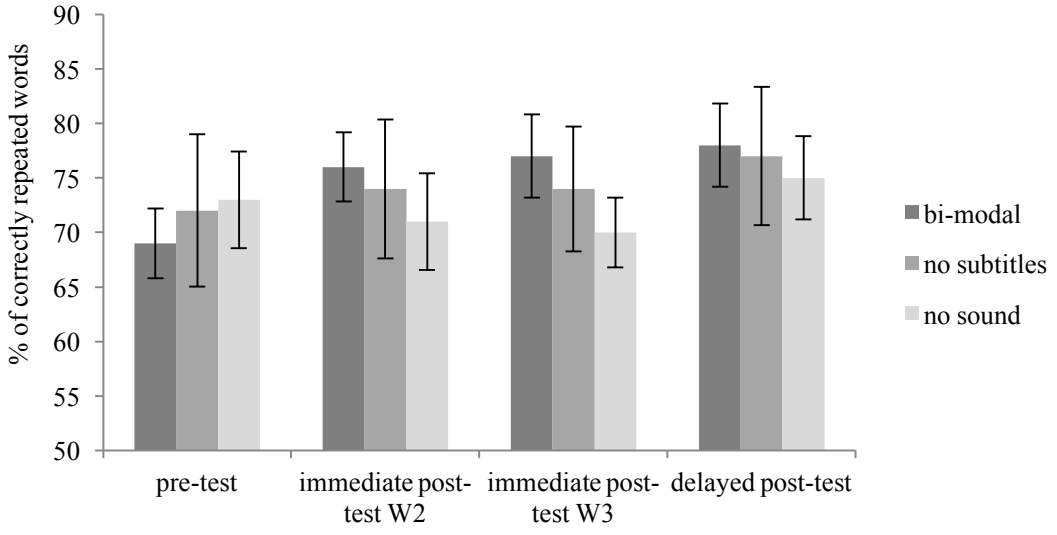

Figure 16. Overall percentage of correctly repeated words on the pre-test, immediate post-tests and delayed post-test across groups in Experiment 2.

The chart shows that the 'bimodal group' scored lower than the other groups on the pre-test in week 1, but thereafter, their scores steadily increased in each week. The 'no subtitles' group made minor improvements on their scores throughout the experiment, and the 'no sound' group had a u-shaped performance, with their highest scores in weeks 1 and 4.

A mixed design ANOVA was conducted with 'group' as the between-subject factor and 'time/test' as the within-subject factor. There was a significant main effect of 'time/test' $F(3,27) = 15.22$, $p < 0.01$, revealing an overall improvement in scores throughout the experiment. However, there was no main effect of group $F(2,9) = 0.17$, $p > 0.05$, indicating that there was no significant difference between groups in their overall performance on the tests.

Having said that, there was a significant interaction between 'group' and 'time' $F(6,27) = 5.71$, $p < 0.05$, suggesting that the groups did not improve at the same rate. More specifically, as Figure 16 shows, the bimodal group's performance improved the most throughout the experiment. Planned contrasts confirmed the 'group' by 'time' interaction for each post-test when compared to the pre-test $F(2,9) = 7.84$, $p < 0.05$ for pre-test vs. immediate post-test in week 2; $F(2,9) = 14.91$, $p < 0.01$ for pre-test vs. immediate post-test in week 3; $F(2,9) = 11.20$, $p < 0.01$ for pre-test vs. delayed post-test. Table 15 summarises groups' performances on the individual utterance types (old, new, unrelated and final) and the increases or decreases in scores for the first immediate post-test, relative to the pre-test.

Table 15. Participants' week 2 scores per item type, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

Group	*Pre-test (week 1)*	*Immediate Post-test 1 (week 2)*		
		old	new	unrelated
Bimodal	66%	69% (+3%)	71% (+5%)	80% (+14%)
No subtitles	69%	70% (+1%)	70% (+1%)	77% (+8%)
No sound	69%	64% (-5%)	65% (-4%)	77% (+8%)

Table 16 summarises groups' performances on the individual utterance types (old, new, unrelated and final) and the increases or decreases in scores for the second immediate post-test, relative to the pre-test.

Table 16. Participants' week 3 scores per item type, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

Group	*Pre-test (week 1)*	*Immediate Post-test 2 (week 3)*		
		old	new	unrelated
Bimodal	66%	80% (+14%)	77% (+11%)	67% (+1%)
No subtitles	69%	75% (+6%)	75% (+6%)	69% (0%)
No sound	69%	73% (+4%)	72% (+3%)	62% (-7%)

Table 17 summarises groups' performances on the individual utterance types (old, new, unrelated and final) and the increases or decreases in scores for the delayed post-test, relative to the pre-test.

Table 17. Participants' week 4 scores per item type, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

Group	*Pre-test (week 1)*	*Delayed Post-test (week 4)*			
		old	new	unrelated	final
Bimodal	66%	71% (+5%)	78% (+12%)	78% (+12%)	72% (+6%)
No subtitles	69%	72% (+3%)	78% (+9%)	78% (+9%)	72% (+3%)
No sound	69%	67% (-2%)	72% (+3%)	77% (+8%)	70% (+1%)

4.2.5.2 Old Items

Recall that the purpose of the first research question was to investigate whether watching captioned videos leads to better speech segmentation of previously heard input, and to test this, 'old' items were included in the experiment. The results of participants' performances on these 'old' items in the week 2 and 3 post-tests can be seen in Figure 17.

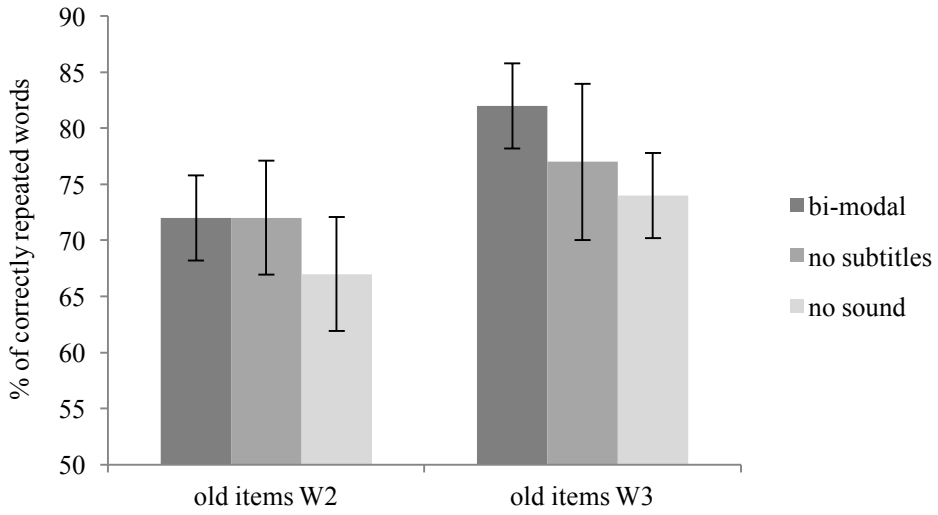

Figure 17. Performance on 'old' items in weeks 2 and 3 in Experiment 2.

The graph shows that all three groups had higher scores in week 3 than in week 2, which is possibly due to the materials used. It seems that Documentary 3, which was used for 'old' and 'new' items in week 3, was considerably easier for participants to segment than other documentaries. Consequently, a repeated measure ANOVA showed a significant difference in participants' scores from week 2 to week 3 on 'old items' $F(1,9) = 38.51$, $p < 0.05$; error bars and a post-hoc analysis revealed that this significant difference was specifically for the bimodal group. However, there were no statistically significant differences between groups nor a time by group interaction $F(2,9) = 1.48$, $p > 0.05$.

4.2.5.3 New Items

The second research question aimed to see whether watching subtitled videos can lead to generalisation of learning to unheard utterances spoken by the same speaker, and this was investigated by using 'new items' in weeks 2, 3 and 4. The results of participants' performances on these 'new' items can be seen in Figure 18.

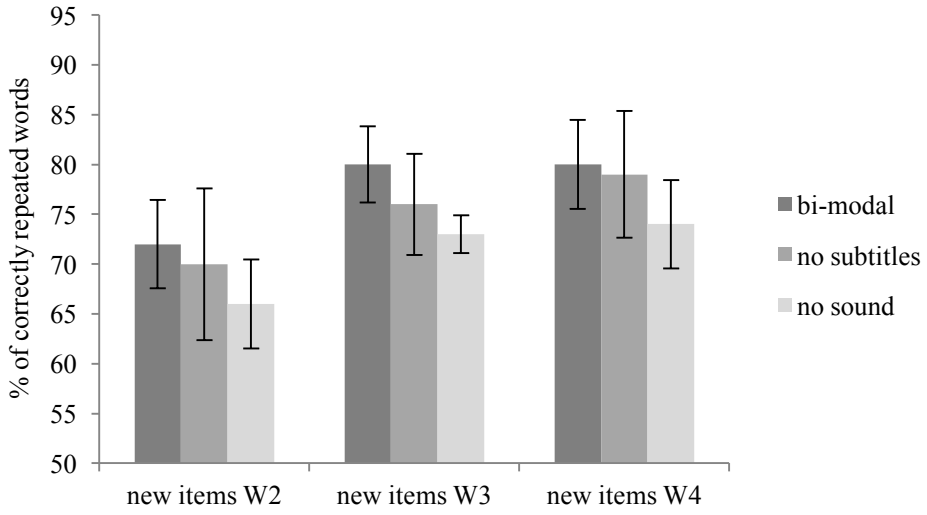

Figure 18. Performance on 'new' items in weeks 2, 3 and 4 in Experiment 2.

Overall, it is clear to see that all groups improved from week 2 to week 3, and the results of a repeated-measures ANOVA show there was a main effect of time $F(2,18) = 28.51$, $p < 0.01$. However, there were no statistically significant differences between groups $F(2,9) = 0.67$, $p > 0.05$, and there was no group by time interaction $F(4,18) = 0.21$, $p > 0.05$.

4.2.5.4 Unrelated Items

Lastly, the third research question sought to determine whether watching captioned videos can lead to the generalisation of learning to segment new utterances spoken by new speakers. Participants' performance was compared on the pre-test with the post-test for items from documentaries that were not part of the training (i.e. 'unrelated' items in weeks 2 and 3, and 'unrelated' and 'final' combined in week 4). The results are summarized in Figure 19.

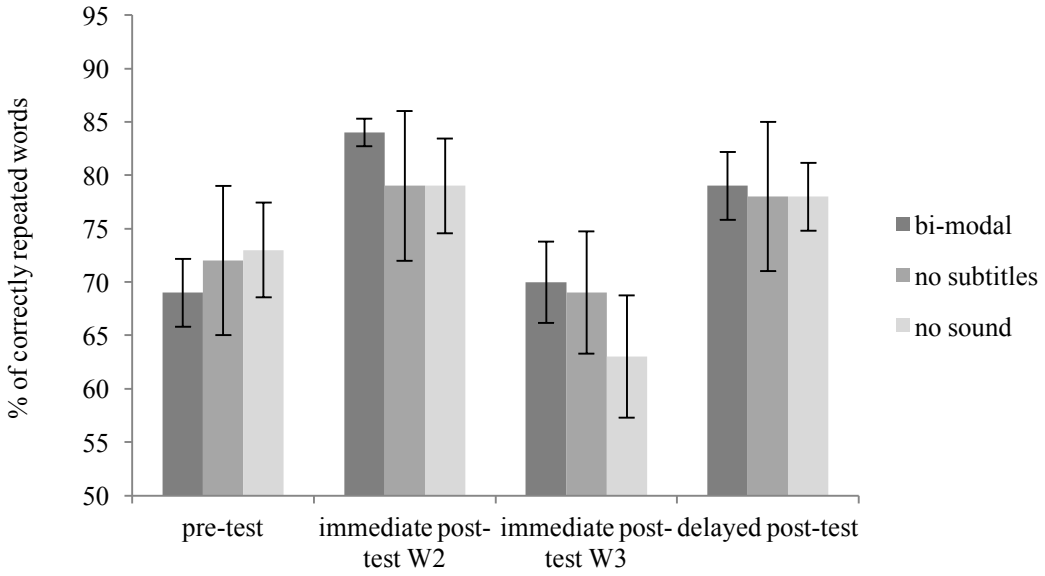

Figure 19. Performance on 'unrelated & final' items throughout the experiment in Experiment 2

The chart shows a distinct difference in participants' scores across weeks, with a clear drop in performance in week 3. A possible explanation for this is that item difficulty was not equally distributed across tests, with some tests being easier or more challenging than others. Despite this problem, the 'bimodal group' started with the lowest score in week 1's pre-test, but outperformed the other groups on all post-tests after the first intervention in week 2.

The mixed design ANOVA with time as the within-subject factor and group as the between-subject factor confirmed the main effect of time ($F = 82.44$, $p < 0.05$). There was no main effect of group ($F = .08$, $p > .05$) but there was a significant group by time interaction ($F = 4.16$, $p < .01$). Planned contrasts confirmed that this interaction was significant in all pre-test vs post-tests comparisons $F = 63.09$, $p < 0.05$ for pre-test vs immediate post-test in week 2; $F = 35.48$, $p < 0.05$ for pre-test vs immediate post-test in week 3; $F = 121.17$, $p < 0.05$ for pre-test vs delayed post-test, confirming that the bimodal group's performance on the post-tests was significantly improved compared to the pre-test, and relative to the control groups.

4.2.6 Discussion

Experiment 2 was designed and conducted to see whether watching captioned videos could help a sample of international students in the UK (like those who participated in Experiment 1) to improve their speech segmentation. The duration of the experiment was 4 consecutive weeks; in week 1 the participants sat a pre-test as a baseline for understanding their performance on subsequent tests throughout the experiment. In week 2 and 3, they watched 30-minutes of a documentary in one of three conditions ('bimodal', 'no subtitles', 'no sound'), then sat an immediate post-test. In week 4, they sat a delayed post-test.

The participants' performances on these tests were compared using a repeated measures ANOVA, and the results indicated a group by time interaction (*see* Figure 16). In week 1, the 'no subtitles' and 'no sound groups' outperformed the bimodal group. But after exposure to the first treatment material, the 'bimodal group' outperformed those groups on the remaining tests. This interaction was statistically significant and supports the hypothesis that watching captioned videos can help in L2 segmentation of continuous speech.

The analysis of participants' scores on previously heard ('old' items) and previously unheard utterances from the same speaker ('new' items) was not fully conclusive. The 'bimodal group' tended to perform better than the control groups on the 'old' items (Figure 17), suggesting that there was a beneficial effect from reading subtitles whilst segmenting the accompanying spoken input. This suggests that watching a programme with subtitles may help second language listeners hear better what is being said. It was also evident that the 'bimodal group' outperformed the control groups on the previously unheard utterances from the documentary they watched (Figure 18). Intriguingly, this suggests that having watched a programme with subtitles, L2 listeners become better able to understand the speaker to whose voice they were exposed, even in the absence of subtitles. Although these results were positive, they were not statistically significant, which could be due to the small sample size.

A more important finding is that bimodal input may lead to the generalisation of learning to unheard utterances by different speakers of a broadly similar accent. The comparison of participants' scores on the pre-test (before being exposed to any training materials) with their post-test performance on utterances from documentaries not included in the training ('unrelated' and 'final' items) again suggested a strong group by time interaction. The 'bimodal group' was initially less able to segment the speech of unfamiliar voices than the other two groups, but after the training was consistently better able than the controls to do so. This finding was statistically significant, which suggests that after watching TV programmes with subtitles, L2 listeners may be better able to understand new non-subtitled programmes and speakers. In other words, this finding suggests that watching programmes with subtitles may be helpful not only for segmenting the spoken input actually accompanied by subtitles, but may have a more far-reaching effect on the development of segmentation abilities in a second language.

4.3 Experiment 3

The findings of Experiments 1 and 2 revealed that intermediate-level, international university students in the UK missed approximately 30% of what they heard on an initial segmentation test. Recall that Experiment 1 was composed of utterances from six different types of TV programme, whereas Experiment 2 consisted only of utterances from documentaries. In both experiments, the mean length of words was similar, however, it appears that some speakers' voices were easier for listeners to comprehend than others. In Experiment 1, participants were able to segment 77.97% of utterances spoken by Russell Brand, compared to only 59.83% of the utterances spoken by Keira Knightly. And in Experiment 2, participants in all groups were able to segment utterances spoken by David Attenborough in week 2, with much greater ease than those spoken by Simon King in week 3 (*see* Figure 19 for participants scores on 'unrelated items').

These findings suggest that (a) exposure to utterances from different types of TV programme could affect participants' scores on the listening test, and (b) even different materials within the same genre may affect participants' performance. Thus, the purpose of Experiment 3 was to investigate this matter further before conducting any further experiments. This section presents research questions, describes participants, and explains the design of the experiment and how it was scored, then closes with a discussion of the results.

4.3.1 Research question

Quite simply, the purpose of this experiment was to replicate Experiment 1, but with focus specifically on the 'documentary' genre, to determine whether different speakers within this genre can have a significant effect on participants' performance in the shadowing task. If certain speakers were found to be too easy (or too difficult) for students to segment, their material would not be included in future experiments.

4.3.2 Participants

A convenience sample of ten international students at a British University, participated in this experiment (*see* Table 18) along with a 20-year old British female, native speaker of English (different from those used in Experiments 1 and 2) who acted as a control, again providing a baseline for understanding the L2 data. The

participants in the sample spoke three L1s between them (7 Chinese, 2 Arabic, 1 Spanish), were all studying at the same university and had a mean IELTS listening test score of 6.0 (*SD* = 2.20), excluding the Spanish participant who had never sat the IELTS exam.

Table 18. Summary of the participants' demographic information in Experiment 3

N	*Age*	*Gender*	*Degree*	*No. of months living in UK*	*No. of years studied English during life*
10	*M* = 23.60; *SD* = 3.13	8 females 2 males	5 Postgrads 5 Undergrads	*M* = 10.70; *SD* = 14.53	*M* = 11.90; *SD* = 6.10

4.3.3 Design and Materials

The design of the experiment was identical to Experiment 1; thus, ten utterances were extracted from 6 documentaries (*see* Table 19), which equated to sixty utterances in total. Again, utterances (*see* Appendix C.3 for the complete list) were selected that (a) had no background music, (b) had no background noises, (c) were spoken clearly in British English, and (d) were of CEFR upper B2-level complexity according to the 'Text Inspector Profile' online tool. The overall average number of words in an excerpt was 6 (*M* = 6.38; *SD* = 1.69), and there were no statistically significant differences in number of words per excerpt between programmes, as determined by a one-way ANOVA $F(1,4) = 1.90$, $p > 0.05$.

Table 19. Summary of the materials used in the listening test for Experiment 3

Documentary	*Title*	*Speaker*
1	'The Bear Family & Me'	Gordon Buchanan
2	'Blue Planet'	David Attenborough
3	'Far Eastern Odyssey'	Rick Stein
4	'How We Built Britain'	David Dimbleby
5	'Meerkats'	Simon King
6	'Wonders of the Universe'	Brian Cox

4.3.4 Procedure and Scoring

As was the case during Experiment 1, participants were informed about the nature of the listening test, signed a consent form (*see* Appendix A), and then completed a questionnaire (*see* Appendix B). They then took the listening test in a quiet, empty room using the same laptop and previously mentioned software, whilst the researcher sat behind them. Again they were presented with sixty excerpts presented in blocks by documentary, which they had to repeat, and the listening test was scored in the same way as it was for Experiment 1. Each participant was paid a nominal fee for their time.

4.3.5 Results

Overall, participants estimated that they would accurately repeat 57.50% (*SD* = 28.21) of excerpts correctly. No participant had previously seen 'The Bear Family & Me', 'Far Eastern Odyssey' nor 'Meerkats'. Two participants had seen 'Blue Planet', another two saw 'How We Built Britain', and one had seen 'Wonders of the Universe'. However, as the viewing had happened some time prior to this study, no verbatim memory for any of the material was expected.

4.3.5.1 Segmentation Test Results

The native speaker who acted as a control accurately repeated all words from all utterances, thus achieving an accuracy of 100%. As for the ten international students, they accurately repeated 70.14% (*SD* = 8.43) of the words they heard during this experiment. Figure 20 presents the percentage of correctly repeated words by participants according to the documentary that utterances were taken from. Participants scored highest in 'How We Built Britain', correctly repeating 89.73% of words (*SD* = 5.06). Conversely their lowest score was in 'Blue Planet', where they correctly repeated just 62.10% of words.

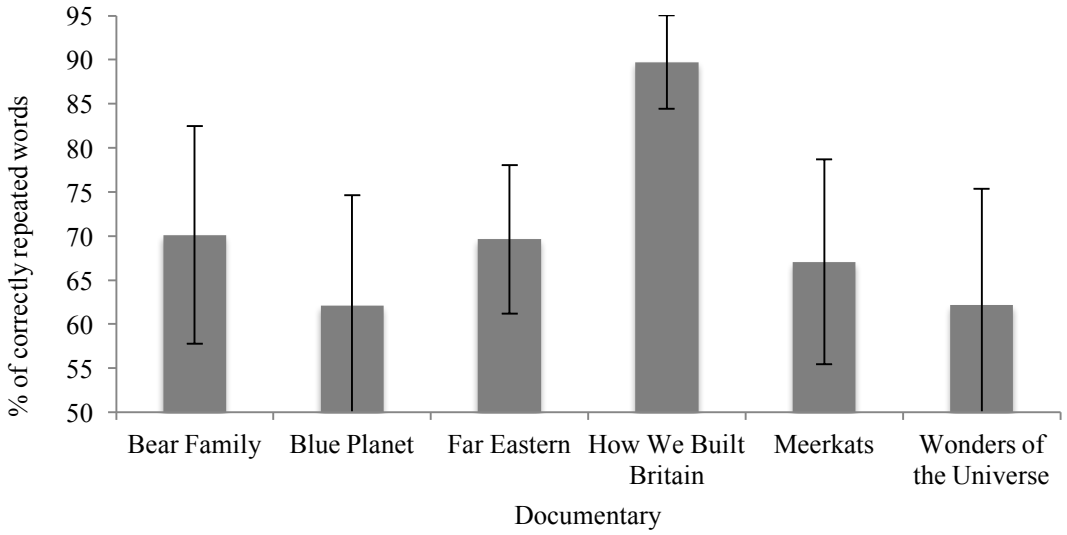

Figure 20. Overall percentage of correctly repeated words by documentary in Experiment 3

A repeated measures ANOVA was performed, with boxplots showing that the data was normally distributed and Mauchly's test of sphericity was not statistically significant, indicating that the assumption of sphericity was met χ^2 (14) = 7.08, p = 0.93. The ANOVA revealed that there was a statistically significant main effect of documentary programme ($F_1(5, 45) = 23.162$, $p < 0.05$, $F_2(9,531) = 10.80$, $p < 0.05$), which suggests that there was a significant difference in participants' performance on the shadowing task, depending upon the documentary they watched. Furthermore, pairwise comparisons suggested that participants' found Documentary 4 to be much easier than the rest. There were no other differences between the documentaries.

4.3.6 Discussion

The aim of this experiment was to replicate Experiment 1, using 6 different documentaries, to find out whether different speakers within this genre can affect participants' performance on the segmentation test. Overall, participants successfully repeated 70% of the utterances they heard, but managed to correctly repeat almost 90% of utterances spoken by David Dimbleby in 'How We Built Britain'. This high score was in conflict with the percentages seen thus far, possibly because he speaks in a distinctively clear way, and at a relatively slow pace. Students were not interviewed, so no qualitative data was collected about their perceptions of the ease or difficulty of listening to a particular speaker, however, it was decided that 'How We Built Britain',

and any other documentaries presented by David Dimbleby, would not be used in future experiments. Conversely, students found difficulty segmenting David Attenborough's speech, with several of them scoring less that 50% on the listening test. Accordingly, it was decided that 'Blue Planet' and other documentaries presented by Sir Attenborough, would not be included in later experiments.

The main lesson learnt from Experiment 3, was that regardless of genre, L2 listeners responded to British speakers differently. As all the speakers had broadly similar accents, and the mean length of excerpts across programmes was similar, it is possible that the pace at which the speakers spoke differed, or the content of what they spoke about had a role to play. In either case, to minimise the effects of these variables on future results, it was evident that counterbalancing the presentation of items in following experiments was essential.

4.4 Summary

The purpose of 'Study 1' was to investigate the speech segmentation ability of intermediate level, international students at British universities, and to determine whether the presentation of captioned videos could help to improve their L2 listening. As discussed in the Literature Review (*see* Chapter 2), previous research claims that watching captioned videos has a beneficial effect on L2 listening comprehension, but many of these studies lack a degree of test-construct validity. 'Study 1' avoided tests that rely on extensive reading or writing, because such tests are not the most reliable way of measuring whether a sentence was accurately heard. As the objective of 'Study 1' was to investigate low-level, bottom-up listening processes at the lexical segmentation level, Mitterer and McQueen's (2009) shadowing task was adapted, as it focuses on a listener's ability to hear individual words in short utterances bounded by pauses, rather than depending heavily on memory.

In Experiment 1, participants were unable to identify approximately 30% of the words they heard. The majority of words presented in the test are familiar to students of CEFR B2 level proficiency, and despite some of the words being unfamiliar, the test only required them to repeat the phonological form, which is possible to do even when words are not meaningful. The important implication of this finding is that potentially, there are international students at British universities who may be failing

to ‘hear’ around 30% of words in lessons, lectures and seminars, which may negatively affect their overall academic performance.

The results of Experiment 1 confirmed that speech segmentation is a genuine problem for international students in the UK, and justified the need for further investigation, specifically, into whether watching captioned videos could aid them. Experiment 2’s results provided evidence that this was certainly possible. Participants in the two control groups outperformed the bimodal group on the pre-test, but the bimodal group outperformed them in all subsequent tests. In particular, the group by time interaction was significant for the overall scores on the 4 tests, and for previously unheard utterances spoken by speakers to which participants had not been exposed. This is an intriguing result as it suggests that watching programmes with sound and subtitles may result in generalized learning that can be applied beyond the programmes watched. The ‘no sound’ group, who watched programmes with subtitles but no sound, consistently scored lower on all post-tests than the other two groups. This suggests that the superior performance of the bimodal group does indeed stem from the simultaneous presentation of sound and subtitles, not from the subtitles alone. In other words, watching programmes with subtitles does not seem to take attention away from listening.

One limitation of Experiment 2 was that a very small sample of only 12 participants was used, so that even though the predicted effects were in the right direction, they did not reach statistical significance. Thus, it was decided that more participants would be recruited in future experiments. Another limitation was the fact that test materials used in some weeks appeared easier than in others. For example, all participants were markedly better on ‘unrelated’ items in week 2 than in week 3. To rule out that this was an effect of participants’ deteriorating segmentation ability from one week to another, Experiment 3 was conducted as an auxiliary study similar to Experiment 1, this time using different documentaries instead of different types of TV programmes as independent variables. The results revealed that it was easier to repeat back utterances from some documentaries than from others. To address this issue in future studies, it was decided that counterbalancing would be used for the presentation of materials across weeks and participants.

CHAPTER 5: STUDY 2 – EFFECTS OF BIMODAL INPUT ON L2 SPEECH SEGMENTATION

5.1 Experiment 4

'Study 1' was composed of three experiments, all of which showed that international students in the UK consistently missed 30% of what they heard during a simple listening test. In Experiment 2, a sample of participants watched captioned videos over two weeks, and showed an improvement on their performance in the listening test, compared to participants who watched the videos without subtitles. Although the results of that experiment were promising, the sample size was limited to only 12 participants and there were several problems with the experimental design.

The purpose of 'Study 2' was to conduct a fourth experiment, which would build and improve upon those conducted in 'Study 1', by using a larger sample size and a more homogenous sample of participants, as well as counterbalanced treatment material and items in the listening tests. This section presents the research questions for Experiment 4, describes the participants who made up the sample, explains how it was designed and conducted, and discusses the results.

5.2 Research Questions

The overarching aim of Experiment 4 was to investigate the extent to which repeated exposure to captioned videos could improve the speech segmentation ability of international students at British universities. Research questions were based upon Experiment 2, and the first of these was to determine whether watching captioned videos could help listeners to segment speech that they had heard previously. Recall that participants were tested on utterances, which they had already heard, defined as 'old items'.

The second research question was to see whether watching captioned videos could help listeners to segment speech that they had not previously heard but that was spoken by a familiar speaker. Again, recall that participants were tested on utterances spoken by a familiar speaker, but with utterances they had not yet heard, and these were defined as 'new items'.

The third and final research question aimed to investigate whether the effects of watching captioned video could lead to generalised learning, such that if an L2 listener is in a 'real world' situation, faced with an unfamiliar speaker (of a broadly

similar accent) who says utterances not previously encountered will she then be able to segment speech effectively? To test this, materials from various documentaries were used, with different speakers and unfamiliar utterances, defined as 'brand new items'. These three utterance types can be visualised in Table 20 below:

Table 20. The three types of items investigated in Experiment 4

	Familiar speaker	*Unfamiliar speaker*
Utterance previously heard	'old item'	
Utterance not heard	'new item'	'brand new item'

5.3 Participants

A sample of 48 international students at two UK universities participated in this study. All were Chinese, and their demographic details are presented in Table 21. Additionally, nine L1 native speakers of English took the listening tests, providing a baseline for understanding the L2 data. Their ages ranged from 25-47 years old, 6 were British and 3 were American.

Table 21. Summary of the participants' demographic information in Experiment 4, divided by the 3 treatment groups.

Group	***N***	*Gender*		*Age*		*No. of months Living in UK*		*No. of years Studied English during life*	
		m	*f*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bimodal	16	3	13	22.75	0.58	3.00	1.71	11.81	3.02
No subtitles	16	1	15	22.56	1.26	3.06	1.61	11.94	2.08
No sound	16	2	14	23.00	1.10	3.75	2.77	10.69	1.14

Overall, the participants constituted a homogenous sample, and were divided into the three groups based upon their IELTS listening score, such that the mean IELTS score for each group was as similar as possible. Table 22 summarises participants' language profiles according to group.

Table 22. Participants' language profiles by group in Experiment 4

Group	***N***	*IELTS overall score*		*IELTS listening score*		*IELTS speaking score*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Bimodal	16	6.50	0.37	6.91	0.64	6.06	0.73
No subtitles	16	6.44	0.54	6.88	0.70	5.97	0.56
No sound	16	6.50	0.52	6.84	0.91	6.31	0.63

5.4 Design and Materials

The fundamental design of this experiment was the same as Experiment 2, in that there were three treatment groups, namely a 'bimodal' group, a 'no subtitles' group and a 'no sound' group. The 'bimodal' group watched a documentary with subtitles turned on and sound playing. The 'no subtitles' group watched the same documentary but with subtitles turned off whilst sound was playing. And the 'no sound' group watched the same documentary with subtitles turned on but sound set to mute. Again, a pre-test – treatment – post-test design was employed over a four-week period. In week 1, participants completed a shadowing task as a pre-test. In weeks 2 and 3 they watched 30-minutes of two different documentaries before sitting an immediate post-test. Finally, in week 4, participants sat a delayed post-test.

5.4.1 Treatment Materials

Six documentaries were used in this experiment (*see* Table 23), all of which were BBC documentaries available in DVD format, based upon a variety of topics.

Table 23. Summary of the materials used in the listening tests for Experiment 4

Documentary	*Title*	*Speaker*
1	Far Eastern Odyssey	Rick Stein
2	Wonders of the Universe	Brian Cox
3	Invisible Worlds	Richard Hammond
4	The Diamond Queen	Andrew Marr
5	Law & Disorder	Louis Theroux
6	Meerkats	Simon King

Counterbalancing is a method of controlling order effects in a repeated measures experiment, whereby separate groups of participants are used, with each group receiving treatments in a different order (Lund and Lund, 2013). After conducting Experiment 2 it was noted that counterbalancing needed to be applied to future experiments because the performance of all participants sharply declined from week 2 to week 3, and Experiment 3 revealed that this happened because participants found some documentaries easier to segment than others. Documentaries 1 and 2 were used as the treatment materials.

In week 2, half of the participants in each group watched the first 30-minutes of Documentary 1, and the other half watched the first 30-minutes of Documentary 2 before sitting the listening test. In week 3, the first half watched the first 30-minutes of Documentary 2, whereas the second half watched the first 30-minutes of Documentary 1 before sitting the listening test. They did not watch any documentaries during week 1 or week 4, as visualised in Table 24.

Table 24. The DVD materials that participants watched in Experiment 4

	Week 1	*Week 2*	*Week 3*	*Week 4*
Documentary	N/A	Doc1 (*or Doc2*)	Doc2 (*or Doc1*)	N/A

5.4.2 Listening Test Materials

Using a counterbalanced approach here meant that the design of Experiment 4 was rather complex (*see* Table 25), and requires a thorough explanation. To begin with, recall that the first research question was to investigate participants' performance on 'old items', which are utterances they were exposed to during the treatments in weeks 2 and 3 when they watched the documentaries. To test 'old items', 40 utterances were included in week 2's test and another 40 in week 3's test, totaling 80 utterances. There were also 80 'old item' utterances included in the delayed post-test. Using a 4x4 matrix design, the 16 participants in each group were presented with test material in different orders.

80 'old items' were extracted from the documentary then split into two lists of 'old items', 'list_A' and 'list_B', thus ensuring that a single participant was not tested on

the same items during the immediate and delayed post-tests. Furthermore, to ensure that there was no effect of ordering, the presentation of items in pre-tests and post-tests were counterbalanced, so that one participant was presented with them in 'A' then 'B' order, whilst the other participant was presented with 'B' then 'A' order.

Taking the bimodal group as an example, Participant 1 watched Documentary 1 in week 2, and was then tested on 40 'old items' from List A in the immediate post-test. Participant 2, on the other hand, watched Documentary 2 in week 2, and was tested on 40 'old items' from List B in the immediate post-test. Participant 3 watched Documentary 1 in week 2, but was tested on 40 'old items' from List B in the immediate post-test. Participant 4 watched Documentary 2 in week 2, and was then tested on 40 'old items' from List A in the immediate post-test. This 4x4 matrix design was also used to counterbalance 'new items', which are utterances they were not exposed to during the treatments in weeks 2 and 3, but came from the same speaker (*see* Table 25).

As for the third research question, which aimed to investigate whether the effects of bimodal input could lead to generalised learning, participants were tested on 'brand new items', which are utterances not previously encountered, and from an unfamiliar speaker. The pre-test was composed of 80 utterances (*see* Appendix C.4 for the complete list), as were the immediate post-tests and the delayed post-test.

Taking the bimodal group as an example again, Participant 1's pre-test was composed of 'brand new items' from Documentary 3. In week 2's immediate post-test, these items were extracted from Documentary 4. Week 3's immediate post-test was composed of items from Documentary 5, and the delayed post-test in week 4 was made up of items from Documentary 6. For Participant 2, the pre-test was composed of 'brand new items' from Documentary 6, followed by an immediate post-test in week 2 with items from Documentary 3. Then in week 3 the immediate post-test was composed of items from Documentary 4, and finally the delayed post-test in week 4 was made up of items from Documentary 5 (*see* Table 25).

Utterances were selected in accordance with the same criteria that were used in 'Study 1', the overall average number of words in an excerpt was 5 (M = 4.88; SD = 1.74), and there were no statistically significant differences in number of words per excerpt between programmes, as determined by a one-way ANOVA $F(1,2) = 6.40$, $p > 0.05$.

Table 25. Design of Experiment 4

Pre-test	*Week 2*	*Week 3*	*Post-test*
	Doc1 [old_A x40] [new_A x40]	Doc2 [old_A x40] [new_A x40]	Doc1 [old_B x40] [new_B x40] Doc2 [old_B x40] [new_B x40]
Doc3 x80	Doc4 x80	Doc5 x80	Doc6 x80
	Doc2 [old_A x40] [new_A x40]	Doc1 [old_A x40] [new_A x40]	Doc2 [old_B x40] [new_B x40] Doc1 [old_B x40] [new_B x40]
Doc6 x80	Doc3 x80	Doc4 x80	Doc5 x80
	Doc1 [old_B x40] [new_B x40]	Doc2 [old_B x40] [new_B x40]	Doc1 [old_A x40] [new_A x40] Doc2 [old_A x40] [new_A x40]
Doc5 x80	Doc6 x80	Doc3 x80	Doc4 x80
	Doc2 [old_B x40] [new_B x40]	Doc1 [old_B x40] [new_B x40]	Doc2 [old_A x40] [new_A x40] Doc1 [old_A x40] [new_A x40]
Doc4 x80	Doc5 x80	Doc6 x80	Doc3 x80

5.5 Procedure and Scoring

Participants attended 4 sessions across 4 weeks. They were trained and tested individually. They all provided informed consent and were paid a nominal fee for their time. The procedure for the shadowing task was identical to the one described for 'Study 1'. During the training phase, participants individually watched the video on a laptop computer, with headphones (if they were in one of the two sound conditions), whilst the researcher sat behind them in the same room. The listening test was scored in the same way as it was for experiments in 'Study 1'.

5.6 Results

The native speakers who sat the test accurately repeated 99.33% (*SD* = 0.51) of the utterances. Regardless of age or nationality, this clearly shows the segmentation task was very easy for all native English-speaking participants, and did not pose any challenges. Consequently their data will not be further considered here. The scores of the 48 international students are presented below.

5.6.1 Old Items

'Old items' were utterances that participants were exposed to during the treatment phase of the experiment, when they watched the documentaries. These 'old items' were only presented after the treatment on immediate and delayed post-test. Based on the results of Study 1, it was expected that the bimodal group would improve on the test throughout the experiment, scoring higher than the control group. This was indeed the case, and Figure 21 shows that the 'bimodal' group outperformed the 'no subtitles' and 'no sound' groups on both the immediate and delayed post-tests.

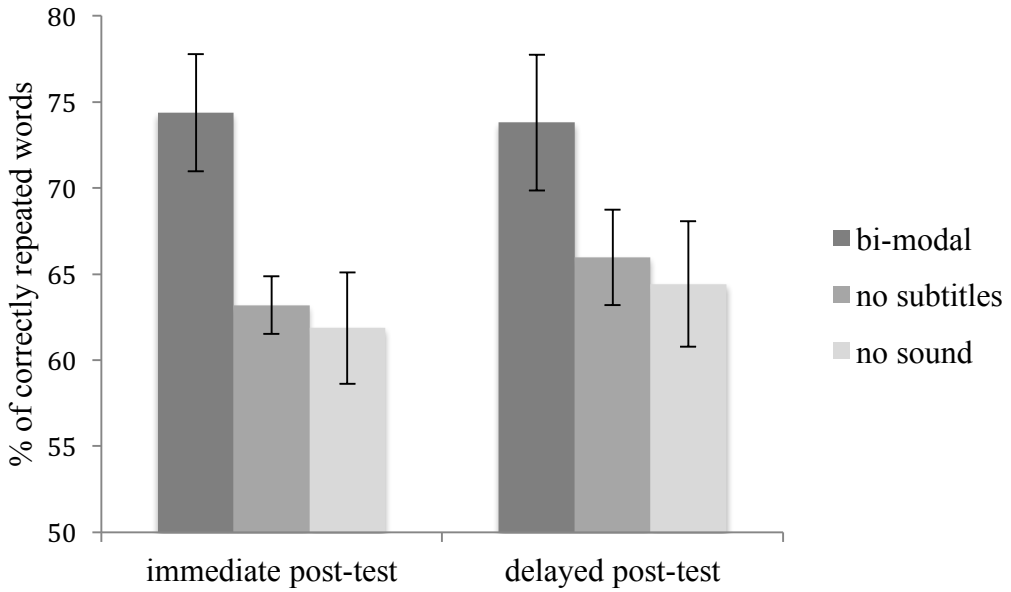

Figure 21. Percentage of correctly repeated words on 'old items' in the immediate and delayed post-tests across groups in Experiment 4.

A mixed design ANOVA was conducted with 'group' as the between-subject factor, 'time' as the within-subject factor, and immediate post-tests' scores collapsed. There was a main effect of 'group' ($F_1(2,45) = 19.21$, $p < 0.01$, $F_2(2,158) = 158.19$, $p > 0.05$). Pairwise comparisons (by-subject) with a Bonferroni adjustment revealed

statistically significant differences between scores on 'old items' for the 'bimodal' vs. 'no subtitles' groups ($M = 9.50$, $SD = 1.91$, $p < 0.01$), and the 'bimodal' vs. 'no sound' groups ($M = 10.94$, $SD = 1.91$, $p < 0.01$), but not for the 'no subtitles' vs. 'no sound' groups ($M = 1.44$, $SD = 1.91$, $p = 1.00$). There was also a main effect of 'time' in by-subject analysis but not in by-item analysis ($F_1(1,45) = 4.72$, $p < 0.05$, $F_2(1,79) = 1.05$, $p > 0.05$), which was confirmed by a planned contrast for immediate post-tests vs. delayed post-test ($F_1(1,45) = 4.72$, $p < 0.05$, $F_2(1,79) = 1.05$, $p > 0.05$). Additionally, there was no 'group' by 'time' interaction in by-subject analysis, but there was in by-item analysis ($F_1(2,45) = 2.15$, $p > 0.05$, $F_2(2,158) = 1.41$, $p < 0.05$).

5.6.2 New Items

'New items' were utterances that participants were not exposed to during the treatment phase of the experiment, but which were spoken by the same speakers they had heard, and presented after the treatment on immediate and delayed post-test. It was presumed that the bimodal group would segment items better but that there would be no effect of time or interaction. Figure 22 shows that the 'bimodal' group outperformed the 'no subtitles' and 'no sound' groups on both the immediate and delayed post-tests.

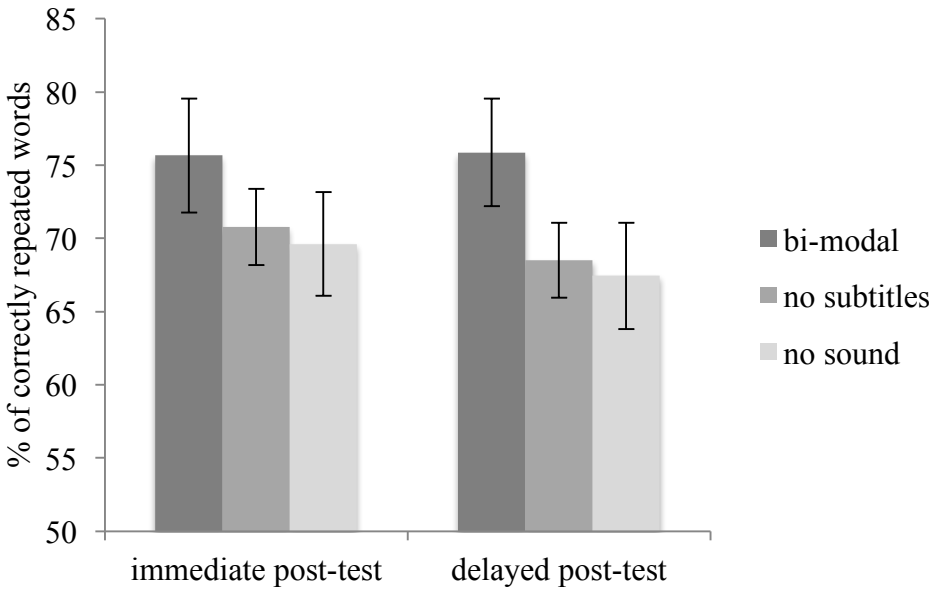

Figure 22. Percentage of correctly repeated words on 'new items' in the immediate and delayed post-tests across groups in Experiment 4

A mixed design ANOVA was conducted with 'group' as the between-subject factor, 'time' as the within-subject factor, and immediate post-tests' scores collapsed. There was a main effect of 'group' ($F_1(2,45) = 7.87$, $p < 0.01$, $F_2(2,158) = 64.83$, $p < 0.01$). Pairwise comparisons (by-subject) with a Bonferroni adjustment revealed statistically significant differences between scores on 'new items' for the 'bimodal' vs. 'no subtitles' groups ($M = 6.13$, $SD = 1.96$, $p < 0.01$), and the 'bimodal' vs. 'no sound' groups ($M = 7.24$, $SD = 1.96$, $p < 0.01$), but not for the 'no subtitles' vs. 'no sound' groups ($M = 1.10$, $SD = 1.96$, $p = 1.00$). There was no main effect of 'time' ($F_1(1,45) = 2.71$, $p > 0.05$, $F_2(1,79) = 1.37$, $p > 0.05$), and there was no 'group' by 'time' interaction in by-subject analysis although there was in by-item analysis ($F_1(2,45) = 0.89$, $p > 0.05$, $F_2(2,158) = 1.26$, $p < 0.05$).

5.6.3 Brand new Items

Figure 23 shows that the three groups were evenly matched on the pre-test in week 1, all scoring approximately 63% on 'brand new items'. After exposure to the first treatment, the 'bimodal' group made a sharp increase in their scores from week 1 to week 2, which then remained fairly constant in the remaining weeks. The 'no subtitles' group made a slight improvement from week 2 to week 3, while the 'no sound' group did not make any improvement throughout the experiment.

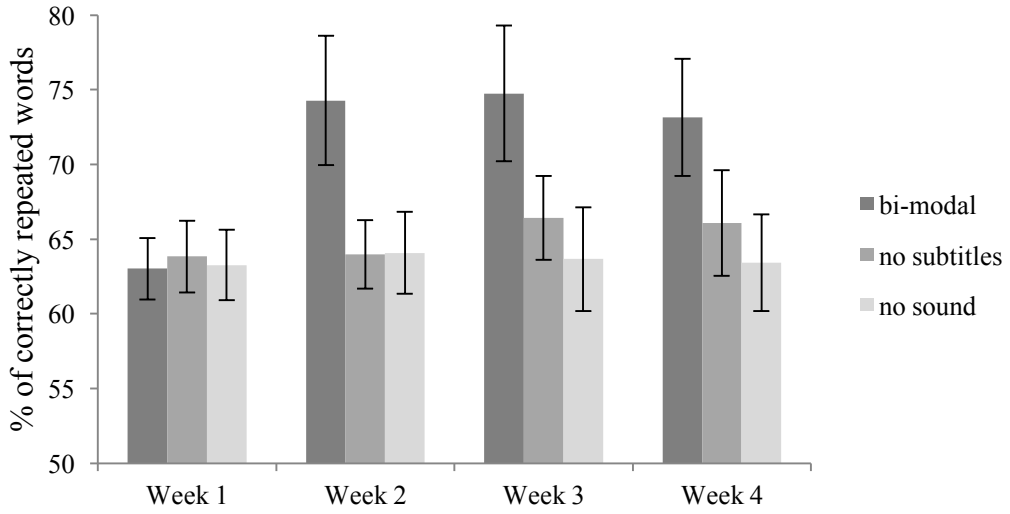

Figure 23. Percentage of correctly repeated words on 'brand new items' in the pre-test, immediate post-tests and the delayed post-tests across groups in Experiment 4

A mixed design ANOVA was conducted again with 'group' as the between-subject factor and 'time' as the within-subject factor. There was a main effect of 'group' ($F_1(2,45) = 12.68$, $p < 0.01$, $F_2(2,158) = 147.52$, $p < 0.01$). Pairwise comparisons (by-subject) with a Bonferroni adjustment revealed statistically significant differences between scores on 'brand new items' for the 'bimodal' vs. 'no subtitles' groups ($M = 6.21$, $SD = 1.62$, $p < 0.01$) and the 'bimodal' vs. 'no sound' groups ($M = 7.69$, $SD = 1.62$, $p < 0.01$), but not for the 'no subtitles' vs. 'no sound' groups ($M = 1.47$, $SD = 1.62$, $p = 1.00$).

There was also a significant main effect of 'time' ($F_1(3,135) = 11.20$, $p < 0.05$, $F_2(3,237) = 5.37$, $p < 0.05$), and it is apparent through the graph that this was mainly driven by the bimodal group. Planned contrasts confirmed this effect of 'time' for each post-test when compared to the pre-test ($F_1(1,45) = 23.29$, $p < 0.05$, $F_2(1,79) = 8.56$, $p < 0.01$) for pre-test vs. immediate post-test in week 2; ($F_1(1,45) = 21.55$, $p < 0.01$, $F_2(1,79) = 15.56$, $p < 0.01$) for pre-test vs. immediate post-test in week 3; and ($F_1(1,45) = 16.81$, $p < 0.01$, $F_2(1,79) = 12.58$, $p < 0.01$) for pre-test vs. delayed post-test.

There was a significant interaction between 'group' and 'time' ($F_1(6,135) = 6.84$, $p < 0.05$, $F_2(6,474) = 9.53$, $p < 0.05$), confirming that the groups did not improve at the same rate. Planned contrasts confirmed this 'group' by 'time' interaction for each post-test when compared to the pre-test ($F_1(2,45) = 18.26$, $p < 0.01$, $F_2(1,79) = 19.29$, $p < 0.05$) for pre-test vs. immediate post-test in week 2; ($F_1(2,45) = 10.85$, $p < 0.01$, $F_2(1,79) = 15.25$, $p < 0.05$) for pre-test vs. immediate post-test in week 3; and ($F_1(2,45) = 8.93$, $p < 0.01$, $F_2(1,79) = 11.545$, $p < 0.05$) for pre-test vs. delayed post-test. Table 26 further highlights the differences between the groups' performances throughout the test. It shows that the 'bimodal' group's performance peaked at a 12% increase in their test scores at week 3. The 'no subtitles' group's highest increase, however, was 3%, and the 'no sound' group's scores were constant at each time point.

Table 26. Participants' weekly scores on 'brand new items' in Experiment 4, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

Group	*Week 1: Pre-test*	*Week 2: Post-test*	*Week 3: Post-test*	*Week 4: Delayed post-test*
bimodal	63.02	74.28 (+11%)	74.75 (+12%)	73.13 (+10%)
no subtitles	63.85	63.99 (+0%)	66.41 (+3%)	66.08 (+2.23%)
no sound	63.27	64.07 (+0.80%)	63.67 (+0.40%)	63.42 (+0.15%)

5.7 Discussion

Experiment 4 set out to investigate whether watching captioned videos could help international students in the UK to improve their speech segmentation. The experiment was conducted over a period of 4 weeks; in week 1 participants sat a pre-test as a baseline for understanding their performance on subsequent tests throughout the experiment. In weeks 2 and 3, they watched 30-minutes of a documentary in one of three conditions ('bimodal', 'no subtitles', or 'no sound'), then sat an immediate post-test. Finally, in week 4, they sat a delayed post-test.

Participants' performances on the tests were compared using a mixed design ANOVA, and the results indicated a group by time interaction. In week 1, all groups performed equally, scoring approximately 63% overall. After being exposed to the first treatment material, the 'bimodal' group proceeded to outperform the control groups in all remaining listening tests. This pattern was also observed in 'Study 1', and again there was a statistically significant interaction, which provides further evidence to support the notion that bimodal input may lead to improvements in L2 speech segmentation.

In Study 1, there were no significant differences between groups in their 'old' and 'new' item scores, but the results of this experiment were different. The 'bimodal' group consistently scored higher on 'old' items after the first treatment compared to the control groups, which suggests that L2 listeners who read at the same time as listening are better able to segment what is being said. Furthermore, the 'bimodal' group also outperformed the control groups on 'new' items, which suggests that when L2 listeners read subtitles while listening to a speaker, it acts as a form of training for them, so that they are in a better position to segment the speech of that same speaker

when hearing her on a later occasion. These findings are consistent with Study 1's results as well as those of Mitterer and McQueen (2009), thus providing a degree of reliability to support that claim.

With regard to 'brand new' items, again the 'bimodal' group outperformed the control groups after the initial treatment, and more importantly, there is a significant improvement in their scores from the pre-test to the delayed post-test. This is an important finding because it suggests that not only does bimodal input help an L2 listener segment speech ('old' items), and help her to quickly adapt to a speaker's voice ('new' items), but it may also lead to the generalisation of learning, meaning that she is better able to segment never-before-heard speech, spoken by different speakers of a broadly similar accent.

Ultimately, this study has provided further evidence to support the claim that watching captioned video, and specifically in this case, bimodal input, contributes toward the improvement of L2 segmentation abilities in a very short period of time, and that these effects seem to persist over an extended period of time. The question that arose next was 'is this segmentation really the bedrock of comprehension?' Would simply being able to hear words on a very artificial task also translate into a real world situation in which somebody better understands what they are listening to?

This claim has often been made. However, in previous experiments, it was never entirely clear whether the participants' comprehension of treatment material came from the orthographic information (i.e. simply reading the subtitles) or the actual listening, or a combination of both? Apart from Bird and Williams (2002), this is the only other study that has included a 'no sound' group, wherein participants read the subtitles without listening to the spoken text. If participants in the 'no sound' group had performed as well as participants in the 'bimodal' group, this could have provided strong evidence against the claims made by proponents of bimodal input, because it would mean that participants in previous studies had focused only on reading the subtitles, and that is how they passed the listening comprehension tests. However, the fact that participants in the 'no sound' group consistently scored lower than the 'bimodal' and 'no subtitles' groups, clearly shows that reading alone is not enough.

Therefore, the claims of previous authors were indeed correct, despite the fact that they never really distinguished between the text-only and sound-only conditions. This study, however, does convincingly show that it is the combination of both listening and reading which contributes towards comprehension. Having established that bimodal input can help L2 listeners with speech segmentation, the next study aimed to determine the extent to which this leads to better listening comprehension.

CHAPTER 6: STUDY 3 – EFFECTS OF BIMODAL INPUT ON L2 LISTENING COMPREHENSION

6.1 Experiment 5

In Study 1, three experiments were conducted that used a shadowing task to measure L2 listeners' speech segmentation ability, which revealed that international students at UK universities were unable to segment approximately 30% of the utterances they heard, when spoken by native speakers of English of a broadly similar British accent. In Study 2, another experiment was conducted, which improved upon the limitations of Study 1. In this experiment, 48 international student participants were initially unable to segment 30 – 40% of the utterances they heard. However, the participants who were exposed to captioned videos significantly improved their speech segmentation ability over a period of four weeks. Compared to the control groups, they were better able to segment (a) utterances they had previously heard from a familiar speaker, (b) utterances they had not previously heard but spoken by a familiar speaker, and (c) utterances they had not previously heard, that were spoken by an unfamiliar speaker.

Considering these findings, a key follow-up question was 'does the improved ability to segment continuous speech on a shadowing task also translate into better L2 listening *comprehension*?' The purpose of Study 3 was to investigate this question by conducting another experiment, which also included listening comprehension tests. This section presents the research questions for Experiment 5, describes the participants who were in it, and explains how it was designed and conducted, as well discussing the results.

6.1 Research Questions

The primary aim of Experiment 5 was to investigate whether repeated exposure to captioned videos could improve the L2 *listening comprehension* ability of international students at British universities. It was also an opportunity to check the reliability of Study 2's findings, by keeping the core experimental design constant, whilst observing how another sample of students performed on the segmentation tests.

The results of Study 2 raised several additional questions, which this experiment set out to answer. First of all, does reading proficiency play a role in the usefulness of subtitles? In Studies 1 and 2, participants sat a pre-test, to determine their

segmentation ability, but they were never tested on their reading ability. For students in the bimodal group (who were required to read subtitles), this information could be insightful, because as noted in Chapter 2, quicker or better readers may benefit more from subtitles than others.

Two secondary research questions arose. First, does vocabulary size play a role in the usefulness of subtitles? It might be assumed that knowing a large number of words would make lexical segmentation much easier. Secondly, does working memory play a role in the usefulness of subtitles for L2 speech segmentation? Essentially, Experiment 5 investigates participants' proficiency skills (in reading, vocabulary and working memory), then their L2 segmentation ability, and finally their L2 listening comprehension ability.

6.2 Participants

A sample of 32 international students from China, who were enrolled at a UK university participated in this study. Table 27 summarises their demographic backgrounds. In this experiment, participants were divided into two groups: 'bimodal' and 'no subtitles'.

Table 27. Summary of the participants' demographic information in Experiment 5

Group	***N***	*Gender*		*Age*		*No. of months Living in UK*		*No. of years Studied English during life*		*Degree Course*	
		m	*f*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*L**	*non-L*
Bimodal	16	2	14	23.63	0.81	9.13	4.46	11.81	1.94	11	5
No subtitles	16	5	11	25.63	2.42	12.88	11.09	12.38	4.03	6	10

**L = linguistics-related degree*

Welch's t-test revealed a statistically significant difference between the groups' ages, with the 'no subtitles' group being slightly older, $t(18.29) = -3.13$, $p < 0.05$. However, there were no significant differences between the number of months they had spent living in the UK, $t(19.71) = -1.25$, $p > 0.05$, nor in the number of years they had spent studying English during their lives, $t(21.59) = -0.50$, $p > 0.05$, meaning they were relatively homogenous groups in this regard.

Participants in the groups also had similar self-reported IELTS scores for reading, listening, and speaking, as summarised in Table 28. Again, t-tests did not reveal any statistically significant differences between IELTS overall scores, $t(30) = 0.34$, $p > 0.05$, IELTS listening scores, $t(30) = -0.29$, $p > 0.05$, IELTS speaking scores, $t(30) = 0.48$, $p > 0.05$, or IELTS reading scores $t(30) = 0.82$, $p > 0.05$. This indicated that the groups were evenly matched in their level of English proficiency, and would presumably perform similarly in the experiment.

Table 28. Participants' IELTS scores by group in Experiment 5

Group	***N***	*IELTS overall*		*IELTS listening*		*IELTS speaking*		*IELTS reading*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bimodal	16	6.63	0.50	6.81	0.95	6.31	0.57	7.06	0.93
No subtitles	16	6.56	0.51	6.91	0.84	6.22	0.52	6.81	0.77

6.3 Design and Materials

At its core, the design of this experiment was identical to Experiment 4. The treatment materials and segmentation tests that were employed in Experiment 4 were used again. However, in this experiment, participants sat 'skills tests' as a baseline measure one week before starting the experiment, and weekly listening *comprehension* tests were added after each shadowing task. Another difference is that a 'no sound' group was used in the previous experiment to determine whether the effect of combining sound with text or simply text alone leads to improvement of scores on the segmentation tests. That experiment convincingly showed that it was indeed the bimodality of sound and text, rather than text alone that helped students improve. Therefore, a 'no sound' group was not used in this experiment. The overall design of Experiment 5 can be seen in Table 29.

Table 29. Design of Experiment 5

Week 1: skills tests	*Week 2: pre-test*	*Week 3: immediate post-test*	*Week 4: immediate post-test*	*Week 5: delayed post-test*
		Treatment: 30 min of Documentary A	Treatment: 30 min of Documentary B	
Vocabulary size Reading rate Reading comprehension Working memory	Shadowing task Listening comprehension test	Shadowing task Listening comprehension test	Shadowing task Listening comprehension test	Shadowing task Listening comprehension test

6.3.1 Listening Comprehension Test

As discussed in Chapter 3, to test listening comprehension, the 'Cambridge English: Advanced (CAE)' exam was selected. The listening section has 30 questions, composed of multiple-choice questions, matching tasks, and sentence completion questions. Each question is worth one point, which means the highest score someone can obtain is 30. Each test took 40-minutes to complete, and four CAE past papers were used (*see* Appendix E) and delivered in a counterbalanced approach across weeks and between participants.

6.3.2 Skills Tests

Four skills that might influence how much somebody can benefit subtitles include (a) vocabulary size, (b) reading rate, (c) reading comprehension, and (d) working memory. Vocabulary size was considered an important factor because it may be the case that participants with large vocabulary sizes benefit more from bimodal input than those with smaller vocabulary sizes. To test vocabulary size, Nation's (2010) online 'Vocabulary Size Test' was used. Seeing as all the participants were Chinese,

they were instructed to take the bilingual (English-Chinese) version of this test, which meant the target English word was presented along with an example sentence, but word definitions were in Chinese (*see* Figure 10).

In addition to vocabulary size, it is possible that a participant's reading rate may affect their comprehension of subtitles in this experiment. On one hand, a fast reader may comprehend the subtitles better than a slow reader (Bell, 2001). On the other hand, being able to read quickly does not necessarily result in increased comprehension of what is being read (Chusing-Weigle and Jensen, 1996). Bearing this in mind, one of the aims of this study was to identify participants' reading rates and determine whether or not it affected their performance on receiving bimodal stimuli.

In addition to vocabulary size, reading rate and comprehension, another important factor to consider was participants' working memory capacity. Since the capacity of working memory is limited and varies from one individual to another, it was decided that this was an important factor to consider in this research study for two reasons. Firstly, during the treatment phase of the experiment, when a participant is exposed to bimodal stimuli, she is required to process two sources of input. It is plausible that participants with larger working memory capacity are better able to process such stimuli (by extracting more data from the bimodal input) than participants with smaller working memory capacity. Secondly, during the listening test, when participants listen to the short pause-bound utterances, again it is possible that those with larger working memory capacity may have an advantage over those with smaller working memory capacity, because they are better able to retain the spoken stimuli for longer and then repeat it. Considering these two points, participants' working memory capacity was measured during the skills test phase.

6.4 Procedure

Participants attended 5 sessions across 5 weeks. They were trained and tested individually, provided informed consent and were paid a nominal fee for their time. During the first meeting, they were given 25-minutes to take the vocabulary size test on a laptop. They were then given a 5-minute break period, before moving onto the reading test, wherein they read the 'History of Chocolate' passage aloud whilst being recorded by the laptop. The comprehension questions were then asked aloud, and

answers recorded as correct or incorrect on a grading sheet, while participants responded. Participants were then given another 5-minute break before starting the working memory test, which was distributed on printed A4 paper. For 'weeks 2 – 5', the procedure for the shadowing task was identical to the one described for 'Study 2'. The only addition was that upon completion of the shadowing task, participants were required to sit the listening comprehension test. Audio files for the listening test were played from a laptop, whilst participants answered the questions on printed paper.

6.5 Results

6.5.1 Skills Tests

In Study 2 (Experiment 4), participants' segmentation ability was measured during a pre-test phase, however, there was no consideration of their reading ability. This study improves upon this experimental design by obtaining baseline measures of reading rate, reading comprehension, vocabulary size and working memory capacity. Reading rate was tested using the York Adult Assessment-Revised (Warmington et al., 2013) and the average reading rate for participants in this study was 97 words per minutes (M = 97.13; SD = 11.99), indicating that they were much slower than the average L1 speaker of English, who generally reads at 164 words per minute (ibid). Reading comprehension was also tested using the YAA-R and the average reading comprehension score was 9 (M = 8.84; SD = 1.89) out of 15 questions, indicating that their ability to understand what they read was close to L1 speakers of English who generally achieve a score of 10 (ibid).

Vocabulary size was measured using the online Vocabulary Size Test (Nation, 2010), and on average, participants in this study knew 7,500 word families (M = 7496.88; SD = 1379.92). This score is significantly lower than the 20,000 word families that L1 speakers of English usually know (ibid), but was high enough to enable them to read the subtitles with ease. Finally, working memory capacity was measured using the 'Phonological working memory span test' (Winke, 2013), and on average participants scored 65% (M = 0.65; SD = 0.08) meaning they were able to remember up to 5 words while processing the grammatical accuracy and plausibility of sentences presented to them. Table 30 below summarises participants' results on these four tests, and the standard deviation scores suggest that their performances were very similar. The results of these baseline measures coupled with the participants'

demographic information and IELTS scores, indicate that they were a reasonably homogenous sample. A regression analysis was used later to determine whether these measures predicted participants' scores on the listening segmentation and comprehension tests.

Table 30. Participants' scores on the 'skills tests' in Experiment 5

N	*Vocabulary size*		*Reading rate*		*Reading comprehension*		*Working memory*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
32	7496.88	1379.92	97.13	11.99	8.84	1.89	0.65	0.08

6.5.2 Shadowing Task

6.5.2.1 Old Items

As was the case in previous experiments, 'Old items' refers to utterances that participants were exposed to during the treatment phase of the experiment, when they watched the documentaries. Figure 24 shows that the 'bimodal' group outperformed the 'no subtitles' group on both the immediate and delayed post-tests.

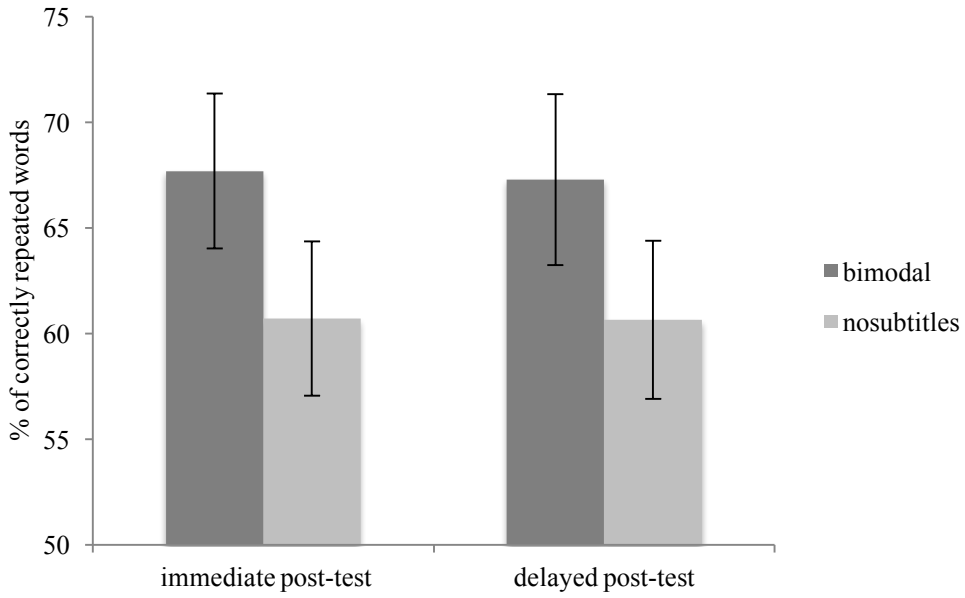

Figure 24. Percentage of correctly repeated words on 'old items' in the immediate post-tests and the delayed post-tests across groups in Experiment 5.

A mixed design ANOVA was conducted with 'group' as the between-subject factor and 'time' as the within-subject factor. There was a main effect of 'group', $F(1,30) = 7.88$, $p < 0.05$, with the 'bimodal' group outperforming the 'no subtitles' group. There was no main effect of 'time' ($F_1(1,30) = 0.12$, $p > 0.05$, $F_2(1,79) = 2.06$, $p > 0.05$). There was also a non-significant 'group' by 'time' interaction ($F_1(1,30) = 0.70$, $p > 0.05$, $F_2(1,79) = 0.08$, $p > 0.05$).

6.5.2.2 New Items

Figure 25 reveals that the 'bimodal' group numerically segmented more words than the 'no subtitles' group on both the immediate and delayed post-tests for 'new items', however this difference was not statically significant. Recall that these were utterances participants were not exposed to during the treatment phase of the experiment, but which were spoken by the same speakers they had heard, and presented after the treatment on immediate and delayed post-test.

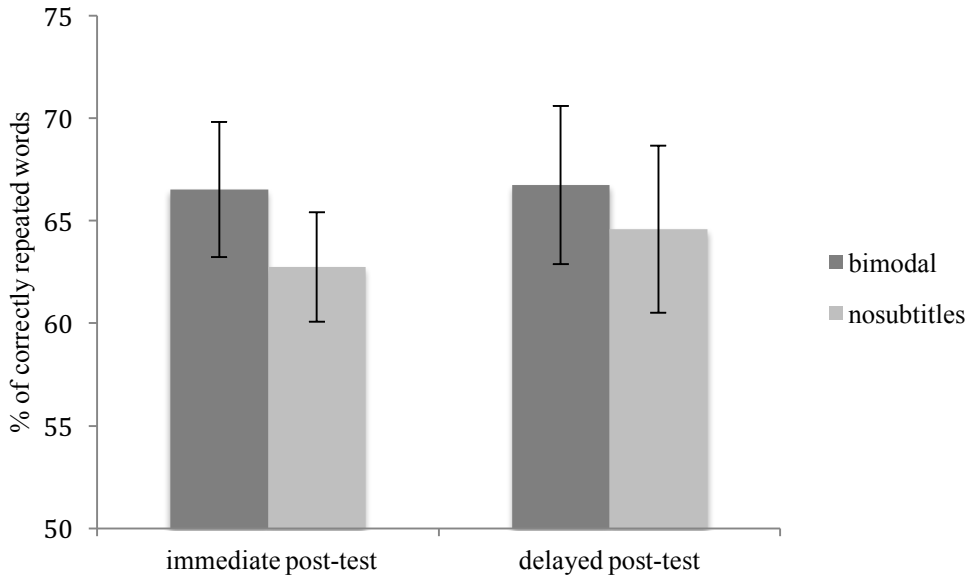

Figure 25. Percentage of correctly repeated words on 'new items' in the immediate post-tests and the delayed post-tests across groups in Experiment 5

Another mixed design ANOVA was conducted with 'group' as the between-subject factor and 'time' as the within-subject factor. There was no main effect of 'group' $F(1,30) = 1.92$, $p > 0.05$, nor a main effect of 'time' ($F_1(1,30) = 1.22$, $p > 0.05$,

$F_2(1,79) = 4.81$, $p < 0.05$). There was also a non-significant 'group' by 'time' interaction ($F_1(1,30) = 0.76$, $p > 0.05$, $F_2(1,79) = 25.32$, $p > 0.01$).

6.5.2.3 Brand new Items

'Brand new items' were unfamiliar utterances, spoken by unfamiliar speakers, which were not presented during the treatment phase of the experiment. As Figure 26 shows, the 'bimodal' group's performance improved from week 1 to week 23, whereas the 'no subtitles' group's scores remained constant throughout the 4-week period.

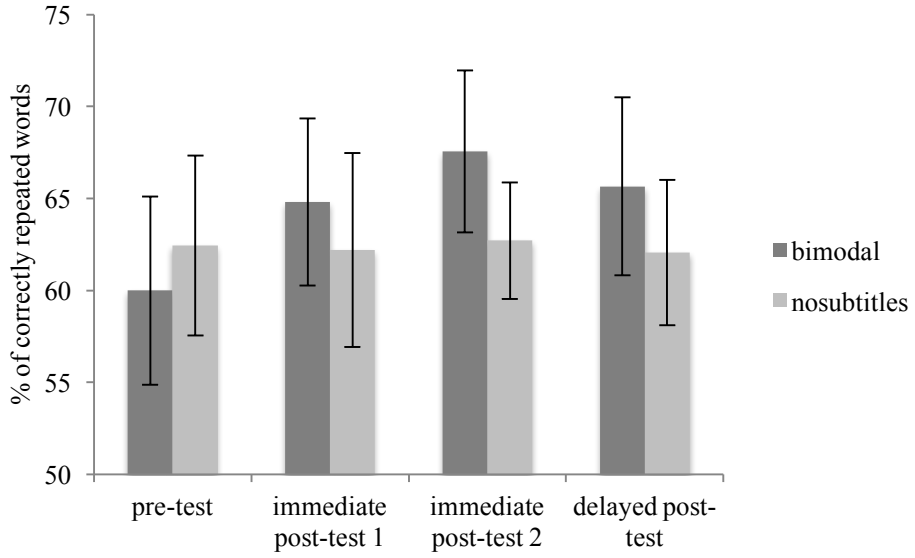

Figure 26. Percentage of correctly repeated words on 'brand new items' in the pre-test, immediate post-tests and the delayed post-tests across groups in Experiment 5

A mixed design ANOVA was conducted with 'group' as the between-subject factor and 'time' as the within-subject factor. There was no main effect of 'group', $F(1,30) = 0.71$, $p > 0.05$, but there was a main effect of 'time', that was marginally significant as determined with a by-item analysis ($F_1(3,90) = 2.59$, $p < 0.05$, $F_2(3,79) = 18.49$, $p < 0.01$). There was also a 'group' by 'time' interaction, which was marginally significant ($F_1(3,90) = 2.51$, $p = 0.06$, $F_2(1,79) = 18.49$, $p < 0.01$). Table 31 further highlights the differences between the groups' performances throughout the test. It shows that the 'bimodal' group's performance peaked at a 7.55% increase in their test scores at week 3. The 'no subtitles' group's highest increase, however, was 0.26%.

Table 31. Participants' weekly scores on 'brand new items' in Experiment 5, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

Group	*Week 2: Pre-test*	*Week 3: Post-test*	*Week 4: Post-test*	*Week 5: Delayed post-test*
bimodal	60.00	64.81 (+4.81%)	67.55 (+7.55%)	65.65 (+5.55%)
no subtitles	62.45	62.20 (-0.25%)	62.71 (+0.26%)	62.07 (-0.38%)

Overall, the findings are consistent with those of Experiment 4, in that the trend is the same, however the results on this occasion were not as strong or robust for the effects of subtitles on listening segmentation. Again, both groups were fairly evenly matched in the pre-test, but after exposure to the treatment materials, the bimodal input group outperformed the no subtitles group on 'old', 'new', and 'brand new' items on the immediate post-tests, and this was sustained at the delayed post-test phase. The next section presents the results of the listening comprehension test, which was the main aim of this study.

6.5.2.4 Regression analysis

Thus far, the analyses conducted indicate that the 'bimodal' group made significantly greater improvements throughout the experiment than the 'no subtitles' group on both the segmentation tasks. However, within the bimodal group itself, there are individual differences between participants that may contribute towards their performances. Table 32 presents a detailed account of each participant within the bimodal group. Initially, by looking at this table, it is interesting to note differences between the participants.

Take participant 2 for example, she has a very high IELTS listening score, the largest vocabulary size, her reading comprehension score is higher than the average native speaker of English, and she reads quite quickly. Despite having a low working memory capacity score, her speech segmentation score was higher than all other participants at the pre-test stage and this increased by over 3% by the delayed post-test stage. Conversely, participant 7 has a lower IELTS listening score, a smaller vocabulary size, a lower reading comprehension score, is a slower reader but has a higher working memory capacity. Her initial segmentation score on the pre-test was extremely low, but this increased by over 19% at the delayed post-test stage. This

raises a question - why did participant 7 benefit more than participant 2 on the segmentation task? Was it because her working memory is better, so she is able to process the utterances better? Or was it simply because participant 2 is already at a higher level of English and did not require much help from captions?

Table 32. Participants of the bimodal group's delayed post-test scores on 'brand new items' in Experiment 5, cross-referenced against their pre-test score to highlight increases or decreases in their overall performance.

P	*IELTS listen*	*Vocab size*	*Reading comp.*	*Reading rate*	*Memory*	*Pre-test*	*Delayed post-test*	*Gains*
1	6.5	8000	6	102	0.54	54.42	53.15	-1.27%
2	8.0	10100	11	109	0.46	81.85	84.38	+2.53%
3	7.0	8600	8	99	0.73	55.24	58.56	+3.32%
4	6.5	5700	6	98	0.70	75.09	76.94	+1.85%
5	7.5	6500	9	96	0.67	73.36	71.28	-2.08%
6	7.0	7600	7	109	0.55	52.71	57.79	+5.08%
7	6.0	7900	9	91	0.73	46.64	66.58	+19.94%
8	5.5	7100	9	68	0.63	58.60	62.91	+4.31%
9	7.0	6200	8	87	0.68	64.52	54.86	-9.66%
10	8.0	9100	11	103	0.47	59.68	77.28	+17.60%
11	7.0	8000	10	114	0.61	58.70	73.79	+15.09%
12	5.5	5500	8	99	0.65	49.76	58.93	+9.17%
13	5.5	5700	7	98	0.71	56.84	60.48	+3.64%
14	7.5	7100	10	77	0.75	53.09	59.33	+6.24%
15	6.0	7000	9	84	0.74	56.01	64.37	+8.36%
16	8.5	10000	13	97	0.64	63.56	69.82	+6.26%
M	6.81	7506.25	8.81	95.69	0.64	60.00	65.65	+5.65%
SD	0.95	1441.97	1.91	12.02	0.09	9.61	9.08	7.46

Note. P=Participant, IELTS listen=IELTS listening scores, Vocab size=Score on Vocabulary Size Test, Reading comp.=Reading comprehension score, Reading rate=Reading rate in WPM, Memory=Score on working memory test.

In order to answer these questions, a multiple regression analysis was conducted to determine whether the baseline measures of 'working memory', 'vocabulary size', 'reading rate', and 'reading comprehension' could predict participants' scores / performance on the shadowing task.

Only participants in the 'bimodal' group were looked at because the aim was to determine which of them benefitted most from reading subtitles whilst listening to an oral narration. It was found that 'reading comprehension' and 'working memory' explain a significant amount of the variance in shadowing task scores for 'brand new' items ($F(2, 13) = 6.46$, $p < 0.05$, R^2 0.49, *adj.* R^2 0.422). The analysis showed that 'reading comprehension' significantly predicted shadowing task scores ($\beta = 0.94$, $t(13) = 3.49$, $p < 0.05$), whilst 'working memory' had a negatively significant prediction ($\beta = -0.81$, $t(13) = -3.02$, $p < 0.05$). 'Reading rate' ($\beta = 0.09$, $t(13) = 0.47$, $p > 0.05$) and 'vocabulary size' ($\beta = -0.13$, $t(13) = -0.46$, $p > 0.05$) were not predictive of participants' scores on the segmentation task (*see* Table 33).

Table 33. Hierarchical multiple regression predicting speech segmentation from vocabulary size, reading rate, reading comprehension, working memory capacity.

	Speech Segmentation					
	Model 1		*Model 2*		*Model 3*	
Variable	B	*B*	B	*β*	B	*β*
Constant	12.72		12.49		18.93	
Vocab	-0.00	-0.24				
Read rate	0.11	0.18	0.06	0.09		
Read comp	4.43*	1.13	3.65*	0.93	3.69*	0.94
Memory	-0.74*	-0.84	-0.70*	-0.79	-0.72*	-0.81
R^2	0.53		0.50		0.49	
F	3.13		4.13*		6.46*	
adj. R^2	0.36		0.38		0.42	

*Note. N=16. * p < .05*

These results suggest that variation in vocabulary size did not make much of a difference for these participants, and that the ability to read quickly did not necessarily help them to understand what they were reading. Rather, participants who were generally better able to comprehend what they read benefitted from subtitles the most. Interestingly, participants with the lowest working memory scores benefitted from the subtitles more than those with high working memory.

6.5.3 Listening Comprehension

The listening comprehension tests were composed of 30 questions, each worth 1-point. As Figure 27 shows, the 'no subtitles' group's scores remained constant throughout the experiment, whereas the 'bimodal' group's scores improved with each test. Recall that non-overlapping error bars (such as those in week 4 on the chart) indicate a significant difference between scores, suggesting that the 'bimodal' group's performance was markedly better by the end of the experiment, compared to the 'no subtitles' group.

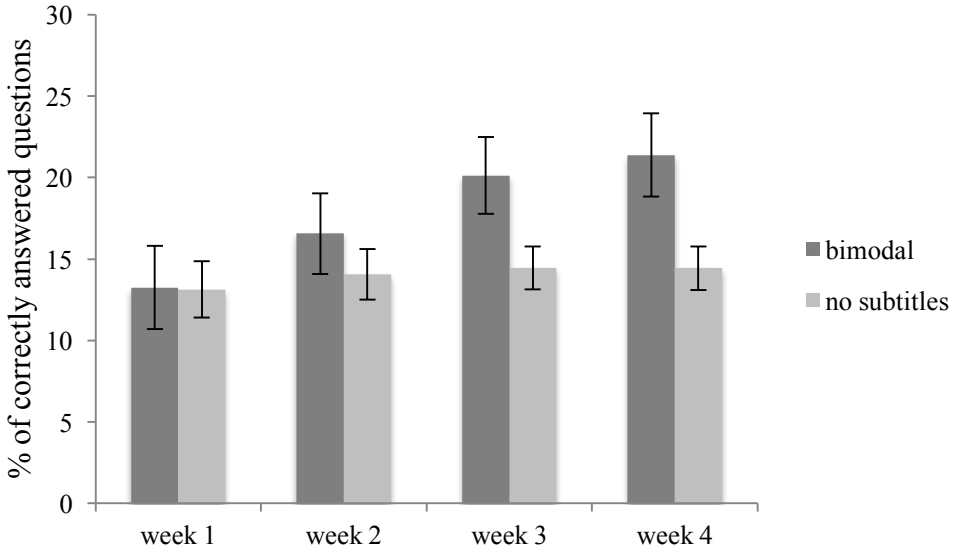

Figure 27. Listening comprehension test scores by group in Experiment 5.

A mixed design ANOVA was conducted with 'group' as the between-subject factor and 'time' as the within-subject factor. There was a main effect of 'group' ($F_1(1,30) = 8.82$, $p < 0.01$, $F_2(1,29) = 210.51$, $p < 0.01$), and a main effect of 'time' ($F_1(3,90) = 69.46$, $p < 0.01$, $F_2(3,87) = 9.34$, $p < 0.01$), showing improved performance over time, which, looking at the chart, was clearly driven by the 'subtitles' group. There was also a significant 'group' by 'time' interaction $F(3,90) = 36.43$, $p < 0.01$ ($F_1(3,90) = 36.43$, $p < 0.01$, $F_2(3,87) = 7.96$, $p < 0.01$), which shows that the groups did not improve equally throughout the experiment.

Pearson's correlations were conducted to assess the relationship between participants' performances on 'brand new' items of the shadowing task and listening

comprehension tests, as they were both measures of generalizability to new speakers (i.e. being able to segment the speech of a new speaker and the ability to comprehend the speech of new speakers). Preliminary analyses showed the relationship to be linear with both variables normally distributed, as assessed by a Shapiro-Wilk test ($p > 0.05$), and there were no outliers.

The results of these analyses revealed there was a moderate positive correlation between the shadowing task and the listening comprehension task at the pre-test stage, $r(30) = 0.38$, $p < 0.05$. There was a strong positive correlation between the shadowing task and the listening comprehension task at week 2, $r(30) = 0.68$, $p < 0.01$; and similarly at week 3, $r(30) = 0.63$, $p < 0.01$. Finally, there was a moderate positive correlation between the shadowing task and the listening comprehension test at the delayed post-test stage, $r(30) = 0.47$, $p < 0.05$. Effectively, at each time point, there was a significant correlation between the shadow task and the listening comprehension task, even though they were completely unrelated.

6.6 Discussion

Experiment 5 aimed to investigate whether repeated exposure to captioned video could improve the L2 listening segmentation and comprehension ability of international students in the UK. The experiment was conducted over a 5-week period; during the first week, participants were tested on four baseline measures, which included vocabulary knowledge, working memory capacity, and reading skills. This data was valuable, in addition to the demographic information collected via a questionnaire, for providing a detailed, holistic view of the 32 international students who participated in this study. In Experiments 1 through 4, only IELTS scores were recorded, and the problem is that these grades could be up to 2 years old, whereas the baseline measures in this experiment were current and all relevant to the study. From weeks 2 to 5, participants sat a pre-test, two immediate post-tests and finally a delayed post-test, with each test measuring both their listening segmentation and listening comprehension abilities.

Participants' performances on the listening segmentation tests were compared using a mixed design ANOVA, and the results indicated a group by time interaction. In week 1, the groups' scores were similar, but after exposure to treatment material, the

'bimodal' group achieved higher scores than the 'no subtitles' group in all subsequent tests. For 'old items', the 'bimodal' group performed better than the 'no subtitles' group, and there was a significant difference between their scores. However, although the 'bimodal' group scored higher than the 'no subtitles' group on 'new items', the group by time interaction was not significant. The group by time interaction for 'brand new items' was significant, however, specifically when comparing group scores on the pre-test and delayed post-test.

Essentially, the results of the listening segmentation test demonstrate the same trend that was seen in Study 1 and Study 2, except that on this occasion, the results were not all statistically significant. This could be a reflection of the sample size, because all three studies revealed similar findings, yet it was only Study 2, with the largest number of participants (48) that consistently yielded statistically significant results on all items. In any case, all three studies have provided evidence to support the claim that watching captioned video can lead to improvements in speech segmentation.

Having conducted this experiment twice already, it was no surprise that the bimodal group outperformed the no subtitles group. On this occasion, closer attention was paid to participants within the bimodal group itself. In an attempt to find out which participants benefitted most from watching captioned videos, a hierarchical regression analysis was run using their 'skills tests' scores taken in week 1. The findings of the regression analysis were interesting because they suggest that students who are better at reading comprehension benefitted more from the subtitles in captioned video than students who are weaker at understanding what they read. Returning to principles of the Dual Coding Theory, it could be hypothesized that the stronger readers are able to form associations cognitively between the spoken utterance (which is usually hard for them to understand) with the written sentence (which is easier for them to grasp).

Reading rate however, was not predictive, meaning someone may be able read the subtitles quickly but it does not necessarily help them. Similarly vocabulary size was not predictive, possibly because it is not definitions of words that students have problems with, it is actually listening to those words when spoken in real-time, which is a challenge for them. Surprisingly, working memory capacity was significantly negatively correlated with shadowing task scores, possibly indicating that students

with the poorest working memory are the ones that need the most help. Again, referring back to research on Dual Coding Theory, some learners process input better when it is received via dual modalities rather than one in isolation.

As for the listening comprehension tests (which were the focus of this experiment), Figure 27 clearly shows that the two groups performed similarly on the first test, but where the 'bimodal' group gradually improved their performance over the duration of the experiment, the 'no subtitles' group did not. A mixed ANOVA analysis revealed that there was a main effect of 'group', and there was a significant 'time' by 'group' interaction, indicating that the 'bimodal' group made statistically significant improvements over time.

This is a particularly interesting finding because of the comprehension test materials used. If the comprehension test materials had been composed of listening materials presented during the experiment, one could argue that participants who segmented speech well were then able to comprehend utterances better. However, there was no direct relationship between the materials used for the listening segmentation tests and the listening comprehension tests, because the focus in this experiment was on the generalisation of learning. In other words, can this bimodal input training improve a person's ability to listen to new speakers (of a similar accent) when talking about unfamiliar topics?

To investigate this, a Pearson correlation analysis was performed to compare participants' performances on 'brand new items' with their scores on the listening comprehension tests. The result of this analysis was statistically significant, and a positive correlation was found, which suggests that a participant who improves her ability to segment unfamiliar utterances spoken by unfamiliar speakers on the segmentation task is then able to comprehend spoken dialogue in a different context.

Ultimately, the results of Experiment 5 introduce new findings into the field of research that claims captioned video helps with L2 listening comprehension. Yet new questions arise from these findings that future studies can address. These will be discussed in the next chapter.

CHAPTER 7: DISCUSSION AND CONCLUSIONS

7.1 Review of findings

This research project was composed of three separate yet connected studies. Study 1 explored the L2 speech segmentation ability of international students in the UK, the findings of which led to the construction of Study 2, which focused on the long-term effects of watching captioned videos to improve L2 speech segmentation. Building upon the results of Study 2, Study 3 was conducted to investigate the long-term effects of watching captioned videos on L2 listening comprehension. This chapter summarises, reviews and discusses the findings of these three studies in the light of existing literature on the role of captioned video in L2 listening segmentation and comprehension.

7.1.1 Study 1

Study 1 was essentially a pilot study, as it was an initial explorative investigation into the L2 speech segmentation abilities of international students in the UK. It was made up of three small-scale experiments, aimed at answering five core research questions. This section will re-state those research questions and discuss the findings for each one.

RQ1: How difficult is L2 speech segmentation for international students?

Inspired by Mitterer and McQueen's (2009) investigation into the effects of captioned videos on L2 speech segmentation for Dutch learners of English, the first experiment aimed to replicate their shadowing task and use it to establish how challenging L2 speech segmentation is for international students living in the UK. The shadow task created was composed of sixty utterances from six DVDs, articulated by L1 speakers of British English. Participants were simply required to sit in a quiet room, listen to these sixty utterances, one at a time and repeat the words they heard. If the participants accurately repeated all of the words in an utterance, they received a full score for that utterance.

A British student acted as a control, sat the test and accurately repeated all utterances. A convenience sample of 10 international students from six different countries sat the test and they were only able to accurately repeat approximately 71.30% of the utterances. This score was quite startling, considering these students had spent an average of 9 years studying English in their home countries, had been living in the

UK for over a year, and had self-reported IELTS listening test scores of 6.5, which is a score good enough for studying in a British higher education institute. It was a startling score because if they were unable to accurately repeat approximately 30% of heard utterances whilst sitting in controlled conditions, how would they cope when sitting in a lecture theatre or any other real-world situation? Students like those who participated in this study were missing, potentially, a significant portion of what their teachers said during academic lectures.

A questionnaire was issued to the participants prior to sitting the shadowing task, which asked them to estimate how accurately they thought they would repeat excerpts from different programmes. Interestingly, the participants correctly predicted that they would only be able to repeat approximately 70% of the utterances heard, indicating self-awareness of their listening segmentation ability. Although the issue of self-awareness is beyond the scope of this thesis, it is an interesting topic in literature on the challenges that L2 international students face when studying abroad (*see* Liu, 2013), specifically on the difficulties they have with listening to L1 speakers of English (*see* Goh, 2000).

Another interesting finding was that participants performed differently on the shadowing task, depending upon the genre of the programme. Surprisingly, pairwise comparisons revealed that participants were better at segmenting utterances from the 'stand-up comedy' genre rather than the 'film' genre. It should be noted however, that although they were able to segment utterances from the stand-up comedy, the task was not designed to test comprehension, thus it is possible that participants were able to clearly hear and repeat words without understanding the meaning of those utterances. Additionally, since only ten utterances were used from each of the six genres, it is possible that these differences in participants' performances were due to differences in the speakers' pronunciation.

The results of this pilot experiment led to three important conclusions that influenced further experiments. First, the modified version of the shadowing task was successfully piloted, the average number of words in an utterance was suitable, the silent period of 3.5 times the utterances' duration was sufficient for participants to verbally repeat what they heard, and the use of mean scores, following Mitterer and

McQueen (2009), was an efficient scoring method. Secondly, rather than using utterances from multiple genres, it was decided that concentrating on one genre would be more appropriate. That genre would be 'documentary' because the monologue presentation style is similar to an academic lecture, and professionally produced documentaries with subtitles were widely available and accessible. Third, and most importantly, participants were incapable of segmenting around 30% of what they heard on this simple shadowing task, meaning that there is a need for them to improve in this area. This pilot study alone contributes to new knowledge in this field of research, because it was the first to identify this speech segmentation problem for international students in the UK, consequently, these findings were shared with the academic community via publication (*see* Charles and Trenkic, 2015).

Having established these findings, a second pilot was conducted in the form of Experiment 2, which aimed to determine whether bimodal input could help international students improve their speech segmentation ability. Previous research indicates that this may be possible (Mitterer and McQueen, 2009), but no previous study has been conducted over an extended period of time. Therefore, a key objective for this pilot study was to administer an experiment over a 4-week period, rather than over a couple of hours on a single day. Crucially, this pilot experiment attempted to address the following three research questions.

7.1.1.2 RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?

Recall that in Mitterer and McQueen (2009)'s experiment participants first watched 25-minutes of a TV show, then sat the shadowing task, which was composed of what they referred to as 'old items' and 'new items'. 'Old items' were utterances heard whilst watching the TV show that were then presented again in the shadow task, whereas 'new items' in the shadow task were utterances spoken by the same actors in the TV show, but to which participants had not been exposed during the 25-minute viewing. Mitterer and McQueen administered this experiment to four groups: a 'no subtitles' group, an 'English subtitles' group, a 'Dutch subtitles' group and a 'control' group. Participants in the 'English subtitles' group were essentially exposed to 'bimodal input' because they were reading English text whilst listening to English

speech, and their overall performance on ‘old items’ during the shadowing task was better than the other groups.

Experiment 2 modified this experimental design slightly by using three treatment groups, namely a ‘bimodal’ group who watched DVDs spoken in English whilst reading English subtitles, a ‘no subtitles’ group (the control group) who watched the same DVD without subtitles, and a ‘no sound’ group who watched the same DVD whilst reading subtitles but without listening to sound. A pre-test – treatment – post-test design was employed over a four-week period, with participants being exposed to the treatment materials during weeks two and three, and shadowing tasks being administered each week. Whereas Experiment 1 used only sixty utterances, this experiment required participants to respond to 100 utterances during the pre-test, 120 utterances in week 2 and again in week 3, and finally 160 utterances at the delayed post-test phase.

A British student (different to the previous one) acted as a control, sat the tests and accurately repeated all of the utterances; again, indicating that the shadowing task is straightforward for L1 speakers of English. A convenience sample of twelve international postgraduate students from China participated in Experiment 2 and their results are now discussed with reference to each of the three research questions.

Regarding RQ2, in weeks 2 and 3 participants watched 30 minutes of a documentary, and then sat a shadowing task, which was composed of utterances from the 30 minutes they were exposed to. Results showed that the bimodal group made larger gains on their ‘old items’ scores from the immediate post-test to the delayed post-test than the ‘no subtitles’ and ‘no sound’. Referring back to section 2.2.3.1 of chapter 2, Bird and Williams (2002) and Mitterer and McQueen (2009), this result would support their claims that repeated watching of captioned video may indeed lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances.

7.1.1.3 RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?

For Mitterer and McQueen (2009), the main focus of their research was the role bimodal input plays on helping participants to adapt their L2 speech perception when listening to utterances they had not yet heard from a speaker they were familiar with. In their study, participants in the bimodal group were better at segmenting 'new items' than the other groups, which led to the claim that:

> "The effects on the new items, however, are the key results. The adaptation effect shows that listeners were able to retune their perceptual categories to characteristics of the exposure speakers, leading to long-term changes in speech perception" (ibid, p.4).

Simply put, this claim suggests that when an L2 learner of English reads subtitles that match the speech they are listening to, her brain somehow adapts to this speaker and future segmentation of this speaker's speech becomes easier for her. No other research has replicated this experiment since Mitterer and McQueen's (2009) publication, so RQ3 aimed to test the reliability of their claim by also investigating how participants perform on 'new items' in a shadowing task. However, this experiment took their research a step further by testing participants on 'new items' multiple times.

In weeks 2 and 3, after participants watched 30 minutes of a documentary, they sat a shadowing task, which was composed of utterances they had not been exposed to during the viewings (i.e. 'new items'). The bimodal group's performance on these items from the immediate post-test to the delayed post-test was numerically (but not significantly) better than the 'no subtitles' and 'no sound' groups. Considering that only 12 students participated in this experiment, it is possible that lack of significant results was due to the small sample size. However the trend was certainly in a direction that supports the findings of Mitterer and McQueen (2009).

7.1.1.4 RQ4: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?

This research question has never been investigated before in the body of literature about the effects of watching captioned videos on L2 speech segmentation. Mitterer and McQueen's (2009) study demonstrated that bimodal input could be used to improve L2 segmentation abilities, when listening to previously encountered utterances and/or when listening to different utterances by the same speaker. Thus far, the preliminary findings of the pilot study provided support for that claim. Potentially, this could be good for L2 learners of English because watching subtitled TV programmes may help them to adapt to a particular speaker's voice, and segment their speech better the next time they listen to that speaker. However, a more useful function of captioned videos would be if they could actually help the L2 learner adapt to other speakers of a similar accent, because this would mean the learner could watch captioned videos as a form of training and then have an easier time segmenting the speech of native speakers.

To investigate this, the shadowing tasks were extended with utterances from a speaker the participants had not been exposed to during any phase of the experiment. Intriguingly, at the pre-test phase all three groups obtained similar scores, however, after exposure to the treatment materials in week 2, the 'bimodal' group outperformed the 'no subtitles' and 'no sound' groups in all subsequent shadowing tasks. This was a critical finding in this pilot study because it suggests that bimodal input can lead to the generalisation of learning, such that L2 learners can listen to different speakers of a broadly similar accent (i.e. standard British English) and be better at segmenting their speech. Again, this finding alone made a meaningful research contribution to field and was published (*see* Charles and Trenkic, 2015).

7.1.1.5 RQ5: How difficult is L2 speech segmentation for international students when listening to different speakers of a broadly similar accent?

As a pilot, Experiment 2 was successful as a trial run for conducting follow-up experiments using this experimental design. Also, the results themselves were insightful because they demonstrated the potential effects of watching captioned videos on L2 speech segmentation over a 4-week period. However, it was noted that

in week 3, every group scored much lower than they did in weeks 2 and 4. Their poor performance in week 3 indicated that perhaps some speakers were more challenging to listen to than others, despite speaking a broadly similar accent. To rule out that this was an effect of participants' deteriorating segmentation ability from one week to another, RQ5 was posed and Experiment 3 designed as an auxiliary study similar to Experiment 1, this time using six different documentaries presented by different speakers as independent variables, instead of different types of TV programme.

A British student acted as a control, sat the test and accurately repeated all utterances. A convenience sample of ten international students from three different countries sat the test and they were only able to accurately repeat 70.14% of the utterances. Even though different materials were used in Experiment 3, this result is almost identical to that of Experiment 1, wherein participants accurately repeated 71.30% of utterances, again providing support for the claim that international students in the UK are unable to segment approximately 30% of the words they hear.

Crucially, it was found that there was a significant difference in participants' performance on the shadowing task, depending on which documentary they had watched. This finding confirmed that regardless of genre, L2 listeners responded to the TV presenters differently. This difference may have been due to differences between speakers voices. Alternatively, it may have been because in some DVDs, the speaker (e.g. Dimbleby) spent a lot of time talking to the camera (as if he were looking at the listener), whereas in other DVDs, the speaker (e.g. Attenborough) was rarely seen, but mostly heard whilst describing imagery of animals. In either case, to minimise the effects of a speaker's voice on participants' performance, it was evident that counterbalancing the presentation of items in following experiments was essential.

Summary

Study 1 in its entirety was a first step toward building an empirically sound test for the long-term effects of watching captioned video on L2 speech segmentation. Five core research questions were posed and the preliminarily findings were as follows:

(a) “RQ1: How difficult is L2 speech segmentation for international students?” – The results of Experiments 1 and 3 appear to indicate that international students in the UK do face a problem with L2 speech segmentation. On average, they appear to miss approximately 30% of the words they hear under controlled laboratory conditions.

(b) “RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?” – The results of Experiment 2 showed that from the immediate post-tests to the delayed post-test, the ‘bimodal’ group made more of an improvement in their scores than the other groups, suggesting that yes, viewing subtitled programmes can indeed lead to long-term improvements in an L2 listener’s segmentation ability for previously encountered utterances.

(c) “RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?” – The results of Experiment 2 did not produce any statistically significant findings to support this claim. However, the overall trend of the data appeared to suggest that it is possible. It was decided that a larger sample size would need to be recruited in subsequent experiments to further investigate this research question.

(d) “RQ4: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?” – The results of Experiment 2 produced a statistically significant improvement in the ‘bimodal’ group’s performance on the shadowing task from pre-test to delayed post-test, suggesting that yes, viewing subtitled programmes may lead to improvements in L2 segmentation when listening to different utterances by different speakers of a broadly similar accent.

(e) “RQ5: How difficult is L2 speech segmentation for international students when listening to different speakers of a broadly similar accent?” – The results of Experiment 3 revealed that international students find it easier to segment the speech of some speakers than others, despite their accent being broadly similar. When listening to David Dimbleby, participants successfully managed to repeat almost 90%

of his spoken utterances, however, they could not repeat 65% of the utterances spoken by David Attenborough, which illuminates how varied their listening ability is for different speakers. Ultimately, although Study 1 was a pilot investigation, it produced new contributions to the field and was published (*see* Charles and Trenkic, 2015).

7.1.2 Study 2

Having conducted the three pilot studies in Study 1, and building upon them, Study 2 aimed to investigate the long-term effects of watching captioned videos on L2 speech segmentation when (a) listening to previously encountered utterances, (b) listening to different utterances by the same speaker, and (c) listening to different utterances by different speakers of a broadly similar accent. Again, this experiment was designed with three treatment groups, a 'bimodal' group, a 'no subtitles' group and a 'no sound' group; and again a pre-test – treatment – post-test design was employed over a four-week period. Treatment materials were extracted from BBC documentaries and counterbalanced in the shadowing tasks.

Nine L1 speakers of English sat the test in order to provide a baseline for understanding the L2 data, and they accurately repeated almost all utterances. A sample of 48 international students from China participated in this experiment, a very homogeneous group, with similar ages, IELTS scores, and time spent living in the UK. Their results will now be discussed according to each research question.

7.1.2.1 RQ1: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?

Based upon Mitterer and McQueen's (2009) original research, and the preliminary findings of Study 1, it was hypothesised that bimodal input would lead to improved L2 speech segmentation when listening to previously encountered utterances. As was the case in Study 1, this was tested by exposing participants to particular utterances during the DVD viewing, and then testing them on these utterances during the shadow task. The results yielded two interesting findings.

Firstly, when comparing the 'bimodal' group's performance on the shadow task with that of the 'no subtitles' and 'no sound' groups, there was no statistically significant

difference between the performance of the 'no subtitles' and 'no sound' groups. However, there were statistically significant differences between them and the 'bimodal group'. This is an important finding because it provides additional support for Dual Coding Theory (Paivio, 1986) and the Cognitive Theory of Multimedia Learning (Mayer and Moreno, 2003), because it supports the proposition that receiving one form of input, whether it is aural or orthographic, is not as beneficial as receiving bimodal input, i.e. the combination of both aural and orthographic data.

The second interesting finding was that the 'bimodal' group's performance on the delayed post-test was consistent with the immediate post-tests. This is noteworthy because 'old items' are utterances that participants were exposed to during the DVD viewings in weeks 2 and 3. When they sat the shadowing tasks immediately after watching the DVDs, these 'old items' may have been stored in short-term memory. However, the delayed post-test was in week 4, so it is unlikely they would have remembered these utterances, yet they were able to segment them equally well. This provides further support for Mitterer and McQueen's (2009) claim that bimodal input can lead to adaptation effects in the perceptual processing of speech.

7.1.2.2 RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?

Study 1 did not produce any statistically significant results to support this claim, however that may have been due to the small sample size of participants. Again, 'new items' were included in the shadowing task in order to investigate this research question; and these items were utterances not heard by participants during the DVD viewings but spoken by the same presenters. The results here were similar to those of 'old items' in that there was no significant difference between the 'no subtitles' and 'no sound' groups, but there were statistically significant differences between them and the 'bimodal' group. Again, this provides support for the claim that reading text whilst listening is better than only reading or only listening (Danan, 1992; Harji et al. 2010). Additionally, the 'bimodal' group exhibited sustained improvement in their speech segmentation because their performance on the shadowing task did not deteriorate at the delayed post-test phase.

7.1.2.3 RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?

This research question was the primary focal point of Study 2. Prior to conducting this research, Mitterer and McQueen (2009) had already demonstrated that watching captioned videos could help L2 learners improve their speech segmentation when listening to previously encountered utterances and/or when listening to different utterances by the same speaker. However, Study 2 is the first to investigate whether bimodal input improves L2 speech segmentation when listening to different utterances by different speakers of a broadly similar accent. Preliminary results from the pilot study were encouraging, however, this would be the first experiment to recruit enough participants to make a decent-sized sample.

The results of this experiment revealed two very interesting findings. First, focusing on differences in performance between the three groups, it is clear that at the pre-test phase, shadowing task scores for all three groups were evenly matched. However, the scores of the 'bimodal' group improved substantially after their first exposure to the treatment material, whereas the other two groups did not. Again, there was no significant difference between the 'no subtitles' and 'no sound' groups. There were, however, statistically significant differences between them and the 'bimodal' group. As was the case with 'old' and 'new' items, this finding suggests that bimodal input is more beneficial for L2 listeners than purely text-only or aural-only input (Huang and Eskey, 1999; Etemadi, 2012).

The second interesting result was that of the 'bimodal' group's progress throughout the experiment. This group's scores increased by a statistically significant 11% in the shadowing task between the pre-test and the first immediate post-test. Thereafter, their performance in week 3 and 4 was sustained, suggesting that a single exposure to bimodal input can have long-lasting effects on L2 speech segmentation ability (Price, 1983), essentially leading to the generalisation of learning (Bird and Williams, 2002), meaning that they are better able to segment never-before-heard speech, spoken by different speakers of a broadly similar accent.

Summary

Within the body of literature on the effects of bimodal input on L2 listening, Study 2 was the first to specifically focus on its effects on L2 speech segmentation over an extended period of time. Using the term 'long-term effects' may be not be the most appropriate description for a four-week experiment, but considering the resources available and the willingness of participants to return on a weekly basis (without missing a session), it could be argued that this was a successful study. Three core research questions were posed as follows:

(a) "RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?" – The results discussed above are certainly in support of the claim that viewing subtitled progammes can lead to long-term improvements in an L2 listeners' segmentation ability for previously encountered utterances.

(b) "RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?" – Again the results of Study 2 are absolutely in support of the claim that viewing subtitled progammes can lead to long-term improvements in an L2 listeners' segmentation ability to listen to different utterances spoken by a speaker they are familiar with.

(c) "RQ4: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?" – Crucially, this study provides evidence to suggest that yes, viewing subtitled programmes may lead to improvements in L2 segmentation when listening to different utterances by different speakers of a broadly similar accent.

7.1.3 Study 3

It is evident from Vanderplank's (2010; 2013) reviews of research on bimodal input that the vast majority of studies have focused on how it may be useful for L2 listening comprehension. As discussed in Chapter 2, most of these studies have found that bimodal input can have a positive effect on L2 listening comprehension, although it

has also been argued that many of these studies lacked test construct validity, as defined by Buck (2001). Study 1 and Study 2 investigated the effects of bimodal input on L2 speech segmentation, but segmenting speech does not necessarily indicate comprehension of speech. Consequently, Study 3 aimed to determine whether improved ability to segment continuous speech on a shadowing task translates into better L2 listening comprehension.

In Study 3, the experimental design and materials employed in Study 2 were used again, with the addition of two extra features, which were (a) listening comprehension tests, and (b) baseline proficiency measures. The 'no sound' group was not included in this experiment because it was evident from Study 2 that just reading text did not benefit the participants in any way. A sample of 32 international students from China participated in this experiment. Again they were a homogeneous group, with similar ages, IELTS scores, and time spent living in the UK. Their results will now be discussed according to each research question.

7.1.3.1 RQ1: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered utterances?

Although this research question had already been posed and answered in Study 2, it was tested again for the purpose of investigating the reliability of the experiment. The results of Study 3 were actually very consistent with the findings of Study 2. The 'bimodal' group outperformed the 'no subtitles' group, and the 'bimodal' group's delayed post-test scores did not deteriorate from the immediate post-tests, which suggests sustained improvement in their segmentation ability.

7.1.3.2 RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?

Regarding 'new items', on this occasion the results of the mixed-design ANOVA were non-significant. However, the overall trend was similar, in that the 'bimodal' group scores were higher than the 'no subtitles' group, and they showed consistent scores at the immediate post-test and delayed post-test phases.

7.1.3.3 RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?

Again, the results here were similar to Study 2, in that both groups were evenly matched at the pre-test phase, with the scores of the 'bimodal' group improving substantially, and remaining consistently higher than the scores of the 'no subtitles' group. Ultimately, both groups were fairly similar at the pre-test stage, but after exposure to the treatment materials, the bimodal input group outperformed the no subtitles group on 'old', 'new', and 'brand new' items in the immediate post-tests, and this was sustained at the delayed post-test phase. Fundamentally, these results suggest that bimodal input can affect L2 speech segmentation when (a) listening to previously encountered utterances, (b) when listening to different utterances by the same speaker, and (c) when listening to different utterances by different speakers of a broadly similar accent.

7.1.3.4 RQ4: Do reading proficiency, vocabulary size and working memory play a role in the usefulness of subtitles for L2 speech segmentation?

In Study 1 and Study 2, participants were asked to report their IELTS listening scores and the time they had spent learning English. Although these were useful indications of their homogeneity, pre-tests were also conducted in order to establish a baseline measure of their L2 speech segmentation ability. The findings of these studies were shared with the academic community at research conferences as well as in a peer-reviewed book chapter (*see* Charles and Trenkic, 2015). Feedback received from the community led to the formulation of RQ5, as a result of questions such as "if bimodal input requires participants to listen and read simultaneously, why does the pre-test only measure listening ability? Why is reading ability not also tested?"; and "do participants with large vocabulary sizes or working memory capacity benefit from subtitles more than other participants?".

These questions raised valid points. Cconsequently, it was decided that prior to the pre-test phase, participants should be measured on their reading proficiency because differences in their reading skills might affect their ability to read subtitles adequately whilst listening to a speaker. Not only might faster readers benefit more than slower readers, but participants with better reading comprehension might also gain more

benefit from subtitles. Thus, all participants were tested on both English comprehension and reading rate. It was also possible that participants with greater knowledge of vocabulary would benefit more from the subtitles, and those with larger working memory capacities would process the bimodal input more easily.

Participants were tested on these skills during week 1, and after the 5-week experiment was completed, regression analyses were conducted specifically on the 'bimodal' group to determine whether any of these variables could be used as a predictor for individual differences in performance on the shadowing and comprehension tasks. The results of these analyses revealed that 'reading comprehension' significantly predicted shadowing task scores, which means that participants who were better at understanding the subtitles, were benefitting more from bimodal input than participants who could not understand the subtitles so well.

Reading rate was not predictive of participants' scores on the segmentation task. This suggests that being able to read the subtitles quickly does not necessarily result in better segmentation. Similarly, vocabulary size was not predictive of participants' scores on the segmentation task, which indicates that having a large vocabulary will not necessarily lead to better speech segmentation. Intriguingly, working memory capacity was a negatively significant predictor of performance on the shadowing task, suggesting that participants with weak working memory were in more need of bimodal input as additional support compared to participants with larger working memory spans.

7.1.3.5 RQ5: Does repeated watching of subtitled programmes lead to long-term improvement in L2 listening comprehension?

This research question was the focal point of Study 3. To test listening comprehension, the 'Cambridge English: Advanced (CAE)' exam was adopted. The listening sections of four past exam papers were used, and each week after sitting the shadowing task participants sat the listening comprehension test. During this listening comprehension test, participants heard short audio recordings of monologues (such as lectures) and interactive speakers (such as interviews). The questions were then based upon their understanding of these recordings. This is important to note because the

shadowing task was created using both 'old items' and 'new items' from the treatment material (documentaries), but 'brand new items' came from a completely different source in order to measure the generalisability of learning. Similarly, the listening comprehension task aimed to test the generalisability of learning (i.e. the ability to comprehend the speech of new speakers), so there was no relationship between the treatment materials and the comprehension tests.

The results yielded two interesting findings. First, the performance of the 'no subtitles' group was constant throughout the experiment. As with the shadowing task, the groups were evenly matched at the pre-test phase, but, after exposure to the initial treatment materials, the 'bimodal' group outperformed the 'no subtitles' group on both the immediate post-tests and the delayed post-test. This suggests that watching TV programmes without the additional help of same-language subtitles may not lead to enhanced L2 listening comprehension (Yoshino et al. 2000; Latifi et al. 2011).

The second interesting finding came from within the 'bimodal' group, because members of the 'bimodal' group made significant improvements in their week-by-week scores throughout the experiment. This finding supports the claim that bimodal input can indeed lead to long-term improvements in L2 listening comprehension (Price, 1983; Bird and Williams, 2002; Zareian et al. 2015). Furthermore, the results of a Pearson's correlation show a direct relationship between shadowing task scores and listening comprehension task scores. Although the utterances used in these two tasks were not related, this result suggests that better L2 segmentation ability leads to better L2 comprehension ability.

Summary

The experiment conducted in Study 3 was by far the most well designed experiment in this research. Reviewing the published literature on bimodal input, it is the only Study that investigates the effects of bimodal input on both L2 speech segmentation and L2 listening comprehension over an extended period of time. Five central research questions were posed and answered as follows:

(a) "RQ1: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to previously encountered

utterances?" – Yes, the statistically significant results of this experiment demonstrated that bimodal input can lead to long-term improvements in an L2 listener's segmentation ability for previously encountered utterances.

(b) "RQ2: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by the same speaker?" – Although results were not statistically significant, the overall trend of the data indicated that bimodal input may lead to long-term improvements in an L2 listener's segmentation ability when listening to different utterances by a speaker they are familiar with.

(c) "RQ3: Does repeated watching of subtitled programmes lead to long-term improvement in L2 segmentation abilities when listening to different utterances by different speakers of a broadly similar accent?" – Yes, the statistically significant results of this experiment showed that bimodal input can lead to improvements in L2 segmentation when listening to different utterances by different speakers of a broadly similar accent.

(d) RQ4: Do reading proficiency, vocabulary size and working memory play a role in the usefulness of subtitles for L2 speech segmentation and comprehension? – The results of this study suggest that vocabulary size and reading rate are not particularly crucial, but that L2 learners who have good reading comprehension ability appear to benefit the most from bimodal input training. It also appears that bimodal input is useful for L2 learners with a limited working memory.

(e) RQ5: Does repeated watching of subtitled programmes lead to long-term improvement in L2 listening comprehension? – The results of this experiment provided additional support to the body of research which proposes that bimodal input may lead to long-term improvements in L2 listening comprehension. Participants in the 'bimodal' group made weekly improvements in their ability to answer the listening comprehension test questions, and there was a significant difference between their pre-test and delayed post-test scores.

7.1.4 Summary

In sum, the three studies that make up this research investigated the L2 speech segmentation abilities of international students in the UK, the extent to which captioned videos could help improve their L2 speech segmentation abilities, and what effects bimodal input may have on their L2 listening comprehension. The overall design and implementation of Study 1 can be seen in Table 34 below.

Table 34. Summary of experimental design for study 1

Study	*Exper-iment*	*Subjects*	*Duration*	*Groups*	*Repeated Measures Design*	*Treatment Materials / Test*
1	1	10 mixed L1 students	1 day	N/A	N/A	Treatment materials: N/A Test: shadow task, 60 utterances from 6 different genres.
	2	12 Chinese students	4 weeks	No subtitles No sound Bimodal	W1: pre-test Week 2: treatment material + immediate post-test Week 3: treatment material + immediate post-test Week 4: delayed post-test	Treatment materials: 2 documentaries Test: 4 shadow tasks composed of 'old', 'new' & 'unrelated' items. 500 utterances in total, average of 5 words per utterance.
	3	10 mixed L1 students	1 day	N/A	N/A	Treatment materials: N/A Test: Shadow task, 60 utterances from 1 genre.

The overall design and implementation of Studies 2 and 3 can be seen in Table 35.

Table 35. Summary of experimental designs for studies 2 & 3

Study	*Exper-iment*	*Subjects*	*Duration*	*Groups*	*Repeated Measures Design*	*Treatment Materials / Test*
2	4	48 Chinese students	4 weeks	No subtitles No sound Bimodal	Week 1: pre-test Week 2: treatment material + immediate post-test Week 3: treatment material + immediate post-test Week 4: delayed post-test	Treatment materials: 2 counterbalanced documentaries Test: 4 counterbalanced shadow tasks composed of 'old', 'new' & 'brand new' items. 640 utterances in total, average of 5 words per utterance.
3	5	32 Chinese students	5 weeks	No subtitles Bimodal	Week 1: baseline measures Week 2: pre-test Week 3: treatment material + immediate post-test Week 4: treatment material + immediate post-test Week 5: delayed post-test	Treatment materials: 2 counterbalanced documentaries Test (a): Nation's (2010) Vocabulary Size Test; Warmington et al.'s (2013) Reading Rate and Reading Comprehension Test; Winke's (2013) Phonological Working Memory Span Test. Test (b): 4 counterbalanced shadow tasks composed of 'old', 'new' & 'brand new' items. 640 utterances in total, average of 5 words per utterance. Test (c): 4 Listening comprehension tests, Cambridge English Advanced exams.

Six key conclusions were drawn from this research:

(a) International students who have lived in the UK for less than 24-months are unable to segment approximately 30% of the speech they hear in controlled conditions.

(b) Bimodal input (in the form of captioned videos) can help these international students improve their L2 speech segmentation ability when listening to previously encountered utterances (*see* Table 36).

Table 36. Summary of findings for shadow tasks on 'old items' across studies 1, 2 and 3

Study	*Experiment*	*Groups*	*Old Items*
1	1	N/A	N/A
	2	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between groups.
	3	N/A	N/A
2	4	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between 'no subtitles' vs. 'no sound' groups. Statistically significant differences between 'bimodal' vs. 'no subtitles' groups and 'bimodal' vs. 'no sound' groups.
3	5	(1) No subtitles (2) Bimodal	Statistically significant difference between 'bimodal' vs. 'no subtitles' groups.

(c) Bimodal input (in the form of captioned videos) can help international students improve their L2 speech segmentation ability when listening to previously unheard utterances by a speaker they are familiar with (*see* Table 37).

(d) Bimodal input (in the form of captioned videos) can help international students improve their L2 speech segmentation ability when listening to previously unheard utterances spoken by any new speaker of a broadly similar accent (*see* Table 38).

Table 37. Summary of findings for shadow tasks on 'new items' across studies 1, 2 and 3

Study	*Experiment*	*Groups*	*New Items*
1	1	N/A	N/A
	2	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between groups.
	3	N/A	N/A
2	4	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between 'no subtitles' vs. 'no sound' groups. Statistically significant differences between 'bimodal' vs. 'no subtitles' groups and 'bimodal' vs. 'no sound' groups.
3	5	(1) No subtitles (2) Bimodal	No statistically significant difference between 'bimodal' vs. 'no subtitles' groups.

Table 38. Summary of findings for shadow tasks on 'brand new items' across studies 1, 2 and 3

Study	*Experiment*	*Groups*	*Brand New Items*
1	1	N/A	N/A
	2	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between groups. Statistically significant difference between the 'bimodal' group's pre-test and delayed post-test scores.
	3	N/A	N/A
2	4	(1) No subtitles (2) No sound (3) Bimodal	No statistically significant differences between 'no subtitles' vs. 'no sound' groups. Statistically significant differences between 'bimodal' vs. 'no subtitles' groups and 'bimodal' vs. 'no sound' groups. Statistically significant difference between the 'bimodal' groups pre-test and delayed post-test scores.
3	5	(1) No subtitles (2) Bimodal	Statistically significant difference between the 'bimodal' groups pre-test and delayed post-test scores.

(e) In order to fully benefit from captioned videos, international students should have good reading comprehension skills, which will enable them to better understand the subtitled programmes.

(f) There is a clear relationship between L2 speech segmentation ability and L2 listening comprehension ability. There appears to be a transitive affect of watching captioned videos on L2 listening, whereby it helps international students improve their L2 segmentation and general L2 listening comprehension ability .

7.2 Original contribution of this research

The present study has made a number of significant contributions to the field of research on bimodal input, which can be categorised under its effects on L2 speech segmentation and L2 comprehension.

7.2.1 Effects of bimodal input on L2 segmentation

To begin with, this is the first study that has used a shadow task under controlled conditions to categorically investigate the L2 speech segmentation abilities of international students studying at universities in the UK. The results indicate that these students miss 30% of what they hear. Secondly, this is the first study since that of Mitterer and McQueen in 2009 that has attempted to replicate their investigation into the effects of captioned video on both 'old' items (previously encountered utterances) and 'new' items (new utterances from the same speaker). Furthermore, their experiment was conducted over only one day, and yet they claimed that bimodal input led to "long-term changes in speech perception" (ibid, p.4). This study tested participants on 'old' and 'new' items multiple times over a 4- or 5-week period to determine whether or not a single exposure to captioned video does indeed lead to long-term changes.

Moreover, this is the first study to build upon Mitterer and McQueen's research by adding a third variable, namely 'brand new' items - previously unheard utterances spoken by any new speaker of a broadly similar accent. This is especially important because 'old' and 'new' items are only relevant to a single speaker's voice, which is not particularly useful in the grand scheme of things. The addition of 'brand new' items has demonstrated that bimodal input can lead to generalised learning, meaning

that L2 listeners can use it as a form of self-training, which will help them understand many speakers with that accent. Finally, this is the first study that has measured the combination of reading rate, reading comprehension, vocabulary size, and working memory capacity all in one experiment, to determine whether these skills could predict an effect of bimodal input on L2 speech segmentation.

7.2.2 Effects of bimodal input on L2 comprehension

As highlighted by Vanderplank (2016), there have been around 130 research studies that investigated the effects of bimodal input on L2 language learning, in a wide-variety of contexts. The present study adds to that body of literature by providing further support for the claim that watching captioned videos leads to long-term improvements in L2 listening comprehension. The unique features of this research are that (a) it is the first to look at the relationship between L2 speech segmentation and L2 comprehension, and (b) it is also the first to use the Cambridge Advanced English exam as measure of L2 listening comprehension over a 4-week period.

Mitterer and McQueen (2009) are oft-cited throughout this discussion because no other study has followed their methodological approach, therefore, it is not possible to compare results with other authors/publications to the same extent. However, the findings that are specific to L2 listening comprehension provide support to Markham (1989), Markham et al. (2001), and Hayati and Mohmedi (2011) because although their research had methodological limitations, their findings appear to be accurate.

7.3 Pedagogical implications

Recent data from the Higher Education Statistics Agency indicate there are over 2.5 million international students enrolled on degree courses at universities throughout the UK (Crawford and Wang, 2014). These universities only admit international students who are able to demonstrate that they are proficient users of English, typically by passing a language test. Despite possessing a good grasp of the English language, both lecturers and students identify poor L2 listening comprehension ability as a major challenge for international students (Yen and Kuzma, 2009). This is particularly pertinent for some students newly arrived in the UK, who struggle to understand even half of what lecturers are saying (Liu, 2013), possibly because they are used to teachers in their home countries speaking English at a slow and clear pace,

unlike the way that L1 speakers converse naturally (Brown, 1977). Essentially, they find it difficult to segment the speech of British lecturers who speak at a 'normal' pace.

The findings of this study may have practical pedagogical implications for these L2 language learners and their teachers. Specifically, most universities in the UK offer intensive English language courses during the summer, aimed at preparing international students for either undergraduate or postgraduate study. These are known as pre-sessional English for Academic Purposes (EAP) courses, and run for periods of between 4 and 10 weeks. A fundamental course-learning outcome for these programs is improvement in students' L2 listening comprehension, specifically for academic lectures. Potentially, EAP teachers could integrate bimodal input theory into their courses by requiring students to watch subtitled documentaries, online lectures, or TV programmes spoken by English presenters. By doing this for 4 weeks, these students may improve their overall L2 listening skills.

7.4 Limitations and future research

It is possible to identify a number of limitations with the present study. Firstly, the sample sizes in these experiments never exceeded 50, so it would have been preferable to recruit more student participants. Secondly, although 4- and 5-week experiments are longer than most previous research in this field, labeling this study a 'long-term' investigation is not entirely accurate; following students over a 3- to 6-month period may have been more insightful. Additionally, since the abilities of international students were a primary focus in this study, it would have been more appropriate to use subtitled academic lectures rather than subtitled documentaries.

Considering the influence an L1 has on L2 language learning, it would be interesting to see future research recruit samples of students from a different L1 background, such as Arabic, to see if similar results are found. Furthermore, as the best readers in this study benefited the most from subtitled input, it would also be interesting to see how lower-level students perform in this experiment – to what extent does bimodal input help them? Finally, the participants in this study were all in their 20s, and it would be interesting to know if younger and/or older language learners benefit from bimodal input in the same way.

7.5 Conclusion

The main contribution of this study is that it is the first to specifically investigate the ability of participants to segment the speech of (a) previously encountered utterances; (b) different utterances by the same speaker; and (c) different utterances by different speakers of a similar accent. A pre-test / intervention / post-test experimental design was performed multiple times on international university students in the UK. Participants who watched captioned video during the treatment phase, consistently outperformed control groups, which suggests that L2 learners of English can improve their L2 listening skills (and more specifically, their L2 speech segmentation ability) by simply watching same-language subtitled TV programmes on a regular basis.

APPENDICES

Appendix A – Consent Forms

A.1 Consent form used in Experiments 1 - 3

Mr. Tendai Charles
PhD Researcher
Langwith College
Department of
Education
University of York
YO10 5DD

02/01/2012

RE: Information Letter / Informed Consent Form

Dear Sir/Madam,

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

Provisional Project Title
The long-term effects of intra-lingual subtitles on listening comprehension

Purpose of the study
The overall aim of this study is to investigate whether or not continual exposure to intra-lingual subtitles (i.e. subtitles in the same language as the spoken audio) can have long-term benefits on a person's listening ability to a foreign language.

Why you have been invited to participate
You have been selected to participate in this research project, as you are an international student at a British university.

Do I have to take part?
Taking part in the research is entirely voluntary. *If you do decide to take part you will be given this information letter to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason*.

What will happen to me if I take part?
As this particular experiment is a 'pre-pilot test', you will simply be asked to complete a short questionnaire and then be tested on your ability to identify words in sentences spoken by British TV personalities. The study will take no more than 15 minutes to complete

Contact information
If you would like to contact the researcher after this study for any reason, please email: tjc513@york.ac.uk

Consent Form

Provisional Project Title: The long-term effects of intra-lingual subtitles on listening comprehension

Name of Researcher: Tendai Charles

Participant Identification Number: _____

	Please initial box
I confirm that I have read and understand the information letter dated 02/01/2012 for the above project and have had the opportunity to ask questions.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.	☐
I understand that my responses will be anonymised before analysis. I give permission to the research to use my responses for anything he needs.	☐
I agree to take part in the above research project; and confirm that I have received a copy of this consent form and the information letter	☐

_____	_____	_____
Name of Participant	Date	Signature

_____	_____	_____
Lead Researcher	Date	Signature

To be signed and dated in presence of the participant

A.2 Consent form used in Experiment 4

Informed Consent Form

Experiment:	DVDs and Listening Comprehension
Lead researcher:	Tendai Charles (PhD Candidate), University of York
Supervisor:	Dr. Danijela Trenkic

The aims and procedures of the experiment

The aim of this study is to investigate how accurately international students listen to English DVDs.

The study will be conducted across four sessions, each lasting approximately 1 hour. In each session, you will be asked to take a listening test, and in 2 sessions you will watch a DVD. Each session will be audio-recorded, and you will be asked to complete a questionnaire.

An experimenter will be present at each session to help you. At the end of the fourth session, you will receive a payment of £10 for your participation.

You are free to stop your participation at any point.

The data from this study will be securely stored by the lead researcher for approximately 3 years, and it will be completely destroyed when he completes his studies. Only the researchers involved with have access to this data. Your identity will not be revealed in any presentation or published version of this research.

Consent

I am happy to participate in this research.

I have been informed of the aims and procedures involved in this study, and given the opportunity to ask questions for further clarification.

I have been informed of what will happen to the information I provide.

I reserve the right to stop my participation at any point.

_____	_____	_____
Signature of participant	*Name of participant*	*Date*
_____	_____	_____
Researcher's signature	*Name of researcher*	*Date*

A.3 Consent form used in Experiment 5

Informed Consent Form

Experiment:	DVDs and Listening Comprehension
Lead researcher:	Tendai Charles (PhD Candidate), University of York
Supervisor:	Dr. Danijela Trenkic

The aims and procedures of the experiment

The aim of this study is to investigate how accurately international students listen to English DVDs.

The study will be conducted across five sessions, totalling 436 minutes (approximately 7 hours). In each session, you will be asked to take a listening test, and in 2 sessions you will watch a DVD. Each session will be audio-recorded, and you will be asked to complete a questionnaire.

An experimenter will be present at each session to help you. At the end of the fifth session, you will receive a payment of £10 for your participation.

You are free to stop your participation at any point.

The data from this study will be securely stored by the lead researcher for approximately 3 years, and it will be completely destroyed when he completes his studies. Only the researchers involved with have access to this data. You identity will not be revealed in any presentation or published version of this research.

Consent

I am happy to participate in this research.

I have been informed of the aims and procedures involved in this study, and given the opportunity to ask questions for further clarification.

I have been informed of what will happen to the information I provide.

I reserve the right to stop my participation at any point.

_____	_____	_____
Signature of participant	*Name of participant*	*Date*
_____	_Tendai Charles_____	*____/_03_/ 2014*
Researcher's signature	*Name of researcher*	*Date*

Appendix B – Questionnaires

B.1 Questionnaire used in experiment 1

Pre-Test Questionnaire

1) How old are you? _____

2) What is your nationality? _____

3) What is your native language? _____

4) What are you studying? _____

5) Where do you study? _____

6) What was your listening score on the IELTS test? _____

7) How long have you been living in the UK? _____

8) How long have you studied English for? _____

9) Please indicate which of the following you find is easiest to listen to for you:

a) Documentary b) Film c) Lecture d) News e) Sitcom f) Stand-up Comedy

10) Please indicate which of the following you find is the most difficult to listen to for you:

a) Documentary b) Film c) Lecture d) News e) Sitcom f) Stand-up Comedy

11) How much do you usually understand when you listen to one of the following:

a) Documentary _____%
b) Film _____%
c) Lecture _____%
d) News _____%
e) Sitcom _____%
f) Stand-up Comedy _____%

12) Please indicate whether or not you have previously seen the following:

a) Meerkats	Yes	No
b) The Duchess	Yes	No
c) A lecture by Peter Millican	Yes	No
d) BBC News about the London riots	Yes	No
e) Blackadder	Yes	No
f) Russell Brand in NYC	Yes	No

B.2 Questionnaire used in experiments 2, 3 and 4

Language Background Questionnaire

1) How old are you? _____

2) What is your gender? Male Female (please circle)

3) What is your native language? _____

4) What are you studying? _____

5) How long have you lived in an English-speaking country? _____

6) For how long have you studied English in an English speaking country? _____

7) For how long have you studied English in your life-time? _____

8) What was your listening score on the IELTS test? _____

9) What was your speaking score on the IELTS test? _____

10) What was your overall score on the IELTS test? _____

11) On a scale of 1-10 (1 meaning 'poor' and 10 meaning 'excellent'), how would you rate your English proficiency in (please circle):

a. Speaking	1	2	3	4	5	6	7	8	9	10
b. Writing	1	2	3	4	5	6	7	8	9	10
c. Listening	1	2	3	4	5	6	7	8	9	10
d. Reading	1	2	3	4	5	6	7	8	9	10

B.3 Questionnaire used in experiment 5

Language Background Questionnaire

1) How old are you? _____

2) What is your gender? Male Female (please circle)

3) What is your native language? _____

4) What are you studying? _____

5) How long have you lived in an English-speaking country? _____

6) For how long have you studied English in an English speaking country? _____

7) For how long have you studied English in your life-time? _____

8) What was your listening score on the IELTS test? _____

9) What was your speaking score on the IELTS test? _____

10) What was your reading score on the IELTS test? _____

11) What was your overall score on the IELTS test? _____

12) On a scale of 1-10 (1 meaning 'poor' and 10 meaning 'excellent'), how would you rate your English proficiency in (please circle):

a. Speaking	1	2	3	4	5	6	7	8	9	10
b. Writing	1	2	3	4	5	6	7	8	9	10
c. Listening	1	2	3	4	5	6	7	8	9	10
c. Reading	1	2	3	4	5	6	7	8	9	10

Appendix C - Segmentation Tests

C.1 List of utterances used in Experiment 1

1 Scans his family's patch
2 the neighbour's nerves break
3 share it with many neighbours
4 but trouble is brewing
5 he's also one of the most diligent guards
6 this group of meerkats
7 although members of the gang may appear identical
8 he reinforces his status by drenching the ground with scent
9 white storks flying down from the North
10 each individual has his or her own special role
11 and what do you suggest
12 well done Mr Gray
13 but it appears that my horse has won
14 a wealth of depth and sentiment
15 I thought he'd be like papa
16 and he never talks to me
17 I must apologise mama were we making too much noise
18 I've only met him twice
19 I know I should've
20 the Duke of Devonshire
21 it's made of ordinary stuff
22 it just continues in a uniform state
23 Galileo wants to say the moon is made of rock
24 left to itself it's inert
25 and why do the planets orbit the sun
26 now explain why the moon orbits around the Earth
27 you'll be reading Descartes meditations
28 well he attacks the Aristotelian tradition
29 he famously comes up with this claim I think therefore I am
30 I clearly and distinctly perceive that it's true
31 his shooting is being investigated by the independent police complaints division
32 our home affairs correspondent June Kelly
33 the scene of the rioting in 1985
34 from Broadwater Farm
35 a police office Keith Blakelock was killed
36 and now the rest of the day's news
37 john thank you John Brain
38 have not yet been disclosed
39 the two other soldiers were killed in the south of the country
40 and steady the markets before they open
41 just use your imagination for heaven's sake
42 I can't believe I've been so stupid
43 I think the obvious point is this
44 possible drawing of German defences from you imagination
45 have you ever visited the planet Earth
46 one thing puzzles me Baldrick
47 I thought I'd get my headstone done
48 get a chisel and some marble
49 here lies Edmond Blackadder
50 unless I can think of some brilliant plan
51 what's the tipping point
52 take murder a bit more seriously
53 hold on to your hats now kids
54 not if the death threats are anything to go by
55 when I emphasise it with volume
56 this is my favourite bit
57 I don't think that the British government's really good
58 I don’t care about the Queen either
59 what I meant to say
60 it's a marketing technique

Text Inspector Profile © for this text is CEFR upper B2

C.2 List of utterances used in Experiment 2

1 I'm going north
2 in the heart of East Anglia
3 tightly packed trees
4 to reach the light
5 perfect for construction
6 the most popular building material
7 nearly five hundred trees
8 was an expensive delicacy
9 was a fiercely guarded privilege
10 in the thirteenth century
11 pursuit for a lord
12 and there are fifteen hundred
13 our journey starts
14 buildings the like of which
15 in the middle ages
16 as a big church
17 designed on the same basis
18 we don't need the height
19 the corn'll be cut
20 it'd be much harder pulling down
21 in our brief moment
22 and that's what we've done
23 I really mean it
24 and when I say
25 how big's that number
26 towards absolute zero
27 well if I were to
28 then there wouldn't be enough atoms
29 representing one year
30 that we can truly understand ourselves
31 barely emit anymore heat
32 that they become so cold
33 and the laws of nature that govern it
34 in to the universe
35 and we can even make
36 the most astonishing wonder of the universe
37 our exploration of the cosmos
38 and we have observatories
39 the language of curiosity
40 means the end of life on our planet
41 to find time to feed themselves
42 they're begging again
43 the moment they head out
44 and with the day cut short
45 to bring food to the babies
46 the older members of the gang
47 just cant get enough of those
48 they get all the water they need
49 meerkats hardly ever drink
50 six hours later
51 with the afternoon starting to cool down
52 even try to remove the sting
53 is dealt with in quick time
54 are a challenging start though
55 but they're already trying to feed themselves
56 come thick and fast to the youngsters
57 little Arnie munches it down without
58 if I were stung by that fella
59 it's one of the huge advantages of living in such a cooperative mob
60 three litters of young in a single season
61 it's a bit steep
62 it reminds me of white latex rubber
63 pounded into a form of dough
64 and readjusting itself in to the modern world
65 coming back from a nightmare
66 its got a long way to go
67 to the baby here
68 pushing this machine
69 and all the chat that's going on
70 who's the happiest
71 we've got a family doing something very much together
72 what's also going through my mind
73 it's got a long way to go
74 obviously getting on with each other
75 something which has always been a mystery to me
76 this is a village that
77 ingredients of Cambodian food
78 barge in southern France some years ago
79 and how he came up with such an idea
80 boiled in water for thirty minutes or so
81 the power of the sun is constantly changing
82 they are the seasonal seas
83 where conditions change
84 water is so viscous
85 that the fighter plankton can't swim against the current
86 that sweep the little algae over the basket like filters around their mouths
87 every spring in vast numbers
88 part of their staple diet
89 to find their prey
90 are able to use these trails
91 rather like jet trails in the sky
92 lighting by lasers
93 the warmth of the shallows is speeding the oaks development
94 these battles continue for the next two months
95 each of these females has been carrying around about
96 if they're to raise their young
97 another kind of sea slug
98 and so it is
99 all navanax gets

100 and since she weighs a hefty seven kilos
1 to say that I first became interested in
2 terrible time when
3 account of that
4 took over the country
5 to the gulf coast the gulf of Thailand
6 and then people started coming back to Cambodia
7 the whole thing was of course destabilised
8 nice looking people
9 was a sense
10 came out of it
11 but one of the things that
12 and broke everything up
13 two to three million
14 and the pepper fields
15 grow rice instead
16 made the farmers
17 products from this region
18 Kampot pepper
19 one of the most famous
20 it's so sweet and so succulent
21 she doesn't approve if you leave anything
22 make sure I've eaten everything
23 check the dish
24 what it means to you
25 tell me about
26 overwhelmed with the fragrance of freshness
27 all away across the world for that
28 it's worth coming
29 sort of for everybody
30 I mean if they're getting fifteen
31 they're not too badly off
32 particularly here they have this
33 any type of chilli crab
34 best crabs possible
35 a glass or two of anchor beer
36 sit there drinking
37 quite sort of
38 and its really nice eating
39 to go with it
40 but they also nip each other
41 or a non-alcoholic beverage
42 it's time for a cold beer
43 and I cant wait for tomorrow
44 one day on this old rice barge
45 I've only had
46 why I love it so much
47 coz we all came out of the
48 all descended from fish actually
49 the site of this is making my mouth water
50 I love the combination of mango and prawns
51 in goes the mango
52 tell me about Vietnamese food
53 good for you I suppose
54 and I suspect it's making yours too
55 I guess it's one of those things
56 with a spoonful of cornflower
57 and he thickens it
58 with salt pepper and lime juice
59 the season the dish
60 it's a different matter
61 in a relatively small country
62 with nearly ninety million
63 to be sceptical about
64 which we in the West can afford
65 I suppose it's a necessity
66 and in a heavily populated country
67 good source of protein
68 and everyone I know
69 these are river prawns
70 and almost immediately turn pink
71 tossed into fried onions and garlic
72 extremely nice and it's smelling
73 the pork should be nice
74 after an hour
75 not just a highway
76 it's the source of everyone's livelihood
77 crowded with houses
78 Vietnam is more populated
79 but just over thirty years ago
80 you don't need a lot of it
81 first or it's last quarter
82 when the moon is either
83 but we do know
84 is a mystery
85 how it's coordinated
86 is called an Arrabaala
87 this mass nesting
88 hundreds of miles away from here
89 where they can dig a nest hole
90 to find a bear patch
91 clamber over one another
92 that they have to
93 get so crowded
94 the top of the beach
95 five thousand are coming and going
96 at the peak time
97 to lay their eggs in the sand
98 will visit this one beach
99 four hundred thousand
100 and over the next six days or so
101 they're all female
102 the rest of the time
103 on a few key nights each year
104 come to this one beach
105 of the world's population
106 all along the beach
107 they are appearing
108 within an hour
109 they come in ones and twos
110 start to emerge from the surf

111 the beach has been empty
112 between full and new
113 exactly half way
114 is in it's last quarter
115 it's just after midnight
116 a very special night
117 coast of Costa Rica
118 fertilised eggs plastered on every rock
119 leaving behind them
120 back out to deeper waters
1 almost nothing about
2 in the desert
3 built this fortified temple
4 the decorations are gone
5 and all traces of its language
6 all but the smallest fragments
7 details of this culture
8 I might build one in my garden
9 it is surely one of the most fascinating
10 it's one of South America's lesser known
11 on the North Western Coast of Peru
12 today's date
13 and covered with painted figures
14 two or three days
15 an accuracy of
16 you have to be here
17 but to do that
18 and you can tell
19 you can measure it's position
20 at any time of year
21 at the left most tower
22 you can see in the distance
23 actually just in-between that mountain
24 the winter solstice the shortest day
25 on June twenty first which is
26 just to the right
27 in the southern hemisphere
28 actually December twenty first
29 now at different times of year
30 is at a different place
31 and the sun rises
32 arose the science
33 and out of those questions
34 should it be something else
35 are concerned with the efficiency
36 back in the nineteenth century
37 around two and a half thousand years ago
38 the day it was built
39 in just the same way as you could
40 if you stand in the right place
41 that make up this pile of sand
42 and buildings collapse
43 not a single thing's
44 and a good way to understand how
45 like the individual grains
46 but has been made up of many
47 is to think of objects
48 left to the mercy of the elements
49 it will fill the entire
50 the sun will have grown so much
51 will become impossible
52 long after life
53 has introduced the concept
54 the second law
55 to the passage of time
56 and that means that there's a direction
57 and in the future
58 the universe was more ordered
59 that there is a difference
60 at its heart
61 evolution of the universe
62 the passage of time
63 concepts like heat
64 for the first time
65 enter the scientific vocabulary
66 along with that
67 the most important law of physics
68 was so profound
69 is called the second law
70 the second law of thermodynamics
71 that everything tend
72 the second law says
73 between the past
74 why is there a difference
75 in the history of science
76 it ended up
77 being able to explain one of the
78 and the efficiency of steam engines
79 talking about how heat moves around
80 see here's a law
81 even trying to remove the sting
82 but little Arnie munches it down without
83 I'd be feeling very sick
84 if I were stung by that fella
85 is dealt with in quick time
86 are a challenging start though
87 but they're already trying to feed themselves
88 bugs come thick and fast to the youngsters
89 just can't get enough of those
90 from juicy grubs and bugs
91 meerkats hardly ever drink
92 to find time to feed themselves
93 it's all the others can do
94 and with the day cut short by the heat
95 to bring food to the babies
96 stimulate the older members of the gang
97 they're begging again
98 so it's not an entirely selfless act

99 they're far to young to dig up their own bugs
100 to start them off on the right foot
101 but they want to explore a wider world
102 popping out of the borough
103 just a few days after
104 trying to keep up with the rest of the family
105 Arnie and his brothers and sisters
106 they're only four weeks old
107 if you can find somewhere to settle
108 these disputes can get quite nasty sometimes
109 status in the group is beginning to take shape
110 even at this tender age
111 by nine thirty in the morning
112 is already in the mid-thirties
113 it's high summer in the desert
114 and one of the male pups is harder than most
115 the response is immediate
116 to the new arrivals without a quibble
117 to an older family member
118 often putting the demands of the babies before their own
119 it's one of the huge advantages of living in such a cooperative mob
120 with other members of the team helping with the kids
1 to say that I first became interested in
2 terrible time when
3 account of that
4 took over the country
5 to the gulf coast the gulf of Thailand
6 and then people started coming back to Cambodia
7 the whole thing was of course destabilised
8 nice looking people
9 was a sense
10 came out of it
11 on the north western coast of Peru
12 today's date
13 and covered with painted figures
14 two or three days
15 an accuracy of
16 you have to be here
17 but to do that
18 and you can tell
19 you can measure it's position
20 at any time of year
21 she doesn’t approve if you leave anything
22 make sure I've eaten everything
23 check the dish
24 what it means to you
25 tell me about
26 overwhelmed with the fragrance of freshness
27 all away across the world for that
28 it's worth coming
29 sort of for everybody
30 I mean if they're getting fifteen
31 and the sun rises
32 arose the science
33 and out of those questions
34 should it be something else
35 are concerned with the efficiency
36 back in the nineteenth century
37 around two and a half thousand years ago
38 the day it was built
39 in just the same way as you could
40 if you stand in the right place
41 or a non-alcoholic beverage
42 it's time for a cold beer
43 and I cant wait for tomorrow
44 one day on this old rice barge
45 I've only had
46 why I love it so much
47 coz we all came out of the
48 all descended from fish actually
49 the site of this is making my mouth water
50 I love the combination of mango and prawns
51 will become impossible
52 long after life
53 has introduced the concept
54 the second law
55 to the passage of time
56 and that means that there's a direction
57 and in the future
58 the universe was more ordered
59 that there is a difference
60 at its heart
61 in a relatively small country
62 with nearly ninety million
63 to be sceptical about
64 which we in the West can afford
65 I suppose it's a necessity
66 and in a heavily populated country
67 good source of protein
68 and everyone I know
69 these are river prawns
70 and almost immediately turn pink
71 that everything tends
72 the second law says
73 between the past
74 why is there a difference
75 in the history of science
76 it ended up
77 being able to explain one of the
78 and the efficiency of steam engines
79 talking about how heat moves around

80 see here's a law
81 first or it's last quarter
82 when the moon is either
83 but we do know
84 is a mystery
85 how it's coordinated
86 is called an Arrabaala
87 this mass nesting
88 hundreds of miles away from here
89 where they can dig a nest hole
90 to find a bear patch
91 meerkats hardly ever drink
92 to find time to feed themselves
93 it's all the others can do
94 and with the day cut short by the heat
95 to bring food to the babies
96 stimulate the older members of the gang
97 they're begging again
98 so it's not an entirely selfless act
99 they're far to young to dig up their own bugs
100 to start them off on the right foot
101 they're all female
102 the rest of the time
103 on a few key nights each year
104 come to this one beach
105 of the world's population
106 all along the beach
107 they are appearing
108 within an hour
109 they come in ones and twos
110 start to emerge from the surf
111 by nine thirty in the morning
112 is already in the mid-thirties
113 it's high summer in the desert
114 and one of the male pups is harder than most
115 the response is immediate
116 to the new arrivals without a quibble
117 to an older family member
118 often putting the demands of the babies before their own
119 it's one of the huge advantages of living in such a cooperative mob
120 with other members of the team helping with the kids
121 it's so beautiful
122 listening to fine music
123 trance like state
124 staring at the ceiling sends me into a kind of
125 to make it look like a dream
126 so you didn’t believe it was real
127 it's just what the builders wanted
128 very tempting if you were a school boy to
129 about a hundred foot
130 right down to the choir stalls
131 if you look through
132 and the holes have been left here
133 communicate with the stone mason on the other side
134 when they built this ceiling they cut
135 tons of stone
136 nearly two thousand
137 almost like an egg shell supporting
138 is only twelve centimetres thick
139 and the stone here
140 but what I'm standing
141 these are the wooden beams that support it
142 the great roof of the chapel
143 this is quite astonishing
144 in the heart of England
145 that have left their mark forever
146 creation of great houses
147 in the world
148 it's the largest fan volt
149 it's sheer scale is breath-taking
150 to the crowning glory
151 dome ribs soar heavenwards
152 from slender columns at the sides
153 with this wonderful lightness and exuberance
154 to this end of the middle ages
155 the final flowering
156 is a miraculous feat
157 elegance of the stone work
158 was unlike anything seen before
159 powerful rather severe buildings
160 from the Normans with their

Text Inspector Profile © for this text is CEFR upper B2

C.3 List of utterances used in Experiment 3

1 the American black bear
2 from the moment they wake up in the Spring
3 I'm a wildlife cameraman
4 I'm a little bit scared
5 near the border with Canada
6 it's one of the last great wilderness areas
7 I'm going to try and shoot them with a camera
8 bears have a big reputation
9 and are almost impossible to find
10 black bears live in thick forest
11 that has ever lived on our planet
12 dwarfed by the vast expanse of the ocean
13 and weighing over two hundred tonnes
14 it's heart is the size of a car
15 it's one of the fastest animals
16 the ocean's largest inhabitant
17 as the sardines travel North
18 thousands of capped gullets track the sardines
19 a whole caravan of predators
20 these are bronze whaler sharks
21 and broke everything up
22 two to three million
23 and what came out of it particularly
24 the whole thing of course was destabilised
25 with a great civilisation behind them
26 I thought I better read up about it
27 and then people started coming back to Cambodia
28 and so here I am
29 every dish that comes my way
30 and I'm very excited
31 to see how we built
32 and through a thousand years of our history
33 they tell us who we are
34 when we reach for the skies
35 this was the richest corner of the country
36 that the foundations of modern Britain
37 but we have a cosy side
38 and a taste for fun
39 and the people who built them
40 this is the story of Britain
41 this family's small by meerkat standards
42 and they're just beginning to stake a territory of their own
43 has already had some incredible adventures
44 he's the head of the family
45 he's an adult male meerkat
46 three years old and in his prime
47 I've had to earn their complete trust
48 exploring the world of the meerkat is quite a challenge
49 and the babies are the size of mice
50 and in a way which has never been seen before
51 it is surely one of the most fascinating
52 around two and a half thousand years ago
53 a civilisation we know almost nothing about
54 it's walls were once brilliant white
55 built this fortified temple
56 covered with painted figures
57 it's one of South America's lesser known
58 if you stand in the right place
59 these towers form an ancient solar calendar
60 actually just in-between that mountain

Text Inspector Profile © for this text is CEFR upper B2

C.4 List of utterances used in Experiments 4 and 5

1 fine British phlegm
2 one of the policemen there
3 turned to the queen mother and said
4 on September the thirteenth
5 King George the sixth
6 at Buckingham palace
7 in the quadrangle
8 a German bomb
9 a workman nearby
10 they would've been
11 by flying glass
12 rather than open
13 had been closed
14 in the room where they were standing
15 if the window
16 if I may say so ma'am
17 decent bombing
18 of a signature
19 on this document
20 and by writing that signature
21 I've ever seen
22 his own hand is on the front of the document
23 signed with his own hand
24 unsuccessful monarch
25 he finishes with the
26 that's a sense of crisis for you
27 in a single day
28 all of its stages
29 that you then get
30 it can only be born in circumstances different
31 shoulders of a sovereign is so heavy
32 that the burden which constantly rests
33 that it should be remembered
34 but I would beg
35 I will not
36 enter now in to
37 in which he says
38 to the house of Lords
39 this is his
40 Edward the eighth
41 misery and crisis
42 by the king himself
43 hadn't been a failed
44 if her uncle
45 still criss crosses Britain
46 hoping to see her
47 pretty substantial crowd
48 she's not due for another hour
49 all over again
50 on the second stage of her north wales visit
51 to see them
52 hoping to be pleased
53 I ask you
54 she loathes being late
55 that she would never have become queen
56 looking like an emperor
57 to mark the coronation that never was
58 designed by him
59 he was bored by duty
60 these are the papers
61 whisky stains on them
62 being in power
63 instinctive feel for monarchy
64 white hall insiders said
65 a vigorous government of self-proclaimed modernisers
66 presented a different problem
67 Tony Blair's new labour
68 personal level they go on
69 Buckingham palace press officer
70 by an enthusiastic
71 this was over briefing
72 considerably younger than you are now
73 she became queen
74 suddenly thrusted in to that role at her age
75 jump that must've been
76 on what an extraordinary
77 do you ever sort of reflect
78 dark days of the monarchy
79 tower at the palace of Westminster
80 carefully stored away
1 a glass or two of anchor beer
2 and contemplating
3 what a great thing a crab is
4 the world going by
5 sauce to go with it
6 green pepper corns
7 particularly here they have this
8 any type of chilli crab
9 but they also nip each other
10 vicious little critters
11 they nip you
12 notice these rubber bands
13 and then sit there drinking
14 they got quite sort of
15 and it's really nice eating
16 they're hard working and
17 there's something very optimistic about
18 probably the best time
19 they've got swimming fins on the back there
20 going on fishing trips making these films
21 they never get the money
22 I actually find
23 cheerful types
24 the farmers the fishermen

25 they're tied up because they're really
26 generous dashes of soy and oyster sauce
27 awful lot of pepper she's putting in there
28 in delicious harmony
29 in delicious harmony
30 fruits of the land and sea
31 crab and pepper corns combine
32 delicate flavours and aromas
33 all they need
34 they don’t use many ingredients
35 and there wasn't a chilli in sight
36 isn't always spicy
37 all the way across the world for that
38 it's so sweet and so succulent
39 she doesn't approve if you leave anything
40 what it means to you
41 coz we all came out of the
42 all descended from fish actually
43 why I love it so much
44 or a non-alcoholic beverage
45 it's time for a cold beer
46 and I can't wait for tomorrow
47 one day on this old rice barge
48 but fish above all
49 tell me about Vietnamese food
50 good for you I suppose
51 very fresh and
52 Vietnamese food is really
53 because I believe
54 and I suspect it's making yours too
55 making my mouth water
56 the sight of this
57 in goes the mango
58 combination of mango and prawns
59 spoon full of cornflower
60 and he thickens it
61 with salt pepper and lime juice
62 the season the dish
63 it's a different matter
64 in a relatively small country
65 with nearly ninety million
66 can afford to be sceptical about
67 I guess it's one of those things
68 I suppose it's a necessity
69 and in a heavily populated country
70 but they're a good source of protein
71 along the banks of the Mekong
72 proliferation of prawn farms
73 everyone I know
74 these are river prawns
75 and almost immediately
76 tossed in to fried onions and garlic
77 I was thinking all the time he was gonna cut his fingers
78 watch the chefs do
79 quite difficult for us
80 working a great delicacy
81 but right in the heart
82 not in the planes
83 the strange lightning that's found
84 to the naked eye
85 something completely invisible
86 these high speed cameras are allowing us to
87 fingers of electricity
88 that launched a thousand horror movies
89 her chest is a massive muscle
90 I feel safe
91 counting down from
92 that means it's now armed
93 so I arm it
94 to be in there
95 that I'm lucky
96 so in that sense
97 twenty thousand tons of rocks
98 to try and shift
99 detonating within half a second of one another
100 twenty separate explosions
101 at first hand
102 but it does mean I'll get to experience
103 from the explosion
104 a hundred metres away
105 so I'll need to be
106 a remote trigger with a short travel
107 at high speed
108 we're gonna need to film it
109 of seeing this
110 so to stand a chance
111 about half a second
112 that's the theory
113 is gonna be over in
114 this whole event
115 by subsequent explosions
116 fragment the rock
117 an open up a face
118 that can then be hit
119 each explosion will
120 I wanna take a closer look at this
121 I need to blow something up
122 it's about a hundred metres long
123 which will mean shifting
124 what we're gonna blow up
125 this new form of lightning
126 as we build more tall buildings
127 in rapid city
128 and it's not just here
129 it's the storm clouds
130 when the upward bolt
131 to the city
132 straight back down
133 away from the town

at twenty five millisecond intervals
explosions need to be choreographed
each one of these
sophisticated than just creating one
our experiment is a bit more
what's going on
begin to understand
keep it airborne
doesn’t get any clearer
even seen in slow motion
apparently random flapping
scientists have struggled to understand
was aerodynamically
calculated that bee flight
the confusion started over seventy years ago
how's that supposed to work?
by tiny wings
that had scientists baffled
of flying finesse
quite so elegant
to gather nectar
it can hover at a complete stand still
should help us see what's happening
now I'm not trying to set fire to them
lurching through the air
here she is
that's why she's so round
before the sun rises
you have to be here
but to do that
the day it was built
in just the same way as you could
you can still experience
if you stand in the right place
and all traces of its language
details of this culture
decorations are gone
all but the smallest fragments
covered with painted figures
its walls were once
built this fortified temple
we know almost nothing about
around two and half thousand years ago
it is surely one of the most fascinating
it's one of South America's lesser known
on the north western coast of Peru
they are never undone
just every now and again you hear this
and you can hear it happening
it's about a million tons a day
blaster in to the lake every year
billion tons of ice
in to the late
about something like that much
and once these changes have happened
that as each moment passes
and that's because the arrow of time
just as in our lives
but that's an illusion
an unchanging
but out there in the universe
joy and tragedy of our lives
I suppose it's
of what it means to be human
or even fifty times
this little read blob
it's the depth
to get anywhere near
in the entire universe
then there wouldn’t be enough atoms
representing one year
start counting with a single atom
if I were to
how big's that number
trillion trillion trillion
I really mean it
an unimaginable period of time
and when I say
or they become a white dwarf
our sun
of its brightness
and a fraction
of its current volume
less than a millionth
of the size of the earth
it will be smaller
of our once magnificent sun
all that remains
the cosmos will
last forever
and all its wonders
that we inhabit
because this structured universe
of the arrow of time
the most profound consequence
in to eternal night
cosmos will be plunged
the same as for all stars
less brightly
eighteen thousand times
they would appear to shine
but to our eyes
the mass of our son
eleven to twelve per cent
star known as a red dwarf star

79 in this photograph
80 the much more distant stars
81 mistakes that your daughter
82 Brett got off track because of
83 in your family
84 meth's been a problem
85 and his mother too
86 she's got a problem with crystal meth
87 the guy they were looking for
88 skinny young man
89 do you know why they've come for you today?
90 from the BBC British broadcasting
91 are you Brett?
92 what kind of drug was it?
93 have large numbers of meth addicts
94 even the small towns
95 using an over the counter cold remedy
96 the sheriff's department
97 I was back with law enforcement
98 where meth is cooked up
99 for hiding clandestine labs
100 the rural parts of Fresno county
101 they can help
102 over your children in a sense
103 sounds like you chose your drug
104 wouldn't you like them to be with you?
105 do you have children?
106 so your track record is not that good
107 you didn't come back for three days did you say?
108 for the whole day
109 you couldn't even stick it out
110 when you get married
111 supposed to be the best day of your life
112 how's that going?
113 is there a man in your life
114 do you think you can do it?
115 was a recovering meth addict
116 one of those taking part
117 so how many times have you been in recovery inpatient?
118 in the hope of repairing their relationships
119 facing off in public
120 the hot seat
121 I've been invited to observe a therapeutic exercise
122 I was heading back to West Care
123 family support to recovery
124 do you feel I'm being judgemental
125 with your children
126 they're relaxing in there with your
127 you know it goes with the lifestyle
128 they could be exposed to some of the chaos
129 talk to Kevin
130 your kids being around
131 fairly sure they hadn't seen any drug use
132 part of the evening had been the presence of children
133 it was close to midnight
134 where did they just come from Kevin
135 damaged by it
136 with some friends
137 contained a woman acting strangely
138 on the other side of the city
139 pulled over for a minor violation
140 the guy in the back
141 really interesting
142 to notify social services
143 in just the conditions
144 kids living here too aren't there
145 well known to the police
146 in a complex
147 sister of a suspected dealer
148 one of the first spots we hit
149 how much of the crime you see
150 where are we now?
151 on the front line of Fresno's meth problem
152 do you feel different now?
153 it's quite weird for me
154 it's not to do it, do you know what I mean
155 Chris had been taking occasional breaks
156 what I'd do when I got there
157 depths of the meth lifestyle
158 Andrew joined him
159 it was an awkward moment
160 where do you get it from?
1 tell me about
2 freshness of this Cambodian food
3 but it's so
4 and a heritage for his children
5 fast food and drink
6 a home for his family
7 a man who plants a coconut
8 from the south pacific
9 I wouldn’t go as far as that
10 preferred coconut milk to wine
11 with pork and pineapple
12 is also the foundation of this lovely dish
13 as well as Asian delis
14 they're not too badly off
15 how much would they catch in a day
16 and broke everything up

two to three million
took over the country murdered about
I thought I better read up about it
but Cambodian food no
but I know a bit about Malaysian food and Thai food
and what came out of it particularly was the food
every dish that comes my way
and I'm looking forward to
and I'm very excited
just go there
the gulf of Thailand
people started coming back to Cambodia
by the war in South East Asia
but one of the things that came out of it
civilisation behind them
nice looking people
terrible time when
one man's account of that
from the hotel I was staying at
he took me out with some local fisherman
without tasting it's famous fresh crabs
no trip to Ket would be complete
five miles off the coast
sort of for everybody
with all the crew
arrived in their village
now if we'd been in a
we're not intruding on them
coming over the water
sound of laughing children
with all that lovely
it's different
but it's the same here
feeling of life going on
you're sitting in this sort of really
fields of sunflowers and vineyards
in a barge
South Western France
going down the canals of
we made a series
three years ago
I was just thinking
I know I do
generally feel happy
I think it's really
people all over the world
it's really nice the way
there were many
before the passing of time
was to scare away the crocodiles
on the front of boats
for painting eyes
serenity and peace
but the eyes live on
all eaten long ago
we're on that barge for about
adding to my sense of
we would've been intruding
interrupt their life
and just observing
just in an easy way
quite an odd shape
because the stone in the middle is
cut up a mango it is difficult
in recent astronomical history
interesting images
it is one of the most
we can see that story
deep in to the cosmos
well what that red blob is
an enormous
is there an arrow
and the future
is there a difference
but what drives this evolution
always be this way
because it's beautiful
it produces a universe
as the arrow of time
and it won't
travelling across lakes
we never see waves
of the entire universe
drives the evolution
change upon change
in the observable universe
is one of a hundred billion
in our galaxy
illuminates the night sky
is just one of two hundred
that we have yet to explore
billions of worlds
with stars surrounded by nebulae
absolutely a wash
and systems of planets
astonishing beauty
an age of
it's almost impossible to relate to them
unimaginably vast
time scales in the cosmos
on these universal scales
where we can begin to encounter
yet there are places on earth
a natural event that's been happening
but the reason it doesn’t stand out
now it's only four point two light years away
of the nearest star to our solar system
this is a picture
type of star will remain

126 but long after they're gone
127 from where the earth is now
128 on a clear night
129 as the full moon
130 the same amount of light
131 in the universe
132 living stars
133 will be the last
134 that means that stars
135 trillions of years
136 so they have life spans
137 incredibly slowly
138 they burn their nuclear fuel
139 they're so small
140 one advantage over their much more
141 but red dwarfs do have
142 it will fill the entire horizon
143 the sun will have grown so much
144 long after life has disappeared
145 will become impossible
146 on this planet
147 the existence of all life
148 perfect day on earth
149 there will be one last
150 it's called the second law
151 evolution of the universe
152 along with that
153 it contained a radically new concept
154 at its heart
155 probably the most important law of physics
156 but eventually
157 a white dwarfs faint glow
158 with no fuel left to burn
159 and it will have profound consequences
160 but in the universe
161 come to link with me
162 that won't scare them
163 a distinctive sound
164 I'm just trying to make
165 I'm not trying to mimic a meerkat's call
166 in the morning
167 is when they first wake up
168 and the best time to spot them
169 but there are meerkats here
170 that live in the sand dunes
171 it's easy to overlook the little guy
172 and stay in the car
173 I keep my distance
174 more nervous than their babies
175 I've decided to get to know this
176 they may never trust me
177 but if I scare them now
178 it is really tempting to try and follow
179 off in to the grass again
180 I just have to watch them
181 but for now
182 I really want to see
183 and that's a part of their lives
184 to escape the Northern winter
185 swifts aren't dangerous to meerkats
186 just arrived in South Africa
187 they've never seen them before
188 but the youngsters don’t know that
189 for the past five months
190 the birds have been in Europe
191 I may have blown it
192 I think they're not ready for this
193 goes in to my throat
194 that it's not me at all they're worried about
195 that no harm can come of it
196 begun to associate me with
197 I might not be in the car anymore
198 start to pay off
199 movements I've been making
200 would scare him
201 I on the other hand
202 looking calm
203 even a simple sneeze
204 one wrong move now
205 the adult male who's up first
206 and sit outside
207 I'm going to leave the vehicle
208 in getting them used to me
209 I'm going to take the next step
210 and today especially
211 pop their heads above ground
212 they sleep through the hours of darkness
213 with the family
214 I need all the time I can get
215 that I have to be out
216 in the South African summer
217 they have to learn to trust me
218 I want the family
219 all part of the same process
220 to be able to recognise me at
221 before they can fend for themselves
222 they're going to need a lot of survival lessons
223 so it does pay to be weary
224 and in a way which has never been seen before
225 to find out what they get up to throughout the day
226 I've had to earn their complete trust
227 and the babies are the size of mice
228 by meerkat standards
229 is quite a challenge
230 a meerkat with real attitude
231 and already he stands out

232 one of the newest arrivals
233 and the stronger the team
234 the bigger the family
235 the more pups they have
236 they'll need reinforcements
237 to do that
238 beginning to stake a territory of their own
239 they're only been together for a year or so
240 digger and mama

Text Inspector Profile © for this text is CEFR upper B2

Appendix D - Skills Tests

D.1 Working Memory Test

WORKING MEMORY TEST - PHONOLOGICAL
Language Learning Study

Student ID # _____
(To be filled in by the researcher)

First Name: _____ Middle Initial: _____

Last Name: _____ Age: _____

Language you are currently studying (write one): _____

University or institute of language study: _____

Directions: This is a test to estimate your working memory capacity. You will hear sentences read out-loud in sets of three, four, or five sentences. For each sentence, you should decide if it makes sense or is nonsense and if it is grammatical or ungrammatical.

A sentence is nonsense if the words used in that order do not make sense in the real world. A sentence such as,
"The box ate the dog with relish,"
would be nonsense, as boxes neither eat dogs nor do things with relish.

A sentence is ungrammatical if grammar rules such as subject-verb agreement, verb tense, or article use are violated. Such sentences are similar to things that children might say, such as the following:
"The dog eating the food with relish."
This sentence is ungrammatical because the verb form "eating" requires a helping verb, as in "is eating". Sentences may have multiple grammar errors or multiple nonsense problems, and may be both ungrammatical and nonsense, so you will need to make two judgments for each sentence. At the end of each group of sentences, you will be asked to recall the last word of each sentence.

Practice with the first three sentences below. After you hear each sentence, indicate if the sentence made sense or not and if the sentence was grammatical or not by filling in the corresponding circles.

	Sense/Nonsense		Grammatical/Ungrammatical	
Sample A	Ⓢ	Ⓝ	Ⓖ	Ⓤ
Sample B	Ⓢ	Ⓝ	Ⓖ	Ⓤ
Sample C	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

A. _____

B. _____

C. _____

That was a practice set. You should have marked sentence one as nonsense and grammatical, sentence two as nonsense and ungrammatical, and sentence three as sense and ungrammatical. (You should have made two decisions for each sentence.) You should have recalled the words "bank," "storage," and "bathrooms".

If you have any questions, please ask the proctor now.

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
1)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
2)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
3)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

1. _____

2. _____

3. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
4)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
5)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
6)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

4. _____

5. _____

6. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
7)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
8)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
9)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

7. _____

8. _____

9. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
10)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
11)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
12)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

10. _____

11. _____

12. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
13)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
14)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
15)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
16)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

13. _____

14. _____

15. _____

16. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
17)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
18)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
19)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
20)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

17. _____

18. _____

19. _____

20. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
21)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
22)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
23)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
24)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

21. _____

22. _____

23. _____

24. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
25)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
26)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
27)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
28)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

25. _____

26. _____

27. _____

28. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
29)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
30)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
31)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
32)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
33)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

29. _____

30. _____

31. _____

32. _____

33. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
34)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
35)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
36)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
37)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
38)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

34. _____

35. _____

36. _____

37. _____

38. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
39)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
40)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
41)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
42)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
43)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

39. _____

40. _____

41. _____

42. _____

43. _____

DO NOT TURN THE PAGE UNTIL TOLD.

	Sense/Nonsense		Grammatical/Ungrammatical	
44)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
45)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
46)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
47)	Ⓢ	Ⓝ	Ⓖ	Ⓤ
48)	Ⓢ	Ⓝ	Ⓖ	Ⓤ

DO NOT TURN THE PAGE UNTIL TOLD.

What was the last word of each sentence that you just heard?
Write them on the lines below.

44. _____

45. _____

46. _____

47. _____

48. _____

The test is finished. Before turning your book in, please answer the question below.

Did you use any specific strategies to perform well on this test? For example, did you do something special to remember the last words of the sentences? Please explain any mental strategies that helped you.

_____

_____

_____

_____

_____

Thank you!

DO NOT TURN THE PAGE UNTIL TOLD.

D.2 Working Memory Test Answer Key

ID	*Correct Answer:*	*Sentence Key:*	*Sentence Class: (for example. 1st sentence in set of 3)*
S/N1	Nonsense	4	1.3
S/N2	Sense	2	2.3
S/N3	Nonsense	3	3.3
S/N4	Nonsense	4	1.3
S/N5	Sense	2	2.3
S/N6	Sense	1	3.3
S/N7	Sense	2	1.3
S/N8	Sense	1	2.3
S/N9	Nonsense	3	3.3
S/N10	Sense	2	1.3
S/N11	Nonsense	4	2.3
S/N12	Sense	1	3.3
S/N13	Sense	2	1.4
S/N14	Sense	1	2.4
S/N15	Nonsense	3	3.4
S/N16	Nonsense	3	4.4
S/N17	Nonsense	4	1.4
S/N18	Sense	2	2.4
S/N19	Nonsense	3	3.4
S/N20	Sense	1	4.4
S/N21	Sense	1	1.4
S/N22	Nonsense	4	2.4
S/N23	Sense	2	3.4
S/N24	Nonsense	3	4.4
S/N25	Nonsense	3	1.4
S/N26	Sense	2	2.4
S/N27	Sense	1	3.4
S/N28	Nonsense	4	4.4
S/N29	Sense	1	1.5
S/N30	Sense	2	2.5
S/N31	Nonsense	4	3.5
S/N32	Nonsense	4	4.5
S/N33	Nonsense	3	5.5
S/N34	Nonsense	3	1.5
S/N35	Sense	2	2.5
S/N36	Nonsense	4	3.5
S/N37	Sense	1	4.5
S/N38	Nonsense	4	5.5
S/N39	Nonsense	3	1.5
S/N40	Sense	1	2.5
S/N41	Nonsense	4	3.5
S/N42	Sense	1	4.5
S/N43	Sense	2	5.5
S/N44	Sense	2	1.5
S/N45	Nonsense	3	2.5

S/N46	Nonsense	4	3.5
S/N47	Sense	1	4.5
S/N48	Nonsense	3	5.5
G/U1	Ungrammatical	4	1.3
G/U2	Ungrammatical	2	2.3
G/U3	Grammatical	3	3.3
G/U4	Ungrammatical	4	1.3
G/U5	Ungrammatical	2	2.3
G/U6	Grammatical	1	3.3
G/U7	Ungrammatical	2	1.3
G/U8	Grammatical	1	2.3
G/U9	Grammatical	3	3.3
G/U10	Ungrammatical	2	1.3
G/U11	Ungrammatical	4	2.3
G/U12	Grammatical	1	3.3
G/U13	Ungrammatical	2	1.4
G/U14	Grammatical	1	2.4
G/U15	Grammatical	3	3.4
G/U16	Grammatical	3	4.4
G/U17	Ungrammatical	4	1.4
G/U18	Ungrammatical	2	2.4
G/U19	Grammatical	3	3.4
G/U20	Grammatical	1	4.4
G/U21	Grammatical	1	1.4
G/U22	Ungrammatical	4	2.4
G/U23	Ungrammatical	2	3.4
G/U24	Grammatical	3	4.4
G/U25	Grammatical	3	1.4
G/U26	Ungrammatical	2	2.4
G/U27	Grammatical	1	3.4
G/U28	Ungrammatical	4	4.4
G/U29	Grammatical	1	1.5
G/U30	Ungrammatical	2	2.5
G/U31	Ungrammatical	4	3.5
G/U32	Ungrammatical	4	4.5
G/U33	Grammatical	3	5.5
G/U34	Grammatical	3	1.5
G/U35	Ungrammatical	2	2.5
G/U36	Ungrammatical	4	3.5
G/U37	Grammatical	1	4.5
G/U38	Ungrammatical	4	5.5
G/U39	Grammatical	3	1.5
G/U40	Grammatical	1	2.5
G/U41	Ungrammatical	4	3.5
G/U42	Grammatical	1	4.5
G/U43	Ungrammatical	2	5.5
G/U44	Ungrammatical	2	1.5
G/U45	Grammatical	3	2.5
G/U46	Ungrammatical	4	3.5
G/U47	Grammatical	1	4.5

G/U48	Grammatical	3	5.5
R/N1	01. office	4	1.3
R/N2	02. major	2	2.3
R/N3	03. uncle	3	3.3
R/N4	04. sweater	4	1.3
R/N5	05. family	2	2.3
R/N6	06. garden	1	3.3
R/N7	07. kitchen	2	1.3
R/N8	08. classes	1	2.3
R/N9	09. foliage	3	3.3
R/N10	10. night	2	1.3
R/N11	11. birds	4	2.3
R/N12	12. boxes	1	3.3
R/N13	13. war	2	1.4
R/N14	14. person	1	2.4
R/N15	15. mountains	3	3.4
R/N16	16. car	3	4.4
R/N17	17. sofas	4	1.4
R/N18	18. weekend	2	2.4
R/N19	19. girl	3	3.4
R/N20	20. picnics	1	4.4
R/N21	21. design	1	1.4
R/N22	22. bracelet	4	2.4
R/N23	23. paint	2	3.4
R/N24	24. firm	3	4.4
R/N25	25. house	3	1.4
R/N26	26. people	2	2.4
R/N27	27. surprise	1	3.4
R/N28	28. library	4	4.4
R/N29	29. summer	1	1.5
R/N30	30. stories	2	2.5
R/N31	31. digestion	4	3.5
R/N32	32. package	4	4.5
R/N33	33. haircut	3	5.5
R/N34	34. display	3	1.5
R/N35	35. mind	2	2.5
R/N36	36. stereo	4	3.5
R/N37	37. area	1	4.5
R/N38	38. smoking	4	5.5
R/N39	39. clubs	3	1.5
R/N40	40. business	1	2.5
R/N41	41. crowds	4	3.5
R/N42	42. piano	1	4.5
R/N43	43. fortune	2	5.5
R/N44	44. news	2	1.5
R/N45	45. organization	3	2.5
R/N46	46. game	4	3.5
R/N47	47. health	1	4.5
R/N48	48. view	3	5.5

Sentence Key:	*Type*	
1	Sense, Grammatical	12
2	Sense, Non-grammatical	12
3	Non-sense, Grammatical	12
4	Non-sense, Non-grammatical	12

D.3 Passage for Reading Rate and Reading Comprehension Tests

Chocolate is now enjoyed all over the world but until the late sixteenth century it was only found in Central and South America. For years the indigenous people of Central and South America, the Aztecs and the Mayans, had been making chocolate from cacao beans and consuming it as a drink. Cacao beans carried great importance for both cultures: the Mayans considered them to be a gift from the gods and the Aztecs associated them with fertility. Both societies even used them as currency. Cacao beans were made into a chocolate drink by roasting them and adding water and chilli spice. The mixture also had medicinal purposes and was an important part of many traditional ceremonies.

Christopher Columbus was introduced to cacao and chocolate on his last journey to the Americas in 1502. He took some beans back to Spain to show the king and queen, however, they were viewed with apathy. Chocolate didn't truly arrive in Europe until 1585, when a shipment of beans came from Mexico to Spain. At this stage chocolate was still served as a drink but the Spaniards replaced the chilli with milk and sugar to sweeten the bitter taste. Cacao beans were in short supply and Spain guarded the secret of chocolate jealously. However, the luxury began to spread across the rest of Europe during the 17th century. Italy was next to appreciate chocolate and it finally arrived in England in 1650. Chocolate was only available to the wealthiest, who consumed it in fashionable 'chocolate houses', much like today's coffee shops. The first chocolate house opened in London in 1657.

Chocolate remained a beverage for almost two hundred more years. It wasn't until 1847 that Joseph Fry created the first solid chocolate bar for eating. Others followed soon after, with John Cadbury adding a chocolate bar to his range. The solid chocolate bar was based on cocoa butter, extracted from cacao beans in a method developed by the Dutch chocolate maker, Casparus van Houten in 1828. Chocolate was still dark at this point and there was not quite the selection that we have today. In 1875 Daniel Peter produced the first milk chocolate bar using powdered milk; he was assisted in his work by Henri Nestle, a name still affiliated with chocolate today.

The world continues to be obsessed with chocolate. We still love to eat chocolate, and Switzerland currently consumes the most at 10kg per person each year. Contemporary chefs not only produce chocolate based deserts and puddings but also combine sweet and savoury by adding chocolate to main course dishes, to stews, meat pies and even brussel sprouts. Modern society has also found other uses for chocolate. Chocolate face masks and chocolate massages are just some of the inventive ways in which chocolate has been used in health spas and beauty salons! It seems that chocolate, though now far removed from that known to the Aztecs, is as relished now as it was then.

D.4 Reading Comprehension Test

1) Prior to the 16th Century, where was chocolate found?

2) The Mayans and the Aztecs enjoyed chocolate as a drink. Name three other purposes that it had for these people?

3) What does indigenous mean (see paragraph 1)?

4) In the context of paragraph 2, what does apathy mean?

5) How do you think Columbus felt about the King and Queen's reaction?

6) Why did the Spaniards replace the chilli with milk and sugar?

7) Give two reasons why it took a long time for other countries to be able to make chocolate drinks?

8) Why did the Spaniards guard the secret of chocolate with jealousy?

9) What formed the foundation of the first solid chocolate bar?

10) Who made it possible for chocolate to evolve from a beverage to a solid bar?

11) What was the difference between chocolate made prior to 1875 and chocolate made after 1875?

12) In the context of paragraph 3, what does affiliated mean?

13) In what ways are the modern uses of chocolate similar to that of the Aztecs and Mayans?

14) Today we eat and drink chocolate, what other modern uses for chocolate are given in this passage?

15) In the context of paragraph 4 what does relished mean?

D.5 Reading Comprehension Test Answer Key

Question	*Correct Response Score 1*	*Incorrect Responses Score 0*
1. Prior to the 16th Century, where was chocolate found?	Central & South America Must have both to score 1	Americas (Q)
2. The Mayans and the Aztecs enjoyed chocolate as a drink. Name three other purposes that it had for these people?	Currency, medical purposes, traditional ceremonies Must have all three to score 1	Fertility
3. What does indigenous mean (see paragraph 1)?	Originating in characteristic particular region or country; native	
4. In the context of paragraph 2, what does apathy mean?	Lacking in interest; indifferent; disinterested; not interested; do not care	Disdain; distaste; negatively; uncertainty; disapproval; not appreciated; disliked
5. How do you think Columbus felt about the King and Queen's reaction?	Disappointed; upset; hurt; gutted	Affronted; not happy; annoyed
6. Why did the Spaniards replace the chilli with milk and sugar?	To sweeten the bitter taste; to sweeten	To taste different **(Q)**
7. Give two reasons why it took a long time for other countries to be able to make chocolate drinks?	Short supply and Spain guarded it with secrecy Must have both to score 1	Cocoa beans were expensive
8. Why did the Spaniards guard the secret of chocolate with jealousy?	To maintain their monopoly on chocolate; to keep the price of chocolate high; it was in short supply so they wanted to keep it to themselves; they did not want anyone else to have it	Short supply (Q)
9. What formed the foundation of the first solid chocolate bar?	Cocoa butter	Cocoa butter

10. Who made it possible for chocolate to evolve from a beverage to a solid bar?	Casparus van Houten; van Houten	Joseph Fry; The Dutch; Dutch method
11. What was the difference between chocolate made prior to 1875 and chocolate made after 1875?	It was still dark before 1875; milk was added to it after 1875	Chocolate made after 1875 was made of powdered milk **(Q)** solid chocolate was made using milk and became lighter in appearance
12. In the context of paragraph 3, what does affiliated mean?	Connected to; associated with; attached to; linked to	
13. In what ways are the modern uses of chocolate similar to that of the Aztecs and Mayans?	Still served as a drink (beverage) ingested; used for purposes; purposes today; consumed; health medical	Used for purposes other than for eating **(Q)** Eating **(Q)**
14. Today we eat and drink chocolate, what other modern uses for chocolate are given in this passage?	Face masks & chocolate massages	Beauty salons and health spas **(Q)** Served as desserts; savoury dishes
15. In the context of paragraph 4 what does relished mean?	Enjoyed; savoured; loved; greatly enjoyed	Popular; appreciated

Appendix E – Listening Comprehension Tests

E.1 Cambridge Advanced English Exam 1

Listening Comprehension Test: **CAE 1**

Part 1
You will hear three different extracts. For questions **1 – 6**, choose the answer (**A**, **B** or **C**) which fits best according to what you hear. There are two questions for each extract.

Extract One

You hear two people on a music programme talking about the singer Nancy Graham.

1 What is the man's opinion of Nancy's second album?

☐ **A** He thinks it is very experimental.

☐ **B** He appreciates the continuity of style.

☐ **C** He wonders if she is lacking inspiration.

2 What do the two speakers agree about?

☐ **A** the freshness of the music

☐ **B** the lack of real emotion in the music

☐ **C** the calming effect of the music on the listener

Extract Two

You hear part of an interview with a woman who trained the winning horse in a top showjumping competition.

3 Why does she compare herself to an Olympic athlete?

☐ **A** to demonstrate how tough she had to be

☐ **B** to explain how she reacted to her victory

☐ **C** to emphasise how fortunate she was to win

4 How did she feel before her horse won the competition?

☐ **A** uncertain of the rider's ability

☐ **B** frustrated with the worsening weather

☐ **C** doubtful whether her horse was fit enough

Extract Three

You hear part of an interview with a food writer called Richard Capstick.

5 Richard decided not to become a chef because he lacked

☐ **A** adequate organisational skills.

☐ **B** a talent for inventive cooking.

☐ **C** the ability to make quick decisions.

6 What did Richard think about food writing before he got involved in it?

☐ **A** He considered himself well suited to it.

☐ **B** He regarded it as a hobby rather than a career.

☐ **C** He imagined a qualification was needed to do it.

Part 2

You will hear a marine wildlife photographer called Bruce Hind talking about his work. For questions 7 ─ 14, complete the sentences.

MARINE WILDLIFE PHOTOGRAPHER

Bruce says that [ ](7) is the most important aspect of his work.

Before going on a trip, Bruce makes [ ](8) of the photographs he hopes to take.

Knowing the type of photographs he wants to take helps Bruce to choose the right [ ](9)

Bruce disagrees with people who say his way of taking photographs is not [ ](10)

It's important to find out whether [ ](11) is needed to photograph in a particular place.

Bruce says that [ ](12) have spoiled several promising shots.

When at sea, Bruce generally keeps his cameras in a container designed for storing [ ](13)

He is particularly pleased when his photographs appear in [ ](14)

Part 3

You will hear part of a radio interview in which the comedian and writer Jane Clarkson is talking about her work. For questions **15 — 20**, choose the answer (**A**, **B**, **C** or **D**) which fits best according to what you hear.

---

15 What did Jane find difficult about writing a book?

☐ **A** She couldn't travel around the country.

☐ **B** She didn't get any instant reaction to her work.

☐ **C** She had to spend time looking after her daughter.

☐ **D** She found the process itself very challenging.

16 According to Jane, why did some critics dislike her novel?

☐ **A** They didn't think the book was funny.

☐ **B** They were dismissive of her initial success.

☐ **C** They thought her male colleagues were better writers.

☐ **D** They thought she should stick to being a comedian.

17 Which aspect of Jane's work as a comedian helped her to write?

☐ **A** her patience

☐ **B** her ability to listen

☐ **C** her habit of watching people

☐ **D** her rational way of thinking

18 According to Jane, how do many people react to female comedians?

☐ **A** They're convinced women can't tell jokes.

☐ **B** They're afraid the women will break down.

☐ **C** They find women's humour too intense.

☐ **D** They find women's jokes embarrassing.

19 What was the disadvantage of the stage image which Jane developed?

☐ **A** It frightened the audience.

☐ **B** It made the audience angry.

☐ **C** People thought it reflected her real personality.

☐ **D** People did not take her seriously any more.

20 Why does Jane prefer being a solo comedian to acting in a play?

☐ **A** She can choose where she works.

☐ **B** There is a greater range of roles.

☐ **C** It's more rewarding financially.

☐ **D** It's a more relaxing way of life.

Part 4

You will hear five short extracts in which people are talking about keeping fit.

TASK ONE

For questions **21 – 25**, choose from the list **A – H** the person who is speaking

TASK TWO

For questions **26 – 30**, choose from the list **A – H** what each speaker is expressing.

While you listen you must complete both tasks.

A an artist			**A** a pride in personal achievements		
B a fitness instructor	Speaker 1	(21)	**B** indifference to current trends	Speaker 1	(26)
C a sales manager	Speaker 2	(22)	**C** an enjoyment of a daily routine	Speaker 2	(27)
D a childminder	Speaker 3	(23)	**D** a commitment to taking regular exercise	Speaker 3	(28)
E a doctor	Speaker 4	(24)	**E** a desire to improve his or her diet	Speaker 4	(29)
F an office cleaner	Speaker 5	(25)	**F** awareness of his or her health problems	Speaker 5	(30)
G a secretary			**G** a reluctance to admit failure		
H a retired person			**H** resentment of another person's attitude		

E.2 Cambridge Advanced English Exam 2

Listening Comprehension Test: **CAE 2**

Part 1

You will hear three different extracts. For questions **1 – 6**, choose the answer (**A, B** or **C**) which fits best according to what you hear. There are two questions for each extract.

Extract One

You hear part of a radio discussion in which two musicians, Alan and Jodie, are talking about their careers.

1 How does Alan compare writing songs for albums and for films?
☐ **A** He prefers the freedom of thinking up his own subjects.
☐ **B** He realises that a film song will provide better publicity.
☐ **C** He welcomes the challenge of writing within certain guidelines.

2 Alan and Jodie agree that the music business has changed in that
☐ **A** some performers are less talented today than in the past.
☐ **B** singers have to know how to make use of the media today.
☐ **C** musicians are now expected to become successful very quickly.

Extract Two

You hear part of an interview with a successful fashion retailer called Jason Pendry.

3 How does Jason feel when he produces a new fashion collection?
☐ **A** critical of his design team
☐ **B** reluctant to give media interviews
☐ **C** apprehensive about losing customers

4 Why does Jason think he is more successful than other retailers?
☐ **A** He makes sure that he keeps pace with the latest trends.
☐ **B** He instinctively knows which clothes are going to be popular.
☐ **C** He provides a wide range of clothes for people to choose from.

Extract Three

You hear a man telling a friend the story of his encounter with a tiger while he was working in India.

5 What did the man do to put himself at risk?
☐ **A** He cycled into a forbidden area.
☐ **B** He ignored instructions he'd been given.
☐ **C** He failed to inform anyone where he was going.

6 How does he feel about his response to the situation?
☐ **A** disappointed by his sense of panic
☐ **B** modest about his own bravery
☐ **C** proud of the way he reacted

Part 2

You will hear a nature conservation worker called Brian Dover talking about his job. For questions **7 – 14**, complete the sentences.

CONSERVATION WORKER

Brian's parents used to have a [ ](7) so he met people who told him about wildlife.

The subject Brian chose to study at university was [ ](8).

Brian's present job involves both [ ](9) and practical skills.

He says dealing with [ ](10) can be difficult.

He says that what he particularly enjoys is seeing the [ ](11) of his work.

The decreasing number of [ ](12) on faming land is a big conservation problem.

His present project aims to join separate [ ](13) together.

Brian says people wishing to work in conservation may have to volunteer initially or do [ ](14) work.

Part 3

You will hear part of a radio interview in which two actors, Patsy Turner and Dale Green, are talking about their careers. For questions **15 – 20**, choose the answer (**A, B, C** or **D**) which fits best according to what you hear.

---

15 According to Patsy Turner, how can actors influence the writers of TV soap operas?

☐ **A** by proposing changes to characters they play
☐ **B** by altering the way they act their parts
☐ **C** by reflecting their characters' history to date
☐ **D** by discussing the success of current storylines

16 What makes Patsy continue acting in soap operas?

☐ **A** the feeling of security it gives her
☐ **B** the irregularity of the filming schedules
☐ **C** the enjoyment of working as part of a team
☐ **D** the challenge of reacting to changes in the plot

17 How has appearing in the popular soap opera affected Patsy?

☐ **A** She finds the level of attention rather difficult to deal with.
☐ **B** She likes the fact that ordinary people feel they know her.
☐ **C** She enjoys certain aspects of a celebrity lifestyle.
☐ **D** She feels the media intrusion has affected her work.

18 According to Dale Green, why are some people attracted to acting?

☐ **A** They long to play romantic roles.
☐ **B** They imagine it is a glamorous life.
☐ **C** They want to be admired by their peers.
☐ **D** They wish to go beyond their normal experience.

19 What is Dale's advice for out-of-work actors?

☐ **A** They should take the initiative to ensure they get goods roles.
☐ **B** They shouldn't worry about the quantity of roles they perform.
☐ **C** They should try to find alternative sources of income.
☐ **D** They shouldn't feel they have to accept sub-standard work.

20 For Dale, what is the most fulfilling part of being an actor?

☐ **A** using skills you have developed
☐ **B** gaining theatre critics' approval
☐ **C** making the audience think
☐ **D** taking part in large-scale projects

Part 4

You will hear five short extracts in which people are talking about the jobs they do now and the jobs they used to do in the past.

TASK ONE

For questions **21 – 25**, choose from the list **A – H**

the job each speaker used to do in the past.

TASK TWO

For questions **26 – 30**, choose from the list **A – H**

the aspect of their new job that each speaker appreciates most

While you listen you must complete both tasks.

A I was a lawyer			**A** dealing with people		
B I was a pilot	Speaker 1	(21)	**B** being my own boss	Speaker 1	(26)
C I was a sales manager	Speaker 2	(22)	**C** travelling abroad	Speaker 2	(27)
D I was a journalist	Speaker 3	(23)	**D** being able to spend more time with family	Speaker 3	(28)
E I was a teacher	Speaker 4	(24)	**E** being able to live in the country	Speaker 4	(29)
F I was a data processor	Speaker 5	(25)	**F** having variety in the work	Speaker 5	(30)
G I was a hotel owner			**G** working at a slow pace		
H I was a bank official			**H** being able to fulfill an ambition		

E.3 Cambridge Advanced English Exam 3

Listening Comprehension Test: **CAE 3**

Part 1

You will hear an archaeologist talking about an ancient civilization in North America. For questions **1-8**, complete the notes.

You will hear the recording twice.

THE PEOPLE OF FOUR CORNERS

Archaeological evidence:

Objects found:
- pots
- [ ](1)

The Region:

Rainfall pattern: [ ](2)

Description of soil: [ ](3)

Farming/Food:

Crops grown: [ ] *and* [ ](4)

Buildings:

Building materials used: [ ] *or* [ ](5)

Shape of meeting rooms: [ ](6)

History:

Wealthiest period: [ ](7)

How goods were moved: [ ](8)

Part 2

You will hear an announcement on the radio inviting people to take part in a tree-planting project. For questions **9-16**, complete the sentences.

Listen very carefully as you will hear the recording ONCE only.

TREE PLANTING

The name of the group organising the event is [ ](9)

The only piece of equipment you are asked to bring is a [ ](10)

The money to pay for the trees has come from [ ](11)

The trees are being planted on what used to be [ ](12) land

It is planned to make a number of [ ](13) among the trees for visitors.

The trees are being planted on the [ ](14) side of the village

The site entrance will be marked by a [ ](15) today

If you are going to help with the digging, you are advised to wear [ ] *and* [ ](16)

Part 3

You will hear a radio interview with Jourdan Kemp, an artist whose work is used on CD covers. For questions **17-22**, choose the answer **A, B, C** or **D**.

You will hear the recording twice.

---

17 Jourdan decided to train as an illustrator because he
☐ **A** knew he could get work in that field
☐ **B** knew other painters were better than he was
☐ **C** felt a painter's lifestyle would be too uncertain
☐ **D** felt he was more suited to illustration than painting

18 How did Jourdan first get involved in designing CD covers?
☐ **A** He made contact with a rock group
☐ **B** He was approached by a company representative
☐ **C** A lecturer put him in touch with the company concerned
☐ **D** A designer put his illustrations in a music magazine

19 Jourdan feels that when he started designing CD covers,
☐ **A** he charged too little for his work
☐ **B** he allowed the company to dictate the fees
☐ **C** he had unrealistic expectations about the fees
☐ **D** he set out to charge less than his rivals

20 Jourdan feels the record company gives him a lot of artistic freedom because he
☐ **A** knows the style of work they want
☐ **B** changes his drawings to suit their taste
☐ **C** gets inspiration from the band's music
☐ **D** produces work at an unusually fast pace

21 Jourdan agrees with Sally that the scenes he creates in his illustration are
☐ **A** joyful
☐ **B** childlike
☐ **C** unrealistic
☐ **D** unwelcoming

22 What does Jourdan say about the photographs he uses?
☐ **A** He discards a lot of them
☐ **B** He travels a long way to find them
☐ **C** He relies on them less than he used to
☐ **D** He is finding them harder to select than he used to

Part 4

You will hear five short extracts in which different people are talking about tourism.

TASK ONE

For questions **23 – 27**, choose from

the list **A – H** each speaker's occupation

TASK TWO

For questions **28 – 30**, choose from

the list **A – H** each speaker's aim for the future

You will hear the recording twice. While you listen you must complete both tasks.

A a travel broadcaster			**A** to increase the amount spent by clients		
B a hotel owner	Speaker 1	(23)	**B** to improve our circulation	Speaker 1	(28)
C a guide book publisher	Speaker 2	(24)	**C** to revive country skills	Speaker 2	(29)
D a tourist board representative	Speaker 3	(25)	**D** to raise standards overall	Speaker 3	(30)
E an environmentalist	Speaker 4	(26)	**E** to restore local transport networks		
F a family executive	Speaker 5	(27)	**F** to refurbish the rooms		
G a manager of a tourist attraction			**G** to attract a new type of client		
H a local government official			**H** to expand tourist accommodation		

E.4 Cambridge Advanced English Exam 4

Listening Comprehension Test: **CAE 4**

Part 1

You will hear a lecturer talking to students at the beginning of their course. For questions **1-6**, fill in the missing information.

You will hear the recording twice.

COURSE INFORMATION

Course name: [ ](1)

People not present will receive a [ ](2)

This will be forwarded by [ ](3)

The two topics for discussion today are the [ ](4) and how to study for the course.

How many TMAs are there? [ ](5)

TMSs must not be [ ](6)

Part 2

You will hear a radio announcement about travel problems on the railway. For questions **7-14**, complete the notes according to the information you hear, using one or two words or a time.

Listen very carefully as you will hear the recording ONCE only.

TRAVEL NEWS

Kind of accident: [ ](7)

Time of accident: [ ](8)

Type of train involved: [ ](9)

Name of the blocked line: [ ](10)

People injured: [ ](11)

Cause of accident: [ ](12)

Trains subject to delay:

Travellers from South Wales will arrive [ ](13)

For travellers from Gloucester to Swindon, everything is [ ](14)

Part 3

You will hear a radio interview with a researcher, Shirley Grainger, who has been investigating the working situation of actresses. For questions **15-20**, complete the statements.

You will hear the recording twice.

ACTRESSES AT WORK

Compared to men, the roles actresses play represent people who are [ ](15)

The survey was commissioned by [ ](16)

It covered three places of work: [ ](17)
[ ](18)
[ ](19)

As well as gender, age and type of role, researchers investigated [ ](20)

Part 4

You will hear **five** short extracts in which different people talk about losing jobs.

TASK ONE

For questions **21 – 25**, match the extracts as you hear them with the professions, listed **A – H**

TASK TWO

For questions **26 – 30**, match the extracts as you hear them with the statements about the speakers, listed **A – H**

You will hear the recording twice. While you listen you must complete both tasks.

A TV star			**A** appreciates a positive approach		
B secretary	Speaker 1	(21)	**B** has made financial changes	Speaker 1	(26)
C bank clerk	Speaker 2	(22)	**C** advocates adopting a routine	Speaker 2	(27)
D therapist	Speaker 3	(23)	**D** was disappointed about monetary arrangements		
E drama producer	Speaker 4	(24)	**E** expects to get a job soon	Speaker 3	(28)
F employer	Speaker 5	(25)	**F** has come to terms with the situation	Speaker 4	(29)
G advertising executive			**G** was given some warning of what was to happen		
H council employee			**H** admits to benefiting from the experience	Speaker 5	(30)

E.5 CAE Tests Answer Key

CAE 1

Part 1		Part 2		Part 3		Part 4	
1	C	7	Planning	15	B	21	B
2	B	8	Drawings	16	D	22	G
3	B	9	Equipment	17	C	23	C
4	A	10	Creative	18	B	24	A
5	A	11	Permission	19	C	25	E
6	C	12	Storms	20	A	26	C
		13	Food			27	H
		14	Magazines			28	F
						29	D
						30	B

CAE 2

Part 1		Part 2		Part 3		Part 4	
1	C	7	Newsagent shop	15	B	21	E
2	C	8	Botany	16	D	22	H
3	C	9	Management	17	A	23	C
4	A	10	Land owners	18	D	24	A
5	B	11	Benefits	19	A	25	F
6	B	12	Population of birds	20	C	26	E
		13	Cities			27	B
		14	Seasonal			28	D
						29	H
						30	A

CAE 3

Part 1		Part 2		Part 3		Part 4	
1	Tools	9	Green partnership	17	C	23	E
2	Irregular	10	Spade	18	B	24	G
3	Shallow	11	Government	19	B	25	D
4	Corn and beans	12	Farm	20	B	26	H
5	Brick or stone	13	Foot paths	21	D	27	B
6	Circular	14	Western	22	C	28	F
7	11th C	15	Yellow notice			29	D
8	Roads	16	Boots and gloves			30	G

CAE 4

Part 1		**Part 2**		**Part 3**		**Part 4**	
1	Art in Italy	7	Derailment	15	Younger	21	B
2	Handout	8	6.55	16	The Actor's Union	22	D
3	Mail / post	9	Goods	17	Theatre	23	G
4	TMAs	10	Wales – London	18	Television	24	F
5	Four	11	None	19	Radio	25	A
6	Late	12	Fallen tree	20	Earnings / pay	26	B
		13	(very) late			27	C
		14	(as) normal / OK			28	D
						29	A
						30	G

*17 – 19 in any order

REFERENCES

Ahn, M. (2013). The impact of subtitled online video clips on incidental vocabulary learning. *STEM Journal*, *14*, 135-151.

Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, *21*(4), 325-358.

Anderson, J. (2005). *Cognitive Psychology and Its Implications*. New York: Worth Publishers.

Attenborough, D. (Writer). (2005). Blue Planet. In BBC (Producer). U.K.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829-839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baharani, B., & Ghafournia, N. (2015). The Impact of Multimodal Texts on Reading Achievement: A Study of Iranian Secondary School Learners. *International Journal of Applied Linguistics and English Literature*, *4*(4), 161-170.

Bahrani, T., & Tondar, S. (2016). The Effectiveness of Exposure to Documentary Films with and without Subtitle As Authentic Language Input on Listening Comprehension of Intermediate Level Language Learners. *Modern Journal of Language Teaching Methods*, *6*(6), 86.

Baltova, I. (1999). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary* (Doctoral dissertation, University of Toronto).

Barekat, B., & Farrokhian, P. (2014). The Effect of Documentary Subtitles on Listening Comprehension of Iraninan Efl Learners at Intermediate Level. *Modern Journal of Language Teaching Methods*, *4*(3), 54.

Başaran, H. F., & Köse, G. D. (2013). The effects of captioning on EFL learners' listening comprehension. *Procedia-Social and Behavioral Sciences*, *70*, 702-708.

Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography, 6*(4), 253-279.

Bean, R. M., & Wilson, R. M. (1989). Using closed captioned television to teach reading to adults. *Literacy Research and Instruction*, *28*(4), 27-37.

Beglar, D. (2009). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*.

Bell, N. D. (2007). How native and non-native English speakers adapt to humor in intercultural interaction. *Humor–International Journal of Humor Research, 20*(1), 27-48.

Bell, T. (2001). Extensive Reading: Rate and Comprehension. *The Reading Matrix*, 1 (1).

Brown, G. (1977). *Listening to spoken English* (pp. 271-83). London: Longman.

Brown, G. (1990). *Listening to Spoken English*. London: Longman.

Brown, G. (1996). *Speakers, listeners and communication: Explorations in discourse analysis*. Cambridge: Cambridge University Press.

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

Burger, G. (2013). Research bibliography: Intralingual subtitles (captions) and foreign/second language acquisition. URL: http://www.fremdsprache-und-spielfilm.de/Captions.htm . Accessed 12.01.2017.

Burley-Allen, M. (1995). Listening: the forgotten skill. Wiley

Cambridge English Language Assessment (N.D.). Cambridge English: Advanced (CAE). Retrieved from http://www.cambridgeenglish.org/exams/advanced/

Campbell, D. T. & stanley, JC (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of experimental child psychology*, *33*(3), 386-404.

Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, *14*(2), 25.

Charles, T., & Trenkic, D. (2015). Speech segmentation in a second language: The role of Bi-modal input. *Subtitles and language learning*, 173-197.

Chen, M. L. (2012). Effects of the order of reading text or viewing a film and L1/L2 captions on reading comprehension. *Perceptual and motor skills*, *115*(1), 18-26.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, *3*(3), 149-210.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. Routledge

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Cox, B. (Writer). (2011). Wonders of the Universe. In BBC (Producer). U.K.

Crawford, I., & Wang, Z. (2015). The impact of individual factors on the academic attainment of Chinese and UK students in higher education. *Studies in Higher Education*, *40*(5), 902-920.

Culbertson, G., Shen, S., Andersen, E., & Jung, M. (2017, February). Have your Cake and Eat it Too: Foreign Language Learning with a Crowdsourced Video Captioning System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York: ACM* (pp. 286-296).

Cushing-Weigle, S., & Jensen, L. (1996). Reading rate improvement in university ESL Classes. *CATESOL Journal*, 9, 55-71.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 105-121). Cambridge: MIT Press.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting, 5*(1), 1-23.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3), 133-142.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. *The neurocognition of language*, 123-166.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of memory and language*, *25*(4), 385-400

Cutler, A., Mehler, J., Otake, T., & Hatano, G. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of memory and language*, *32*(2), 258.

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.

Cutler, A., & Shanley, J. (2010). Validation of a training method for L2 continuous-speech segmentation. In *11th Annual Conference of the International Speech Communication Association (Interspeech 2010),* (pp. 1844-1847).

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274-294.

Dimbleby, D. (Writer). (2007). How We Built Britain. In BBC (Producer). U.K.

Ebrahimi, Y., & Bazaee, P. (2016). The Effect of Watching English Movies with Standard Subtitles on EFL Learners' Content and Vocabulary Comprehension. *Journal of Applied Linguistics and Language Research*, *3*(5), 284-295

Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. *Individual differences and instructed language learning*, 181-209.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex.

Etemadi, A. (2012). Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International journal of English linguistics*, *2*(1), 239.

Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal, 75*(2), 173-180.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Flowerdew, J., & Miller, L. (2005). *Second Language Listening*. Cambridge: Cambridge University Press.

Ghasemboland, F., & Nafissi, Z. (2012). The effects of using English captions on Iranian EFL students' listening comprehension. *Procedia-Social and Behavioral Sciences*, *64*, 105-112.

Gilman, R. A., & Moody, R. L. M. (1984). What practitioners say about listening: Research implications for the classroom. *Foreign Language Annals, 17*(4), 331-334.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55-75.

Goldman, M., & Goldman, S. (1988). Reading with close-captioned TV. *Journal of Reading*, *31*(5), 458-461.

Gorjian, B. (2014). The effect of movie subtitling on incidental vocabulary learning among EFL learners. *International Journal of Asian Social Science*, *4*(9), 1013-1026.

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

Greenberg, S., & Ainsworth, W. (Eds.). (2006). *Listening to Speech : An Auditory Perspective*. London: Lawrence Erlbaum Associates.

Harji, M. B., Woods, P. C., & Alavi, Z. K. (2010). The effect of viewing subtitled videos on vocabulary learning. *Journal of College Teaching and Learning*, *7*(9), 37.

Harmer, J. (2008). How to teach English. *ELT journal*, *62*(3), 313-316.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, *42*(1), 181-192.

Hirsh, D. & Nation, P. (1992), What Vocabulary Size Is Needed to Read Unsimplified Texts for Pleasure?, *Reading in a Foreign Language*, **8** (2), pp. 689–96.

Hoogendyk, F. W. C., Arifin, Z., & Novita, D. (2014). Increasing students'vocabulary by using bimodal subtitling in short movie. *Jurnal Pendidikan dan Pembelajaran*, *3*(11).

Hu, J., & Huang, L. (2013). Subtitles and Chinese EFL Learners' Vocabulary Acquisition. *International Journal of Digital Content Technology and its Applications*, *7*(4), 359

Hu, M., & Nation, P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403-430.

Huang, H. C., & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems*, *28*(1), 75-96.

Hui, W. A. N. G. (2005). The Effects of Captions on Chinese EFL Students' Incidental Vocabulary Acquisition [J]. *Media in Foreign Language Instruction*, *2*, 011.

Karakas, A., & Sariçoban, A. (2012). The impact of watching subtitled animated cartoons on incidental vocabulary learning of ELT students. *Teaching English with Technology*, (4), 3-15.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.

Kothari, B., Pandey, A., & Chudgar, A. R. (2004). Reading out of the "idiot box": Same-language subtitling on television in India. *Information Technologies & International Development*, *2*(1), pp-23.

Kothari, B., Takeda, J., Joshi, A., & Pandey, A. (2002). Same language subtitling: a butterfly for literacy?. *International Journal of Lifelong Education*, *21*(1), 55-66.

Koumoundouros, T. (2015). What can pictures do for you? URL: https://thinkypictures.com/2015/08/16/what-can-pictures-do-for-you-2/. Accessed 27.03.2017.

Latifi, M., Mobalegh, A., & Mohammadi, E. (2011). Movie subtitles and the improvement of listening comprehension ability: Does it help?. *The Journal of Language Teaching and Learning*, *1*(2), 18-29.

Lin, C. Y. (2013). The use of segmentation cues in second language learners of English. PhD Thesis.

Liu, C. (2013). From language learners to language users: A study of Chinese students in the UK. *International Journal of Applied Linguistics, 23*(2), 123-143.

Lund, A., & Lund, M. (2013). Laerd statistics. *Lund Research Ltd. Available: statistics. laerd. com Accessed 03/09/2016.*

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501-533.

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge

Markham, P. L. (1989). The Effects of Captioned Television Videotapes on the Listening Comprehension of Beginning, Intermediate, and Advanced ESL Students. *Educational Technology*, *29*(10), 38-41.

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

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477.

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and instruction*, *13*(2), 125-139.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of educational psychology*, *93*(1), 187.

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

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of verbal learning and verbal behavior*, *20*(3), 298-305.

Meyer, C. F. (2010). *Introducing English Linguistics International Student Edition*. Cambridge University Press.

McClelland, J., & Rumelhart, D. E. (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*. Cambridge, MA: The MIT Press.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language, 45*(1), 103-132.

Mitterer, H., & McQueen, J. (2009). Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. *Plos One, 4*(11), 1-5.

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

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156.

Naghizadeh, M., & Darabi, T. (2015). The Impact of Bimodal, Persian and No-Subtitle Movies on Iranian EFL Learners' L2 Vocabulary Learning. *Journal of Applied Linguistics and Language Research*, *2*(2), 66-79.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society,* **62,** 714–719.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59-82.
Nation, P. (2010). Vocabulary Size. Retrieved from http://my.vocabularysize.com/

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The language teacher*, *31*(7), 9-13.

Neuman, S. B., & Koskinen, P. S. (1990). Using captioned television to improve the reading proficiency of language minority students. *Falls Church, VA: National Captioning Institute*.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191-243.

Paivio, A. (1986). Mental Representation: A Dual-Coding Approach Oxford Univ. *Press, New York*.

Paivio, A. (2006, September). Dual coding theory and education. In *Draft paper for the conference on "Pathways to Literacy Achievement for High Poverty Children," The University of Michigan School of Education*.

Paivio, A. (2007). Mind and Its Evolution: A Dual Coding Theoretical Debate.

Paivio, A. (2013). *Imagery and verbal processes*. Psychology Press.

Paivio, A. & Begg, I. (1981). The Psychology of Language. New York: Prentice-Hall.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding?. *Cognitive psychology*, *5*(2), 176-206.

Pak, H. H. (2012). The Effect of Films with and without Subtitles on Vocabulary Acquisition Scores of Foreign Language Learners: Using a Korean Sit-com, Unhindered High-Kick. *STEM Journal*, *13*, 17-31.

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

Press, C. U. (2012). English Vocabulary Profile. Retrieved from http://www.englishprofile.org/

Price, K. (1983). Closed-captioned TV: An untapped resource. *Matsol Newsletter, 12*(2), 1-8.

Quene, H., & Koster, M. (1998). Metrical Segmentation in Dutch: Vowel Quality or Stress? *Language and Speech, 41*(2), 185-202.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies*, 73-79.

Rodgers, M. P. (2013). English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions.

Rodgers, M. P., & Webb, S. (2017). The Effects of Captions on EFL Learners' Comprehension of English-Language Television Programs. *CALICO Journal*, *34*(1), 20-38.

Rost, M., & Wilson, J. J. (2013). *Active listening*. Routledge.

Rostam Shirazi, R., Hesabi, A., & Simin, S. (2015). Effects of pedagogical movie English subtitles on EFL learners' vocabulary improvement. *International Journal of Research Studies in Language Learning*, *5*(2).

Saed, A., Yazdani, A., & Askary, M. (2016). Film Subtitles and Listening Comprehension Ability of Intermediate EFL Learners. *International Journal of Applied Linguistics and Translation*, *2*, 29-32.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, *35*(4), 606-621

Sharif, M. R., & Ebrahimian, M. (2013). A Comparative Study of "Subtitled" vs."Auditory" Documentaries Comprehension among Persian Speakers of English. *International Journal of Applied Linguistics and English Literature*, *2*(2), 185-192.

Stein, R. (Writer). (2010). 'Far Eastern Odyssey'. In BBC (Producer). U.K.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50-73.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, *38*(3), 422-427.

Tsai, F. H. (2010). *Integrating Feature Films with Subtitles to Enhance the Listening Comprehension of Students Attending College in Taiwan*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of memory and language*, *28*(2), 127-154.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America,* **58,** 434 – 445.

Ur, P. (1984). *Teaching listening comprehension*: Cambridge University Press.

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from online education videos. *Journal of Media Psychology*.

van Heuven, V. J. J. P., & Hagman, P. J. (1988). Lexical statistics and spoken word recognition in Dutch.

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied Linguistics*, ams074.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3-25.

Vandergrift, L. (2011). Second Language Listening. *Handbook of research in second language teaching and learning, 2*, 455.

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

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System*, *18*(2), 221-234.

Vanderplank, R. (1992). Non-native speaker viewers' use of teletext subtitles in English: Some exploratory studies. *New technology in language learning. Frankfurt am Main, Germany: Peter Lang*, 89-102

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

Vanderplank, R. (2013). '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*, 1-16.

Vanderplank, R. (2014) Listening and understanding in Driscoll, P., Macaro, E., & Swarbrick, A. (Eds.). (2014). *Debates in Modern Languages Education*. Routledge.

Vanderplank, R. (2016). Captioned Media in Foreign Language Learning and Teaching.

Vroomen, J., & De Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human perception and performance*, *21*(1), 98.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*(2), 133-149.

Vroomen, J., Van Zon, M., & De Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, *24*(6), 744-755.

Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: the York adult assessment battery-revised. *Journal of Research in Special Educational Needs, 13*(1), 48-56.

Warner, Natasha 2005. "Examples of Reduced Speech", from <http:// www.u.arizona.edu/~nwarner/reduction_examples.html>.

Weber, A. (2000, May). The role of phonotactics in the segmentation of native and non-native continuous speech. In *Proceedings of the Workshop on Spoken Access Processes* (pp. 143-146).

Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it?. *Frontiers in psychology*, *4*, 433.

Wilson, J. (2008). *How to Teach Listening*. Essex: Pearson Education.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, *97*(1), 109-130.

Winke, P., Stafford, C., & Adams, R. (2003). Assessing working memory capacity in SLA research. *American Association of Applied Linguistics, Arlington, VA*.

Winke, P., S. M. Gass & T. Sydorenko (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology* 14.1, 66–87.

Yen, D., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*, (3), 1-7.

Yoshino, S., Kano, N., & Akahori, K. (2000). The effects of English and Japanese captions on the listening comprehension of Japanese EFL students. *Language Laboratory*, (37), 111-130.

Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *TOJET: The Turkish Online Journal of Educational Technology*, *8*(2).

Zarei, A. A. (2009). The effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall. *Pazhuhesh-e Zabanha-ye Khareji*, *49*(1), 65-85.

Zarei, A., & Saddeghi, M. (2011). The effects of synchronous and asynchronous interlingual and intralingual transcript presentation on L2 vocabulary comprehension and production. *Journal of Teaching English Language and Literature Society of Iran*, *5*, 101-123.

Zareian, G., Adel, S. M. R., & Noghani, F. A. (2015). The Effect of Multimodal Presentation on EFL Learners' Listening Comprehension and Self-Efficacy. *Academic Research International*, *6*(1), 263.

Zeng, J. J. (1997). *Ethnic minority students' experiences in British higher education* (Doctoral dissertation, University of Leeds).

--- End of document ---