Listening in a second language: a pupillometric investigation of the effect of semantic and acoustic cues on listening effort

Giulia Borghini

Department of Speech, Hearing and Phonetic Sciences

University College London

A thesis submitted for the degree of

Doctor of Philosophy

2019

I, Giulia Borghini, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Giulia Borghini

Abstract

Non-native listeners live a great part of their day immersed in a second language environment. Challenges arise because many linguistic interactions happen in noisy environments, and because their linguistic knowledge is imperfect. Pupillometry was shown to provide a reliable measure of cognitive effort during listening. This research aims to investigate by means of pupillometry how listening effort is modulated by the intelligibility level of the listening task, the availability of contextual and acoustic cues and by the language background of listeners (native vs non-native).

In Study 1, listening effort in native and non-native listeners was evaluated during a sentence perception task in noise across different intelligibility levels. Results indicated that listening effort was increased for non-native compared to native listeners, when the intelligibility levels were equated across the two groups.

In Study 2, using a similar method, materials included predictable and semantically anomalous sentences, presented in a plain and a clear speaking style. Results confirmed an increased listening effort for non-native compared to native listeners. Listening effort was overall reduced when participants attended to clear speech. Moreover, effort reduction after the sentence ended was delayed for less proficient non-native listeners.

In Study 3, the contribution of semantic content spanning over several sentences was evaluated using lists of semantically related and unrelated stimuli. The presence of semantic cues across sentences led to a reduction in listening effort for native listeners as reflected by the peak pupil dilation, while non-native listeners did not show the same benefit.

In summary, this research consistently showed an increased listening effort for non-native compared to native listeners, at equated levels of intelligibility. Additionally, the use of a clear speaking style proved to be an effective strategy to enhance comprehension and to reduce cognitive effort in native and non-native listeners.

Acknowledgements

My first and most heartfelt thank you goes to Valerie Hazan. I didn’t have a clear idea of what a PhD entailed when a few years ago Paul Iverson knocked at her door to introduce her to me. It was the first time of many others to come when we talked about second language and pupillometry (I had no idea that pupils could be measured at the time!). Since then, her door was always open for discussion. I am grateful for always encouraging me, even in pursuing those activities that had little to do with the topic of our research.

Thank you to Mark Huckvale, for his support and his valuable insights on my work at different stages of my research.

I am also very grateful to the UK Economic and Social Research Council (ESRC) whose funding, which was linked to grant number ES/J500185/1, made it possible for me to pursue this PhD.

I am thankful to Paul Iverson, for his valuable advice during the application process for my PhD, and for giving me the opportunity to test my teaching skills for the first time during his course.

Thanks to Outi Tuomainen, for the never-ending stats questions she answered, and for always being so generous in helping me. And also for comforting me when I realised how stupid it was to rent a very expensive mansard with no heating, in the middle of winter.

Thank you to Andrew Clark, for rescuing me several times when Matlab stopped working, usually shortly before a participant arrived.

I am extremely grateful to Matthew Winn for welcoming me to his lab at the University of Washington. The three months spent working together are one of the best memories of my PhD journey. I learned a lot from him, and he remains for me an example of passion for his work, as well as great intelligence and kindness. Thanks to Ashley Moore, for making me feel at ease from my day one in Seattle, for helping me carry out my research while I was there, and for always having a smile in the morning.

A mention is also due to those who participated in my tests, without whom this thesis would have been impossible. A special thank you is for those who lied, telling me that it wasn’t that boring!

Thanks to Anna, Max, Julie, Gwijde, Shiran, Chengxia, Gwen, Vanessa and Jieun, for the chats, coffees, lunches and chocolate biscuits around Chandler House. Thanks to my new desk mate Florian, for the painted Easter eggs, and his unbeatable formatting tips.

Thank you to those people who, despite being outside my university life, helped me to be happy during these years.

Thanks to Ambra, for always being so enthusiastic to meet me all around Europe for weekend adventures during the years I spent in London.

To Emanuela, for being interested in the details of what I was doing here, without ever seeming bored while listening to my stories about experiments and analyses.

Thank you to Jacopo, for letting me find a little piece of home in Seattle. And to be, after all, a piece of home himself.

Thank you to Lorenzo, for those times when he asked at the entrance of Chandler House to let him in to give me a surprise. And also for the other times when he cooked me a delicious dinner when I came home, even if he was a bit annoyed when I was late.

Finally, thank you to my mum and dad, for never having made me doubt myself.

Impact Statement

This project led to a better understanding of how listening effort during speech perception is differentially modulated in native and non-native listeners. Findings shed light on the effects of linguistic knowledge on speech communication, and on the various contributing factors to challenges faced by the non-native population during everyday communicative interactions. It has been previously shown that non-native listeners are more greatly affected by background noise when attending to speech, and are less able to take advantage of semantic context. However, these findings were mainly based on intelligibility performance, or only considered words in isolation. This project is novel in considering performance together with the evaluation of listening effort by means of pupillometry in native and non-native listeners during sentence processing. Results from this research contributed to our basic understanding of speech perception and related listening effort in the non-native population, and factors affecting it. These benchmarks will be of use for institutions or professionals working with, or to the benefit of, individuals that live in a country where a language different from their native one is spoken. This concerns different sectors, including educational staff working with second language learners, or with individuals that are acquiring new knowledge and skills by means of a second language. Indeed, implementing communicative strategies that are proved to lead to a lower cognitive effort can facilitate the success of both the communication itself, and also of activities carried out in parallel. It also includes the planning and delivery of public announcements in places such as airports or train stations, where non-native listeners are likely to be the majority of the message's receivers. Being able to efficiently deliver messages or instructions in situations where a multitude of people occupy a limited space (e.g. concerts, sports events and airports) can significantly contribute to maintain public safety. An additional desirable outcome includes the development of communication training programmes designed for medical staff and health and care professionals, particularly for those working in areas with a consistent number of non-native speakers. The reciprocal communication of sensitive information about personal health or treatment is indeed a crucial moment to assure that the medical staff can access to the needed information about patients, and to ensure the patients' understanding of medical recommendations. Therefore, easing the information exchange between medical staff and patients will potentially result in an improved patients' compliance, and importantly in a reduced level of communication-related stress.

Contents

List of Figures

List of Tables

Chapter 1

1. Introduction

Listening to and understanding a spoken foreign language is a vital skill for those who live their life in a country where a language different from their mother tongue is used. Being able to comprehend spoken language is indeed essential for everyday interactions, which often occur under complex or non-ideal circumstances. Therefore non-native listeners have to cope with the double challenge of both an imperfect signal (as for example during a telephone call or a conversation in a noisy bar) and an imperfect knowledge of the language (Lecumberri, Cooke, & Cutler, 2010). This is true not only for beginner learners, but it persists also after years of exposure, even after they have gained experience, practice and confidence in the non-native language. It is indeed possible for listeners to achieve an excellent level of comprehension in a second language, but this may require a greater cognitive effort.

Understanding the nature and extent of problems faced by non-native listeners in noise is important in developing and perfecting general theories of speech perception. Therefore, comparing the performance and the different strategies used by native and non-native listeners allows us to explore when and how linguistic knowledge is used to identify, segregate and understand the target speech we aim to comprehend. Practical applications are particularly relevant given increased world mobility, which means that an increasing number of people work or study on a daily basis in a second language environment. Knowing the factors leading to an increased cognitive effort for non-native listeners can inform us in optimising information transfer such as announcements in public transport or communication in safety-critical situations involving emergency procedures, in developing training material for teachers and educators, and in designing classroom and workspaces.

Most research on second language perception focuses on performance in terms of correctly understood words or sentences, but few explore the cognitive mechanisms behind the comprehension outcome. The recently introduced technique of pupillometry can be useful in clarifying those aspects, as the task-evoked pupil response has been demonstrated to be a reliable measure of cognitive effort (for overviews, see Laeng, Sirois, & Gredebäck, 2012b; Sirois & Brisson, 2014; Zekveld, Koelewijn, & Kramer, 2018). Additionally, because pupillometry consists of a time-series measurements, information gathered with this technique can provide us with meaningful insights on the deployment of listening effort over time. The purpose of this thesis is to investigate by means of pupillometry how cognitive effort for non-native listeners is modulated by the intelligibility of the listening task and by the availability of contextual and acoustical cues in the spoken language attended.

This chapter reviews the available literature on speech perception in a native and non-native language, and on the application of pupillometry in language perception research. First, section 1.1 gives an overview of the most accepted models in speech understanding, and discusses the main factors impacting on speech perception. Section 1.2 focuses on the specific features of second language speech perception, including the effect of speaking style and contextual information on comprehension performance. Section 1.3 explores the concept of listening effort and experimental methods used to measure it. It also introduces the technique of pupillometry, its advantages and disadvantages when applied to listening effort, and it reviews previous pupillometry literature investigating cognitive effort during speech perception. Finally, section 1.4 summarises the findings of this chapter, and gives an overview of the research questions and methodologies of this thesis.

1.1 Speech perception

Speech perception refers to the process by which the sounds of language are heard, interpreted and understood, namely mapped into a linguistic representation. Therefore, the aim is to extract meaning from a continuous stream of acoustic information. Two types of processing are implicated in the understanding of a verbal message. On one hand, the reception and processing of the acoustic information via the auditory system, and on the other hand, the exploitation of the linguistic and environmental context (Kalikow, Stevens, & Elliott, 1977). From an acoustic perspective, although spoken phonemes and words persist in the acoustic world for short fraction of a second, a reasonably skilled native listener is able to gather a consistent amount of information from these short lived acoustic signals such as words and sentences. Seemingly effortlessly , listeners are normally able to understand the linguistic message by decoding the key acoustic patterns, and also to infer a great deal of additional information, such as the speaker's gender, approximate age, the geographical area of origin, the emotional state and physical location in the space (Abercombie, 1967). The speech signal contains a number of spectral and temporal acoustic cues that help the listener to differentiate speech sounds belonging to different phonetic categories. A wide range of acoustic cues, that signal features such as voicing and manner and place of articulation, are exploited by listeners in order to recognise the phonetic properties of the speech sounds. Moreover, each phoneme normally contains redundant acoustic cues that mark its presence and identity. This redundancy becomes crucial when the complexity of a listening scenario reduces the availability and therefore the potential exploitation of individual acoustic cues. In such circumstances, the sum of the residual degraded cues, may still enable the listener to gain sufficient information to recognise a specific phoneme (Hsia, 1977).

However, as highlighted by Lecumberri et al. (2010), the phonemic inventory of all languages (31 phonemes on average) is exceptionally small compared to the vocabulary that it supports (hundreds of thousands of words). As a consequence, many words resemble each other, and the acoustic input the listener is exposed to, temporarily supports a range of possible alternatives. Additionally, important sources of variability in the acoustic characteristics of the signal also include:

- differences in the speakers’ anatomy and physiology (Fant, 1966);

- differences in the speaking rate (Gay, 1978);

- the effect of the surrounding phonetic context, i.e., coarticulation (Kent & Minifie, 1977);

- sociophonetic factors, such as regional, social and interactional markers (Docherty, 2003);

- the effect of the acoustic environment, such as the distance between speakers, and the place where the conversation takes place.

On the other hand, the semantic context in which a sentence is heard also has an impact on speech intelligibility. Indeed, word intelligibility has been shown to be a function of their predictability level when embedded in a sentence (Duffy & Giolas, 1974). More recently, research comparing native and non-native listeners revealed the differential use of contextual cues in L1 and L2 speech perception, and the close interaction with acoustic cues, particularly when the signal available to listeners is degraded (Bradlow & Alexander, 2007). Further details in this regard will be discussed in a later section.

Moreover, visual cues such as being able to see a speaker’s visible articulatory movements, the speaker identity and location in the space, significantly contribute to improve the listener’s ability to understand speech (Grant & Seitz, 2000), particularly when speech perception happens in noisy conditions. Specifically, a window of maximal integration has been shown to exist at an intermediate level of environmental noise, that is when the enhancement of speech recognition gained thanks to the visual cues is maximal (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

However, the multisensory integration of auditory and visual cues in speech understanding is even more sophisticated. The reciprocal weighting of auditory and visual cues is highly flexible depending on the availability of meaningful cues from the two channels (i.e. when only auditory or visual signal is degraded), and on the individual abilities and linguistic background of listeners and speakers (Hazan, Kim, & Chen, 2010).

All in all, the nature of spoken language perception presents listeners with a speech signal that is highly variable, fast, continuous, non-unique and redundant (Lecumberri et al., 2010). It follows that listeners cannot simply rely on the presence or absence of certain acoustic cues, but instead have to accomplish a complex categorisation task within a highly multidimensional space (Holt & Lotto, 2010). In the next section, some of the most influential models developed to account for the complex process of speech perception are briefly presented.

1.1.1 Models of speech perception

Any theory of speech perception must be able to explain how speech is decoded despite the high degree of variability, due to differences in speakers' characteristics, and to different phonetic contexts in which the speech sounds occur. In addition, theoretical models must account for the need of the listener to resolve a stream of speech that does not contain obvious discontinuities or evident cues for word segmentation, and that is delivered at a high rate of transmission in terms of sounds per second.

One group of theories, generally referred to under the umbrella name of *Motor theory*, highlights the link between speech perception and production. According to the Motor theory, information about articulation, such as place and modality, is extracted from the acoustic signal. Listeners maps the perceived signal to articulatory gestures used to produce sounds in order to perceive and understand spoken words (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

A different approach is supported by the group of theories generally labelled as *Auditory theory*, that interprets the process of speech perception as primarily auditory. In this thesis I will generally refer to the auditory approach, according to which the listener identifies acoustic patterns or features using the hearing and perceptual system also employed for any type of sounds, and matches them with learned acoustic and phonetic feature of the spoken language that are stored in the memory. What differentiates various theories of the auditory approach are different views around the relative weight given to recognising abstract and episodic representations of words and phonemes. The two extremes interpretations are the prototype and the exemplar theories.

The prototype theory suggests that when listeners hear a stimulus, this will be compared with a single abstract prototype in a category (Kuhl, 1991; Samuel, 1982), while exemplar theory suggests that a stimulus just perceived is compared to multiple exemplars of a given category stored in memory (Johnson, 2008). However, currently the focus of the debate in the field has moved towards how to combine contributions from abstract and episodic representations, and therefore in the formulation and perfection of hybrid models (Goldinger, 2007; Weber & Cutler, 2004). As an example, the Complementary Learning System (CLS) is an attempt to solve the abstract-episodic dilemma, by postulating that episodic and abstract representations are inter-dependent, and rely on different neural networks (Goldinger, 2007). On one hand, a fast-learning contribution from the hippocampus, which is specialised in memorizing specific events, and on the other hand a more stable integration from the cortex, that slowly extracts and learn statistical regularities form the single representations perceived. The CLS proposes that these two systems work in synergy, with cortical abstract representations created thanks to single accumulated episodes stored in the hippocampus and gradually consolidated. This view also allows the exclusion of the "catastrophic interference" of extreme single episodic events, such as exposure to regional accents. In this regard, the Filtering Listener hypothesis (Denby, 2013) argues that listeners do not store phonetically ambiguous percepts in their phonetic memory, therefore these ambiguous realisations do not contribute to update the phonemic category of the listener.

1.1.2 The role of working memory in speech processing

The working memory (WM) capacity of individuals is measured in terms of their ability to simultaneously store and process information (Rönnberg et al., 2013). The speech perception process can be seen as a real time competition of rival lexical alternatives, which is solved in favour of any one of them as soon as further evidence is available to the listener's auditory system that provides strong support for one of the alternatives. Evidence of this on-line processing comes from experimental results. When listeners are asked to make a lexical decision, that is, to decide if a presented stimulus is a word or not, they take longer if the non-word formed the beginning of meaningful lexical item rather than not (Taft, 1986). Working memory is therefore a crucial component of this process as the audio signal has to be held in memory while potential lexical alternatives are being compared, until the best match is chosen (Lecumberri et al., 2010). Differences in WM are a key factor in understanding individual differences in speech perception abilities (Heald & Nusbaum, 2014; Wingfield & Tun, 2007). Moreover, working memory is important for on-line language processing during conversation, for maintaining relevant semantic information, inhibiting the processing of irrelevant stimuli, and for attending to a specific audio stream selectively.

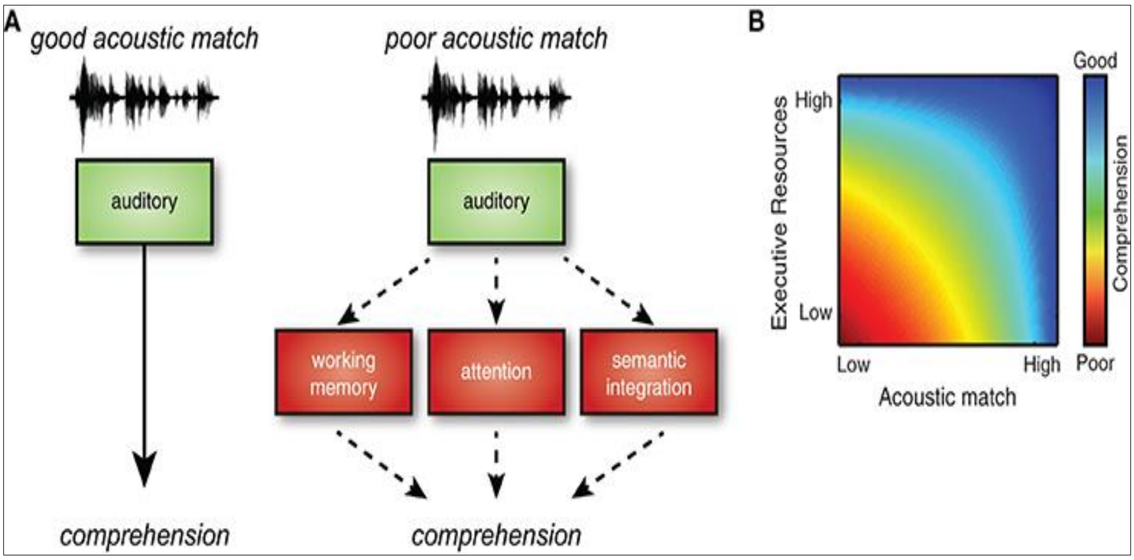

Figure 1.1 Speech processing in case of good and poor acoustic match. (Flege, Yeni-Komshian, & Liu, 1999)

The Ease of Language Understanding (ELU) model in its original and revised version (Rönnberg, Holmer, & Rudner, 2019; Rönnberg et al., 2013) also highlights the central role of the on-line comparison process between the input signal and the lexical representations stored in long term memory. Further to that, it claims that additional explicit WM is required if there is a mismatch between the speech signal input and the long term memory representation this is compared with (see Figure 1.1). In this regard, the demand on WM capacity is considered to be related to communication needs.

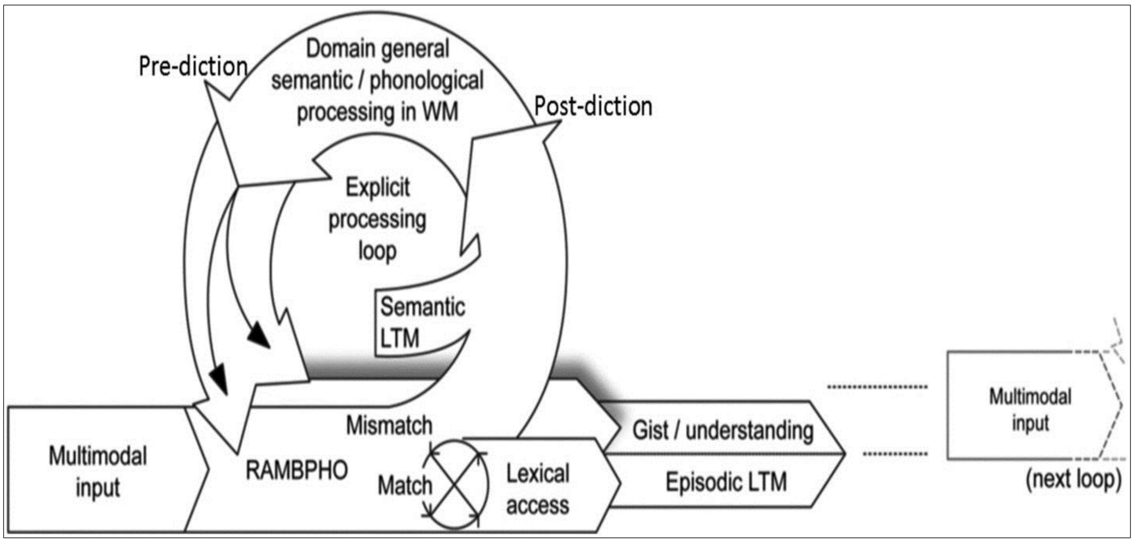

Figure 1.2 The Ease of Language Understanding model, from Rönnberg et al. (2019)

As shown in Figure 1.2, according to the ELU model, when communication is taking place in ideal listening conditions, the linguistic input (uni- or multi-modal) rapidly and automatically matches the mental lexical representation with enough degree of precision, and lexical access proceeds quickly and without additional explicit effort. On the other hand, when communication happens in sub-optimal listening conditions (i.e. signal distortion, background noise, non-native listener/speaker), an explicit contribution of WM is necessary to support listening and resolve the mismatch occurred. The aim of this explicit additional processing loop is to help in filling the gap made by missing information, using both phonological and semantic knowledge stored in long term memory. According to the authors of the ELU model, explicit and implicit processes run in parallel, the former being rapid, the latter being slower and more demanding in terms of cognitive resources, and together modulating the WM demand during speech perception (Rönnberg et al., 2019). Although the ELU model was specifically developed by researchers working on ageing and on the impact of hearing loss on speech perception, its components are also relevant to L2 learners who experience a different of difficulties.

1.1.3 Semantic context contribution to language perception

As noted back in the '50s by Miller, Heise, and Lichten (1951), a word is harder to understand if it is heard in isolation than if it is heard in a sentence. This is because while attending to speech, listeners make continuous predictions based on what they have already heard. The same mechanisms apply both to the limited context available within the sentence itself, and to the wider context derived from the communication settings. When listening is challenged by the presence of background noise, or other adverse conditions, contextual information and prior knowledge play an even greater role in assisting perception (Drager & Reichle, 2001; Wingfield & Tun, 2007). Zekveld, Rudner et al. (2011) showed that listeners performing a speech in noise perception task improved their performance when presented with semantically related text cues relative to non-words and mismatching cues. The benefit was largest for more difficult listening conditions, corresponding to 29% intelligibility level. Using a visual world paradigm, measurements of eye gaze also reveal the role of predictions in speech perception (for an overview see Huettig, Rommers, & Meyer, 2011). For example, when attending to a sentence beginning with "the woman will drink…" listeners will focus their gaze more quickly towards objects compatible with the action of drinking, such as wine, and away from other objects like a plant (Altmann & Kamide, 1999). Context also improves basic word intelligibility. The word "bowl" is easier to understand even if masked with noise when presented within the highly predictable sentence "The soup was served in a bowl", as opposite to when embedded in the neutral context "You're glad she called about the bowl" (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Kalikow et al., 1977).

Working memory is likely to play a role when listeners are asked to integrate the incoming information (the auditory stream) with the information stored in memory (prior knowledge or the context outlined by the preceding sentence segment), and to continuously evaluate prediction hypotheses (Kane & Engle, 2000). Speech perception abilities, particularly in challenging conditions, are indeed associated with working memory capacity and with the ability to read partly masked text (Daneman & Carpenter, 1980; Zekveld, George, Kramer, Goverts, & Houtgast, 2007). Individual differences in working memory abilities are also related to differential activation patterns in a speech sensitive network. A study from Zekveld, Rudner, Johnsrude, Heslenfeld, and Rönnberg (2012) showed that higher working memory capacity is associated with a greater ability in exploiting contextual cues during a speech perception task, and with a diminished activation in the left anterior inferior frontal gyrus and the left superior temporal gyrus. These findings support the neural efficiency hypothesis as formulated by Neubauer and Fink (2009) and the ELU framework previously outlined (Rönnberg et al., 2013), by confirming that individuals with higher cognitive abilities display a lower, and therefore more efficient, brain activation when performing cognitive tasks.

Brain imaging techniques can also be used to help clarify processing mechanisms on the basis of neural activation pathways. Two classes of account have been proposed to explain the effect of the semantic context on successful comprehension of degraded speech. One proposes that contextual benefit is driven by top-down processes that allow higher level content processing to influence peripheral perceptual mechanisms via top-down feedback excitatory connections (McClelland & Elman, 1986). Conversely, in feedforward accounts processing is seen exclusively as bottom-up, and the influence of context is interpreted as an integration of perceptual and semantic information that happens in higher-level brain regions. This integration mechanism does not imply a backward interaction between regions supporting higher cognitive and lower perceptual processes (Norris, McQueen, & Cutler, 2000). Davis, Ford, Kherif, and Johnsrude (2011) designed an elegant study to test the proposal that top-down neural processes could contribute to the perception of speech under challenging listening conditions, driven by coherent sentence-level context. In order to do so, location and timing of the blood-oxygen-level dependent signal (BOLD) were considered. Speech clarity and semantic context led to additional activation of both superior temporal and inferior frontal regions. These areas are primarily linked to low-level perceptual processing (the former) and high level semantic processing (the latter). Therefore, this provides initial evidence that relatively low-level perceptual areas show neural activity modulated by sentence content. However, when assessing the timing of responses in the two key areas, an earlier BOLD response from the temporal lobe preceded a later inferior frontal activation. Hence, this temporal pattern of neural activation does not support the idea that semantic context can aid comprehension via top-down mechanisms. Rather, it is compatible with the possibility that a context benefit when attending to degraded speech may arise at a later stage of processing, when multiple sources of uncertainty in the speech signal are combined. The bottom-up lexical decision between competing perceptual interpretations may therefore be delayed until lower- and higher-level representations can be integrated (Davis et al., 2011).

1.1.4 Language perception in adverse and real-life conditions

In everyday life, speech perception often occurs in non-ideal conditions. These include all the situations where there are added energy sources to the acoustic speech signal, such as reverberant energy, channel distortion, other conversations happening simultaneously in the same place, environmental noise in general. It is very common to have more than one element acting at the same time. As an example, when speaking over the telephone while walking on the street we will have a signal distortion due to the sound transmission over the telephone, and the added noise coming from the road. The majority of linguistic exchanges take place in presence of other sound sources. Factors affecting the ease of language understanding include the signal-to-noise ratio (SNR, which is the ratio of the intensity level of the signal to that of the noise), the capacity of the non-target noise to attract the attention of the listener (e.g. a competing speech will be more distracting than the noise produced by the washing machine, due to its informational content), and the sound source distance (for a review see Mattys, Davis, Bradlow, & Scott, 2012).

The presence of other sound sources results in the masking of the attended speech at two different levels, conventionally known as energetic and informational masking (Cooke, Lecumberri, & Barker, 2008). Energetic masking refers to the situation where competing signals overlap in time and frequency in such a way that portions of the target signal are inaudible (Brungart, Simpson, Ericson, & Scott, 2001). Therefore, energetic masking has to do with the physical acoustic interaction between signal and masker, and it affects speech perception by rendering unavailable to the listener potential cues for identifying phonemes and their boundaries. To a smaller extent, energetic masking can also interfere with the perception of prosodic cues. However, due to their long-term nature, suprasegmental information is usually easier to retrieve (Lecumberri et al. 2010). On the other hand, the definition of informational masking has more facets and, by including everything that reduces intelligibility once energetic masking has been accounted for, it is less clear (Cooke et al., 2008). Indeed, in the portions of the audio stream signal that are dominated by the masker, the information conveyed by it can interfere with the speech understanding process at various higher levels. A definition proposed by Brungart et al. (2001) refers to informational masking as a difficulty for the listener to segregate the target and the distracting signals that are both clearly audible. If it is easy to imagine how we can identify a picture based on a partial and puzzled view in a situation where our visual panorama is disrupted, the same metaphor is not so intuitive when dealing with acoustic signals. Indeed, contributions from different sound sources combine together additively, and the signal reaching the ears is a mixture of those different competing sounds.

Two important features of spoken language should be considered when studying speech perception in noise (Cooke, 2006). First, speech is a highly modulated signal in time and frequency, meaning that even at particularly adverse listening conditions (with a low SNR) there will be time windows dominated by the target speech signal. Second, there is a high degree of redundancy in the information conveyed by the speech signal, as proved by several studies on speech intelligibility after spectral filtering, temporal gating or spectrotemporal impoverishment (Drullman, 1995; Gustafsson & Arlinger, 1994; Kasturi, Loizou, Dorman, & Spahr, 2002). Redundancy enables listeners to identify phonemes and words even on the basis of an impoverished signal. Those two characteristics lead to the interpretation of speech perception in noise as based on the exploitation by listeners of those "glimpses" of the speech signal that are audible because are located in spectrotemporal regions that are least affected by the background noise. In this regard, Cooke (2006) conducted research showing that audible glimpses of speech in a variety of different masking conditions contain more than sufficient information to allow speech recognition in a computational model. This evidence suggests that a similar process may occur in human spoken language perception in noise. In addition to those acoustic features, native listeners also derive a significant benefit from semantic and contextual information during speech perception in noise (Bradlow & Alexander, 2007), meaning that they are able to take advantage of higher-level, semantic cues in order to compensate for an impoverished speech signal. The contextual benefit during speech perception will be discussed in greater detail in the next paragraph.

There is ample research investigating factors that impact on native speech perception in noise, even though this research typically used energetic masking only for investigation. An advantage for frequently used over uncommon words has been shown when listeners are choosing from an open list (Pollack, Rubenstein, & Decker, 1959; Savin, 1963). Moreover, a voice familiarity effect results in a greater ease in understanding speech pronounced by voices with which the listener has previously been familiarised (Nygaard, Sommers, & Pisoni, 1994). The same also applies for familiar versus unfamiliar dialects (Clopper & Bradlow, 2006). Further to that, Van Wijngaarden (2001) showed that the intelligibility of speech in noise is lower if the speaker is non-native instead of native for the given language. The overall difference in sentence intelligibility for native listeners of the language considered was found to correspond to a difference in SNR of approximately 3 dB. Interestingly, the main factor contributing to the degradation of speech intelligibility appeared to be the confusion of vowels, particularly those that do not occur in the speaker's native language. This again points for the need of a speech perception model that takes into account the episodic words and phoneme realisations that a listener is exposed to, as a factor contributing to the creation of a mental phonetic repertoire (e.g. the exemplar and hybrid models, as opposite to prototype models). By considering the temporal aspect of the speech signal, it is evident that under optimal listening conditions the beginning of words is heard and therefore processed before their end. However, evidence indicates that in noisy conditions native listeners are flexible in exploiting available cues, and that word recognition processing is equally sensitive to word-initial and word-final cues (Van der Vlugt & Nooteboom, 1986).

Taken together, the evidence shows that for native healthy listeners perceiving speech in a noisy environment is a challenging task, but that they are able to cope with it efficiently and rapidly.

1.1.5 Use and effects of clear speech speaking style

One strategy used to maintain a high level of intelligibility in difficult listening conditions is modifying the speaking style in order to make acoustic cues more readily available and more resistant to noise. The umbrella term “clear speech” is generally used to indicate a speaking style that is spontaneously adopted by talkers when they are aware they may have difficulty being understood, such as in a noisy environment or when their conversation partner is hearing impaired or does not share the same first language (Ferguson, 2004; Payton, Uchanski, & Braida, 1994). Clear speech is generally characterised by a decrease in the speaking rate, a wider dynamic pitch range, a larger vowel space, an increase in the frequency and duration of pauses (Smiljanić & Bradlow, 2009). These modifications have been extensively investigated in infant-directed speech, and appear to be relatively robust across different languages and across genders, although some minor differences were also reported (Fernald et al., 1989; Kuhl et al., 1997). Importantly, clear speech elicited under different circumstances significantly differs in its characteristics; its features are indeed primarily driven by the communication needs defined by the listener type and the communicative situations. As an example, it was found that, compared with British adult-directed speech, vowels were equivalently hyperarticulated in infant- and foreigner-directed speech. On the other hand, pitch was higher in speech to infants than to foreigners or adult British controls; and positive affect was highest in infant-directed and lowest in foreigner-directed speech (Uther, Knoll, & Burnham, 2007).

The use of a clear speech style has been shown to enhance intelligibility; however, the size of the clear speech benefit significantly varies across different populations of talkers and listeners. In order to investigate the production of clear speech in the most ecological way it is essential to consider spontaneous samples of speech produced in an interactive environment; this can be a challenge in a controlled research setting. The Diapix task (Van Engen et al., 2010) has been developed for this purpose, and consists of a method of eliciting dialogue-based speech recordings. The Diapix test material was then further extended with additional pictures with equal difficulty across the pictures pairs, and showing no learning effect when more than one picture task was completed (Baker & Hazan, 2011). By designing interactive studies, it has been shown that since a specific acoustic cue does not have the same facilitation effect for all listeners and all adverse conditions, talkers are able to adapt their clear speech style to the specific type of communication barrier that their listener is experiencing (e.g. in presence of multi-talker noise, or when dealing with manipulated speech), showing therefore a high degree of flexible adaptation to the listener needs. Hazan, Grynpas, and Baker (2012) found that listeners were faster in identifying consonants in the presence of a babble background noise when presented with keywords that were extracted from recordings produced in the same type of background noise, as compared to keywords produced via a vocoder even though both type of stimuli were rated as similarly clear.

To better understand their relative contribution, the beneficial effects of clear speech and semantic context was also examined in combination with other intelligibility-enhancing cues and under different masking conditions. Van Engen, Phelps, Smiljanic, and Chandrasekaran (2014) reported semantic context, clear speech and audio-visual presentation of the speech input as all improving intelligibility. In addition to that, authors found that the clear-speech benefit was greater when listeners attended to semantically anomalous sentences. This could suggest that enhancing acoustic-phonetic cues may be particularly useful when relying on semantic cues is not an option. Clear speech was also found to improve intelligibility to a greater extent in speech presented in audio-visual modality. Indeed, the articulatory gestures needed to produce clear speech may improve lip-reading. Alternatively, the visual support may help in segregating the target audio-stream, making in turn easier for the listeners to benefit from the acoustic enhancement available in the target speech. Lastly, results from this study suggested that the availability of semantic context only enhanced comprehension in the presence of energetic masking (not in the case of a competing speech stream). One possible interpretation suggested by the authors is that when inhibiting the processing of linguistic information available in the masker (as in the case of several competing speech streams), listeners have to deal with the unwanted consequences of a reduced ability to benefit from the semantic context of the target speech.

Experimental evidence so far accumulated has revealed that not only do listeners utilise a variety of intelligibility-enhancing cues to cope with challenging listening situations, but also that the advantage gained thanks to such cues interact with one another and with the type of environment in which they are immersed. This leads us to need to design studies that are able to capture the complexity and the interconnected nature of the topic.

1.1.6 Listening effort

The studies and results discussed so far addressed different aspects of speech perception by considering comprehension performance as the main index of a beneficial or detrimental effect. However, an experimenter or clinician might be interested not only in the ultimate accuracy in a listening task, but also in the mechanisms used to accomplish the task (as discussed when considering brain imaging studies), and how effortful it was for the listener to complete the task (Winn, Wendt, Koelewijn, & Kuchinsky, 2018; Zekveld, Kramer, & Festen, 2010). Because two listeners might be able to achieve the same performance (for example the same intelligibility score or the same SNR to achieve a certain level of comprehension) by exerting different amounts of effort, it follows that by only considering behavioural parameters we are not always able to capture all relevant aspects of the listening process. These considerations lead to a recently growing interest towards the concept of listening effort.

The expression "listening effort" can be intuitively understood as the feeling of being mentally tired when we need to work very hard in order to comprehend what is being said. Slightly different is the concept of mental fatigue, a more long term condition, resulting from effortful listening, which is instead involved in real time listening. Despite their close and intuitive link, there is currently little empirical evidence for this connection (McGarrigle et al., 2014). In the last decade, few formal definitions of listening effort have been proposed. McGarrigle et al. (2014) regarded it as the mental exertion required to attend to, and understand, an auditory message. The Framework for Understanding Effortful Listening (FUEL) defined listening effort as "the deliberate allocation of metal resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016). A highlight of this definition is the central role given to the active and voluntary engagement of mental capacity that the individual allocates in order to succeed in a task. The concept of listening effort and alternative ways in which it can be investigated is discussed in greater detail in section 1.3.

1.2 Non-native speech perception

The challenges faced during a listening task, such as the degree of similarity between words, inter-speaker variability and coarticulation, are in principle unchanged in any language, and occur by different degrees whether it is a native language, a relatively new or an unknown one. However, in addition to the factors listed above, additional variables are known to affect speech perception for non-native listeners. By broad categories those factors are the listener type, the influence of the first language (L1) and differences in the input to which listeners are exposed to (Lecumberri et al., 2010).

1.2.1 Second language spoken word recognition

As discussed above, word recognition has a central role during spoken language processing. This is true considering either first or second language perception. Indeed, the process of word recognition is not language-dependent, and the same system implying the activation of competing words and a following resolution of that competition will be in place regardless of the language (L1 or L2) that the listener is attending to (Lecumberri et al., 2010). However, an impoverished language knowledge as in the case of L2 listeners will be detrimental for speech perception in different ways. First of all, the availability of candidate words and their online selection while the speech signal becomes available strictly depends upon an accurate phonemic perception and representation. As a consequence, words candidates might be erroneously activated if the listener's impoverished L2 phonemic perception fails to rule them out, and this can lead to a delayed resolution of words competition for L2 compared to L1 listeners (Cutler, Weber, & Otake, 2006).

In addition to that, the lexical inventory of an L2 listener can be extremely reduced compared to an L1 listener. Therefore the target word may not even be available for selection. Furthermore, the competitor set for L2 listeners may contain words from their L1 vocabulary, which could make the selection process more effortful. Indeed it has been demonstrated by studies using eye-tracking techniques that even experienced L2 listeners during a word recognition task often activate words from their L1 vocabulary in parallel with L2 words (Spivey & Marian, 1999). This added competition has also been shown to be hard to overcome for L2 learners (Broersma & Cutler, 2011).

Finally, as discussed in previous paragraphs, higher-level processes are in place for helping in efficiently resolving lower-level (perceptual) ambiguity. However, in the case of L2 listeners, those higher-level processes are asked to perform a more complex task, because of the increased uncertainty due to impoverished vocabulary and perception ability, and to the simultaneous activation of L1 words. Moreover, the higher-level ability to resolve the word conflicts is itself less efficient for non-native listeners. Indeed this capacity relies on the listeners experience of the lexicon and syntax, and on the subtle pragmatic and contextual knowledge which guide lexical decisions for native speakers. Obviously, this deep and readily available linguistic knowledge is deficient in L2 listeners. It follows that, while non-native listeners have to rely more heavily on higher-level context to compensate for poorer perceptual abilities, on the other hand their higher-level resources are less efficient than in native listeners (Lecumberri et al., 2010). Further details will be discussed in the next sections.

1.2.2 Non-native listener proficiency level

In reviewing the plentiful literature dedicated to second language (L2) perception, a first significant difficulty emerges from the heterogeneity of the L2 learner grouping across published papers, and from the lack of clarity in the terminology adopted for describing the participants' linguistic background. Indeed, precise and well defined distinctions in the participants' language proficiency, both in their native and non-native language, are often hard to infer from written experimental records. A common distinction used in language perception research is between native and non-native listeners. Nonetheless, within this distinction the terms "second language" and "foreign language" are often used indiscriminately, thus creating ambiguity about the real nature of the listeners' linguistic background. The label "second language" applies to a language that is learned after full acquisition of a distinct mother tongue, and that is widely spoken in the geographical area and community where the listener/speaker lives during second language acquisition (e.g. in the case of immigrants). On the other hand, term "foreign language" defines a language typically learned through explicit instruction, and without the massive native exposure to the language (e.g. a foreign language learned in a class room) (Ringbom, 1980).

The lack of homogeneity of competence in the native language group, and a lack of standard in reporting the linguistic competence of non-native subjects across studies, are also important factors to consider when conducting and reviewing speech perception research. Differences in hearing acuity, in the familiarity with different regional varieties of the same language and in phonetic and phonological awareness can impact on experimental test performance. An extreme example of variable native and non-native competence is bilingualism, that can be seen as a continuous scale ranging from strict simultaneous bilingual to true monolinguals (Garland, 2007). Inconsistencies in the literature regarding the terminology used for describing native and non-native language proficiency are also exacerbated by the scarcity of language assessment tools that are comprehensive and reliable across languages. Attempts to fill this gap led to the development and validation across languages of self-assessment questionnaires, such as the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAP-Q has been shown to be an efficient tool for assessing the language profiles of multilingual healthy adults in research settings. However, due to a still limited agreement, it is difficult to conduct a solid comparison and generalisation of the experimental results reported by different authors. Moreover, the language background and experience of any individual is so multi-faceted that it would be extremely difficult to capture with any degree of precision by using a questionnaire.

Overall, previous research examined to what extent proficiency in a non-native language influences speech perception in noise. Francis, Tigchelaar, Zhang, and Zekveld (2018) showed that a higher level of English proficiency predicted a better ability to cope with less favourable levels of background noise for Dutch listeners. This trend emerged for both native and non-native speech perception (when the target speech was either English or Dutch), therefore reflecting a broad tendency for individuals with better English levels to perform better in all conditions. Similarly, Kilman, Zekveld, Hällgren, and Rönnberg (2014) explored how English proficiency affected native (Swedish) and non-native (English) speech perception under different type of energetic and informational maskers. The level of English proficiency was found to be a decisive factor in predicting performance for speech intelligibility in noise in all conditions, but only when the target language was English. Interestingly, listeners with a lower English proficiency benefitted from the largest release of masking when the target speech was masked by English babble noise, possibly because of a reduced understanding of the masker speech, which made easier to suppress its processing. Therefore, the effects of non-native proficiency level might affect speech perception in noise at multiple levels depending on the complexity of the listening environment.

1.2.3 First language influence

Although theories which see the first language as a "filter" to second language perception are now considered too strict because they exclude other processes such as developmental and universal mechanisms (Wode, Burmeister, & Ufert, 1980), the influence of L1 phonology on second language sound perception is still considered more influential than in other areas of L2 acquisition (Ellis, 1994). Indeed, the most accepted models of second language speech perception attribute a core role to L1 influence, and they interpret the level of difficulty in correctly perceiving and discriminating a L2 phoneme as closely related to its distance from, or similarity to, the learner's L1 phonology. The Native Language Magnet Theory (Kuhl, 1993) hypothesises the existence of native language phonetic prototypes that act as magnets during the perception of a second language. Thus, second language phonemes are harder to discriminate the closer they are to native prototypes. Authors of this theory also infer that previous language experience in a native language causes a distortion of the perceptual space, with a reduced sensitivity near L1 phonetic prototypes (Iverson & Kuhl, 1995).

Similarly, the Perceptual Assimilation model (Best, McRoberts, & Goodell, 2001) predicts the degree of difficulty in L2 phonemes discrimination on the basis of their similarity to L1 phonemes, but relying on the patterns of articulatory phonology that characterise each phoneme. According to this model, listeners compare the gestures, including timing, used by speakers for the production of their L1 sounds to those of the L2 they are attending to. As an example, in the case of "single category assimilation" (that is both of the L2 phonemes falling under the same L1 phonemic category), the discrimination between the two non-native phonemes would be extremely challenging, because they would be equally perceived as different realisations of the same L1 phoneme. Also Flege (1995), in his Speech Learning model, claims that the acquisition of L2 sounds can be more or less challenging depending on whether a sound is perceived as totally different, similar or identical to one of the learner's L1 phonemes. Specifically, when a new sound is perceived as similar to an L1 phoneme, this would be the most difficult case for the non-native listeners, because they would be more likely to erroneously assimilate the new phoneme to an already existing L1 category, and as a consequence the creation of a new L2 phonetic category would be hampered. The ongoing influence of L1 on the acquisition of a second language was also considered in a study from (Flege, Frieda, & Nozawa, 1997) analysing speech production of

Italian immigrants living in Canada. Results indicated that a greater use of L1 during everyday communication may also influence L2 production: individuals who spoke Italian relatively often had significantly stronger foreign accents than those who seldom spoke Italian.

A further consideration on the role of L1 influence during the course of a second language acquisition is about its dynamic contribution and its interaction with latent universal learning phenomena. It has been hypothesised that while at the beginning of the L2 acquisition process the influence of L1 has a crucial and dominant role, its relative importance decreases as the L2 exposure and acquisition proceeds (Major, 2001). Along similar lines, communicative efficiency between two talkers is influenced both by the alignment of talkers to the target language (whether the language they are attending is their native or non-native language), and by the alignment of talkers to one another in terms of language background, that is whether they share the same first language when communicating in a non-native language (Van Engen et al., 2010).

1.2.4 Quantity and quality of the second language input

Important differences between L1 and L2 acquisition include the different linguistic input that learners receive, in terms of both quantity and quality, in addition to the different age and neural maturation levels when L1 and L2 are acquired at different stages in life. The most recent and accepted models of second language acquisition propose a twofold contribution of the age factor, arguing that later exposure contributes both to the creation of different learning patterns if compared to first language acquisition, and to a variable quantity and quality of the linguistic input (Kuhl, 2000). Moreover, recent research in the field of language acquisition has highlighted the essential role of the quality and quantity of the linguistic input received as a result of a diverse range of social interactions even during first language acquisition in children (Kuhl, 2004). When dealing with foreign language learners, we can speculate that the impoverished quantity and quality of the language input compared with the one received by second language learners might be a dominant factor in their speech perception ability.

In general, the age of arrival in the host country for L2 learners conditions the quantity and quality of L2 input they are likely to receive, due to variable occupational and social environments, and to a different social motivation (Flege & Liu, 2001). Indeed adult immigrants are more likely to keep frequent contacts with other speakers sharing their same L1, and are likely to take part in less and less diverse interactions with L2 speakers if compared with a young native learner (e.g. through the school and other social activities). As well as the amount of exposure, the input quality is also a crucial factor during L2 acquisition (Lecumberri et al., 2010). The quality and diversity of L2 input can indeed range from a single and foreign-accented source in the case of a classroom context, to a native, diverse and variable speech which is the case of a natural linguistic environment.

A study evaluating the validity of the existence of a critical period for second language acquisition showed that while the degree of foreign accent ratings for L2 Korean learners of English was higher as a function of the age of arrival in the United States, various parameters of second language proficiency, such as different aspects of morphosyntax, varied as a function of the quantity of education received in English, and depending on how much L2 was used on a daily basis (Flege et al., 1999). Even when considering formal acquisition of L2 in a classroom setting, a recent large-scale 5-year longitudinal study conducted in Switzerland on secondary school students learning English, revealed that age-related attainment effects are overshadowed by other effects (Pfenninger & Singleton, 2018). An earlier start of English exposure was found to yield linguistic advantages only for children raised as bilateral simultaneous bilinguals (even if English was not one of the languages spoken at home). Substantial parental support (measured via a questionnaire considering direct parental involvement, frequency of reading and attitudes and beliefs of parents) was one of the main factors positively impacting on L2 acquisition, and it remained a robust predictor throughout adolescence. Taken together, this evidence highlights the crucial role of the quantity and quality of exposure in second language acquisition.

Further to that, in studies considering intensive phonetic training, providing highly-variable linguistic input during training was been shown to be essential for the formation of phonetic categories in L2 learners (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994). However it is still debated which specific effect intensive phonetic training has on the abstract phonemic representations of L2 learners. Iverson, Pinet, and Evans (2012) showed that both experienced and inexperienced English speakers are able to gain a similar benefit from high-variability auditory training, even though the exposure to English stimuli during their study was less compared to what experienced listeners would hear in their daily interactions with L1 English speakers. This supports the authors' conclusion that it is not the exposure alone to natural variable speech that improves performance in auditory training.

On a final note, evaluating with a good degree of accuracy the contribution of the quantity and quality of L2 input is one the greatest challenge in the field. First, because at present we are not able to quantify (not only to estimate) precisely what kind of input, and how much, learners receive in their L2. Second, it is not clear to what extent this input can modify long term memory representations of L2 phonemes, and how in turn these abstract representations guide L2 production (Flege, 2012).

1.2.5 Second language perception in real world conditions

It is well known that the detrimental effect of noise on speech perception is much stronger for non-native compared to native listeners, and this is true also when considering highly proficient second language learners. Indeed, several studies showed that highly competent non-native listeners were significantly less accurate than native listeners at speech recognition in presence of noise or reverberation, even if their performance was native-like under more favourable listening conditions (Mayo, Florentine, & Buus, 1997; Takata & Nábělek, 1990). Also bilinguals from infancy, that are often assumed by their peers and educators to have perceptual abilities identical to monolingual children, showed greater difficulties in speech recognition under suboptimal listening conditions (Rogers, Lister, Febo, Besing, & Abrams, 2006).

Interestingly, when considering energetic masking, the increased detrimental effect of background noise for non-native listeners is evident in tasks involving words or sentence recognition, but not when the potential help derived from high-level linguistic processing is minimised or eliminated, e.g. when phonemic units are considered (Cutler, Weber, Smits, & Cooper, 2004; Flege & Liu, 2001). As suggested by Lecumberri et al. (2010), the finding that word and sentence processing is more adversely affected for non-native listeners might indicate that a less effective use of phonotactic and contextual/semantic knowledge is responsible for the non-native disadvantage. On the other hand, attending to a meaningful speech stream when more than one intelligible speech sources are available requires additional cognitive load, and this can affect non-native listeners more adversely, since understanding a second language is already more challenging in itself. Moreover, under cognitive load, second language listeners show an inverted pattern of acoustic and lexical cues reliance compared to subjects attending to their first language. Indeed, Mattys, Carroll, Li, and Chan (2010) showed that while native speakers rely more on contextual plausibility than on acoustic cues when an enhanced cognitive effort is required simultaneously with the speech recognition task, non-native listeners do not. These results further suggest an increased difficulty for non-native listeners in exploiting lexical information, particularly under high cognitive load, presumably due to deficient lexical and semantic knowledge (Mattys et al., 2010).

Additionally, Brouwer, Van Engen, Calandruccio, and Bradlow (2012) examined whether the native and non-native listeners performance in background speech noise masking is also sensitive to the degree of similarity between the target and the distracting speech stream, other than to the listener’s familiarity with the masker speech. The investigation reported two key findings. First, the more the target speech matches the masker along linguistic dimensions (same or different language, meaningful or anomalous semantic content), the greater the difficulty listeners have to face, because of an increased linguistic interference. Second, subjects performed worse when they were familiar with the competing speech, arguably because this enables listeners to access to more easily available linguistic information about the masker speech stream, that is therefore more likely to act as a stronger distractor compared to when a less familiar language is on the background.

A further study from Kilman, Zekveld, Hällgren, and Rönnberg (2015) evaluated how hearing-impaired listeners perceive native and non-native speech in the presence of noise (stationary or fluctuating) and speech maskers. The speech maskers proved to be more detrimental than the noise maskers during both native and non-native speech perception. Additionally, a better hearing acuity and a larger working memory capacity were associated with a better performance when attending to a non-native language masked with fluctuating noise, suggesting that both individual characteristics are crucial when listening in highly taxing conditions.

Overall, it appears clear that the interplay between listeners' knowledge of the language, their individual characteristics and the linguistic release from masking is complex, and depends on how listeners differentially allocate their cognitive and attentional resources during speech recognition.

1.2.6 Use and effect of clear speech style in second language speech perception

It has been found that, compared to native listeners, non-native individuals exhibit only a small clear speech benefit, that is they improved significantly less than native listeners when presented with naturally produced clear speech as opposed to plain speech (Bradlow & Bent, 2002). In discussing these results authors highlighted the fact that non-native listeners showed a significant, even if smaller, clear speech benefit. This is due to the fact that clear speech production involves some modifications that are likely to benefit all listeners, such as a slower speaking rate and wider pitch range. Nevertheless, many of the variations made during clear speech are essentially "native listener oriented", and therefore being able to take full advantage of them requires a sound knowledge and experience of the phonology and phonetics of the target language.

The interaction between language background and speaking style adaptation was further investigated by testing high proficiency non-native listeners (Smiljanić & Bradlow, 2011). Fluent non-native listeners showed a large clear speech benefit when attending to sentences in noise pronounced by native talkers using a clear speaking style. In addition, proficient non-native talkers implemented speech modifications that enabled both native and non-native listeners to significantly improve their intelligibility performance. Interestingly, while intelligibility scores were improved, the accentedness ratings obtained for native and non-native speech remained constant in the conversational and clear speaking styles.

Overall, evidence shows that a greater proficiency and experience in L2 speech processing leads to improved intelligibility, higher tolerance to background noise and better ability to take advantage form speaking style modifications.

In parallel, it has been highlighted that a wide range of adaptations during communicative exchanges can enhance speech understanding for second language learners, and modifications of the speech itself only constitutes a fraction of those adaptations. Modifications in the interactional structure of conversation are indeed more consistently found, and are prone to vary with the prior experience and proficiency of the L2 speaker. As an example, native speakers interacting with L2 learners are likely to select conversational topics that are salient for their interlocutor in order to facilitate comprehension. Furthermore, some tactics are used to repair the discourse when misunderstandings occur, such as accepting unintentional topic-switches, repeating own and other's utterances and requesting clarifications (Long, 1983).

1.2.7 Semantic context contribution to second language speech perception

The observations reported in the previous sections, taken together with evidence of a less effective use of higher linguistic processing levels to compensate for a degraded speech signal for non-native listeners (Cutler et al., 2004; Mattys et al., 2010), prompted the design of further studies focused on understanding whether the non-native disadvantage in language perception could be overcome or relieved by enhancements at different processing levels. To this end, native and non-native subjects were presented with sentences where semantic and phonetic enhancements were available either alone or in combination (Bradlow & Alexander, 2007). Semantic enhancements consisted of the last word being highly predictable based on the sentence semantic context, while phonetic enhancements implied the use of clear speaking style. Results showed that non-native performance improved only when both acoustic and semantic enhancements were available at the same time. Conversely, native listeners were able to take advantage from both the facilitating conditions presented separately and in combination.

Studies using electroencephalography recordings (EEG) constitute an additional contribution to the understanding of semantic integration during second language speech processing. The N400 response is a negative event-related potential that has been linked to the ease of lexical access and integration, with greater response for more difficult words, or for words that are incongruent given the previous sentence context (Lau, Phillips, & Poeppel, 2008). Hahne (2001) found a delayed N400 response for coherent sentences in L2 compared to L1 listeners, suggesting that the semantic integration of the sentence final word was more difficult for listeners attending to a non-native language. Whereas the semantic benefit (intended as the difference in the N400 response between the semantically coherent and incoherent sentences) was significant in both native and non-native listeners, it was significantly smaller in the non-native group. This may indicate that when considering the ease of processing, non-native listeners do not benefit from the semantic context to the same extent as native listeners do. Similarly, Song and Iverson (2018) reported native listeners having a significantly greater N400 context-related differences between high and low predictable sentences compared to non-native listeners.

The extent to which non-native listeners are able to benefit from context seems to also depend on the age of second language acquisition. Mayo et al. (1997) indicated that early bilinguals (who learned the second language before the age of 6) could take better advantage of context than listeners who acquired their second language after puberty. Considered together these observations suggest that non-native listeners do not have a specific inability in taking advantage of semantic information, but rather need either increased audio quality of the signal or a greater employment of their speech processing strategies to be able to efficiently exploit those semantic cues.

1.3 Pupillometry and listening effort

Pupillometry is the measurement of pupil diameter and its change. Just out of curiosity, this research domain was originally named *pupillography*, because of the graphic trace that the machine recording pupil changes used to draw. However, with the spreading of more sophisticated electronic measuring devices, such as the modern eye trackers, the “graphic” part fell out of use, and therefore the term pupillometry replaced the old fashioned pupillography.

Towards the end of the nineteenth century, researchers first documented changes in pupil size not due to illumination conditions. As an example, the pupil was shown to enlarge when subjects were performing mental multiplication tasks (Heinrich, 1896). However, this early body of research did not reach the scientific world community and remained confined to the German neurological literature for longer than half a century (Beatty & Lucero-Wagoner, 2000), until the publication in Science of the first seminal paper showing a pupil response to emotionally relevant stimuli in 1960. This paper represents a milestone for establishing the method within the scientific psychological community. Hess and Polt (1960) showed that the pupil size of male and female adults increased when they viewed images of half-naked members of the opposite sex; in addition, female participants also showed a significant pupil dilation to pictures of babies. In the following years, further research showed that changes in pupil size could also express other kinds of cognitive processing, and do not only constitute a response to arousal or emotions. In particular, working memory load has been linked with greater pupillary responses in tasks of different nature, which include for example remembering strings of digits (Beatty & Kahneman, 1966), performing difficult mathematical calculations (Landgraf, Van der Meer, & Krueger, 2010) and attending to speech in difficult listening conditions (Zekveld et al., 2010).

In this section, a brief overview of different methods used in the literature to study listening effort is given. Afterwards, the physiological basis of pupil response are illustrated, and finally research using pupillometry to investigate language processing and listening effort is presented and discussed.

1.3.1 Listening effort: what it is, and how do we measure it?

As mentioned in a previous section, one of the most accepted definitions of listening effort was developed within the FUEL model, and refers to listening effort as the "deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a listening task" (Pichora-Fuller et al., 2016). According to this model, the available mental capacity is modulated by the arousal status, the allocation policy and the possible simultaneous activities to which cognitive resources are allocated. In turn, the allocation policy is affected by both involuntary (e.g. a sudden sound) and intentional (e.g. instruction to attend a target stimulus) attention, and by the general arousal level. The evaluation of the input-related demand is complex and affected by a range of factors. Overall, it reflects the individual's evaluation of the potential benefits derived from the successful completion of the task relative to the effort required to achieve that performance. In addition to that, fatigue, displeasure and low motivation may also result in task disengagement (with a consequent drop in performance and related effort) even if the available mental resources would have been adequate to perform the target task (Pichora-Fuller et al., 2016; Richter, Gendolla, & Wright, 2016). Despite and because of the complex interplay of factors affecting its modulation, listening effort is increasingly recognised as a valuable addition to the assessment of speech perception abilities, especially when it is likely that underlying differences in processing strategies do not become apparent when only accuracy is recorded.

While the definition and its experimental appeal are quite straightforward, the methodologies used for measuring it are still under debate among clinical audiologists and research experts in the field. Generally speaking, we expect a good measurement tool to be reliable and consistent, and to be sensitive in assessing differences in listening effort across different groups of participants and across different conditions.

Overall, the approaches available today can be grouped in 3 categories: (1) subjective, or self-report methods, (2) behavioural measures and (3) physiological measures.

Self-reported measures of listening effort consist in asking participants (either individuals taking part in an experiment, or patients being assessed by a clinician) questions about their perceived effort during the listening task. This can be done via closed-set questionnaires or rating scales. As an example, the Speech, Spatial and Qualities Hearing Scale (Gatehouse & Noble, 2004) includes some questions about listening effort in various real-world settings. Some researchers in the field have used subjective ratings as the main means to explore listening effort (Larsby, Hällgren, Lyxell, & Arlinger, 2005), but often these ratings are considered in conjunction with other behavioural or physiological measures (Gosselin & Gagne, 2011; Holube, Haeder, Imbery, & Weber, 2016; Mackersie & Cones, 2011; Zekveld, Kramer, & Festen, 2011). Advantages of the self-reported measure of effort are their ease and speed with which they can be administered, without requiring any specific expertise. On the other hand, their subjective nature itself is the cause of the main constraints. Indeed, the individual effort "threshold" can be different from person to person, and those individual differences cannot be measured. The interpretation of what listening effort is can also vary across participants, and some individuals could confuse it with task difficulty. Indeed, Moore and Picou (2018) showed that when participants are asked to rate mental effort, they are likely to judge the perceived performance instead, in order to simplify the complex task of evaluating the multidimensional concept of mental effort.

Participants likely substitute an easier question when asked to rate the multidimensional construct of mental effort. The results presented here suggest that perceived performance can serve as a ready heuristic and may explain the dissociation between subjective measures of listening effort and behavioural and physiological measures.

For those reasons, self-reported measurements are not an ideal index of listening effort.

Behavioural measures of listening effort have been widely used in hearing and language research. For simplicity, they can be categorised in 2 groups: single-task and dual-task paradigms. In **single task paradigms** experimenters record participants' response to stimuli (e.g. words or sentences played) by means of accuracy (Gatehouse & Gordon, 1990) or speed of response (Houben, van Doorn-Bierman, & Dreschler, 2013), and use these indexes as a measure for listening effort. As an example, a study found that response times in an identification task were slower in more challenging listening conditions (i.e. with a more unfavourable signal-to-noise ratio), even though the intelligibility level was optimal across all conditions (Houben et al., 2013). Authors of the study interpreted the increased time needed for responding as uncovering an increased mental effort required to understand speech in challenging acoustic backgrounds. Importantly, a diminished speed in language processing can result in an impaired communication due to the high rate of spoken language, and response times therefore represent an important factor to consider. However the relationship between the effort required to understand an auditory stimulus and the time needed to respond to it is not to be given for granted. Indeed, quicker answers might be on the contrary the consequence of a greater focused attention, as required by a more challenging listening environment; in that case the relationship would indeed be inverted (McGarrigle et al., 2014).

Multi-tasking methodologies capitalise on the assumption that each individual has a finite amount of cognitive resources, that has to be efficiently distributed between simultaneous mental operations (Kahneman, 1973). Therefore it is assumed that when a subject is asked to perform two mental tasks at the same time, if one of them (regarded as the primary task) becomes more demanding, this will lead to an impoverished performance in the other one (regarded as the secondary task). According to this logic, variations in the secondary task's performance can be considered as reflecting changes in the amount of effort required for performing the primary task. When specifically studying listening effort, various secondary tasks have been employed, ranging from memory tasks (Howard, Munro, & Plack, 2010) to tactile or visual recognition tasks (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Fraser, Gagné, Alepins, & Dubois, 2010; Gosselin & Gagne, 2011). A plus side of using a dual-task paradigm is its ecological validity; indeed having to efficiently allocate mental energies to multiple activities is a very common scenario in our everyday life. Nevertheless, a downside is that the validity of the methodology is based on the strong assumption that subjects are constantly fully utilising their mental capacity for both tasks. However it is very difficult to precisely estimate resources dedicated to each task. As a consequence, while dual-task paradigms can successfully be employed for evaluating the behavioural costs of an increasingly demanding listening environment, it remain unclear whether secondary tasks performances can constitute a reliable index of listening effort (McGarrigle et al., 2014). Recently, McGarrigle, Gustafson, Hornsby, and Bess (2018) compared primary and secondary tasks' performance in children with normal hearing or hearing impairments, within the same experimental design. The verbal response time at the primary task (a word recognition task in multi-talker babble noise) appeared to be a better measure of listening effort than the response time at the secondary visual task. The verbal response time at the speech perception test was indeed more sensitive to the detrimental effect of lower SNR, and to hearing loss.

Physiological measures refer to the recording of variations in central or autonomic nervous system during the execution of a task. Those include the use of functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) for the central nervous system, and changes in heart rate, skin conductance and pupil diameter when referring to the autonomic nervous system.

Using **fMRI** to assess the role of attention in effortful language, an increased activation of the left inferior frontal gyrus was found when participants attended to degraded speech compared to normal speech (Wild et al., 2012). Authors interpreted this response as a neural marker of effortful listening.

When considering **EEG studies**, alpha band activity has received attention from language researchers since it is thought to be involved in our ability to inhibit the processing of distracting information, such as a competing speech stream or a noisy background as in the case of listening in adverse conditions (Sauseng & Klimesch, 2008). However, the alpha oscillatory network seems to have a broader functionality within the cognitive processing. Indeed, an enhanced alpha activity was found both when increasing the memory load and the auditory stimuli degradation (Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012). The same EEG technique can also be used to look at **event-related potentials (ERPs)**. Specifically, the N1 ERP component (a large, negative polarity potential occurring approximately 100 milliseconds after a stimulus onset) appears to have an earlier peak and a greater amplitude when subjects are asked to process more degraded speech (Obleser & Kotz, 2011). N1 phase synchronisation (how well the single event-related potentials are time locked with the stimuli presented) has also been considered as a marker for cognitive effort, since it increases systematically with increasing task difficulty in a syllable detection task (Bernarding, Strauss, Hannemann, Seidler, & Corona-Strauss, 2013). Further to that, the N400 ERP component (a negative-going deflection that peaks around 400 milliseconds post stimulus onset) has been linked to the ease of lexical access, with a greater response for more difficult or unexpected words (Federmeier, 2007). Song and Iverson (2018) suggested that N400 can thus be regarded as a marker of effort at the lexical level, since any factor affecting lexical access, such as context, word frequency or repetition had been linked with N400 amplitude (see Lau et al., 2008 for a review). One advantage of the ERP technique is its high temporal resolution, which allows the researcher to strictly link stimuli presentation with cognitive processing.

Recently, **neural entrainment** has been proposed as an additional measure to investigate listening effort. It has been demonstrated that low-frequency (1-8 Hz) neural oscillation in the auditory cortex are phase locked to the speech envelope during listening (Luo & Poeppel, 2007). Nevertheless, the relationship between listening effort and neural entrainment is not unambiguous. Indeed, neural entrainment to the target speech can be selectively enhanced when a competing speech stream is also present (Ding & Simon, 2012), indicating therefore a greater effort needed to focus on the target speech. However, there is evidence too that the degree of cortical entrainment is linked to higher intelligibility (Peelle, Gross, & Davis, 2012), and it may therefore be greater when listening is less effortful.

One other physiological measure is **skin conductance**. The capacity of the skin to conduct electricity indeed can momentarily change in response to an external or internal stimulus, because our sweat glands are influenced by the sympathetic nervous system. An increase in the skin conductance has been shown to correlate with more demanding listening conditions (Mackersie & Cones, 2011). However, results are inconsistent, since recent research reported skin conductance, together with heart rate, to lack sensitivity to distinguish between different levels of background noise (Cvijanović, Kechichian, Janse, & Kohlrausch, 2017). Other physiological measures such as **skin temperature** and **heart rate** were also included in the study, but weren't found to consistently change across conditions. Due to the fact that it is a non-invasive procedure and it is relatively easy to administer, skin conductance measurement could be extremely useful for studying cognitive effort in those individuals from whom it is difficult to obtain reliable behavioural measures such as accuracy and response times, due for example to a physical disability. Even so, currently this approach in the field of listening effort research is very recent, and more replications with more complex experimental designs are still needed to test its reliability.

Finally, as mentioned earlier, **measurements of the pupil diameter** have been used in language research in last 50 years approximately. Further details about the physiology underlying it, and its applications in the listening effort research are discussed in the next sessions.

1.3.2 Physiological basis of the pupillary response

The pupil is the opening located in the centre of the iris of the eye; its function is to control the amount of light reaching the retina for the process of visual perception. It is therefore a contractile structure, and its diameter is controlled by two opposing groups of smooth muscles of the iris: a circular group called the *sphincter pupillae*, responsible for the pupil constriction and innervated by the Sympathetic nervous system, and a radial group called the *dilator pupillae*, which is in turn responsible for the pupil dilation and it is innervated by the parasympathetic nervous system.

Under normal conditions, in a healthy individual, the pupil can vary its diameter approximately from 3 to 8 mm due to changes in the lighting conditions (Laeng, Sirois, & Gredebäck, 2012a). However, changes that are cognitively driven are smaller, reaching approximately 0.5 mm, with a peak commonly in the time window from 0.7 and 1.5 seconds after the target stimulus presentation (Beatty, 1982; Verney, Granholm, & Marshall, 2004).

Additionally, changes in pupil diameter have been associated with the activity of the *Locus Coeruleus Norepinephrine* (LC-NE) system, which is involved in cognitive tasks as well as in general functions, such as the sleep-wake cycle and arousal (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010). The Locus Coeruleus is a neuromodulatory nucleus situated in the dorsal pons and it is involved with physiological responses to stress and panic, being the main site for the synthesis of norepinephrine (also called noradrenaline), a hormone and neurotransmitter. The Locus Coeruleus appears to have two modes of functioning: phasic and tonic, where the tonic activity is characterised by an elevated firing rate and a degraded task performance, while the phasic activity typically occurs in response to a task relevant event, and is associated with a lower LC activity. Gilzenrat et al. (2010) explored the relationship between the LC activity and changes in pupil size, showing that the baseline pupil diameter appears to be linked with the LC tonic activity, and with a cognitive state of exploration of the environment, which corresponds to a task disengagement. On the contrary, the LC phasic activity has been correlated with task-evoked pupil dilations, and with an increased task performance/engagement (Gilzenrat et al., 2010). Because of this link, pupillary responses have also been used to test the so called Adaptive Gain Theory, that predicts that the cognitive control state is driven by a continuous on-line assessment of task utility, with the aim of determining whether task engagement (exploitation) or disengagement (exploration) is most adaptive to the environment (Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999). Nevertheless, recent research suggested that pupil dilation is comodulated with cortical activity in general (Reimer et al., 2016). Overall, evidence indicates that the Locus Coeruleus might act as a hub that coordinates attentional related brain activity, rather than directly controlling pupil dilation (Winn et al., 2018).

This curious state of strict correlation but non-causality between pupil size and central cognitive processes, has been deemed not to be a problem, but rather a potential resource for neuroscience research. Beatty and Lucero-Wagoner (2010) proposed a fascinating parallelism between task-evoked pupillary responses and the use of reporter genes in molecular cell biology. A reporter gene is a DNA sequence that encodes for a protein of no interest per se rather than being easily detected when expressed. In molecular biology experiments, reporter genes are artfully attached adjacent to the gene of interest, thus providing a rapid and convenient way of measuring genetic events (Wood, 1995). According to the same logic, it has been suggested that cognitive research can profit by exploiting the naturally occurring pupillary response as a psychophysiological reporter variable, which is reliably correlated with central cognitive processing (Beatty & Lucero-Wagoner, 2000). Remarkably, measuring pupillary responses is a relatively inexpensive method compared to other physiological techniques, and is absolutely safe and non-invasive for participants taking part in the research. Together with eye movements (e.g. eye fixations and saccades), changes in pupil size provide us with real time insights into the structuring of cognitive processing. Additionally, because they occur from birth and without the need of overt responses or participants' collaboration, they are an optimal candidate for research with experimental subjects who are either preverbal, or not able to cooperate (Laeng et al., 2012a).

As a final consideration, task-evoked pupillary responses also fulfil the three criteria that an indicator is meant to accomplish (Kahneman, 1973). It is indeed deemed to be sensitive to:

- within-task variations in task demand due to the experimental manipulation of the task's condition (Zekveld & Kramer, 2014);

- between-task differences in processing load caused by qualitative differences in the cognitive operations required (Kramer et al., 2012);

- between individual differences in processing load, due to group differences in the tested population (Schmidtke, 2014).

As any other technique, pupillometry also comes with some limitations. First of all, because the main function of the pupil is to modulate the amount of light reaching the retina, its diameter is extremely sensitive to variations in lighting condition. Second, changes in pupil size are believed to represent a summative measure. They could therefore reflect the combined effect of any cognitive activation, such as anxiety, interest, age, intelligence, motivation, illness, medication (Siegle, Steinhauer, Stenger, Konecky, & Carter, 2003). Hence, when designing pupillometric studies it is essential to ensure that the experimental manipulation of interest affects the process which is the focus of the investigation, without affecting other processes that are not within the primary research scope. This is because there may be an unwanted confounding effect in the case of other simultaneously ongoing processes also affecting cognitive resources allocation (Zekveld et al., 2018). A clear example of a confounding factor for language research is represented by the loudness of stimuli, which elicits pupil dilation but might not be main object of interest (Liao, Kidani, Yoneya, Kashino, & Furukawa, 2016). Lastly, choosing a reliable baseline against which to compare changes in pupil size can be challenging, due to a varying noise level in pupil size. However, these factors can be minimised when rigorous research methods and design are carefully planned, in order to control for every potentially interfering variable. Finally, because individual pupil data tent to be extremely noisy (similar to EEG data), collecting a good amount of data by including enough trials and testing a good number of participants is essential in order to draw reliable conclusions (Winn et al., 2018).

In summary, the task-evoked pupillary response is believed to provide a reliable and sensitive measure of cognitive load when experiments are carefully designed. In the next paragraphs I will review the available literature that has applied pupillometry to study cognitive load in language processing, with a particular focus on listening effort in adverse listening conditions and in a non-native language.

1.3.3 Pupillometry in language and listening effort research

Some intrinsic features of pupillometry are particularly advantageous when considering listening effort. Pupillometry is indeed a time-series measurement, since multiple measures of the pupil diameter are recorded during the task execution. Timing is an essential component of the evaluation of listening effort, because attending to speech demands both a rapid auditory encoding, as well as a higher cognitive processing distributed over time during and after stimulus presentation (Winn et al., 2018). Effort might not be evenly distributed over a perceptual event, and pupil data have the benefit of showing changes in the dilation over an extended period, time-locked with significant landmarks in the experimental task. Moreover, pupillometry is suitable for testing individuals who use assistive devices, such as hearing aids and cochlear implants, because problematic interference of the device with electrical or magnetic techniques are avoided.

The use of pupillometry in language processing research has been quite productive in the last 40 years. Alongside studies on letter perceptions (Beatty & Wagoner, 1977), many investigations have focused on language processing load and listening effort, and utilised measures of speech perception performance and effort as complementary to each other (see Zekveld et al., 2018 for an extensive review). Overall, we can distinguish between studies focusing on input-related demands (characteristic of the experimental task that makes its completion less or more challenging), and research investigating internal factors, namely listeners' characteristics such as hearing loss or language knowledge. Input-related and internal factors might be manipulated either alone or in combination, and may affect different kinds of processes (such as linguistic, auditory or memory processing) at different stages of language comprehension.

Pupillometry has been widely applied to investigate effects of intelligibility on listening effort, and several studies have led to a good body of converging evidence. One of the focuses is to clarify what the influence is of the task characteristics on the pupil response. Kramer et al. (2012) considered a variety of listening tasks differing in the complexity of the auditory and linguistic information presented, ranging from passive listening to identification of meaningful words in background noise. Results showed that the pupil response was maximally sensitive to the task involving the processing of linguistic information, proving that task evoked pupillary responses are a robust and reliable measure of linguistic related listening effort. Performance has been commonly measured in terms of correctly understood sentences or words, across different listening conditions. The two most common ways of manipulating speech intelligibility are applying a masker signal (a noise masker or interfering speech) or degrading the quality of the auditory stimuli. Studies that consider the effect of spectral degradation often have the aim of reproducing signal distortion caused by hearing aids and cochlear implants, and overall indicate that more degraded sentences lead to lower intelligibility levels and to larger pupil dilation (Bernarding, Strauss, Hannemann, Seidler, & Corona-Strauss, 2014; Winn, Edwards, & Litovsky, 2015). Speech intelligibility has also been controlled by adaptively varying the SNR level while targeting different intelligibility levels. In a study conducted by Zekveld et al. (2010), participants were asked to perform three speech reception threshold tests. When listeners were presented with sentences in stationary noise, the peak dilation amplitude of the pupil, mean pupil dilation and peak latency consistently increased when intelligibility level of the speech presented decreased, indicating a greater cognitive effort required for processing speech with a reduced intelligibility. Moreover, regardless of SNR level, the pupil response was higher for incorrectly reported sentences, and an order effect was also found in the blocks presented, with the pupil response being greater in the experimental condition presented first. Because changes in pupil size are recognised as being related to a more intense and effortful utilisation of working memory for mental processing, these results also support the hypothesis made by the Ease of Language Understanding model: when listening conditions are more challenging, speech comprehension is more reliant on explicit and effortful exploitation of working memory capacity (Rönnberg et al., 2013).

These findings were later expanded by considering a wider range of intelligibility conditions (from 0% to 99% intelligibility), and the relationship between pupil dilation and SNR has proved to be more complex than just a linear correlation (Zekveld & Kramer, 2014). Overall pupil dilation has been revealed to be at its maximum for intermediate levels of intelligibility, and lower for easy intelligibility conditions. However, at very low level of SNR (i.e. sentences are very difficult to comprehend because of very high background noise) results are less straightforward to interpret. Indeed, at very demanding levels, inter-individual differences in the ability to read masked sentences and in the level of task engagement also played an important role in modulating pupil response. In general however, accumulating evidence shows that when intelligibility is very low the pupil response is smaller compared to intermediate intelligibility levels (Ohlenforst et al., 2017; Zekveld & Kramer, 2014). This may reflect the listeners' tendency to give up when it is very hard to achieve a successful result, thereby reflecting the impact of low motivation on pupil response. The reduced pupil dilation for very low levels of performance also corroborates a possible impact of the *evaluation of demands on capacity* mechanism on the allocation policy, as included in the FUEL model (Pichora-Fuller et al., 2016). Indeed, motivation has been shown to affect the pupillary response. For example, the pupil dilation was found to be greater when individuals were more curious to find out the answer to trivia questions (Kang et al., 2009). On the contrary, if the experimental task is not engaging enough for participants, changes in the pupil size are likely to be uninformative for research purposes, because not aligned with stimulus presentation (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013).

It is important to acknowledge that different studies have used different types of background noise, but since we know (see section 1.1.4 – Language perception in adverse and real-life conditions) that different types of noise can affect speech perception to varying degrees, it is reasonable to think that listening effort will be differently affected too. To test the effect of different masker types on listening effort, Koelewijn, Zekveld, Festen, and Kramer (2012) measured pupil dilation during a speech perception task where spoken sentences were presented in stationary noise, fluctuating noise and in the presence of a single-talker masker at comparable levels of intelligibility. Crucially, results showed a larger pupil response for the single-talker masker compared with both the other masker types, although behavioural results indicated a slightly better performance for the single-talker masker compared to the fluctuating noise. This pattern of results clearly indicates how the presence of informational masking (as opposed to purely energetic masking) is responsible for an increased listening effort, and that mental effort is not only affected by intelligibility level. Instead, the linguistic interference of a competing speech seems to have a main role in determining the mental effort required to understand a target speech stream. A greater pupil response during speech perception masked by a single talker when compared to fluctuating noise and stationary noise masking was then confirmed by a further study, corroborating the hypothesis that both auditory and cognitive processes are reflected by pupil dilation during speech perception in challenging conditions (Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014). Target-masker similarity has also been explored, revealing that when speech is masked by a single talker, a more dissimilar masker may facilitate target-masker segregation and therefore ease the cognitive processing of the target speech. Interestingly, while both gender and location differences between target and masker speech aided comprehension, only gender differences resulted in a reduction in the pupil response (Zekveld, Rudner, Kramer, Lyzenga, & Rönnberg, 2014). Additionally, pupillometric results from Francis et al. (2018) suggested that when target and masker languages are matched effort is greatest. This pattern was found when listeners were presented both with their native and non-native language as a target.

Alongside studies considering manipulation in intelligibility and masker type, a large body of research has focused on the pupil dilation response elicited by changes in the linguistic complexity of spoken stimuli. Larger pupil responses have been reported when attending to semantically difficult words, and for low-frequency relative to high frequency words, particularly for high storage-load conditions (Chapman & Hallowell, 2015; Elshtain & Schaefer, 1968). Additionally, lexical manipulations such as increasing word frequency, reducing lexical competition, facilitating semantic processing and presenting stimuli with a sparser neighbourhood density also led to a reduction of the pupil dilation response (Kuchinsky et al., 2013; Kuipers & Thierry, 2011; Schmidtke, 2014). Moreover, pupil dilation was found to be a good predictor of learning outcome for unknown words, being correlated with memory accuracy and confidence during a word-learning task (Papesh, Goldinger, & Hout, 2012). Converging evidence indicates that the pupil response is also sensitive to various types of manipulation at the sentence level. Pupil diameter was found to increase when the syntactic complexity of sentences presented was higher, and interestingly correlated more strongly with grammar complexity than with subjective rating of sentence comprehensibility, suggesting therefore that pupillometry may successfully track the on-line cognitive load imposed by the level of grammar complexity (Schluroff, 1982). This finding was replicated by examining cognitive load when participants were attending to sentences with centre-embedded relative clauses, which are known to impose a high demand in terms of working memory. Results showed greater pupillary responses, and a delayed peak, for object-relative compared to subject-relative centre-embedded clause items, showing how grammar complexity results both in greater and longer cognitive processing (Just & Carpenter, 1993). Sentence ambiguity was also found to impact on pupil dilation: sentences containing ambiguous locutions elicited larger pupil responses compared with unambiguous sentences (Ben-Nun, 1986). Similar findings supported the impact of sentence complexity on pupil response (Kramer et al., 2012; Piquado, Isaacowitz, & Wingfield, 2010; Wendt, Dau, & Hjortkjær, 2016). A further study also investigated the contribution of prosody in the modulation of cognitive load by using pupillometry, showing that when prosodic cues conflicted with the syntactic structure in sentences containing a temporary ambiguous syntax, pupil diameter and thus cognitive load increased (Engelhardt, Ferreira, & Patsenko, 2010). Finally, attending to predictable sentences led to a more rapid reduction of listening effort compared to unpredictable sentences (Winn, 2016).

Despite the large body of research in the field of non-native language perception, and the increased interest in measuring listening effort, few studies have used pupillometry to investigate non-native speech comprehension. A study considering the complex task of simultaneous translation, showed among other results that repeating back words in a non-native language entailed an increased pupil dilation compared to the same task performed in the speaker's native language (Hyönä, Tommola, & Alaja, 1995). More recently, a pupillometry study investigating spoken word recognition considered the performance of three groups of participants: monolingual English speakers, early and late Spanish-English bilinguals (Schmidtke, 2014). Pupil response was delayed for bilingual compared to monolingual listeners, and a larger neighbourhood effect was obtained for bilingual compared to monolingual listeners. Neighbourhood density is a measure of the number of competing words temporarily matching the speech signal, and is therefore an index strictly linked with working memory demands, since the competing words have to be held in working memory until the incoming speech stream is disambiguated. Since the competitor word set of L2 learners may include words from their L1 (Spivey & Marian, 1999), this additional competition it is likely to be a source of increased mental effort during spoken language comprehension for non-native compared to native listeners. Researchers also reported a greater word frequency effect for late bilingual compared to monolingual and early bilingual individuals, with an increased mental effort required to retrieve less common words. Interestingly, within bilingual participants, higher English proficiency was associated with an earlier pupil response, and with a smaller effect of word frequency and neighbourhood density. However, these previous studies only considered single word recognition in quiet, without therefore directly addressing the challenges of everyday communication (Hyönä et al., 1995; Schmidtke, 2014). Another study combining eye-tracking and pupillometry investigated the added cognitive load needed for bilingual individuals to process language switches within a sentence. Bilinguals, both at the beginning of development and in adulthood, are affected by language switches in terms of increased cognitive load, even when listening to simple sentences (Byers-Heinlein, Morin-Lessard, & Lew-Williams, 2017). Although providing interesting insights on the mind's ability to cope with complex language environments, this study does not address the additional challenges faced by non-native listeners who acquired a second language later in life, and the frequent need to deal with suboptimal listening conditions. Finally, Kruger, Hefer, and Matthew (2013), although they did not directly compare native and non-native speech perception, designed a very ecologically valid study. University students with Sesotho as a first language were asked to watch a recorded English-spoken lecture with or without English subtitles. The subtitled condition created lower cognitive load in terms of changes in the pupil diameter compared to the non-subtitled condition, which in turn engendered a higher level of frustration in students.

In summary, although there is not a large number of pupillometric studies exploring speech perception in a non-native language, evidence from the literature suggests that an increased effort is required for non-native compared to native listeners. The additional effort seems to be related not only to the intelligibility levels of the speech signal, but also to specific factors affecting second language comprehension in noise, such as a reduced linguistic experience in the non-native language. Moreover, preliminary evidence seems to suggest that listening effort in second language learners might be also modulated by individual characteristics, such as linguistic experience and proficiency in the L2, and other cognitive abilities (working memory and phonological memory abilities). Given its relatively early stage, and the number of aspects to be investigated further, pupillometry research in the field of second language perception promises to be an interesting breeding ground to pursue.

1.3.4 Pupil outcome measures

Within the literature applying pupillometry to investigate listening effort, it is customary to consider four main pupil outcome measures:

- ***Pupil baseline:*** the average pupil diameter in the time window (generally 1 or 2 seconds) preceding the stimulus onset.
- ***Mean pupil dilation*** relative to baseline pupil diameter, calculated over a time window of variable duration, depending on the experimental procedure, and on the type and length of stimuli used.
- ***Peak pupil dilation:*** the maximum positive deviation from the baseline during a time window of variable duration, often the same as for the mean dilation calculation.
- ***Latency*** of the peak dilation amplitude.

These outcome measures are visually displayed in Figure 1.3, where the dotted and the solid line represent two distinct experimental conditions. Their interpretation is strictly linked to the experimental stimuli and design. Generally speaking, mean pupil dilation is believed to provide a more reliable and stable measure of cognitive resource allocation compared to the peak pupil dilation (Ahern & Beatty, 1979; Verney, Granholm, & Dionisio, 2001). Interpretation of the baseline measure in relation to the LC activity has been discussed in section 1.3.2. In some cases, when the deployment of cognitive effort over time within certain time windows is of specific interest, the evaluation of the pupil response could also be approached with time-series methods within these carefully selected time windows (Winn et al., 2018).

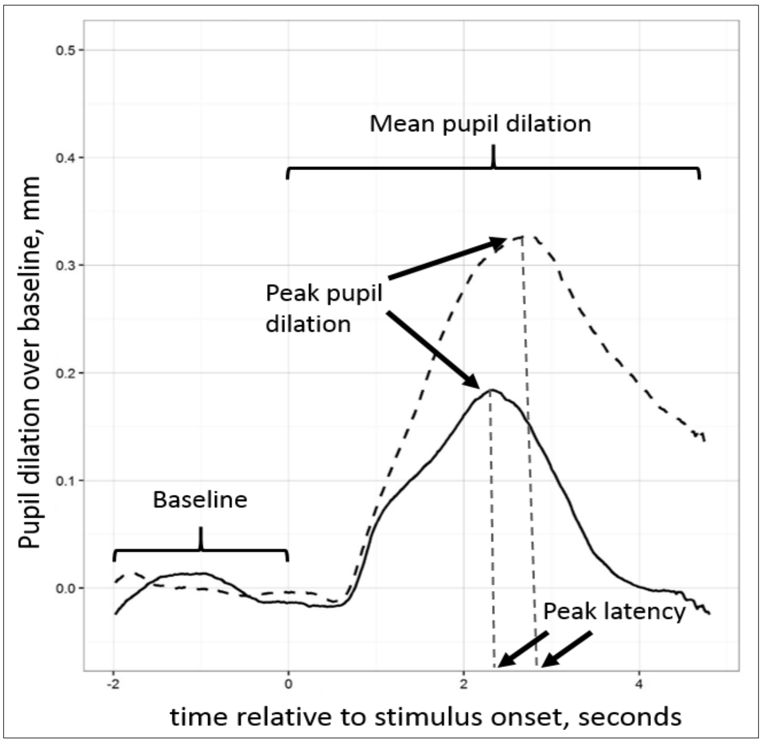

Figure 1.3 Visual display of the most commonly used measures of pupil response: mean and peak pupil dilation, baseline and peak latency **(adapted from Borghini & Hazan, 2018).**

In the literature, two methods for baseline correction have been mainly used: proportional (or divisive) baseline correction (corrected pupil size = pupil size/baseline) and subtractive baseline correction (corrected pupil size = pupil size - baseline). The subtractive baseline correction will be used for the three studies presented in this thesis. This is because it has been shown to be less strongly affected by data distortion (e.g. blinks, data loss) compared to proportional baseline correction (Mathôt, Fabius, Van Heusden, & Van der Stigchel, 2018). Additionally, subtractive baseline correction is believed to be a simple but powerful technique for normalising changes from 0 mm. This can in turn make experimental replication easier, by allowing comparisons between experiments regardless of individual differences in anatomy or discrepant luminance levels across laboratories (Reilly, Kelly, Kim, Jett, & Zuckerman, 2018). For these reasons, pupil response for the three studies will be shown in millimetres, and not in percentage change.

1.4 The current research

The current research is novel in applying pupillometry to investigate differences in listening effort between native and non-native listeners during a sentence processing task in noise, at equated levels of intelligibility. The general purpose of the present investigation is to gain insights into the factors affecting listening effort in non-native listeners, by comparing native and non-native listeners' pupil response during a speech perception in noise task. Specifically, I compared the listening effort experienced by native and non-native participants when the intelligibility level and the availability of acoustic and semantic cues were manipulated. The proficiency level of non-native listeners was also taken into consideration. The research presented in this thesis consisted of three Studies.

In Study 1 (Chapter 2), listening effort for native and non-native listeners was compared in quiet and in noise at two matched levels of speech intelligibility. An adaptive procedure was used to match the intelligibility levels across participants and conditions. I aimed to investigate whether native and non-native listeners performing at the same accuracy level differed in terms of cognitive effort required, and whether intelligibility level differentially modulated the listening effort for native and non-native participants (e.g., if the same increase in task difficulty led to a greater increase in listening effort for non-native individuals). It was predicted that the listening effort reflected by the pupil response would be higher for non-native listeners when compared to native listeners for a given intelligibility level. This is because listeners were expected to allocate a greater amount of cognitive resources when attending to a second language compared to their native language. It was also hypothesised that increases in task difficulty would cause pupil response to change at a steeper rate for non-native compared to native listeners, because of the previously documented increased detrimental effect of noise on non-native compared to native speech perception. Additionally, it was expected that the listening effort reflected by the pupil response would be higher at lower relative to higher intelligibility levels, in line with previous research.

Study 2 (Chapter 3) explored how the presence of acoustic and semantic cues affected listening effort in native and non-native listeners. The research design featured predictable and semantically anomalous sentences, both presented with a plain and a clear speaking style. An adaptive procedure was again used to equate the intelligibility level across listeners and conditions. A wide range of proficiency levels for non-native listeners was included in this study. It was predicted that the availability of semantic cues in the stimuli presented would overall reduce listening effort.

Lastly, the long-term contribution of semantic content was evaluated in Study 3 (Chapter 4). In order to manipulate the presence of a semantic context, participants were presented with lists of semantically related and unrelated sentences. It was predicted that listening to semantically related sentences would result in a reduced listening effort compared to unrelated sentences.

Overall, five main research questions were formulated and addressed across the three Studies:

1. Do native and non-native listeners performing at the same accuracy level differ in terms of cognitive effort required?
2. How does the presence of acoustic enhancements modulate listening effort in native and non-native listeners?
3. How does the availability of semantic cues within a sentence affect listening effort in native and non-native listeners?
4. Is long-term context beneficial for native and non-native listeners in terms of listening effort reduction?
5. How does proficiency level affect listening effort in non-native listeners?

In the final chapter, findings from the three Studies are discussed in light of the ELU model. Additionally, suggestions are made on how the model might be expanded to better account for challenges faced by non-native listeners during speech perception. Limitations of my research are also discussed, together with potential improvements to overcome them. Lastly, future research directions and some conclusive remarks on the value of listening effort evaluation by means of pupil response are presented.

Chapter 2

2. Study 1

2.1 Aim of the study

The purpose of the first study was to compare listening effort experienced by native and non-native listeners during a speech perception in noise task when their performance in the test is matched. Specifically, native and non-native listeners' pupil responses at two matched levels of speech intelligibility were compared. The present study was designed to answer the following research questions:

1. Do native and non-native listeners performing at the same accuracy level differ in terms of cognitive effort required?
2. Does intelligibility level differentially modulate the listening effort for native and non-native participants (e.g., does the same increase in task difficulty lead to a greater increase in listening effort for non-native individuals)?

With respect to the research questions reported above, the following predictions were formulated:

1. It was predicted that the listening effort reflected by the pupil response would be higher for non-native listeners when compared to native listeners for a given intelligibility level. This is because it is expected that listeners will allocate a greater amount of cognitive resources when attending to a second language compared to their native language.
2. It was expected that the listening effort reflected by the pupil response would be higher when the intelligibility level was lower compared to when it was higher, in line with previous research (Winn et al., 2015; Zekveld et al., 2010).
3. It was hypothesised that increases in task difficulty would cause pupil response to change at a steeper rate for non-native compared to native listeners, because of the previously

documented increased detrimental effect of noise on non-native compared to native speech perception (Mayo et al., 1997; Takata & Nábělek, 1990).

2.2 Materials and Method

2.2.1 Participants

Fifty adults from two different language backgrounds took part in the experiment. The first group included 27 participants (18 women and 9 men) with Italian as L1 and English as L2, aged 20–35 years (M = 28.4, SD = 4.1). The second group included 23 native British English participants (15 women and 8 men), aged 18–32 years (M = 23.3, SD = 4.2 years). All participants had been living in the UK for at least 10 months. Participants were recruited from the UCL Psychology subject pool and from social media. They reported not to suffer from cataracts or diabetes, and to not have used drugs or medications in the 48 h prior the experiment. Moreover, they were able to fixate the cross appearing on the screen without glasses or contact lenses. These selection criteria were chosen because of their potential impact on pupil dilation. All participants provided written informed consent to participate and received monetary compensation for their participation. The study was approved by the Ethics Committee at University College London.

2.2.2 Stimuli and Tests

Background Tests

All participants were screened using pure tone audiometry to ensure that their hearing thresholds were 20 dB HL or better at octave frequencies between 250 and 8,000 Hz. At the beginning of the experimental session, all participants carried out a set of background tests. The aim of these tests was to obtain a cognitive profile for each participant including measures which previous research suggested to be related with the ability to perform a speech perception task in noise (Besser, Koelewijn, Zekveld, Kramer, & Festen, 2013; Flege et al., 1999). Specifically, for each participant, the following tests were administered:

Digit span, forward, and backward (Wechsler, Coalson, & Raiford, 2008). This is commonly used as a measure of verbal working memory storage capacity. The test was administered in the participant's first language (either English or Italian). Both the forward and backward memory tests consisted of 7 pairs of items, with increasing number of digits. The score corresponded to the maximum number of digits the participant was able to repeat back without mistakes.

Phonological short term memory test: the Children's Test of Non-word Repetition (CN-Rep) (Gathercole, Willis, Baddeley, & Emslie, 1994). This consists of 40 non-words from 2 to 5 syllables length (e.g., “diller,” “defermication”) preceded by 2 practice items. Non-words were recorded from a native English speaker, and played from a loudspeaker. Answers were recorded and evaluated *post-hoc*, the score obtained corresponded to the number of non-words correctly repeated.

In addition, non-native participants were asked to complete an online linguistic background questionnaire designed to collect information about their level of self-reported English proficiency, their language usage, and their perceived cultural identity (see Appendix A). The questionnaire was designed by adapting questions from two different sources: the Language History questionnaire (Li, Zhang, Tsai, & Puls, 2014) and the Language Experience and Proficiency Questionnaire (Marian et al., 2007). Participants were also recorded while reading aloud a short story, "Arthur the rat" (MacMahon, 1991). A British English native speaker (without TEFL training) not involved in the study subsequently rated the degree of foreign accent of their speech on a scale from 1 (= native-like) to 7 based on a sentence extracted from the speech recorded (*Arthur stood and watched them hurry away. "I think I'll go tomorrow," he calmly said to himself, but then again "I don't know; it's so nice and snug here.")*. Given that all non-native participants were from the same L1 background (Italian) and that the same sentence was used for the rating, the rating provided us with a measure of relative accent within the L2 participant group. The aim of these tests was to obtain an accurate linguistic profile for the non-native participants included in this study, in order to be able to explore any correlation between listening effort and language use and proficiency.

Experimental Stimuli

Sentences presented in the study were taken from the Basic English Lexicon (BEL) sentence materials (Calandruccio & Smiljanic, 2012) which include 20 lists of 25 sentences. BEL sentences were specifically developed to test speech recognition for various listener populations, and contain lexical items and syntactic structures appropriate for use with non-native listeners. BEL sentences were reported by authors to be equivalent in difficulty across lists for native English listeners with normal hearing. The evaluation was conducted on native speakers of English in order to gain insights into the relative difficulty across lists on a more homogeneous subject pool compared to non-native listeners. The BEL sentence corpus is reported in Appendix B.

Each sentence has four keywords, which were used to score comprehension. Examples of the sentences are: "The PARK OPENS in ELEVEN MONTHS," "My DOCTOR WORKS in that BUSY HOSPITAL" (keywords in capital letters). Sentences were recorded in an anechoic chamber with a sampling rate of 44.1 kHz, 16-bit resolution, and produced by four native Southern British English speakers (two females) at a natural self-paced rate. The mean sentence duration was 1.9 s, (range: between 1.6 and 2.6 s). The four speakers had a similar speaking rate, the mean sentence duration as produced by individual speaker was 2.1 and 1.9 s for the female speakers, and 1.9 and 1.8 s for the male speakers. Recordings were root-mean-square (RMS) normalised to an average amplitude of 65 dB. Overall, each participant was presented with 8 experimental blocks of 15 trials each (120 sentences in total). For each experimental block, a list was randomly selected. From the selected list, only 15 sentences per block were randomly chosen and presented to the participant. Each sentence was only played once during the entire experimental session (including the practice trials) for a given participant.

Experimental task

The experimental task was a speech intelligibility test: participants were asked to listen to sentences and repeat them back to the experimenter. A loudspeaker was used for the presentation of auditory stimuli in order to ensure the participants' comfort and avoid pupil measurement being affected by discomfort that could be caused by wearing headphones. The experimental task consisted of three speech perception tests: a first one performed in quiet, and the remaining two performed in noise. The background noise used as a masker consisted of an 8-talker babble noise, obtained from recordings of spontaneous speech from 4 female and 4 male English native speakers. The main purpose of the test in quiet was to obtain a measure of intelligibility for each participant. The test in quiet was always presented at the beginning of the experimental session. This is so that the measure of speech perception in quiet would not to be affected by any learning effect due to previous exposure to the speech perception task in noise, particularly for non-native listeners. The presentation order of the two conditions in noise was randomised: 24 participants were presented with the high intelligibility condition first, 26 with the low intelligibility condition first. Therefore, the order of presentation should not affect the comparison across the two conditions in noise. During the three conditions, the speech level was constant at ~67–69 dB, as measured by a sound level meter. The speaker order during the test was randomised across the sentences presented within each block, and the randomisation was performed individually for each participant. This was done to avoid habituation and to increase the ecological validity of the task.

Speech perception in quiet

Participants were presented with five practice items followed by two blocks of 15 sentences each. All the stimuli were presented in quiet.

Speech perception in babble background noise

For each condition, three experimental blocks were presented. For the first block, an adaptive procedure was used to estimate the signal-to-noise (SNR) level required for reaching the target intelligibility level (Levitt, 1971). Levels of 40% (“low”) and 80% (“high”) intelligibility were chosen as targets to cover a considerable range in listening effort, but without resulting in extreme conditions where perception would be either effortless or too difficult. This is because when the processing demands of a task exceed available resources, pupil responses decline, reflecting task disengagement (Granholm, Asarnow, Sarkin, & Dykes, 1996). During the adaptive block, the SNR was manipulated by adapting both the speech and the masker levels so that the overall intensity level of the compound signal was fixed at 67–69 dB. The rationale for this was to avoid any confounding effects on pupil dilation of variations in overall sound intensity. The first sentence of the adaptive block was always presented at 20 dB SNR; subsequently, the SNR was manipulated to target the level at which 40 or 80% of keywords were understood. The changes in step size were defined by an algorithm taking into account the participant's performance and test stage; 9 dB SNR changes were applied during the initial stage and smaller 3 dB steps subsequently. The adaptive test terminated when either there had been five reversals or 15 trials had been presented. From this adaptive procedure, the SNR values corresponding to the reversals were averaged to obtain a single SNR value. In the two following blocks, audio stimuli were presented at that fixed SNR level. The same procedure (1 adaptive + 2 fixed blocks) was repeated twice for tracking both the high and low intelligibility levels. The experimental design adopted is visually displayed in Figure 2.1.

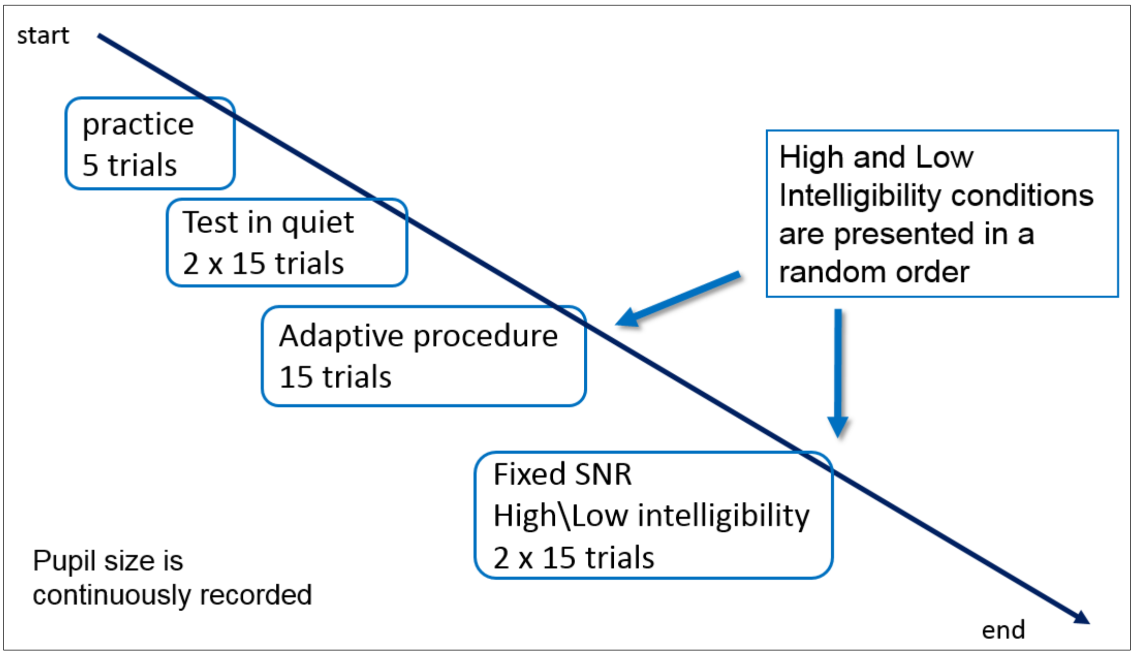

Figure 2.1 Experimental design for Study 1.

2.2.3 Procedure

The test was administrated in a sound-attenuated booth, with the participant seated on a comfortable chair. First, the audiometric assessment and background tests were performed. For the intelligibility tests, participants placed their chin in the head stabiliser in front of a screen positioned 70 cm away. The luminance of the room was individually adjusted so that the pupil of the participant was approximately in the middle of its dynamic range, in order to prevent ceiling and floor effects, as in Zekveld et al. (2010). The illumination ranged from 65 to 110 lx. A 9-point calibration procedure was initiated and validated. Then, the experimental task was initiated and participants were instructed to maintain their gaze and focus at a fixation cross positioned in the middle of the screen, in order to maximise the accuracy of the pupil data recorded. Each trial started with the fixation cross on the participant's screen turning black, signalling participants to fixate the screen in order to properly record their baseline pupil size. After 2 s, the sentence was played, and the fixation cross remained black for 3 additional seconds following the sentence offset, in order to allow enough time for the pupil to reach its maximum dilation. For the speech in noise conditions, the babble noise started 2 s before sentence onset (corresponding to the beginning of the baseline) and ended 3 s after sentence offset, which signalled the end of the trial. After the fixation cross had turned green, participants repeated the sentence back to the experimenter who was simultaneously scoring keyword accuracy on another screen. Participants were told that they could close and rest their eyes, and move their gaze while the fixation cross was green. After the sentence was scored, the experimenter initiated the following trial, after making sure that the participant was ready to continue. A break was taken preferably at the end of each section, but pauses at any time between trials were also allowed in case participants felt tired or needed to rest their eyes.

2.2.4 Pupillometry

The pupil size and location of the left eye were measured during the speech perception tasks using an EyeLink 1000 eye-tracker. The left and right pupils should show congruent dilation patterns (Purves et al., 2004), therefore monocular tracking can be considered reliable. The system uses infrared video-based tracking technology, with a spatial resolution of ~0.01 mm (value calculated for a pupil diameter of 5 mm), and was positioned at a horizontal distance of 55 cm from the participant. A headrest supporting the forehead and chin of the participant was used in order to reduce movement artefacts while performing the experiment. Pupil data were collected at the sampling rate of 500 Hz, and were stored in a connected PC. During data collection, the experimenter was able to visually inspect the video recording from a monitor, and to take action if needed (e.g., reminding the participants to fixate the centre of the screen, asking them to move in order to have the pupil in the eye-tracker searching area). The experimental task and data collection were controlled using MATLAB version R2015a. Pupil diameter was recorded during the entire duration of the three experimental conditions; event messages were included in the experimental script, so that the onset and end of each trial and each sentence was time locked to the pupil data. The pupil data were pre-processed using the steps reported below.

Pupil diameters below three standard deviations of the mean pupil diameter for the trial were considered as blinks. Linear interpolation was performed using the 50 data points preceding and following the blink. When more than 20% of the blinks for one experimental block happened in one trial, the trial was excluded. A smoothing first-order 10 Hz low-pass filter was applied in order to reduce the high frequency noise in the data, that were then down-sampled to 50 Hz. Lastly, the pupil data were visually inspected for artefacts. After exclusions, an average of 96% of trials per participant were included. From the continuous stream of pupil diameter data points, the section starting from 2 s prior to sentence onset (which was regarded as baseline) and ending 6.8 s after the beginning of the trial was included in the analysis. The rationale for excluding any data point beyond 6.8 s from sentence onset was that these measurements were only available for a small number of sentences and therefore any average would be calculated over very limited data.

Following the pre-processing, pupil data were averaged separately for each participant per conditions: quiet, high, and low intelligibility level. Four pupil outcome measures were obtained from the average trace of each participant and condition:

I. ***Pupil baseline:*** the average pupil diameter in the 2 s preceding the sentence's onset.

II. ***Mean pupil dilation*** relative to baseline pupil diameter between 0 and 6.8 s after the stimuli onset.

III. ***Peak pupil dilation***, as the maximum positive deviation from the baseline during the 6.8 s following stimuli presentation.

IV. ***Latency*** of the peak dilation amplitude.

2.2.5 Statistical Analyses

One way repeated-measures analyses of variance (ANOVAs), mixed design ANOVAs and t-tests were conducted to test whether order of tests presentation, test condition (high and low intelligibility levels) and linguistic background of participants (native or non-native listeners) affected behavioural and pupillometric data.

Mixed-effect regression models were performed to analyse the effect of language background, intelligibility level, order of tests presentation and individual performance in the background tests on pupil measures. Additionally, mixed-effect regression models were also used to investigate the effect of length of residence, self-reported English knowledge and foreign accent ratings on the pupil measures for non-native listeners only.

2.3 Results

2.3.1 Background Tests Results

Means and standard deviations for cognitive/phonological tests and language background information are shown in Table 2.1. Independent-sample t-tests with Bonferroni correction were conducted in order to compare the performance of native and non-native listeners on the forward and backward digit span test, and the phonological short term memory test. Despite the tests were administered in the participant's first language (either English or Italian), non-native participants performed more poorly than native participants on the forward digit span test, $t_{(41.2)} = -3.47$, $p = 0.003$. A marginally significant difference, with again lower performance for non-native participants, was also obtained for the backward digit span test, $t_{(39.7)} = -2.43$, $p = 0.06$, and for the phonological short term memory test, $t_{(48)} = -2.55$, $p = 0.04$. The two digit span tests results were additionally corrected for the violation of the assumption of variances' equality.

Table 2.1 Background tests results.

Background tests		**Native listeners**		**Non-native listeners**	
		M	***SD***	**M**	***SD***
Digit span	Forward	7.5	*1.4*	6.3	*1.1*
	Backward	6.2	*1.5*	5.3	*1.1*
Short term phonological test		37.7	*3*	35.4	*3.2*
(Non-native only)	Accent rating (1-7, 1=native-like)	N/a		5.1	*1.1*
	Length of residence (years)	N/a		3.6	*2.6*
	Overall English use	N/a		50%	*0.1*
	Self-reported English knowledge (0-6, 6=excellent)	N/a		4.5	*0.9*

2.3.2 Behavioural Results

Intelligibility scores in quiet are summarised in Table 2.2. The reported means are averaged across the two experimental blocks, excluding the practice trials, across participants. There was a significant difference in the percentage of correctly reported words in the speech in quiet task between native and non-native participants, $t_{(48)} = -4.80$, $p < 0.001$. However, the effect size for this analysis (d = 0.14) was found not to reach Cohen's convention for a small effect (Cohen, 1988). The comprehension level for non-native listeners was nearly at ceiling (approximately

90%) during the test in quiet, indicating that they were highly proficient in English when listening in optimal conditions.

Table 2.2 Descriptive statistics of the behavioural results for speech perception in quiet.

Behavioural results in quiet						
	All participants		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Performance (% correct)	94.7	*8.5*	90.2	*9.6*	99.9	*0.3*

Table 2.3 summarises results from the speech perception task in noise, reporting intelligibility levels averaged over the two blocks run for each intelligibility condition (high and low), and the corresponding SNR levels. The adaptive block used to set SNR level is not included in the analysis. A mixed design ANOVA with condition (high and low intelligibility) as within-subjects factor, and language (native and non-native) as between-subjects factor showed a significant difference in performance across intelligibility levels [$F_{(1, 48)} = 76.45$, $p < 0.001$], showing a significantly higher accuracy for the high compared to low intelligibility condition, as expected. The effect size for this difference (d = 1.87) was found to exceed Cohen's convention for a large effect (Cohen, 1988). The main effect of language group and the interaction were both found not to be significant, showing therefore that intelligibility levels did not vary across the native and non-native participants, showing that the adaptive procedure was successful in achieving matched intelligibility across groups. As expected, for each intelligibility level, the SNR levels for native listeners were significantly lower (i.e. more background noise) than those required by non-native listeners: $t_{(48)} = 5.95$, $p < 0.001$ for the high intelligibility condition, $t_{(48)} = 5.97$, $p < 0.001$ for the low intelligibility condition. SNR levels are displayed in Figure 2.2.

Table 2.3 Descriptive statistics of the behavioural results for speech perception in noise.

	Babble masking / high intelligibility					
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Performance (% correct)	71.3	*14.2*	70.3	*13.4*	72.5	*15.3*
SNR	-4.5	*4.4*	-1.9	*3.9*	-7.6	*2.5*
	Babble masking / low intelligibility					
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Performance (% correct)	43.8	*15.2*	42.1	*15.6*	45.7	*14.9*
SNR	-8.8	*3.7*	-6.6	*3.1*	-11.4	*2.4*

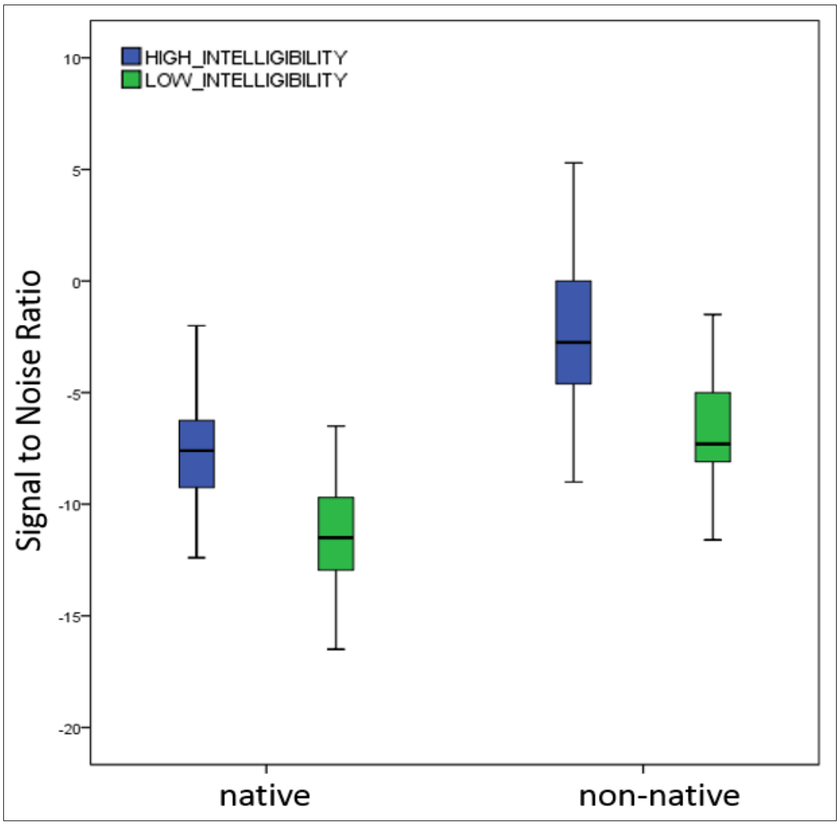

Figure 2.2 SNR values for high and low intelligibility conditions for native and non-native listeners.

It is worth noting that, although 80% intelligibility level was targeted for the high intelligibility condition, the average keyword intelligibility level was closer to 70%. This is likely to be due to a relatively small number of trials presented in the adaptive procedure block. Importantly however, as reported above, performance levels did not vary significantly across language groups for both intelligibility conditions. Although large standard deviations were obtained, reflecting within-group variability, this was the case for both the native and non-native groups.

2.3.3 Pupil results in quiet

Descriptive statistics for the pupil data in quiet are reported in Table 2.4. These include mean and peak pupil dilation over the baseline, latency of the peak and baseline pupil diameter.

Table 2.4 Descriptive statistics of the pupil measures in quiet.

	Pupil data in quiet					
	All participants		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.20	*0.18*	0.30	*0.17*	0.08	*0.10*
Peak dilation, mm	0.38	*0.26*	0.52	*0.26*	0.22	*0.14*
Latency of peak, sec	2.64	*0.79*	2.49	*0.29*	2.82	*1.11*
Baseline, mm	5.17	*0.71*	5.13	*0.66*	5.22	*0.77*

The test in quiet had some specific features that contrast with the two conditions in noise. It was always presented first, the 2 s baseline was in silence (as opposite to babble noise) and because of the nature of the test itself, the performance level was not matched between language groups. For these reasons, pupil data from the condition in quiet has been analysed separately in order to rule out potential confounding factors, and has been excluded from the subsequent analyses. An independent-sample t-test was conducted to compare the pupil response in native and non-native participants. The mean and peak pupil dilation were found to be significantly greater for non-native compared to native listeners [$t_{(48)} = 5.52$, $p < 0.001$ and $t_{(48)} = 4.93$, $p < 0.001$ respectively]. The effect sizes for these comparisons ($d = 1.60$ for the mean value and $d = 1.43$ for the peak dilation) were both found to exceed Cohen's convention for a large effect (Cohen, 1988). It is worth noting that the behavioural performance in quiet did significantly differ between native and non-native listeners, without however reaching Cohen's convention for a small effect (Cohen, 1988). Nevertheless, this yielded a large difference in the mean and peak pupil dilation between native and non-native listeners. No statistically significant differences in the baseline and in the latency of the peak were observed between the two listeners' groups. The pupil curves dilation for native and non-native listeners during the test in quiet are displayed in Figure 2.3.

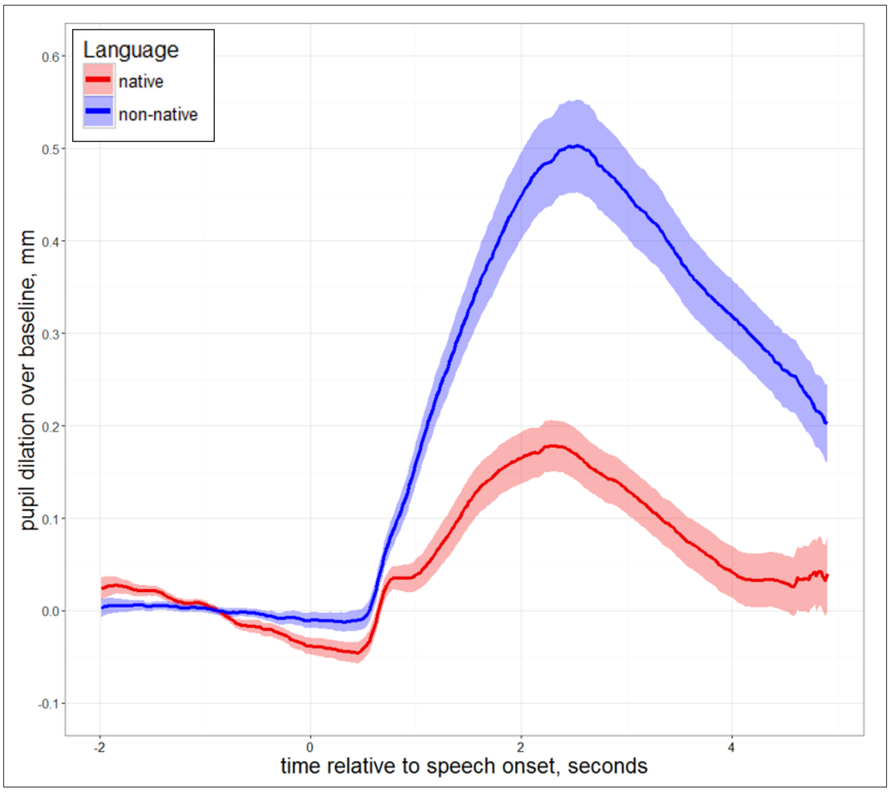

Figure 2.3 Mean pupil response over time during speech perception in quiet for native and non-native listeners.

2.3.4 Pupil results in noise

Descriptive statistics for the pupil data in noise are reported in Table 2.5 (measures per condition), and Table 2.6 (measures per presentation order). The pupil data presented and entered in the analyses are those collected during the blocks with SNRs previously fixed via adaptive procedure.

Table 2.5 Descriptive statistics of the pupil measures in noise.

Babble masking – high intelligibility						
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.13	*0.16*	0.18	*0.17*	0.06	*0.13*
Peak dilation, mm	0.29	*0.22*	0.36	*0.24*	0.21	*0.18*
Latency of peak, sec	2.66	*0.72*	2.79	*0.71*	2.50	*0.72*
Baseline, mm	5.37	*0.79*	5.24	*0.69*	5.52	*0.89*
Babble masking - low intelligibility						
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.18	*0.20*	0.24	*0.19*	0.10	*0.18*
Peak dilation, mm	0.34	*0.27*	0.42	*0.27*	0.26	*0.25*
Latency of peak, sec	2.63	*0.76*	2.64	*0.83*	2.61	*0.68*
Baseline, mm	5.44	*0.80*	5.35	*0.69*	5.55	*0.92*

Table 2.6 Descriptive statistics of the pupil measures in noise sorted by presentation order.

Babble masking – first session						
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.18	*0.18*	0.23	*0.16*	0.11	*0.17*
Peak dilation, mm	0.35	*0.25*	0.42	*0.24*	0.26	*0.25*
Latency of peak, sec	2.69	*0.73*	2.68	*0.84*	2.70	*0.60*
Baseline, mm	5.39	*0.76*	5.31	*0.68*	5.49	*0.85*
Babble masking – second session						
	All		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.13	*0.18*	0.19	*0.20*	0.06	*0.13*
Peak dilation, mm	0.29	*0.25*	0.36	*0.28*	0.21	*0.18*
Latency of peak, sec	2.60	*0.75*	2.75	*0.71*	2.42	*0.76*
Baseline, mm	5.42	*0.84*	5.28	*0.71*	5.58	*0.96*

Analyses using mixed-effect modelling were performed in order to investigate the effects of language background, intelligibility level, order of presentation and individual factors on the pupil measures during the condition in noise. Those analyses were performed using the lme4 package in the R environment (Bates, Mächler, Bolker, & Walker, 2014; R Core Team, 2017). For each dependent variable (mean and peak pupil dilation, peak latency, and baseline), I began with a saturated model that included interaction terms for all independent variables considered as fixed effects with random intercepts and slopes. Due to non-convergence, I simplified the models hierarchically from most complex to least complex. The resulting converged maximal models for all four variables included the following fixed effects: intelligibility level (2: high and low), language background (2: native and non-native), presentation order (2: first and second), forward digit span, backward digit span and short term phonological memory test. Participant was included as random effect but no random slopes. The maximal models also included the following two-way interactions: intelligibility x language background, language background x presentation order, intelligibility x each of the background measures (forward and backward digit span, phonological memory), and language background x each of the background measures. Model residuals via chi-square tests ($\alpha = .05$) were compared from the most complex models (containing the largest interaction term) to the least complex models (containing only single terms). If an interaction term was significant, all lower level effects involved in the interaction were included in the final model. Results for each dependent variable considered are reported below. The pupil curves dilation for native and non-native listeners during the high and low intelligibility conditions are displayed in Figure 2.4. Figure 2.5 shows the pupil curves dilation for all participants for the two intelligibility conditions.

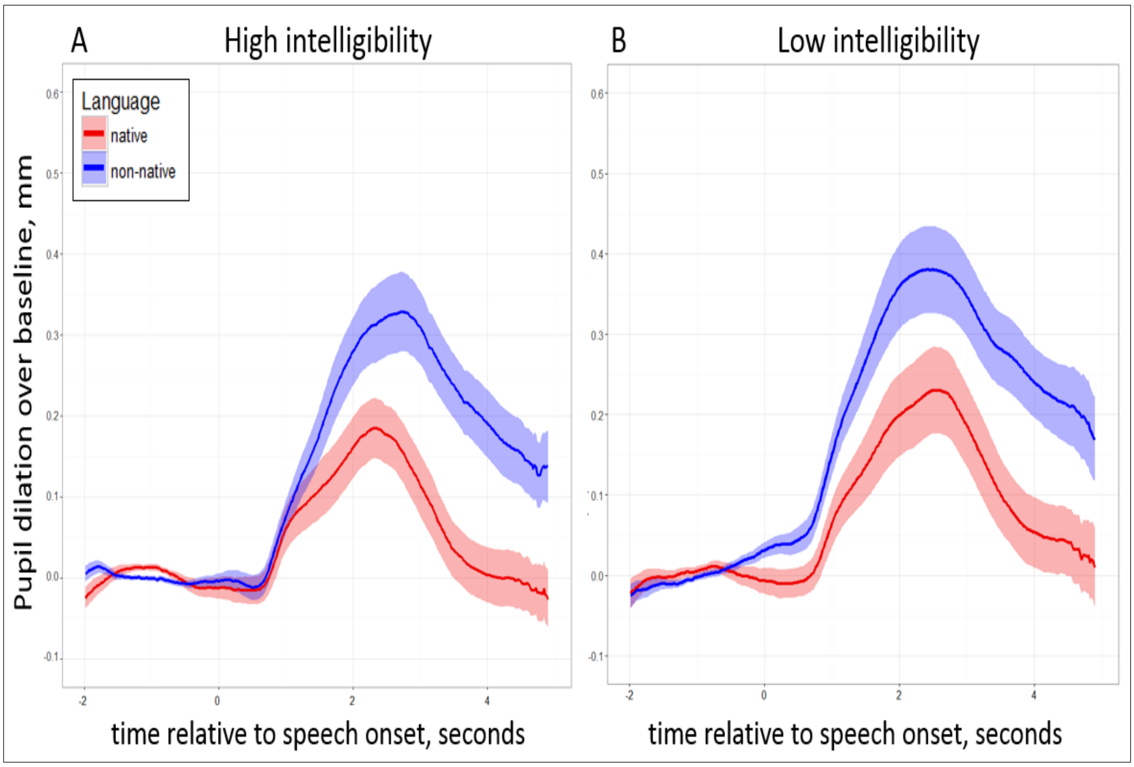

Figure 2.4 Mean pupil response over time during speech perception in noise for high (A) and low (B) intelligibility conditions, for native and non-native participants.

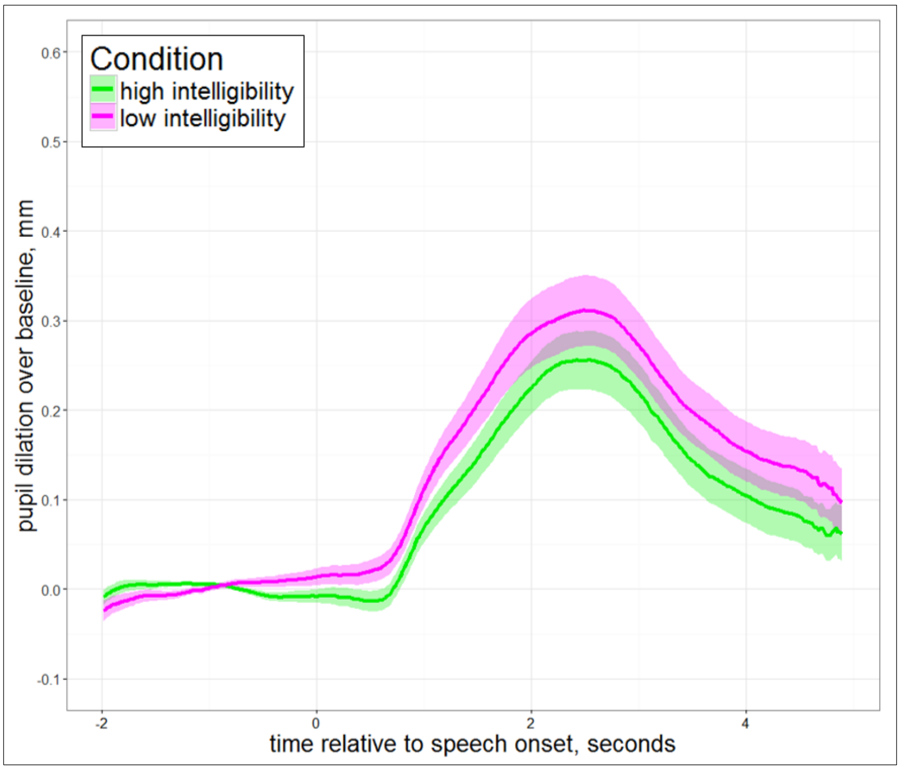

Figure 2.5 Mean pupil response over time during speech perception in noise for all participants in high and low intelligibility conditions.

Mean pupil dilation

The final model included fixed effects of intelligibility level, language background, and order of presentation (see Table 2.7 for coefficients and estimated *p* values). This indicated that overall the mean pupil dilation was greater for the low compared to the high intelligibility condition, for non-native compared to native listeners, and for the first compared to the second session in noise.

Table 2.7 Fixed and random effects in a mixed-effects model of mean pupil dilation for all listeners.

Mean ~ Language + Intelligibility + Order + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.13	0.03	3.81	< 0.001
Language	Non-native	0.13	0.04	2.81	0.005
Intelligibility	high	-0.05	0.01	-3.66	< 0.001
Order	second	-0.04	0.01	-3.46	< 0.001
Random effects		**Variance**			
Participant	(intercept)	0.02			
Residual		0.00			
***Note.* Number of observations = 100; participants (N) = 50.**					

Peak pupil dilation

The final model included fixed effects of intelligibility level, language background, and order of presentation (see Table 2.8 for coefficients and estimated *p* values). Similarly to the mean, the peak pupil dilation was overall greater for the low compared to the high intelligibility condition, for non-native compared to native listeners, for the first compared to the second session in noise.

Table 2.8 Fixed and random effects in a mixed-effects model of mean pupil dilation for all listeners.

Peak ~ Language + Intelligibility + Order + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.28	0.05	5.93	< 0.001
Language	Non-native	0.15	0.06	2.32	0.020
Intelligibility	high	-0.05	0.01	-3.41	< 0.001
Order	second	-0.05	0.01	-3.57	< 0.001
Random effects		**Variance**			
Participant	(intercept)	0.04			
Residual		0.00			
***Note.* Number of observations = 100; participants (N) = 50.**					

Latency

None of the terms were found to be significant.

Baseline

The final model only included fixed effects of intelligibility level (see Table 2.9 for coefficients and estimated *p* values). The baseline pupil diameter was greater for the low compared to high intelligibility condition.

Table 2.9 Fixed and random effects in a mixed-effects model of baseline pupil diameter for all listeners.

Baseline ~ Intelligibility + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		5.44	0.11	48.66	< 0.001
Intelligibility	high	-0.07	0.03	-2.19	0.028
Random effects		**Variance**			
Participant	(intercept)	0.06			
Residual		0.00			
***Note.* Number of observations = 100; participants (N) = 50.**					

None of the three memory measures available for all participants (forward and backward digit span and short term phonological memory test) improved the fit of the model for any of the dependent variables.

<u>Non-native participants</u>

Additionally, in order to investigate the effect of accent rating, length of residence, overall English use and self-reported English knowledge on the pupil measure, the same type of analysis was run with non-native listeners only for all four dependent variables (mean, peak, latency of the peak and baseline). The resulting converged models for all four variables included the following fixed effects: intelligibility level (2: high and low), accent rating, length of residence, English use and self-reported English knowledge. Participant was included as random effect but no random slopes. The maximal model only included up to two-way interactions between condition and background measures.

Results for each dependent variable considered are reported below.

Mean pupil dilation

The final model only included the fixed effect of intelligibility level, confirming that overall the mean pupil dilation was greater for the low compared to the high intelligibility condition for non-native listeners (see Table 2.10 for coefficients and estimated *p* values).

Table 2.10 Fixed and random effects in a mixed-effects model of mean pupil dilation for non-native listeners.

Mean ~ Intelligibility + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.24	0.03	7.02	< 0.001
Intelligibility	high	-0.06	0.02	-3.49	< 0.001
Random effects		**Variance**			
Participant	(intercept)	0.03			
Residual		0.00			
***Note.* Number of observations = 54; participants (N) = 27.**					

Peak pupil dilation

The final model included fixed effects of intelligibility level and order of presentation (see Table 2.11 for coefficients and estimated *p* values). The peak pupil dilation in non-native listeners was overall greater for the low compared to the high intelligibility condition and for the first compared to the second session in noise.

Table 2.11 Fixed and random effects in a mixed-effects model of peak pupil dilation for non-native listeners.

Peak ~ Intelligibility + Order + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.43	0.05	8.78	< 0.001
Intelligibility	high	-0.05	0.02	-2.58	0.010
Order	second	-0.04	0.02	-2.43	0.015
Random effects		**Variance**			
Participant	(intercept)	0.06			
Residual		0.00			
***Note.* Number of observations = 54; participants (N) = 27.**					

Latency

As for all listeners, none of the terms considered were found to be significant predictors of the latency of the peak for non-native listeners.

Baseline

The final model for baseline pupil dilation for non-native listeners included the interaction between the fixed effects of intelligibility and self-reported English knowledge, and a random effect of listener (see Table 2.12 for coefficients and estimated *p* values). Because of the significant interaction between intelligibility and self-reported English knowledge, the relative fixed factors were also included in the final model. However, follow-up regressions did not reveal any significant effect of the level of self-reported English knowledge on the pupil baseline diameter, either for the high or the low intelligibility condition.

Table 2.12 Fixed and random effects in a mixed-effects model of baseline pupil diameter for non-native listeners.

Baseline ~ Intelligibility + English_knowledge + Intelligibility:English_knowledge + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		4.86	0.68	7.16	< 0.001
Intelligibility	high	0.28	0.14	1.96	0.050
English_knowledge		0.11	0.15	0.72	0.468
Intelligibility x English_knowledge	high	-0.9	0.03	-2.73	0.006
Random effects		**Variance**			
Participant	(intercept)	0.44			
Residual		0.01			
Note. **Number of observations = 54; participants (N) = 27.**					

In summary, results showed that when considering all participants, the mean and peak pupil dilations were greater for non-native compared to native listeners, and for the high intelligibility compared to the low intelligibility condition. In addition, the mean pupil dilation was also significantly greater during the first compared to the second session in noise. The pupil baseline was overall greater for the low compared to the high intelligibility condition. When only non-native listeners were considered, the mean and peak pupil dilation were greater for the low compared to the high intelligibility condition. The peak pupil dilation was also greater for the first compared to the second experimental session in noise. The latency of the peak did not significantly change across the variable considered. Lastly, none of the background measures considered were found to be significant predictors of the pupil measures considered for both groups of listeners.

2.4 Discussion

Study 1 assessed the effect of speech intelligibility levels and language background on listening effort, as measured by means of pupil response. The main findings of the experiment are:

1. Pupil response is greater for non-native compared to native listeners during speech perception in quiet, and in noise when intelligibility levels for the two groups of listeners are matched.
2. Pupil response is greater for low compared to high intelligibility levels.
3. Pupil response is not differentially modulated by intelligibility level for native and non-native listeners.
4. The order of test presentation modulates pupil response in native and non-native listeners.

The first and second findings are in line with predictions, while the third is not. As hypothesised, pupil response (mean and peak dilation relative to baseline) was greater for non-native compared to native participants. This is in line with previous research in the field of second language perception (Schmidtke, 2014); and it also expands the limited literature about second language perception using pupillometry, by directly addressing the challenge of non-native sentence perception in adverse listening conditions.

I argue therefore that the overall increased listening effort reflected in the greater pupil response for non-native compared to native listeners might be a result of an increased difficulty arising at multiple levels, as extensively discussed in Chapter 1. First, at a perceptual level because of the less accurate phonetic-perceptual discrimination (Best et al., 2001; Iverson & Kuhl, 1995). Second, at a lexical level due to the increased word competition deriving from L1 words activation (Spivey & Marian, 1999), and third because of a generally lower L2 linguistic proficiency. Therefore, in order to achieve a performance level similar to native listeners, non-native individuals need to rely more heavily on working memory capacity, which results in more effortful listening. One additional factor that might have contributed to the differences in listening effort between the two listeners' groups is the observed difference in the cognitive abilities, as shown by the cognitive tests results.

As predicted, the listening effort reflected by the mean and peak pupil dilation was higher for the low compared to high intelligibility condition. This result is in line with previous research in native listening, also using individual speech reception thresholds, showing that the pupil response during listening to sentences systematically varied as a function of speech intelligibility if extremely low intelligibility levels are excluded (Zekveld & Kramer, 2014; Zekveld et al., 2010). The growing body of evidence in this direction corroborates the idea that speech perception in difficult listening condition is more heavily reliant on the explicit and effortful exploitation of cognitive resources, particularly working memory. Together with our first finding, an increased pupil response for low compared to high intelligibility conditions, also supports the predictions made by the ELU model (Rönnberg et al., 2019; Rönnberg et al., 2013).

Contrary to predictions, pupil response was not differentially modulated in the two different listeners' groups across intelligibility conditions. That is, the additional amount of listening effort required to non-native compared to native individuals was not greater for lower intelligibility levels relative to higher levels. Interestingly, the difference in SNR levels also remained constant across the high and low intelligibility conditions (approximately 5 dB), and its magnitude was comparable with the difference reported by Van Wijngaarden (2001) when comparing speech perception in native and non-native listeners at 50% intelligibility level. This result might change if a wider range of intelligibility levels is considered. Along the same lines, previous research also did not report a differential effect of noise for native and non-native listeners on the number of simultaneously activated candidate words during speech perception (Scharenborg, Coumans, & van Hout, 2017). Other individual factors more subtle than the mere linguistic background in terms of native vs. non-native might also contribute to modulate the relationship between intelligibility level and listening effort, as suggested by previous pupillometry research. For example, the ability to read partially masked speech has been regarded as being the visual analogue to speech reception threshold in a previous study, and was found to play a role in the modulation of pupil response together with the tendency to give up listening in particularly challenging conditions (Zekveld & Kramer, 2014).

Additionally, an effect of presentation order across the two tests in noise was found, with a mean and peak pupil dilation higher in the first compared to the second session in noise. This is in line with findings from previous research (Hyönä et al., 1995; Koch & Janse, 2016; Zekveld et al., 2010). The influence of presentation order on the pupil response once again confirms the importance of a careful experimental design when investigating listening effort using pupillometry. Indeed, if a randomisation of conditions is not appropriately performed, the presentation order might constitute a confounding variable that may invalidate the results.

Lastly, the effect of individual differences on the behavioural performance and pupil response was explored. Nevertheless, results showed that none of the individual measures considered significantly contributed to predict pupil measures. This could be due to the large individual variability that is typically observed in pupillometric data (Winn et al., 2018), that makes pupillometry a non-ideal candidate for the investigation of individual differences. In addition to that, individual measures collected for each participant might have been not fully reliable. The working memory measure collected (forward and backward digit span) was not sensitive enough to show great individual variability in a population of healthy participants, so a potential correlation between cognitive abilities and listening effort was difficult to establish based on the available data. Additionally, when considering non-native listeners, an objective measure of English proficiency was not available for the present study, and it was not possible to divide participants into balanced groups based on proficiency or length of stay criteria. Lastly, the accent rating entered in the analyses was based on a single sentence. Although the sentence considered was the same for all participants, and all non-native listeners shared the same L1 background, this might have been not sufficient for an accurate judgement of the degree of the listener's foreign accent. Additionally, it could be interesting to evaluate differences between native and non-native pupil response at an intermediate level of understanding (e.g., 50 or 60% of intelligibility). Indeed, the maximum peak pupil dilation has been observed at around 50% correct sentence recognition performance (Ohlenforst et al., 2017), signalling that this might be the intelligibility threshold where listeners engage the most with the speech perception task, and where the maximum amount of resources are actively employed. An alternative approach that might allow for a greater precision in the adaptive SNR setting would be to assess pupil size while performing the adaptive procedure itself, instead of adopting a two-step approach (adaptive block followed by blocks with fixed SNR). Previous studies successfully implemented the one-step approach (Zekveld & Kramer, 2014; Zekveld et al., 2010), which has the advantages of enhancing precision (because only sentences with the exact target intelligibility level are considered for the analyses) and reducing the duration of the experimental session.

In conclusion, Study 1 corroborates pupillometry as a sensitive investigation technique to uncover listening effort differences both within and between participants. This measure was sensitive to differences in intelligibility levels and different listener types; this gives the possibility to quantify differences in listening effort even when listener groups are performing at near-ceiling level, as was the case in the quiet condition. Importantly, the present study showed a greater pupil response in non-native compared to native participants, proving that a greater listening effort is required when trying to understand speech in noise even when intelligibility levels are matched. This was the case for proficient non-native listeners who were achieving around 90% intelligibility for speech comprehension in quiet. Therefore, maintaining a good level of performance when understanding speech in noise comes at a much higher cost for non-native listeners. This is likely to have considerable subsequent effect on the ability to perform more than one task simultaneously and to efficiently and quickly recall information in typical communicative environments. As documented for individuals suffering from hearing loss (McGarrigle et al., 2014), it is reasonable to speculate that a prolonged increase in the listening effort needed to attend speech will result in a greater mental fatigue also for all non-native listeners.

Chapter 3

3. Study 2

3.1 Aim of the study

While discussing findings from Study 1, it was argued that the overall increased listening effort for non-native compared to native listeners might be the result of the combined effects of difficulties arising at multiple levels of the speech perception process. However, given the type of experimental design, it was not possible to disentangle the relative contribution of each individual factor, namely the acoustic and lexical contribution. Therefore, Study 2 was designed with the aim of expanding findings from the first study, by considering the effect of semantic context and acoustic cues enhancement, both isolated and in combination, on listening effort during non-native speech perception. The enhancement of acoustic cues was operationalised with the use of clear speaking style. Moreover, by administering to non-native listeners an international standardised test of English proficiency, I specifically addressed the issue of the lack of a reliable measure of L2 proficiency in the first experiment.

Behavioural studies evaluating the performance of native and non-native listeners indicate that individuals tend to adopt different strategies when attending to their first language in challenging listening conditions, as compared to their second language. Mattys et al. (2010) suggested that native listeners generally rely more on lexical plausibility than on acoustic cues compared to non-native listeners both when attending to intact speech and in the presence of a competing talker. Moreover, under cognitive load conditions (i.e. when simultaneously performing a visual search task), native listeners seem to further increase the weight given to contextual plausibility, while non-native listeners do not appear to perform the same lexical drift. Additionally, Bradlow and Alexander (2007) showed that non-native listeners are able to benefit from modification in the speaking style even when presented in isolation (i.e. not combined with semantic cues). However, the same is not true for semantic cues, indeed authors suggested that non-native listeners require a higher signal clarity in order to successfully exploit contextual cues (Bradlow & Alexander, 2007). Taken together, the evidence seems to indicate that lexical information is less readily exploitable for non-native compared to native listeners, and that its availability also depends on the listening environment the speaker is immersed in. However, little is known about the implications of those speech perception strategies for listening effort. How does the presence of one or more cues (acoustic and contextual cues) modulate cognitive effort during speech perception in adverse conditions in native and non-native listeners? Does the availability of semantic cues reduce listening effort by allowing an easier access to sentence meaning? Or conversely is the "contextual strategy" more effortful because of the need for higher level processing? Do those cues affect listening effort differentially in native compared to non-native listeners?

The second study reported here has been designed to investigate these research questions. To this purpose, semantic predictability and speaking style were manipulated both separately and in combination during a speech perception task in noise, which was administered to native and non-native listeners. This was done to provide either semantic enrichment, acoustic enrichment, or both simultaneously. An adaptive procedure was used to match intelligibility levels across participant groups and conditions. Further to that, this second study aimed at exploring how L2 proficiency level modulated listening effort in relation to the availability of semantic and acoustic cues. To achieve this, non-native listeners who varied widely in their length of residence in an English speaking country were recruited, and their proficiency in the L2 was measured by means of a standardised test created for testing speech understanding ability in non-native learners of English.

In summary, the present study has been designed to answer the following research questions:

1. How does the presence of semantic cues affect listening effort in native and non-native listeners, at equated levels of intelligibility?
2. How does the presence of enhanced acoustic cues by means of clear speaking style modulate listening effort in native and non-native listeners, at equated levels of intelligibility?
3. How does the combined presence of both clear speaking style and semantic cues impact on listening effort in native and non-native listeners?
4. How does proficiency level affect listening effort in non-native listeners?

With respect to the research questions stated above, two predictions were formulated. First, it was predicted that the availability of semantic cues in the stimuli presented would overall reduce listening effort. This is based on previous findings showing that facilitations in semantic processing provided both at word and sentence level are associated to a reduced pupillary response in native listeners (Ben-Nun, 1986; Chapman & Hallowell, 2015; Kuchinsky et al., 2013; Kuipers & Thierry, 2011; Winn, 2016). Second, it was predicted that the reduction in the pupil response due to semantic context would be smaller for non-native listeners, since they have been previously shown to be less able to exploit contextual cues at a sentence level (Bradlow & Alexander, 2007; Mattys et al., 2010). In addition, I expected the pupil response to be overall larger for non-native compared to native listeners, as found in Study 1 (Borghini & Hazan, 2018). It was not formulated any hypothesis regarding the effect of clear speaking style on listening effort, because there was no literature available on which justified prediction could be based.

3.2 Materials and Method

3.2.1 Participants

Fifty five adults from two different language backgrounds participated in the study. One non-native participant was subsequently excluded due to the performance level during the speech perception test in noise exceeding two standard deviations of the mean performance in two out of four experimental conditions. After the exclusion, the non-native group included 35 participants (8 men and 27 women) with Italian as L1 and English as L2, aged 19-36 years (M = 26.3, SD = 4.6). All participants had been living in the UK for at least 3 months at the time of testing. The native group consisted of 19 native British English participants (10 men and 9 women), aged 19-34 years (M = 25.8, SD = 4.8). Three native participants who took part in this study also participated in Study 1, approximately one year earlier. The study was advertised via the UCL Psychology subject pool and via social media. None of the participants reported to suffer from cognitive or neural disorder, cataracts or diabetes, and to have taken drugs or medications in the 48 hours prior the experiment. These criteria were considered in order to exclude the potential effect of any confounding variable on pupil dilation. Participants were able to maintain the visual focus on the cross presented on the screen without contact lenses; glasses were allowed for those participants who needed them, and their usage did not interfere with the eye-tracker recording. All participants provided written informed consent before participating in the study. They received a monetary compensation for their time under a protocol approved by the Ethics Committee at University College London.

3.2.2 Stimuli and Tests

Background tests

All participants had pure tone thresholds in both ears at 20 dB HL or better at octave frequencies between 250 and 8000 Hz. Non-native participants completed an online linguistic background questionnaire which collected information about their overall English usage and their level of self-reported English knowledge. To this purpose, the same questionnaire as in Study 1 was used. In addition, non-native participants were presented with the Listening module of the International English Language Testing System (IELTS), an international standardised test of English proficiency for non-native speakers ("https://www.ielts.org/," 2017). The module administered comprises four sections, with ten questions in each section and a final score ranging from 0 to 40. As an example, listeners were asked to fill gaps in a sentence (e.g. “Express train leaves at ________ “) or to answer multiple choice questions according to what they heard. The total duration of the test was 30 minutes. The score obtained by non-native participants who took part in the experiment ranged between 9 and 38, therefore covering a wide range of language proficiency. The aim of this test was to obtain an accurate picture of the proficiency level of the participants, with a focus on their speech understanding skills. Participants’ background data are summarised in Table 3.1.

Table 3.1 Demographics and background tests results.

Background information		**Native listeners**		**Non-native listeners**	
		M	**SD**	**M**	**SD**
Age		25.8	4.8	26.3	4.6
(Non-native only)	Length of residence (years)	N/a		2.3	1.9
	Overall English use	N/a		45%	0.2
	Self-reported English knowledge (0-6)	N/a		4.3	1.2
	IELTS Listening (0-40)	N/a		26.7	8.8

Experimental Stimuli

Semantic predictability and speaking style of the stimuli were manipulated for this study. This resulted in the construction of four different sets of stimuli: semantically predictable and anomalous sentences produced with either plain or clear speaking style. The Basic English Lexicon (BEL) sentence materials (Calandruccio & Smiljanic, 2012), including 20 lists of 25 sentences, were used as predictable stimuli, because the semantic context of each sentence was always consistent within the sentence itself. This testing material was the same as used in Study 1, and a detailed description is given in 2.2.2. Anomalous sentences were constructed based on BEL sentences with the following procedure. The keywords contained within each list of BEL sentences were categorised as nouns, verbs, adjectives and adverbs, and subsequently shuffled across sentences within the same category and list. Sentences created following these steps were subsequently manually checked one by one by the experimenter and one additional native English speaker in order to ensure their grammaticality. Twenty lists of 25 sentences were thus created and were used as anomalous stimuli (see Appendix C). These do not have a meaningful semantic context, but they retained the same grammatical structure of the BEL sentences. Examples of anomalous sentences are “The VEGETABLES OPEN a DIFFICULT HAT”, “The PHONE TRAVELLED the FAITHFUL MEAT” (keywords are in capital letters). Each predictable and anomalous sentence had four keywords, which were used to score comprehension.

Four native Southern British English speakers (two females) were recorded producing these sentences in a sound-attenuated booth in the Department of Speech, Hearing and Phonetic Sciences at UCL, with a sampling rate of 44.1 kHz, 16-bit resolution. Four speakers were recorded in order to increase the ecological validity of the study and improve the generalisability of findings, by ensuring that listeners were not unduly influenced by the individual characteristics of a single speaker, particularly because plain and clear speaking style were being compared. Speakers read the sentences from a monitor and were asked to produce each set of sentences (predictable and anomalous) twice. For the plain speaking style they recorded the sentences in a conversational self-paced rate, without any particular focus on clarity. To obtain stimuli in a clear speaking style, talkers were instructed to pronounce sentences as if they were speaking to a person with hearing impairment, or with a second language learner, similarly to the procedure used by Bradlow and Alexander (2007) to elicit a clear speaking style. Recordings were root-mean-square (RMS) normalised to an average amplitude of 65 dB. The average sentence duration was 1.9 seconds (sd = 0.2) for plain stimuli and 4.7 seconds (sd = 1.6) for clear stimuli.

Experimental task

The experimental task consisted of a speech intelligibility test: participants were asked to listen to sentences and to repeat them back to the experimenter. Sentences were presented by means of a loudspeaker placed in front of the participant at a distance of approximately 1 metre. This was done to ensure participants’ comfort and avoid pupil measurement being affected by discomfort potentially caused by headphones. The experimental task consisted of two sections: one performed in quiet, and the other with background noise. Each of the two sections included four parts where semantic predictability and speech clarity were manipulated, resulting in four conditions tested: predictable and semantically anomalous sentences pronounced both in plain and clear speaking style. The four experimental conditions were labelled as predict_plain, predict_clear, anom_plain and anom_clear for clarity purposes. The test in quiet was always presented first, while within each of the two sections the presentation order of the four conditions was randomised. The main purpose of the test in quiet was to check that all non-native participants had a nearly native-like performance when attending to speech in optimal listening conditions.

Pupil data from this section were not considered for analyses. The sound level for the entire duration of the experiment was constant at ~67–69 dB, as measured by a sound level meter, to avoid any confounding effect on the pupil dilation due to changes in the sound intensity. Each participant was presented with sentences pronounced by a single speaker across all experimental conditions, randomly selected among the four speakers recorded. This was done to reduce performance variability within the same participant, so as to ensure that the adaptive procedure adopted for testing in noise could function at its best. During the entire experimental session each participant was overall presented with 220 sentences, and each sentence was only played once for a given participant. The same list was never repeated across the semantically predictable and anomalous conditions for each participant, to minimise the repetition of the same lexical item.

Speech perception in quiet

Participants were presented with six practice items, followed by one block of 10 sentences each per experimental condition. As mentioned above, experimental conditions included predictable and semantically anomalous sentences pronounced both in plain and clear speaking style. All stimuli were presented in quiet. The purpose of this speech perception test is twofold. First, it is useful to get an evaluation of the participants' performance level in quiet. Additionally, these data were used to check whether the use of stimuli recorded from different speakers had an effect on speech intelligibility.

Speech perception in background noise

For each of the four conditions, two experimental blocks were presented. In the first block, an adaptive procedure (Levitt, 1971) was used to estimate the signal-to-noise ratio (SNR) required for reaching the target intelligibility level of 50% of correctly reported keywords. 50% was chosen as target intelligibility level because it allowed to elicit a significant pupil dilation without resulting in participants giving up the task due to extremely poor performance (Zekveld & Kramer, 2014). As for Study 1, an 8-talker babble noise obtained from recordings of spontaneous speech from 4 female and 4 male English native speakers was used as a masker for the speech. The SNR was manipulated by adapting both the target and the masker levels, so that the overall intensity of the compound signal remained unchanged at 67-69 dB. The first sentence was always presented at 20 dB SNR, and in the following sentences the SNR was manipulated to target the level at which half of keywords were correctly reported. The changes in step size were defined by an algorithm according to the participant's performance and test stage; steps of 9 dB were applied during the initial stage and smaller 3 dB steps closer to the target SNR. In order to improve the accuracy of the adaptive procedure compared to the previous study described in chapter 2, the number of trials presented in the adaptive block was increased to 25. Following this adaptive procedure, the SNR values corresponding to the reversals were averaged to obtain a single SNR value. In the following block, 20 sentences were presented using the fixed SNR level previously determined. The same structure consisting of 1 adaptive (25 trials) + 1 fixed (20 trials) block was repeated four times for tracking the four experimental conditions. The experimental design for the present study is displayed in Figure 3.1.

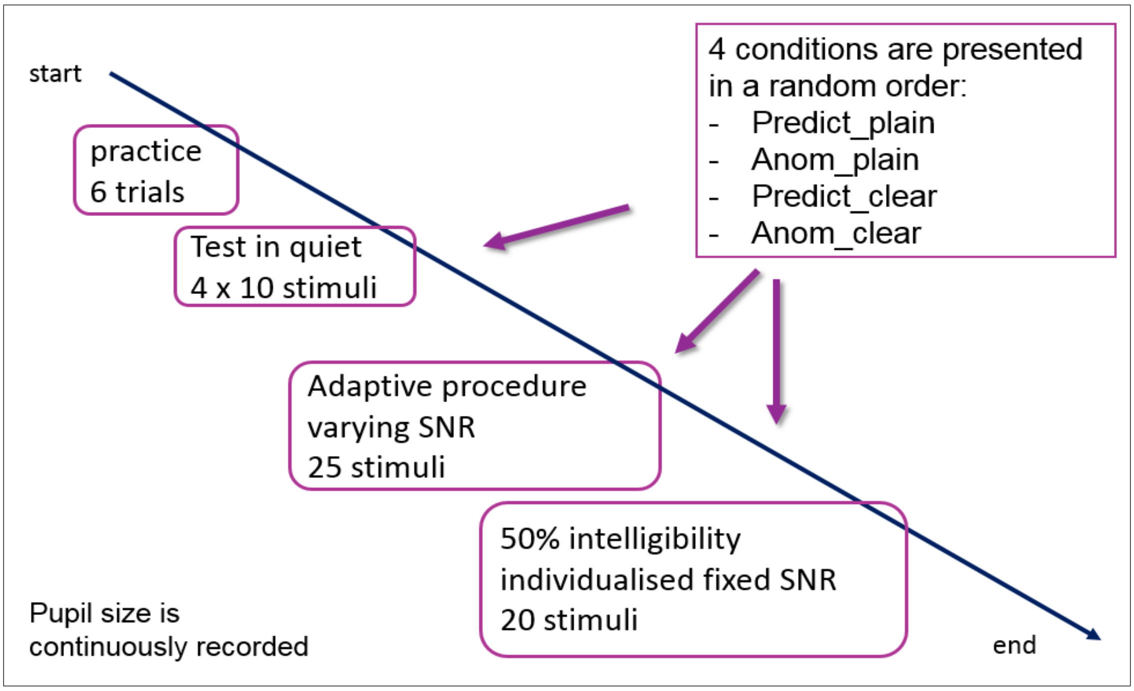

Figure 3.1 Experimental design for Study 2.

3.2.3 Procedure

The experimental procedure for the present study followed the same steps and arrangements used in Study 1 (see section 2.2.3). The only exceptions were that no background tests were administered at the beginning of the session, while at the end of the speech perception test, non-native participants were presented with the Listening section of the IELTS ("https://www.ielts.org/," 2017). The four audio tracks were presented from the loudspeaker, and participants were asked to register their answers on printed forms.

3.2.4 Pupillometry

Pupil diameter was recorded during the entire duration of the three experimental conditions using the same equipment and following the same exact procedure as in Study 1 (see section 2.2.4 for details). The pupil data were pre-processed following a multistep procedure. Series of missing values corresponding to blinks were expanded asymmetrically such that the time window including 80ms prior to and 160ms following a missing data stretch was omitted. The rationale for this was to exclude local pupil size disturbances caused by partial eyelid closures (Winn, 2016; Zekveld et al., 2010). Afterwards, missing data replacement and smoothing were performed as described in section 2.2.4 for Study 1. On average, missing data including loss of pupil track, blinking and gap expansion before and after blinks resulted in 19% of the total measurements collected. The one second period preceding the stimulus onset was regarded as baseline. A minimum of 25 samples (corresponding to 50ms) was defined as a requirement for a valid baseline estimation: otherwise the baseline window was extended backward until the criteria was met. Following the pre-processing, pupil data were averaged separately for each participant and condition (predict_plain, predict_clear, anom_plain and anom_clear).

The change in the pupil diameter was quantified in mm relative to the baseline. The time window considered for the analyses started 1 second prior the sentence onset, and ended 4.9 and 7.7 seconds after the sentence offset for plain and clear stimuli respectively. This was done to consider solely the time window before participants were prompted to repeat back the sentence they heard (the prompt always appeared 3 seconds after the sentence offset). The rationale for this choice was to exclude changes in the pupil diameter caused by movement planning and execution (Richer & Beatty, 1985). Unlike in Study 1, the latency of the pupil was not included in the analyses for Study 2 and 3. This decision was taken because from visual inspection of the plotted pupil data, the peak pupil dilation appeared to precisely align across conditions and participants groups. It was also supported by the fact that Study 1 did not show any effect of the variables considered on the latency of the peak. Three pupil outcome measures were obtained from the average trace of each participant and condition:

I. ***Pupil baseline:*** the average pupil diameter in the 1 s preceding the sentence's onset.

II. ***Mean pupil dilation*** relative to baseline pupil diameter. Time windows considered: between 0 and 4.9 s after the stimuli onset for plain stimuli, between 0 and 7.7 s after the stimuli onset for clear stimuli.

III. ***Peak pupil dilation***, as the maximum positive deviation from the baseline during the 4.9 and 7.7 s following stimuli presentation, respectively for plain and clear stimuli.

3.2.5 Statistical analyses

One way repeated-measures ANOVAs, one-way MANOVA and mixed design ANOVAs were conducted to test whether test condition (semantic context and speaking style) and linguistic background of participants (native or non-native listeners) affected behavioural data (intelligibility and SNR levels).

Mixed-effect regression models were performed to analyse the effect of semantic context, speaking style and language background on pupil measures. Additionally, mixed-effect regression models were also used to investigate the effect of length of residence, overall English use and IELTS score on pupil measures for non-native listeners only.

Stepwise regressions were performed to test whether background information about second language proficiency and usage were correlated with listeners' ability to tolerate background noise. Lastly, exploratory analyses using Growth Curve Analyses were performed with the aim of investigating the impact of second language proficiency level on the effort release after speech offset in non-native listeners.

3.3 Results

3.3.1 Behavioural results

Intelligibility scores in quiet for each of the four conditions are summarised in Table 3.2 (results for native, non-native and all participants) and Table 3.3 (results sorted by speaker and condition, non-native participants only). The reported means refer to the block in quiet (10 trials per condition) presented at the beginning of the experiment. The values are averaged across participants, excluding the practice trials.

Table 3.2 Descriptive statistics of the behavioural results for speech perception in quiet.

Behavioural results in quiet - Performance (% correct)						
Condition	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
predict_plain	93.2	*10.5*	89.8	*11.8*	99.5	*2.3*
predict_clear	96.5	*6.9*	94.6	*8.1*	99.6	*1.3*
anom_plain	86.4	*15.6*	80.0	*16.1*	98.2	*4.2*
anom_clear	93.3	*9.4*	90.1	*10.4*	98.8	*2.3*

Table 3.3 Descriptive statistics of the behavioural results in quiet for non-native listeners sorted by speaker.

Behavioural results in quiet – Non-native performance (% correct)								
Condition	**Speaker**							
	Female 1 **n = 12**		**Female 2** **n = 14**		**Male 1** **n = 14**		**Male 2** **n = 14**	
	M	***SD***	**M**	***SD***	**M**	***SD***	**M**	***SD***
predict_plain	83.4	*14.9*	91.1	*9.8*	92.2	*10.7*	91.7	*11.4*
predict_clear	92.5	*6.4*	94.7	*8.5*	95.0	*9.7*	96.1	*8.1*
anom_plain	73.7	*14.4*	75.6	*18.1*	84.2	*16.0*	85.8	*14.8*
anom_clear	89.1	*10.2*	89.7	*9.7*	91.4	*11.7*	90.3	*11.5*

A one-way MANOVA was run to determine whether there was an overall difference in performance between language groups, across the four conditions. Pillai's criterion was used to assess the statistical significance, because it is considered more robust for unequal sample sizes compared to the standard Wilks' λ (Tabachnick, Fidell, & Ullman, 2007). There was a statistically significant difference between native and non-native listeners on the combined dependent variables [F(4, 51) = 6.00, $p < 0.001$; Pillai's Trace = 0.616, partial $\eta^2 = 0.32$]. Overall, native listeners had a better performance in quiet compared to non-native listeners. Further analyses for the speech perception test in quiet focused on non-native participants. This choice was determined by the fact that native listeners' performance was at ceiling level (>98%) for each of the four conditions in quiet, as confirmed by a preliminary 2x2 repeated measures ANOVA with semantic context and speaking style as within-subjects factors performed on native listeners only. Indeed, neither factor was found to affect native listeners' intelligibility levels in quiet.

A 2x2 repeated measures AVOVA was conducted to examine the effect of semantic context and speaking style on non-native listeners' performance in quiet. There was a statistically significant interaction between semantic context and speaking style [F(1, 34) = 9.93; $p = 0.003$]. Main effects of semantic context and speaking style were also found to be significant [F(1, 34) = 52.71; $p < 0.001$ and F(1, 34) = 40.41; $p < 0.001$ respectively]. Follow-up Bonferroni-adjusted pairwise comparisons indicated that both in predictable and anomalous sentences, clear speech led to higher intelligibility rates compared to plain speech ($p<0.001$). Additionally, in both speaking style conditions, predictable sentences corresponded to a better performance compared to anomalous sentences ($p<0.001$). However, visual inspection of the plotted values seemed to indicate a greater benefit of clear compared to plain speaking style for anomalous sentences than for predictable ones. Additionally, contextual cues seemed to enhance intelligibility in quiet to a greater extent in the plain compared to the clear speaking style.

Further to that, for each condition (predictable and anomalous sentences pronounced in a plain and clear speaking style) an independent one-way ANOVA was performed on non-native listeners' performance in order to assure that different speakers used for the experiment did not affect intelligibility levels. In all conditions in quiet the effect of the speaker on the percentage of keywords correctly reported by non-native listeners did not reach significance level, hence individual characteristics of the speaker did not affect intelligibility levels in quiet.

Tables 3.4 and 3.5 summarise results from the speech perception test in noise, reporting intelligibility levels for each experimental condition and the corresponding SNR levels at which the fixed procedure blocks were run. As for Study 1, the adaptive block used to set SNR level is not included in the analyses.

Table 3.4 Intelligibility results in noise (50% target level) for all, native and non-native listeners.

Performance in noise (% correct)						
Condition	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
predict_plain	51.1	*15.8*	49.0	*16.1*	54.8	*14.9*
predict_clear	51.3	*11.9*	53.1	*10.5*	48.0	*13.7*
anom_plain	52.4	*10.1*	51.7	*9.7*	53.8	*10.8*
anom_clear	52.1	*11.4*	53.1	*9.8*	50.2	*14.0*

Table 3.5 SNR levels targeting 50% intelligibility for all, native and non-native listeners.

SNR levels						
Condition	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
predict_plain	-7.2	*5.2*	-5.1	*5.1*	-11.1	*2.7*
predict_clear	-14.8	*4.7*	-12.6	*4.3*	-18.8	*2.3*
anom_plain	-3.3	*7.4*	-0.5	*7.9*	-8.5	*2.1*
anom_clear	-12.1	*5.2*	-10.0	*5.0*	-16.0	*3.1*

A mixed design ANOVA performed on intelligibility levels with semantic context (predictable and anomalous) and speaking style (plain and clear) as within-subjects factors, and language (native and non-native) as a between-subjects factor was performed in order to verify whether the performance levels across conditions and participants were successfully equated. Results showed a significant interaction between speaking style and participants' linguistic background [$F_{(1, 52)}$ = 6.12, p = 0.017]. Follow-up Bonferroni-adjusted pairwise comparisons indicated that only native listeners performed marginally better when presented with clear speech as compared to plain (p = 0.050), and that native slightly outperformed non-native participants when attending to clear speech (p = 0.052). Despite the implementation of an adaptive procedure which included 10 additional trials compared to Study 1, behavioural results are not perfectly equated. Importantly however, only a small degree of variation is left across language groups and conditions. As expected, the SNR levels corresponding to 50% intelligibility were significantly lower for native participants than those required by non-native participants [$F_{(1, 52)}$ = 29.58, $p < 0.001$], for predictable compared to anomalous stimuli [$F_{(1, 52)}$ = 63.52, $p < 0.001$] and for sentences pronounced with a clear compared to a plain speaking style [$F_{(1, 52)}$ = 337.04, $p < 0.001$]. None of the interactions were found to be significant. SNR levels are displayed in Figure 3.2.

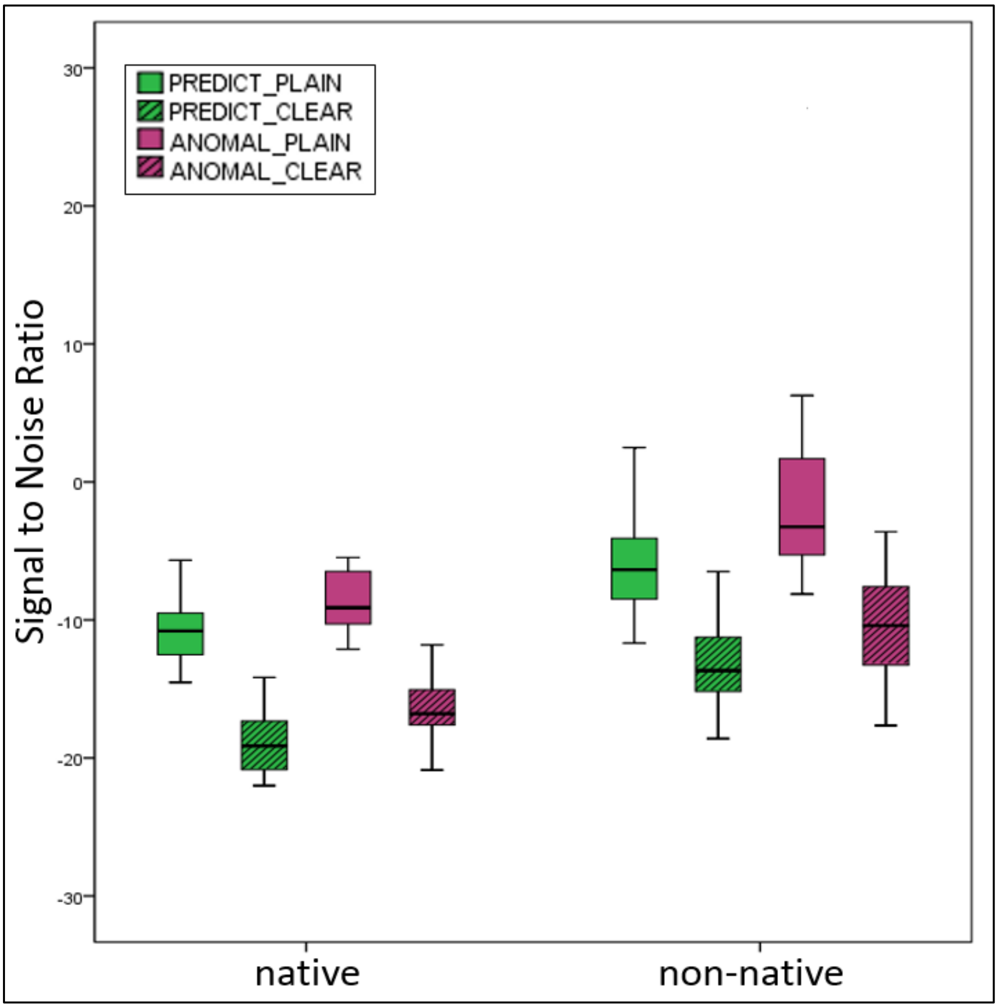

Figure 3.2 SNR values targeting 50% intelligibility for each experimental condition for native and non-native listeners.

3.3.2 Pupil data

Descriptive statistics for the pupil data collected during the blocks in noise with individually-fixed SNR are reported in Table 3.6 (measures for plain speaking style) and Table 3.7 (measures for clear speaking style). Measures include baseline pupil diameter, mean and peak pupil dilation over the baseline following stimuli presentation.

Table 3.6 Descriptive statistics of the pupil measures for the plain speaking style condition.

Pupil data in noise – plain speaking style						
Predictable stimuli						
Pupil outcome	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.12	*0.17*	0.15	*0.18*	0.07	*0.14*
Peak dilation, mm	0.28	*0.24*	0.33	*0.24*	0.20	*0.22*
Baseline, mm	4.96	*0.83*	5.14	*0.86*	4.65	*0.69*
Anomalous stimuli						
Pupil outcome	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.09	*0.16*	0.14	*0.14*	0.00	*0.16*
Peak dilation, mm	0.25	*0.22*	0.31	*0.21*	0.13	*0.19*
Baseline, mm	5.00	*0.85*	5.12	*0.90*	4.78	*0.73*

Table 3.7 Descriptive statistics of the pupil measures for the clear speaking style condition.

Pupil data in noise – clear speaking style						
Predictable stimuli						
Pupil outcome	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.05	*0.17*	0.10	*0.16*	-0.05	*0.16*
Peak dilation, mm	0.19	*0.20*	0.25	*0.20*	0.08	*0.14*
Baseline, mm	4.96	*0.84*	5.12	*0.88*	4.67	*0.67*
Anomalous stimuli						
Pupil outcome	**All**		**Non-native**		**Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.05	*0.21*	0.09	*0.18*	-0.03	*0.23*
Peak dilation, mm	0.21	*0.23*	0.26	*0.23*	0.13	*0.20*
Baseline, mm	5.05	*0.86*	5.22	*0.87*	4.73	*0.75*

Mixed-effects modelling

The effect of semantic context, speaking style and language background on pupil measures was modelled with a series of mixed-effect regression models, implemented using the lme4 package in the R environment (Bates et al., 2014; R Core Team, 2017). Separate models were built for each dependent variable (mean and peak pupil dilation and baseline). The initial saturated model included interaction terms for all independent variables as fixed effects with random intercepts and slopes (Barr, Levy, Scheepers, & Tily, 2013). Due to non-convergence, I simplified the models hierarchically from most complex to least complex. The resulting converged maximal models for all three variables included the following fixed effects: language background (2: native and non-native), semantic context (2: predictable and anomalous) and speaking style (2: plain and clear). Participant was included as random effect but no random slopes. The maximal model included up to three-way interactions between language background, semantic context and speaking style. Model residuals via chi-square tests (α = .05) were compared from the most complex models (containing the largest interaction term) to the least complex models (containing only single terms). If an interaction term was significant, all lower level effects involved in the interaction were included in the final model. Results for each of the dependent variables considered are reported below. Pupil curves dilation for native and non-native listeners in all experimental conditions are displayed in Figure 3.3.

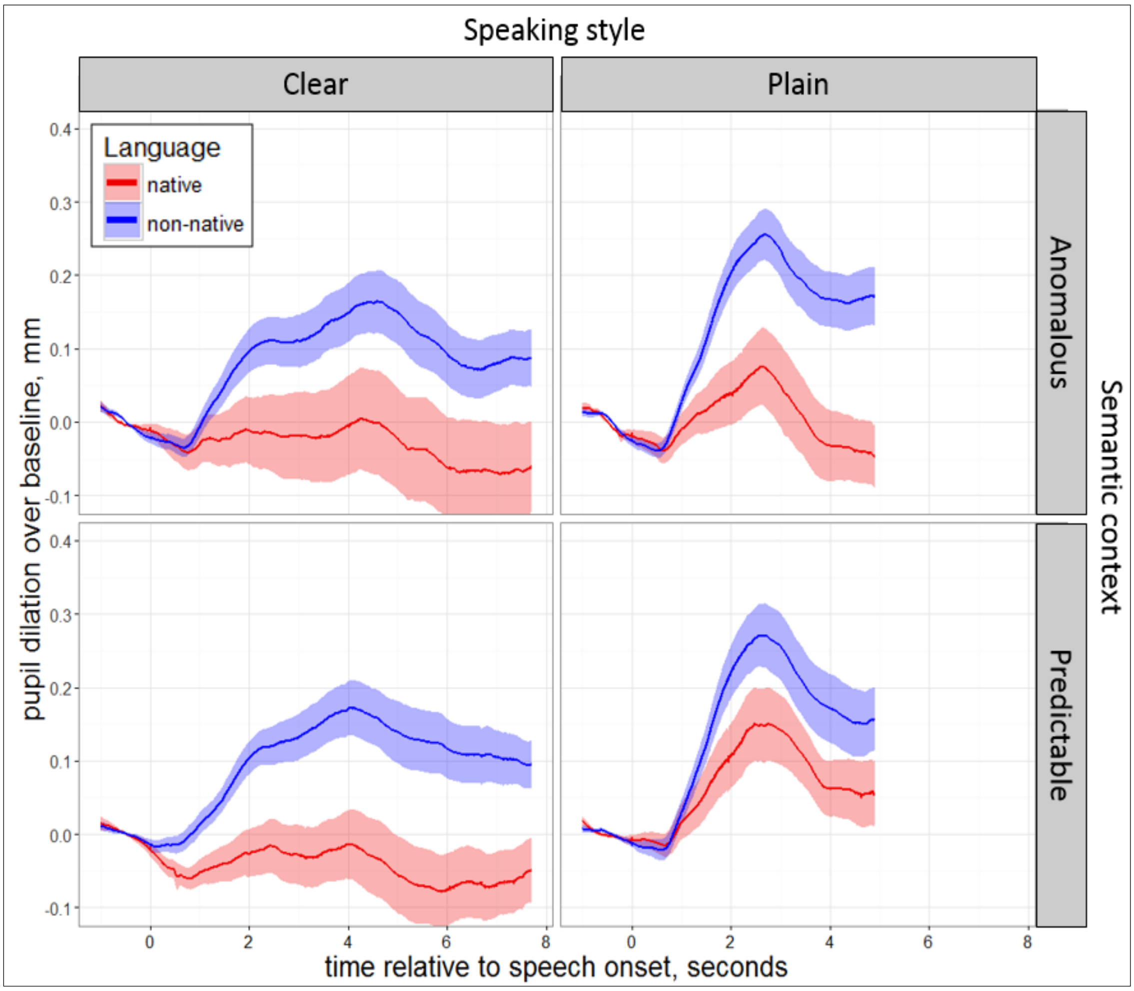

Figure 3.3 Mean pupil response over time during speech perception in all experimental conditions, for native and non-native listeners.

Mean pupil dilation

The final model included fixed effects of language background and speaking style, as well as a random effect of listener (see Table 3.8 for coefficients and estimated *p* values). This indicates that the overall mean pupil dilation was greater for non-native compared to native listeners, and sentences pronounced in plain compared to clear speaking style. However, there was no significant effect of semantic context on the mean pupil dilation.

Table 3.8 Fixed and random effects in a mixed-effects model of mean pupil dilation for all listeners.

Mean ~ Language + Style + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		-0.03	0.03	- 0.85	0.40
Language	Non-native	0.12	0.04	2.85	0.004
Style	plain	0.06	0.01	4.38	< 0.001
Random effects		**Variance**			
Participant	(intercept)	0.02			
Residual		0.01			
***Note.* Number of observations = 216; participants (N) = 54.**					

Peak pupil dilation

Similarly to the mean pupil dilation, the final model for peak values included fixed effects of language background and speaking style, as well as a random effect of listener (see Table 3.9 for coefficients and estimated *p* values). Overall, the peak pupil dilation was greater for non-native compared to native listeners, and for sentences in plain compared to clear speaking style.

Table 3.9 Fixed and random effects in a mixed-effects model of peak pupil dilation for all listeners.

Peak ~ Language + Style + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.10	0.04	2.46	0.014
Language	Non-native	0.15	0.05	3.00	0.003
Style	plain	0.06	0.02	3.99	< 0.001
Random effects		**Variance**			
Participant	(intercept)	0.03			
Residual		0.01			
***Note.* Number of observations = 216; participants (N) = 54.**					

Baseline

The final model for baseline dilation included the fixed effects of language background, speaking style and semantic context, as well as the interactions language background x speaking style, language background x semantic context, speaking style x semantic context, language background x speaking style x semantic context, and a random effect of Listeners (see Table 3.10 for coefficients and estimated *p* values). It was decided not to take the analyses of the three way interaction further, because there were no clear hypotheses behind the variation of the pupil baseline measure.

Table 3.10 Fixed and random effects in a mixed-effects model of baseline pupil diameter for all listeners.

Baseline ~ Language * style * context + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		4.73	0.19	25.48	< 0.001
Language	Non-native	0.48	0.23	2.10	0.036
Context	Predictable	-0.60	0.04	-1.33	0.183
Style	Plain	0.04	0.04	1.06	0.029
Language x Style	Non-native, Plain	-0.14	0.06	-2.54	0.011
Language x Context	Non-native, Predictable	-0.03	0.06	-0.61	0.539
Context x Style	Predictable, Plain	-0.07	0.06	-1.14	0.254
Language x Context x Style	Non-native, Predictable, Plain	0.18	0.08	2.28	0.023
Random effects		**Variance**			
Participant	(intercept)	0.03			
Residual		0.02			
***Note.* Number of observations = 216; participants (N) = 54.**					

Non-native participants

Additionally, in order to investigate the effect of length of residence, overall English use and IELTS score on the pupil measure, the same type of analysis was run with non-native listeners only for all three dependent variables (mean, peak and baseline).

The resulting converged models for all three variables included the following fixed effects: semantic context (2: predictable and anomalous), speaking style (2: plain and clear), length of residence, self-reported English use and IELTS score. Participant was included as random effect but no random slopes. The maximal model only included up to two-way interactions between conditions (semantic context and speaking style), and the background measures considered (length of residence, self-reported English use and IELTS score). Results for each dependent variable considered are reported below.

Mean pupil dilation

The final model only included the fixed effect of speaking style, confirming an overall reduction of the mean pupil dilation for non-native listeners for sentences produced with clear compared to plain speaking style (see Table 3.11 for coefficients and estimated *p* values).

Table 3.11 Fixed and random effects in a mixed-effects model of mean pupil dilation for non-native listeners.

Mean ~ Style + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.09	0.02	3.74	< 0.001
Style	plain	0.05	0.02	2.92	0.003
Random effects		**Variance**			
Participant	(intercept)	0.02			
Residual		0.01			
***Note.* Number of observations = 140; participants (N) = 35.**					

Peak pupil dilation

The final model only included fixed effect of speaking style (see Table 3.12 for coefficients and estimated *p* values). As for the means, the peak pupil dilation in non-native listeners was overall smaller for sentences produced with clear compared to plain speaking style.

Table 3.12 Fixed and random effects in a mixed-effects model of peak pupil dilation for non-native listeners.

Peak ~ Style + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.26	0.03	7.63	< 0.001
Style	plain	0.06	0.02	3.16	0.002
Random effects		**Variance**			
Participant	(intercept)	0.03			
Residual		0.01			
***Note.* Number of observations = 140; participants (N) = 35.**					

Baseline

The final model for baseline pupil dilation for non-native listeners included the interaction between the fixed effects of speaking style and semantic context, and a random effect of listener (see Table 3.13 for coefficients and estimated *p* values). Because of the significant interaction between speaking style and semantic context, the relative fixed factors were also included in the final model. Results showed that the pupil baseline diameter was greater for the anomalous compared to predictable sentences only when considering the clear speaking style condition. Additionally, the baseline pupil dilation was greater for clear compared to plain speaking style only for anomalous sentences.

Table 3.13 Fixed and random effects in a mixed-effects model of baseline pupil diameter for non-native listeners.

Baseline ~ Style + Context + Style x Context (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		5.22	0.15	35.66	< 0.001
Style	plain	-0.09	0.03	-2.85	0.004
Context	Predictable	-0.09	0.03	-2.86	0.004
Context x Style	Predictable, Plain	0.10	0.05	2.31	0.021
Random effects		**Variance**			
Participant	(intercept)	0.73			
Residual		0.02			
***Note.* Number of observations = 140; participants (N) = 35.**					

Overall, analyses on the non-native pupil data did not reveal any significant effect of length of residence, self-reported English use and IELTS score on the pupil measures.

Proficiency effect and Growth Curve Analysis

Individual differences in SNR levels were investigated in order to test whether the background information about second language proficiency and usage were correlated with listeners' ability to tolerate background noise. Four stepwise regression analyses were performed to this purpose, considering each of the four experimental conditions (predict_plain, predict_clear, anom_plain and anom_clear). In each of those regressions, length of residence, overall self-reported English use and result at the Listening section of the IELTS were entered as predictors. The variance of inflation factor was smaller than 2 for each regression coefficient considered, therefore we can assume that the regression results were not affected by multicollinearity. Results showed that the score obtained for the IELTS Listening test significantly contributed to predict the estimated SNR level for all the experimental conditions: a higher score predicted a lower SNR (i.e., better performance). Table 3.14 shows the results for all the significant predictors reported above.

Table 3.14 Effect of individual differences on the SNR levels for non-native listeners, stepwise regression results.

Dependent variable	**Predictor**	**R^2**	**B**	**Std. error**	**Std. beta**	**F**	**t**	**Sig.**
SNR predict_plain Non-native	IELTS Listening	.695	- .485	.056	- .834	75.239	- 8.674	<.001
SNR predict_clear Non-native	IELTS Listening	.514	- .349	.059	- .717	34.842	- 5.903	<.001
SNR anom_plain Non-native	IELTS Listening	.593	- .691	.100	- .770	48.025	-6.930	<.001
SNR anom_clear Non-native	IELTS Listening	.545	- .422	.067	- .738	39.558	-6.290	<.001

Although IELTS score was a significant predictor of SNR levels, results from the mixed-effect models showed that the IELTS score did not predict any of the pupil measures. This was not surprising, given that while pupillometry has been proved to be a reliable technique to uncover between subjects differences at a group level, pupillometric data from a single person are not considered as reliable due to a wide range of individual variability (Winn et al., 2018). Visual inspection of the pupil data collected for this study also confirmed the large individual variability of the pupil measures. Therefore, to further explore whether the non-native listeners' L2 proficiency level had an impact on pupil dilation, non-native participants were divided in two groups (labelled as High and Low Proficiency) based on their score at the IELTS Listening test. To obtain approximately the same number of participants in each group, the median IELTS Listening score was calculated (median score = 28) and considered as the upper limit of the low proficiency group. The high and low proficiency groups included 17 and 18 participants respectively. Table 3.15 summarises relevant background data and performance results for high and low proficiency groups.

Table 3.15 Background data and performance results for high and low proficiency groups.

	High proficiency		**Low proficiency**	
	M	***SD***	**M**	***SD***
Intelligibility in quiet (mean value across conditions)	95.8%	*4.3*	81.9%	*10.8*
Intelligibility in noise (mean value across conditions)	52.0%	*5.0*	51.5%	*5.5*
SNR (mean value across conditions)	-10.0	*1.7*	-4.3	*5.6*
Length of stay	3.0	*1.7*	1.6	*2.0*
English use (%)	46.0%	*0.19*	44.6%	*0.17*
IELTS Listening score	33.8	*2.7*	20.7	*7.2*

The investigation of the impact of proficiency level on pupil dilation focused on the recovery from effort following sentences' offset. This choice was driven by visual inspection of the pupil dilation over time for the two non-native proficiency groups (see Figure 3.4). High and low proficiency listeners seemed to differ in the overall effort deployment, and particularly on the rate of decrease of the pupil diameter after the peak dilation.

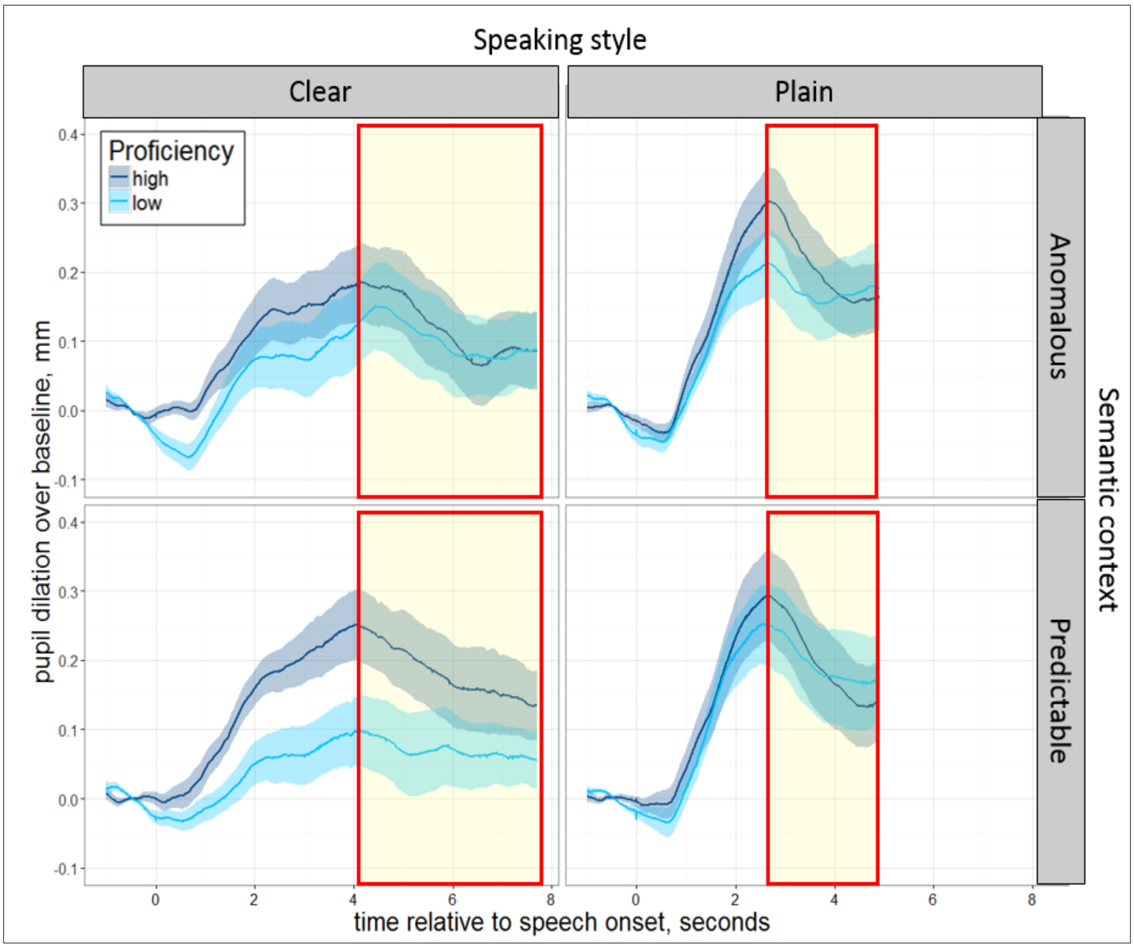

Figure 3.4 Mean pupil response over time during speech perception in all experimental conditions, for high and low proficiency non-native listeners.

To this purpose, growth curve analysis (GCA) (Mirman, 2014) was used to analyse the effort release after speech offset in non-native listeners for each of the experimental conditions (predict_plain, predict_clear, anom_plain and anom_clear). GCA is particularly well suited to test differences across multiple time-component (e.g. linear, quadratic, cubic) for specific time windows of interest. Consistent with the approach adopted by Verney et al. (2004) and Winn (2016), two different time windows were identified for the plain and clear speaking style condition, respectively starting at 2.6 and 4 seconds, and ending at 4.9 seconds and 7.7 seconds after sentence onset. The use of two different time windows for the plain and clear speaking style was justified by a slower speaking rate used in the clear condition. The rationale for these time windows' selection was to include the section of the pupil curve starting just after the peak, and ending just before the response prompt, in order to explore the effort release trajectory. Figure 3.4 displays the pupil curves dilation for high and low proficiency non-native listeners for all the experimental conditions, and the relative time windows considered for the GCA. The pupil curves for non-native listeners for each condition during the release time windows were modelled with second-order orthogonal polynomials and fixed effects of proficiency level on all time components (intercept, linear and quadratic). The intercept was intended to gauge the overall level of pupil dilation, and it can be considered as analogous to a standard analysis of variance. The linear component (slope, time1) was used to estimate growth rate in pupil dilation (or constriction as is the case for the time window selected) over time. Lastly, the quadratic term (time2) was intended to capture the inflection of the growth function, and it was primarily included to improve the model fit, rather than corresponding to a specific experimental prediction. The high proficiency group was treated as the baseline and parameters were estimated for the low proficiency group. The model also included random effects of listeners on all time terms. The fixed effect of second language proficiency (high vs low proficiency level) was added individually and their effects on model fit were evaluated using model comparisons. Improvements in model fit were evaluated using -2 times the change in log-likelihood, which is distributed as χ^2 with degrees of freedom equal to the number of parameters added. All analyses were carried out in R using the lme4 package (R Core Team, 2017).

The effect of proficiency level on the intercept and on the quadratic term did not improve the model fit for any of the experimental conditions. The effect of proficiency on the linear term, however, did improve the model fit for anomalous sentences pronounced in a plain speaking style ($\chi^2_{(1)} = 4.40$, $p = 0.036$). As can been seen in Figure 3.5, the model provided a good fit for the data within the time window selected for the anom_plain condition. These results suggest that, even though the overall pupil dilation during the effort release time window did not significantly differ between highly and low proficient non-native listeners, the rate of effort release after stimuli presentation is significantly lower for less proficient non-native speakers when they are attending to anomalous sentences produced in plain speech (i.e. the most difficult condition), as compared to highly proficient non-native speakers. Nevertheless, the effect was significant but fairly weak, and the same result was not replicated for sentences pronounced with a clear speaking style, and for semantically predictable sentences pronounced with plain speaking style. Table 3.16 shows the fixed effect parameter estimates and their standard errors along with *p*-values estimated using the normal approximation for the *t*-values.

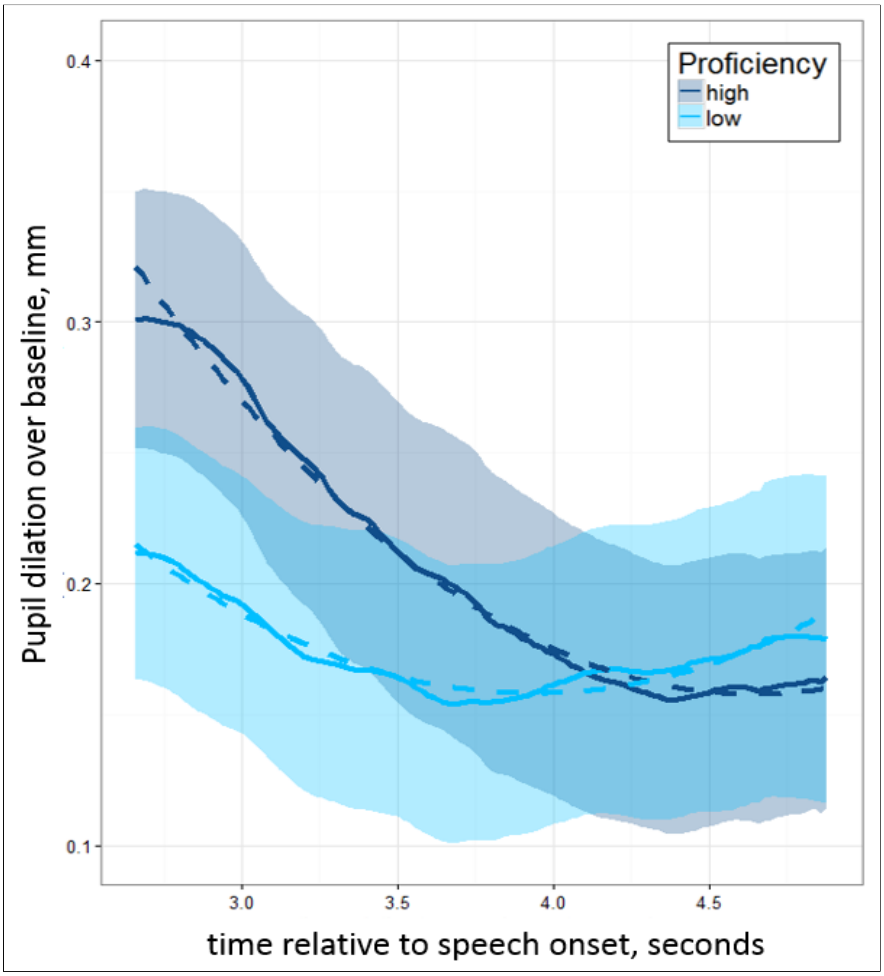

Figure 3.5 Mean pupil response over time during effort release for high and low proficiency non-native listeners for the anom_plain condition (solid line), and second order polynomial model fit (dashed line).

Table 3.16 Generalised linear mixed-effects model formula code and summary output for growth curve analyses on non-native listeners' pupil data in noise.

Formula code: PupilDilation ~ (time1 + time2) * Proficiency + (time1 + time2 | Listener)								
Non-native pupil data in noise – plain speaking style								
	Predictable stimuli				**Anomalous stimuli**			
Term	Est.	St. error	t value	*p*	Est.	St. error	t value	*p*
Intercept	.20	.06	3.17	0.002**	.21	.05	4.07	<0.001***
time1	-.59	.13	-4.48	<0.001***	-.50	.14	-3.63	<0.001***
time2	.10	.06	1.72	0.085	.16	.07	2.47	0.014*
ProficiencyLow	-.00	.09	-.05	0.964	-.03	.07	-0.47	0.639
time1:Proficiency Low	.30	.18	1.63	0.103	.42	.19	2.19	0.029*
time2:Proficiency Low	.00	.08	.04	0.967	-.03	.09	-0.28	0.780
Non-native pupil data in noise – clear speaking style								
	Predictable stimuli				**Anomalous stimuli**			
Term	Est.	St. error	t value	*p*	Est.	St. error	t value	*p*
Intercept	.18	.05	3.59	<0.001***	.12	.06	2.16	0.031*
time1	-.48	.14	-3.40	<0.001***	-.51	.18	-2.76	0.006**
time2	.09	.11	.81	0.416	.15	.09	1.67	0.096
ProficiencyLow	-.11	.07	-1.58	0.113	-.02	.08	-0.22	0.826
time1:Proficiency Low	.34	.20	1.74	0.082	.20	.26	0.77	0.444
time2:Proficiency Low	-.03	.15	-.22	0.829	-.05	.13	-0.36	0.721

3.4 Discussion

Study 2 was designed to investigate how the presence of semantic and acoustic enhancements affected listening effort in native and non-native listeners when levels of intelligibility are equated.

Behavioural results showed that overall native listeners were able to tolerate a less favourable SNR compared to non-native listeners, when the intelligibility level across the two groups was equated. Additionally, overall listeners were able to cope with a lower SNR level when attending to semantically predictable compared anomalous stimuli, and when sentences were produced with a clear compared to a plain speaking style.

The two groups of listeners hence differed in their overall speech perception ability in noise. Importantly however, native and non-listeners were equally able to take advantage of sentence-level contextual information, and of acoustically enhanced speech by means of a clear speaking style produced by the talker. The results complement data from previous research, and further establish that clear speech is an effective means to enhance speech perception for both native and non-native listeners, and that they are able to derive a considerable benefit from signal clarity enhancement strategies (Bradlow & Bent, 2002; Grynpas, Baker, & Hazan, 2011; Van Engen et al., 2014). In addition to that, the availability of contextual information has proved to be effective in improving the intelligibility of speech in noise for both listener groups, as already shown for native listeners (Drager & Reichle, 2001; Wingfield & Tun, 2007; Zekveld, Rudner, et al., 2011). Nevertheless this pattern of results is somehow in contrast with previous research that reported a greater clear speech benefit for native compared to non-native listeners (Bradlow & Bent, 2002), and a contextual benefit for non-native listeners only when speaking clarity was simultaneously enhanced (Bradlow & Alexander, 2007).

However, the perspective taken here is slightly different from that in previous studies. Indeed, rather than measuring how acoustic and semantic cues improved intelligibility, in the current experiment intelligibility remained fixed at 50%, and changes in the SNR levels were compared. However, because lowering the SNR means enhancing the level of energetic masking, it can be argued that performing the task with a less favourable SNR further reduced the availability of relevant acoustic cues, and that this further reduction would act in addition to the experimentally planned manipulation on signal clarity. Therefore, while the speaking style benefit showed by native and non-native listeners was about the same magnitude (approximately 8 dB), the native group might have experienced a greater degradation of the target speech due to the combined effect of a louder background noise and of the speaking style experimental manipulation. It might thus remain true that native listeners are able to gain a greater advantage from clear speaking style compared to non-native listeners, as previously reported (Bradlow & Bent, 2002).

Taken together, results from pupil data confirmed that listening effort during sentence perception is higher for non-native compared to native listeners when intelligibility levels are equated, and this was found overall regardless of which contextual and acoustical enhancements were made available to the listener. These findings are in line with previous research evaluating listening effort in non-native listeners during spoken word recognition (Schmidtke, 2014) and sentence perception (as in Study 1).

Surprisingly, the presence of a coherent semantic context within a sentence was not found to impact on the associated listening effort. This was unexpected, based on previous findings reporting a consistent reduction in cognitive effort for predictable versus unpredictable sentences in native listeners (Winn, 2016). However, the speech material used to manipulate context predictability in the present study was inherently different from the material used in the study mentioned above. Indeed, the listening task adopted from Winn (2016) featured sentences with high and low semantic context (e.g. "*Stir your coffee with a spoon*" versus "*Jane thought about a spoon*"), while in the study presented here the predictable sentences were contrasted against semantically anomalous sentences (e.g. "*The talented artist drew a picture*" vs "*The vegetables open a difficult hat*"). Unlike in Winn (2016), sentences were presented in blocks of either semantically predictable or anomalous sentences, therefore listeners had the chance to anticipate if a semantically coherent sentence was to be expected. This may have resulted in listeners being aware of the necessity to not rely on semantic cues, given the lack of a coherent semantic context. By ruling out this strategy, they may have consequently inhibited the processing itself of the semantic information available in the sentence. This could explain why a theoretically more effortful meaning integration did not elicit a greater pupil response, as found by Kuipers and Thierry (2011) when considering unrelated versus related word/picture pairs. Therefore, the stimulus selection adopted in the present study might not have been ideal to the purpose of investigating whether non-native listeners are able to benefit from semantic context in terms of listening effort reduction. Study 3 will address this research question by designing better tailored testing material to allow an improved experimental comparison between the presence and the lack of a semantic context during speech perception.

This study also showed reduced listening effort attributable to a speaking style benefit (i.e. lower effort for sentences produced with a clear compared to a plain speaking style). There is limited availability of research addressing the effects of speaking style on pupil response.

Simantiraki, Cooke, and King (2018) similarly found reduced listening effort for sentences pronounced with a clear (author referred to this condition as "Lombard", since stimuli were recorded in the presence of background noise) compared to a plain speaking style when presented at –3 and –5 SNR (but not a -1 SNR). However, intelligibility levels across speaking styles were not equated, and may be responsible for the difference observed in the pupil response. Along a related line of investigation, Koch and Janse (2016) did not find any speech rate effect on pupil response. To enhance the ecological validity of their study, the authors did not perform any artificial time-compression of speech, but instead used conversational materials with natural variation in speech rate. Although changes in the speaking rate are an essential component of speaking style adaptations that fall under the term of "clear speech", other factors that are intrinsic to a clear speaking style may have contributed to the significant reduction in listening effort for clear compared to plain speaking style reported in the present study. The combined contribution of a diminished speaking rate, a wider dynamic pitch range and a larger vowel space may have affected the ease of processing and the cognitive demand in a way that solely a reduction in the speech rate did not. Additionally, it is worth noting that a reduction of listening effort linked to the use of clear speaking style was found despite a more challenging SNR level. Indeed, listeners were able to tolerate a less favourable SNR when attending to clear compared to plain speech at equated levels of intelligibility.

One additional focus of the present study was the investigation of the effects of L2 proficiency on non-native listening effort and performance. Non-native participants were divided into high and low proficiency groups based on their IELTS Listening performance, which was the only measure that significantly predicted the SNR levels in all the experimental conditions for non-native listeners. The rate of listening effort release after speech offset was slower for listeners with a low compared to high proficiency level when attending to anomalous sentences pronounced in plain speaking style, even though the proficiency level did not affect the overall pupil dilation in any of the experimental conditions. This indicates that the difference between high and low proficiency non-native listeners is more evident in more challenging listening conditions (when acoustic cues are more degraded and semantic cues are not available), and that the listener's proficiency level mainly affects the rate of recovery from effort after speech offset. Previous research has shown that highly proficient non-native listeners are better at ignoring L1 interferences during a listening task, demonstrating that they were better able to "zoom in" to the target language compared with less proficient listeners (Elston-Güttler & Gunter, 2009). One possibility is that an increased interference from L1 vocabulary contributed to the prolonged increase in listening effort found for less proficient non-native listeners. A less rapid effort release was also found for people who use a cochlear implants compared to normal hearing listeners (Winn, 2016). Nevertheless, results from the present study are somehow in contrast with findings reported by Francis et al. (2018), that reported a positive correlation between pupil dilation and English proficiency score in a study exploring speech perception in Dutch listeners attending to sentences both in Dutch and English. However, authors found a significant correlation only in few of the experimental conditions examined (2 out of 8), also including trials where native speech perception was considered. It was therefore difficult to formulate a straightforward interpretation of the trend reported. Authors suggested that results may simply reflect the influence of a more general cognitive capacity, which also affects susceptibility to distractions.

In order to clarify the relationship between listening effort and second language proficiency, further research might better tailor the distinction between high and low proficiency in non-native listeners. Indeed, a median split based on the IELTS score may not be ideal given the continuous scale of proficiency levels. A participant recruitment process targeted at selecting two groups of participants with distinct proficiency levels may be a better solution for exploring the impact of proficiency on listening effort.

Lastly, it is worth mentioning that since the stimuli used for the predict_plain condition in the present study were also used in Study 1 (the BEL sentence corpus), a cross-study comparison was made possible. Both SNR levels and pupil outcome measures for native and non-native listeners recorded in Study 2 (targeting 50% intelligibility) are comparable to those measures collected in Study 1 for the low intelligibility condition (where 40% intelligibility was targeted). A small difference in the values reported is compatible with the 10% difference in intelligibility levels, therefore a slightly lower SNR and a greater pupil response was recorded for Study 1 when targeting 40% intelligibility, compared to Study 2 targeting 50% intelligibility, when the same set of sentences was used.

In conclusion, Study 2 confirmed that pupil response is a sensitive measure to uncover listening effort differences between native and non-native listeners, and established further that a greater listening effort is required when trying to understand a second compared to a native language in noise even when overall intelligibility is matched. In addition, this study uncovered potential effects of speaking style on listening effort. Findings suggested that attending to sentences pronounced with a clear speaking style has a reduced cost in terms of cognitive effort. Lastly, findings from the present experiment did not show any impact of semantic context manipulations on the cognitive demand required, even though attending to semantically predictable rather than anomalous sentences allowed listeners to tolerate a greater background noise. However, a more tailored contrast between experimental conditions (a comparison between high and low availability of semantic context, as opposite to the use of predictable and anomalous stimuli) would help to clarify whether a reduction in listening effort can be achieved through the exploitation of semantic content, by allowing an easier access to the sentences' meaning.

Chapter 4

4. Study 3

4.1 Aim of the Study

Study 3 was designed to clarify the contribution of semantic context during non-native speech perception. Contrary to predictions, results from Study 2 did not indicate any effect of semantic context on listening effort. To explain this, it was hypothesised that listeners might have inhibited the processing of the semantic context during the blocks of anomalous sentences, since it was not helpful to formulate useful predictions. It should also be recalled that the target intelligibility level in Study 2 was fixed at 50% of correctly reported keywords. As a result, when semantic processing of stimuli is attempted by listeners, it is reasonable to assume that this will come at a considerable cognitive cost, because of the sparse semantic context available. Therefore there were shortcomings in the design chosen to evaluate whether non-native listeners are able to benefit from contextual information and to reduce the amount of effort deployed. As mentioned earlier, a comparison between experimental conditions featuring high vs low availability of semantic content would be better suited to test the hypothesis. Additionally, in real life communications, the availability of semantic information does not lie exclusively at a sentence level, but it is likely to span several sentences within a dialogue or a communicative unit.

For these reasons, in Study 3 a new set of experimental stimuli was developed to investigate whether the availability or lack of semantic information spanning several sentences would result in a reduction in listening effort for native and non-native participants when intelligibility levels were equated. Semantic context was manipulated across sentences, by creating lists of 12 semantically related or unrelated sentences; for example, 12 sentences all referring to a specific animal or fruit, as opposed to 12 sentences each about a different topic. I will refer to this experimental manipulation of semantic context as "long-term context", as opposed to sentence-level context. To investigate any reduction in listening effort due to the presence of long-term semantic context, the analyses of pupil outcome measures focused on the last section of each block. The rationale for it was to compare the related and unrelated conditions at a stage where substantial semantic context was available to listeners, if present, and where its potential benefit could have reached its maximum. A higher target intelligibility level (80%) compared to Study 2 (50%) was chosen in order to ensure that participants had a greater availability of acoustic and contextual (when available) cues across sentences.

Additionally, subjective ratings of listening effort were collected in order to investigate whether perceived effort differed between native and non-native listeners, and to verify whether any benefit gained from the semantic context (either at a behavioural or listening effort level) would also be reflected in a reduced perceived effort. Previous research found that in native listeners with or without hearing impairment, a higher intelligibility score was overall related to lower listening effort as reflected by both pupil response outcomes and subjective ratings of effort. However, differences in the subjective ratings between conditions were generally not related to differences in the pupil measures (Zekveld, Kramer, et al., 2011).

Specifically, the study aimed to answer the following research questions:

1. Is long-term context beneficial for listeners in terms of listening effort reduction?
2. Are non-native listeners able to take advantage of the long-term semantic context within a block of sentences?
3. Is the effect smaller compared to native listeners?
4. Is the long-term context benefit modulated by proficiency level?

In relation to the research questions, it was predicted that listening to semantically related sentences would result in a reduced listening effort compared to unrelated sentences. This is based on previous research showing a rapid reduction of listening effort due to semantic context at a sentence level for listeners with normal hearing (Winn, 2016). The study showed that listeners performed better when words were preceded by relevant semantic context, and that this advantage was also reflected in a reduction in listening effort. However, the study from Winn (2016) only considered semantic manipulation at a sentence level, while the aim of the present experiment is to expand these findings by considering semantic context over several sentences, and by comparing native and non-native listeners.

4.2 Materials and Method

4.2.1 Participants

Forty three participants from two different language backgrounds participated in the study. Three non-native participants were then excluded from analyses because their performance during the experimental task in noise was below two standard deviations of the mean performance in at least one of the two experimental conditions. One non-native participant was further excluded because the adaptive procedure failed to work resulting in an extremely high SNR level, probably due to a low proficiency level in English. Following exclusions, the non-native group included 21 participants (4 men and 17 women) with Italian as L1 and English as L2, aged 20-33 years (M = 26.7, SD = 3.9). All participants had been living in the UK for at least 3 months at the time of testing. The native group consisted of 18 native British English participants (7 men and 11 women), aged 19-35 years (M = 26.4, SD = 4.8). Among non-native participants, one of them previously took part in Study 1, while three of them also participated in Study 2. Among native participants, two of them also took part in Study 1 and 2, four of them in Study 2. Study 1 and 2 were run approximately 1 and 2 years earlier respectively. The study was advertised via the UCL Psychology subject pool and via social media. The same exclusion criteria as in Study 1 and 2 were considered. All participants provided written informed consent before participating in the study. They received a monetary compensation for their time under a protocol approved by the Ethics Committee at University College London.

4.2.2 Stimuli and Tests

Background tests

All participants were screened using pure tone audiometry to verify that their hearing threshold were 20 dB HL or better at octave frequencies between 250 and 8,000 Hz. Prior to the start of the experimental session, non-native participants were asked to complete the same on-line linguistic background questionnaire as used in Studies 1 and 2. In addition, as for Study 2, non-native participants were presented with the Listening module of the International English Language Testing System (IELTS), an international standardised test of English proficiency for non-native speakers ("https://www.ielts.org/," 2017). For Study 3, a version of the test using American accent was presented. This was decided because the stimuli used for the present experiment were recorded from an American English speaker. This was due to a research visit at the Department of Speech and Hearing science of the University of Washington, Seattle, to work in collaboration with Dr. Matthew Winn, which coincided with the experiment design and stimuli recording phase. Only one speaker was used for the present study. Although this may weaken the generalisability of results as compared to multiple speakers, only plain speaking style was used here, therefore reducing the potential variability across speakers. The IELTS module administered comprises four sections, with ten questions in each section and a final score ranging from 0 to 40. The total duration is 30 minutes. The score obtained by non-native participants who took part in the experiment ranged between 17 and 38. The aim of this test was to obtain an accurate picture of the proficiency level of the participants, with a focus on their speech understanding skills. Participants' background data are summarised in Table 4.1.

Table 4.1 Demographics and background tests results for native and non-native participants.

Background information		**Native listeners**		**Non-native listeners**	
		M	***SD***	**M**	***SD***
Age		26.4	*4.8*	26.7	*3.9*
(Non-native only)	Length of residence (years)	N/a		3.0	*1.8*
	Overall English use	N/a		46%	*0.2*
	Self-reported English knowledge (0-6)	N/a		4.8	*0.9*
	IELTS Listening (0-40)	N/a		29.6	*6.4*

Experimental Stimuli

Stimuli used for this experiment were an adaptation of the stimuli from the Connected Speech Test developed by Cox, Alexander, and Gilmore (1987). This test was originally developed as a test of intelligibility of everyday speech for the investigation of hearing aid benefit in native speakers with hearing impairment. Each passage of the test was of equal intelligibility for normal hearing listeners, as reported by authors. The corpus used for this study includes 45 lists of 12 sentences, each list consists of a passage of connected speech about a familiar topic such as common plants, animals and household objects (see Appendix D). A total of 540 sentences were recorded in a sound-attenuated booth at the Speech and Hearing Sciences Department at the University of Washington with a sampling rate of 44.1 kHz, 16-bit resolution, and pronounced by one male native Western American speaker of English at a natural self-paced rate. Sentence duration was between 2.1 and 4.3 seconds (M = 2.9). Recordings were RMS normalised to an average amplitude of 65 dB. Each sentence had four keywords, which were used to score comprehension. In order to manipulate the long-term context, two sets of stimuli were created for each participant before the testing session. First, 6 lists of 12 related sentences were randomly selected to be presented. Afterwards, 72 sentences were randomly drawn from the remaining material not previously selected, and arranged in 6 lists of 12 unrelated sentences each, so that one list of unrelated stimuli never included more than one sentence on a given topic. Lastly, 30 additional sentences were randomly selected from the remaining stimuli, and used during the practice and adaptive blocks. A sentence was never played more than once during the entire experimental session for a given participants. An example of one related and one unrelated list is presented in Table 4.2.

Table 4.2 Example of related and unrelated lists of sentences presented (keywords are in capital letters).

RELATED SENTENCES	UNRELATED SENTENCES
1. a CARROT is a VEGETABLE RELATED to PARSLEY	1. one END of a NAIL is VERY SHARP
2. the PLANT of CARROT probably ORIGINATED in PERSIA	2. this ALLOWS the BIRD to CLING to TREES
3. the long STEM of the CARROT GROWS UNDERGROUND	3. VEGETABLES can be EATEN RAW or COOKED
4. it is this STEM that MOST PEOPLE EAT	4. in the PAST GOLD was used as a MONETARY CURRENCY
5. the LEAVES of the CARROT are ALSO EATEN	5. later VIOLIN MAKERS IMPROVED their CRAFT
6. they are OFTEN USED to FLAVOR FOODS	6. the CABBAGE PLANT can live through SEVERAL FREEZES
7. the ROOTS CONTAIN high AMOUNTS of VITAMINS	7. ICE was once USED to COOL REFRIGERATORS
8. SPRING CROPS are GROWN in the western STATES	8. DONKEYS are often USED for HARD LABOR
9. the CROP is HARVESTED in one HUNDRED DAYS	9. a POPULAR STYLE of GUITAR has a FLAT top
10. fall CROPS are GROWN in the NORTHERN STATES	10. PAPER ENVELOPS were DEVELOPED in CHINA
11. WINTER HARVESTS USUALLY come from CALIFORNIA	11. they can JUMP MANY TIMES their own LENGTH
12. CARROTS can be STORED for SEVERAL MONTHS	12. LOCUSTS have much SHORTER FEELERS than GRASSHOPPERS

Experimental task

As for the previous experiments presented, the experimental task consisted of a speech intelligibility test where participants heard sentences and were prompted to repeat them back to the experimenter. The mode of presentation of stimuli was as described for Study 1 and 2. Testing started with five practice items presented in quiet. Following that, one block of 25 trials was presented, using an adaptive procedure. During the adaptive block, unrelated sentences were presented. This was decided because only 12 related sentences were available for each topic, while 25 trials were needed for an accurate SNR level estimation. The SNR level targeting 80% of correctly reported keywords was therefore calculated using unrelated stimuli. For this reason, I expected the intelligibility levels not to be perfectly equated across conditions, but still to be comparable across language groups. This had the disadvantage of a potential ceiling effect in intelligibility levels for the related condition, leaving less room for improvement attributable to long-term context. However, a relatively high intelligibility level was targeted to ensure a substantial availability of long-term contextual cues in the related condition. The adaptive procedure implemented here followed the exact same steps described in details for the previous study, with the same 8-talker babble noise used to mask speech. Following the adaptive block, the SNR values corresponding to the reversals were averaged to obtain a single SNR value. In the following 12 blocks, audio stimuli were presented at that fixed SNR level. Six related and six unrelated blocks of 12 sentences each were then presented in an alternate order, with the starting condition (related or unrelated) randomly decided for each participant. The procedure for the testing session closely followed the steps described for Study 2.

A novel aspect used in this study was that subjective measures of effort were collected during the experiment. The purpose of it was to investigate whether perceived effort differed between native and non-native listeners, and between the related and unrelated condition. Additionally, I was interested in exploring whether subjective measures would correlate with pupillometric measures of listening effort. Following the practice trails, and before the start of the experimental blocks, participants were informed that during the course of the experiment they would be asked to rate their perceived effort during listening. An adaptation of the definition of listening effort proposed by McGarrigle et al. (2014) was given to them as a reference: “the mental effort required to attend to, and understand, the sentences you heard”. The subjective listening effort measurements were collected using a simplified version of the scaling method used by Luts et al. (2010). The scale consisted of 7 labelled categories, to each of them was assigned an effort scaling numerical value from 1 to 7 (correspondent labels were “no effort”, “ very little effort”, “little effort”, “moderate effort”, “considerable effort”, “very much effort”, “extreme effort”). At the end of blocks 5 and 6 (that corresponded at approximately half of the experimental session), participants were presented with a printed copy of the rating scale and listening effort definition, and were asked to give a rating of subjective listening effort referring to the last block of sentences they heard, using the scale provided. Since related and unrelated sentences were presented in an alternate order, one subjective measure for each of the two conditions was collected for each participant. The experimental design adopted is visually displayed in Figure 4.1.

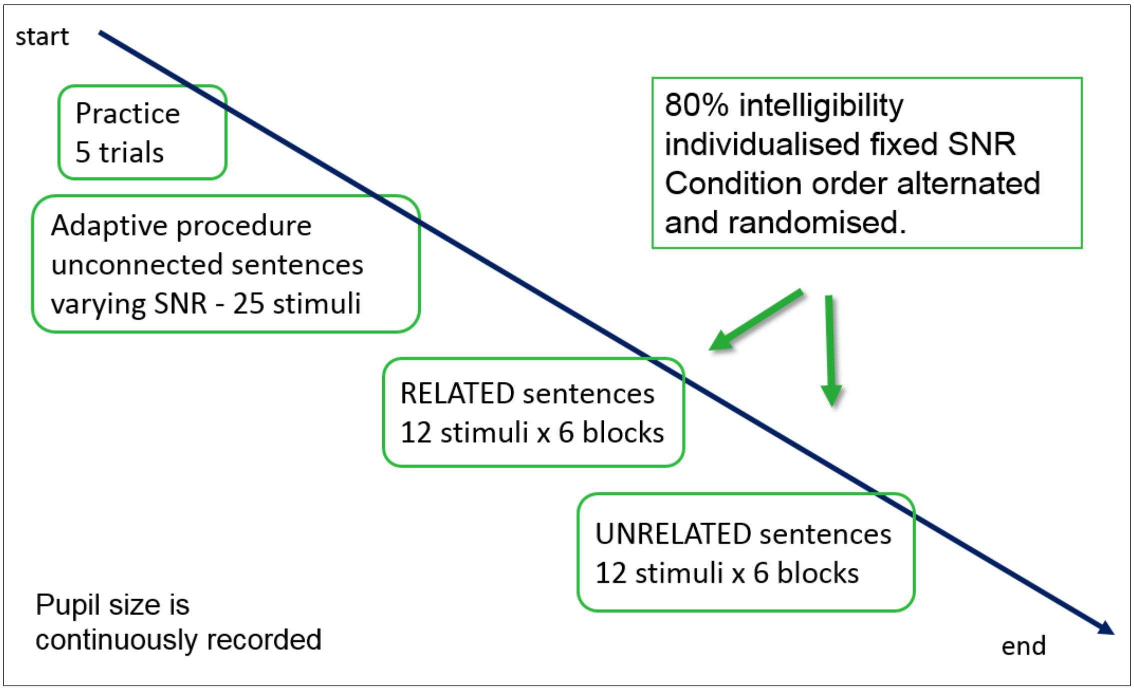

Figure 4.1 Experimental design for Study 3.

4.2.3 Procedure

As for the previous studies, non-native participants were asked via e-mail to complete an on-line language background questionnaire before the experimental session. The same procedure as in Study 2 was followed, and the same precautions were taken in order to ensure both a comfortable experience for the participant, and a reliable data collection (see sections 2.2.3 and 3.2.3). Participants were not informed of the purpose of the study before or during the experimental session. The experimenter did not mention to participants that there would be an alternation of related and unrelated blocks.

4.2.4 Pupillometry

During the speech perception test, the pupil size and location of the left eye were measured using an EyeLink 1000 eye-tracker. The exact same settings as in the previous studies were used for data collection. The multistep procedure for pupil data pre-processing described in Study 2 was implemented for the present study. On average, missing data including loss of pupil track, blinking and gap expansion before and after blinks resulted in 18% of the total measurements collected.

The 500 milliseconds period preceding the stimulus onset was regarded as baseline. A shorter time window compared to Study 1 and 2 was selected to be used as baseline. This decision was taken because from a visual inspection of the data it appeared that the participants' pupil was still constricting following the dilation in the previous trial during the 2 seconds prior to sentence onset. This was likely to be caused by the experimenter using too fast a pace when initiating trials. The selection of a shorter baseline window therefore minimised the impact of the previous trial on pupil dilation as calculated over the baseline. A minimum of 25 samples (corresponding to 50ms) was defined as a requirement for a valid baseline estimation: otherwise the baseline window was extended backward until the criterion was met.

Following the pre-processing, pupil data were averaged separately for each participant and condition (related and unrelated). Additionally, each block of sentences was divided into 3 sections consisting of 4 sentences each, and each containing 16 keywords. Sections were labelled as "beginning" (sentences 1-4), "middle" (sentences 5-8) and "end" (sentences 9-12). The justification for it was to explore how intelligibility score would change within a block of sentences, and whether any change would be modulated by the availability of a consistent semantic context. The change in pupil diameter was quantified relative to the baseline. The time window considered for the analyses started 500 milliseconds prior the sentence onset, and ended 5.9 seconds after the sentence offset. This was done to consider solely the time window before participants were prompted to repeat back the sentence they heard (the mean sentence duration was 2.9 seconds, and the response prompt always appeared 3 seconds after the sentence offset). As for previous studies, this choice was made to exclude from analyses changes in the pupil diameter caused by movement planning and execution (Richer & Beatty, 1985). Three pupil outcome measures were obtained from the average trace of each participant and condition:

I. ***Pupil baseline:*** the average pupil diameter in the 500 milliseconds preceding the sentence's onset.

II. ***Mean pupil dilation*** relative to baseline pupil diameter between 0 and 5.9 s after the stimuli onset.

III. ***Peak pupil dilation***, as the maximum positive deviation from the baseline during the 5.9 s following stimuli presentation.

4.2.5 Statistical analyses

An independent sample t-test and mixed design ANOVAs were conducted to investigate whether the availability of long-term semantic context and the linguistic background of participants affected the estimated SNR level, the intelligibility score across blocks and the subjective ratings of listening effort.

Mixed-effect regression models were performed to analyse the effect of long-term semantic context and language background on pupil measures. Additionally, mixed-effect regression models were also used to investigate the effect of length of residence, overall English use and IELTS score on pupil measures for non-native listeners only. Exploratory analyses using Growth Curve Analyses were performed with the aim of investigating the impact of second language proficiency level on the effort release after speech offset in non-native listeners.

Correlation analyses were performed to check the consistency of subjective ratings within participants, and to investigate whether SNR levels and subjective ratings of listening effort were correlated. Lastly, a regression analysis was used to investigate whether the subjective ratings of listening effort for the related and unrelated conditions were able to predict the peak and mean pupil dilation for the corresponding experimental condition.

4.3 Results

4.3.1 Behavioural results

An independent-sample t-test was conducted in order to compare the SNR level required to target 80% of intelligibility for native and non-native listeners for unrelated sentences. Native participants were able to tolerate a significantly less favourable SNR (M = -6.0, SD = 2.1) compared to non-native participants (M = 1.6, SD = 5.4), $t_{(26.5)} = -5.91$, $p < 0.001$.

Intelligibility scores in noise for each of the three sections (beginning, middle and end), and benefit across sections (end – beginning) for related and unrelated conditions are summarised in Table 4.3. The adaptive block used to determine the SNR level targeting 80% of intelligibility was not included in the analyses.

Table 4.3 Descriptive statistics of the behavioural results in noise for all participants, and for native and non-native listeners.

Behavioural results in noise – Performance (% correct)												
	All participants				**Native**				**Non-native**			
	Related		**Unrelated**		**Related**		**Unrelated**		**Related**		**Unrelated**	
Part	**M**	***SD***	**M**	***SD***	**M**	***SD***	**M**	***SD***	**M**	***SD***	**M**	***SD***
1_Beg	83.5	*9.2*	76.4	*7.6*	85.0	*9.1*	76.6	*8.1*	82.3	*9.4*	76.2	*7.3*
2_Mid	85.8	*6.5*	78.9	*7.0*	86.4	*5.4*	80.1	*7.5*	85.3	*7.5*	77.8	*6.6*
3_End	87.1	*6.1*	81.9	*7.0*	87.7	*7.6*	83.4	*7.5*	86.6	*4.5*	80.6	*6.4*
Overall	85.5	*6.5*	79.0	*6.0*	86.3	*6.6*	80.0	*6.7*	84.7	*6.6*	78.2	*5.3*
Benefit (3-1)	3.6	*6.1*	5.5	*6.8*	2.8	*5.3*	6.8	*7.0*	4.3	*6.8*	4.3	*6.6*

In order to analyse intelligibility score, each block of sentences was divided in 3 sections consisting of 4 sentences each. A mixed design ANOVA performed on intelligibility levels with long-term context (related and unrelated) and part (beginning, middle and end) as within-subjects factors, and language (native and non-native) as between-subjects factor was performed in order to verify whether the performance levels differed across conditions and participants' groups. Results showed that overall participants achieved a better performance when attending to related compared to unrelated sentences [$F_{(1, 37)} = 68.26$, $p < 0.001$]. A main effect of part was also found to be significant [$F_{(2, 74)} = 18.02$, $p < 0.001$]. Follow-up Bonferroni-adjusted pairwise comparisons indicated that each part was significantly different from all the others, with a better performance toward the end of the block (beginning < middle < end). The analyses did not reveal any significant effect of linguistic background on the performance. No interaction between the variables considered was found to be significant. Behavioural results are displayed in Figure 4.2.

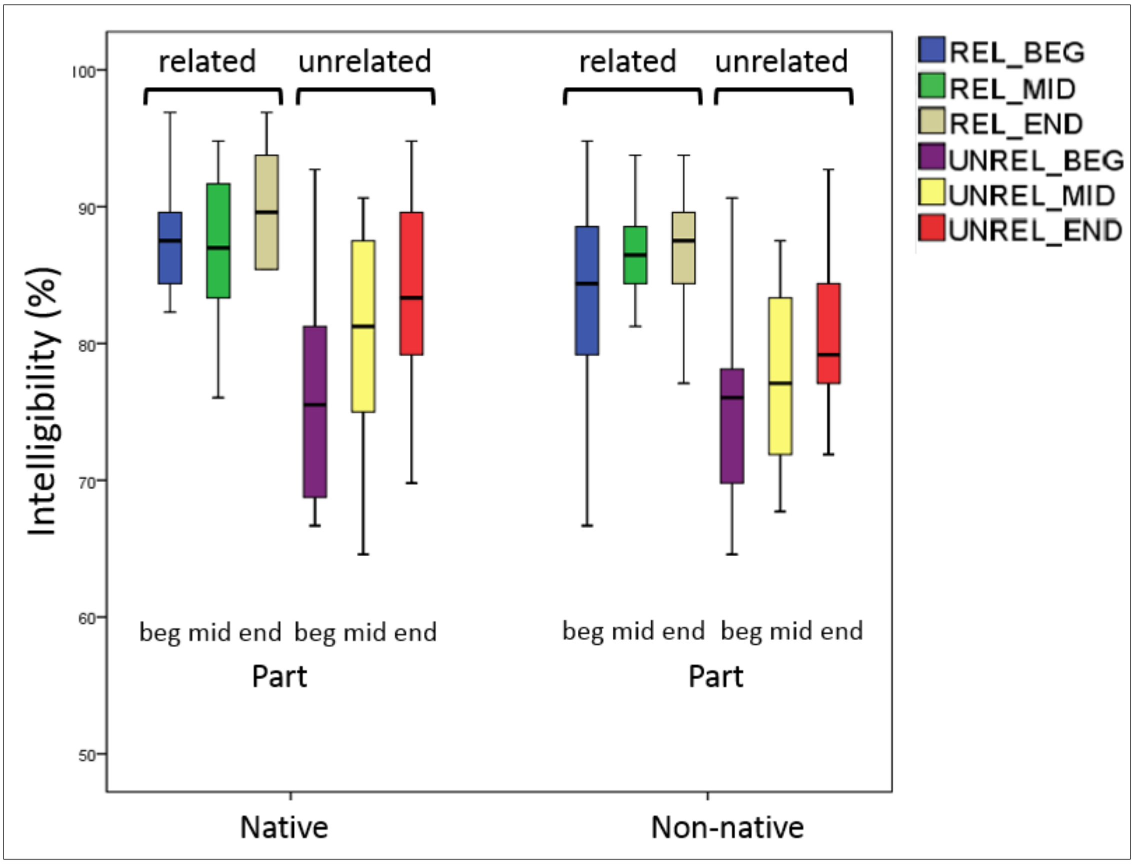

Figure 4.2 Intelligibility results for each experimental condition for native and non-native listeners.

As anticipated, behavioural results were not perfectly equated across conditions (semantic context and part), because the SNR level was set using unrelated sentences. However, the intelligibility level for related compared to unrelated sentences was already higher in the beginning section of blocks, and the difference between the two conditions did not increase across sections. This was surprising, since any potential benefit deriving from long-term context was expected to build up across sentences, rather than being present from the beginning part, and remaining constant across sections. A factor that may have contributed to this result is the repetition of recurrent keywords across sentences in related blocks (as an example, in the related block reported in Table 4.2, the keyword “carrot” is repeated 3 times in the first 4 sentences). Another potential explanation is that the effect of a consistent semantic context across sentences may have come into play within the first 4 sentences of related blocks. In the attempt to verify whether keywords repetition significantly contributed to a higher intelligibility in related compared to unrelated sentences, intelligibility levels from a sample of 4 participants were re-evaluated excluding keywords that were repeated within a block. Even after accounting for keywords repetition, a difference was still present from the first section between intelligibility levels for related and unrelated sentences for the subsample considered (M = 80.2% and M = 73.7% for related and unrelated sentences respectively when accounting for keywords repetition, M = 81.8% and M = 73.7% for related and unrelated sentences when not accounting for keywords repetition). Therefore, the unexpected pattern of results is likely to be due to the combined effect of keywords repetition and of an early onset of long-term context semantic benefit. Additionally, it has to be noted that the difference between conditions cannot be due to differences in the difficulty level of the sentences presented, since the same stimuli arranged in different manner were used for both conditions across participants, even though different sentences were used for the related and unrelated condition for each participant. Despite this, the implementation of the adaptive procedure allowed to target matching levels of intelligibility for native and non-native listeners.

4.3.2 Pupil data

As previously justified, analyses of the pupil data focused on the last section of each list of sentences presented, consisting of the last 4 sentences. Descriptive statistics for the pupil data collected during the blocks in noise with fixed SNR for the final section of the related and unrelated condition are reported in Table 4.4. Measures are displayed for native, non-native and all participants, and include mean and peak pupil dilation, and baseline pupil diameter.

Table 4.4 Descriptive statistics of the pupil measures for the clear speaking style condition.

Pupil data in noise – End section						
Related						
	All		**Native**		**Non-Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.04	*0.14*	-0.03	*0.11*	0.10	*0.13*
Peak dilation, mm	0.18	*0.17*	0.10	*0.14*	0.25	*0.17*
Baseline, mm	6.16	*1.00*	6.02	*1.19*	6.28	*0.83*
Unrelated						
	All		**Native**		**Non-Native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Mean dilation, mm	0.06	*0.15*	0.01	*0.13*	0.09	*0.17*
Peak dilation, mm	0.21	*0.17*	0.18	*0.18*	0.24	*0.16*
Baseline, mm	6.17	*1.01*	6.02	*1.19*	6.29	*0.83*

Mixed-effects modelling for all participants

The effect of long-term semantic context and language background on pupil outcome measures was modelled using a series of mixed-effect regression models. Analyses were performed using the lme4 package in the R environment (Bates et al., 2014; R Core Team, 2017). Three separate models were built for the mean, peak and baseline pupil dilation. Similarly to Study 2, the initial saturated model included interaction terms for all independent variables as fixed effects with random intercepts and slopes (Barr et al., 2013). Due to non-convergence, I simplified the models hierarchically from most complex to least complex. The resulting converged maximal models for all three variables included the fixed effects of language background (2: native and non-native) and long-term semantic context (2: related and unrelated). Participant was included as random effect but no random slopes. The maximal model included up to two-way interactions between language background and long-term semantic context. Model residuals via chi-square tests (α = .05) were compared from the most complex models (containing the largest interaction term) to the least complex models (containing only single terms). If an interaction term was significant, all lower level effects involved in the interaction were included in the final model. Results for each of the dependent variables considered are reported below. The pupil curves dilation during the end section of blocks (sentences 9-12) for native and non-native listeners in the related and unrelated conditions are displayed in Figure 4.3.

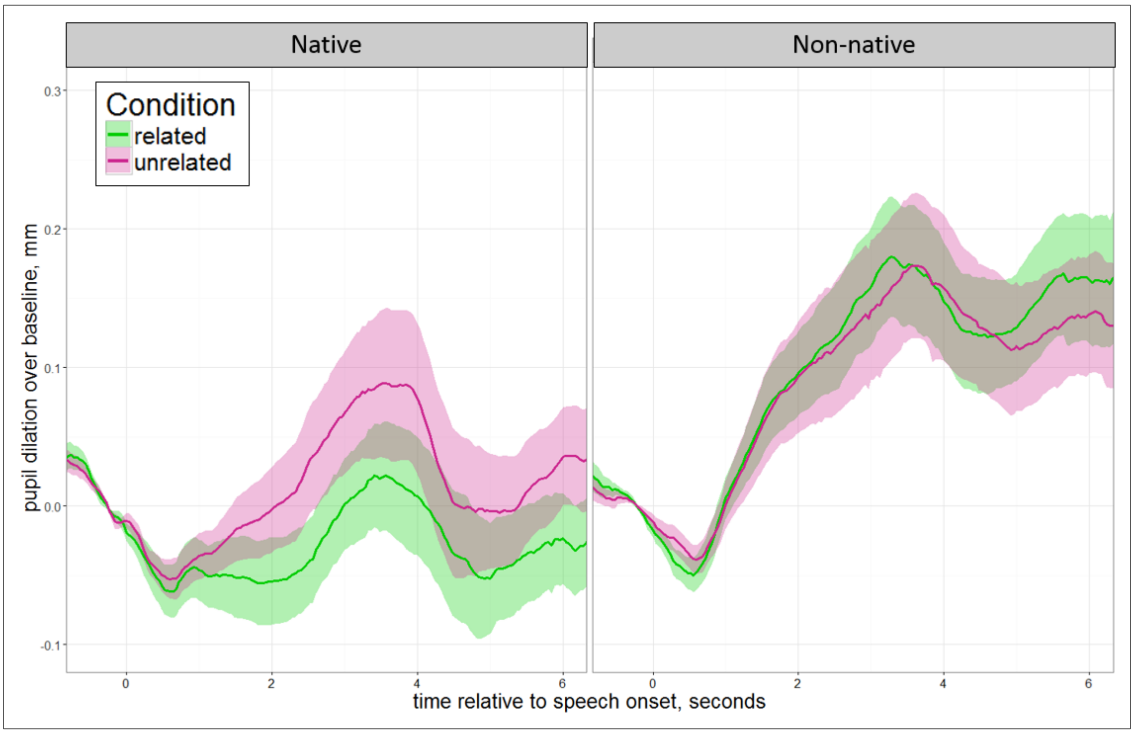

Figure 4.3 Mean pupil response over time during speech perception in the related and unrelated, for native and non-native listeners during the end section of blocks.

Mean pupil dilation

The final model only included fixed effects of language background, as well as a random effect of listener (see Table 4.5 for coefficients and estimated *p* values). This indicates that the overall mean pupil dilation was greater for non-native compared to native listeners.

Table 4.5 Fixed and random effects in a mixed-effects model of mean pupil dilation for all listeners.

Mean ~ Language + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		-0.01	0.03	- 0.28	0.78
Language	Non-native	0.10	0.04	2.56	0.010
Random effects		**Variance**			
Participant	(intercept)	0.01			
Residual		0.00			
Note. **Number of observations = 78; participants (N) = 39.**					

Peak pupil dilation

The final model for peak values included the interaction between the fixed effects of language background and context and a random effect of listener. Because of the significant interaction between language and context, the fixed factor of language background and context were also included in the final model (see Table 4.6 for coefficients and estimated *p* values). Results indicated that attending to related sentences elicited a greater peak pupil response compared to unrelated sentences only for native listeners. Non-native listeners did not show a significant difference in the peak pupil response for related and unrelated sentences.

Table 4.6 Fixed and random effects in a mixed-effects model of peak pupil dilation for all listeners.

Peak ~ Language * context + (1|Participant)					
Fixed effects		**Estimate**	**SE**	***t***	**Estimated *p***
(intercept)		0.10	0.04	2.64	0.008
Language	Non-native	0.15	0.05	2.88	0.004
Context	Unrelated	0.08	0.02	3.43	0.001
Language x Context	Non-native, Unrelated	-0.09	0.03	-2.95	0.003
Random effects		**Variance**			
Participant	(intercept)	0.02			
Residual		0.00			
***Note.* Number of observations = 78; participants (N) = 39.**					

Baseline

None of the terms entered in the model were found to be significant.

Mixed-effects modelling for non-native participants

To investigate the effect of length of residence, overall English use and IELTS score on the pupil measure, the same type of analysis was run with non-native listeners only for all three dependent variables (mean, peak and baseline). The resulting converged models for all three variables included the following fixed effects: long-term semantic context (2: related and unrelated), length of residence, self-reported English use and IELTS score. Participant was included as random effect but no random slopes. The maximal model only included up to two-way interactions between conditions (semantic context and speaking style), and the background measures considered (length of residence, self-reported English use and IELTS score). Results indicated that none of the terms entered in the model had a significant effect on any of the pupil measures for non-native listeners.

Proficiency effect and Growth Curve Analysis

The impact of proficiency level in L2 on SNR levels and pupil measures was investigated for non-native listeners. Similarly to Study 2, individual differences in SNR levels were investigated in order to test whether the background information about second language proficiency and usage were correlated with listeners' ability to tolerate background noise. A stepwise regression analysis with SNR as outcome variable was performed. Length of residence, overall self-reported English use and results at the Listening section of the IELTS were entered as predictors. Results confirmed that only the score obtained for the IELTS Listening test significantly contributed to predict the estimated SNR level (see Table 4.7). A higher score obtained at the IELTS test predicted a lower SNR (i.e., better performance).

Table 4.7 Effect of individual differences on the estimated SNR level for non-native participants, stepwise regression results.

Dependent variable	Predictor	R^2	B	Std. error	Std. beta	F	t	Sig.
SNR non-native	IELTS Listening	.615	- .681	.127	- .784	28.805	- 5.367	< .001

As in Study 2, non-native listeners were further divided in two groups based on their score on the IELTS Listening test, with the aim of investigating at a group level the impact of L2 proficiency on the pupillary response. To obtain two balanced groups, the median IELTS Listening score (median score = 31) was used as the upper limit of the low proficiency group. The high proficiency group included 10 participants, while 11 individuals were included in the low proficiency group. Table 4.8 summarises relevant background data and performance results for high and low proficiency groups.

Table 4.8 Background data and performance results for high and low proficiency groups.

	High proficiency		**Low proficiency**	
	M	***SD***	**M**	***SD***
Intelligibility Related sentences	86.2%	*5.9*	83.4%	*7.0*
Intelligibility Unrelated sentences	79.8%	*5.31*	76.6%	*5.3*
SNR	-2.6	*1.9*	5.5	*4.6*
Length of stay	2.6	*1.2*	3.3	*2.1*
English use (%)	41.7%	*0.2*	51.0%	*0.2*
IELTS Listening score	35.0	*1.9*	24.7	*4.9*

Growth curve analysis (Mirman, 2014) was used to analyse the effort release after speech offset in non-native listeners for the related and unrelated experimental condition. Similarly to the approach adopted in Study 2, a time window corresponding to the effort release phase was identified starting at 3.4 and ending at 5.9 seconds after stimuli onset. The rationale for choosing these time windows was to include the section of the pupil curve starting just after the peak, and ending just before the response prompt, in order to explore the effort release phase. Figure 4.4 displays the pupil curves dilation for high and low proficiency non-native listeners for the related and unrelated conditions during the final section of blocks, and the relative time windows considered for the GCA. The pupil data of interest were modelled using a second-order orthogonal polynomials and fixed effects of proficiency level on all time components (intercept, linear and quadratic). The intercept captured the overall level of pupil dilation, the linear component (slope) was used to estimate change rate in pupil dilation, and the quadratic term was intended to capture the inflection of the growth function. The high proficiency group was treated as the baseline and parameters were estimated for the low proficiency group. Random effects of listeners on all time terms were also included. The fixed effect of second language proficiency (high vs low proficiency level) were added individually and their effects on model fit were evaluated using model comparisons. Improvements in model fit were evaluated using -2 times the change in log-likelihood, which is distributed as χ^2 with degrees of freedom equal to the number of parameters added. As for Study 2, analyses were performed using the lme4 package in the R software environment (R Core Team, 2017). Results revealed that the effect of proficiency level on the intercept, on the linear and on the quadratic terms did not improve the model fit neither for the related nor for the unrelated condition. Therefore, the proficiency level in L2 did not appear to modulate the overall listening effort after speech offset, or the rate of effort release for non-native listeners.

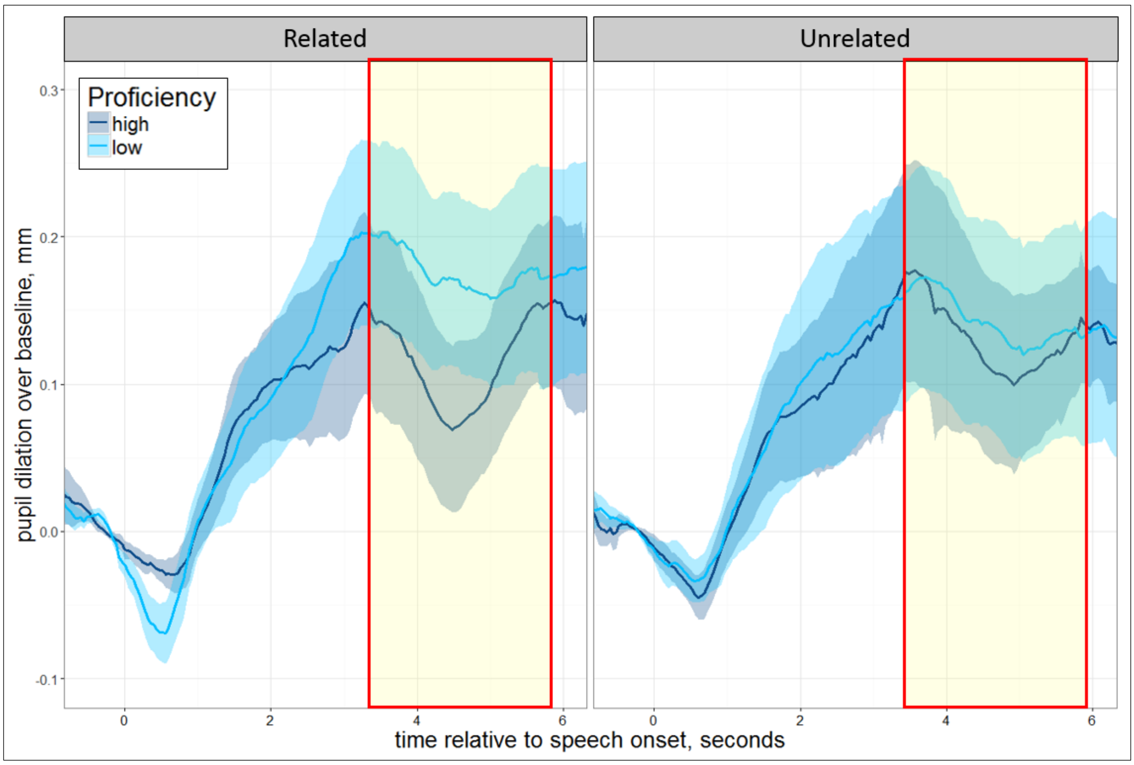

Figure 4.4 Mean pupil response over time during speech perception in the related and unrelated condition during the end section (sentences 9-12), for high and low proficiency non-native listeners.

4.3.3 Subjective ratings of listening effort

Table 4.9 summarises subjective ratings of listening effort expressed by native and non-native listeners for the related and unrelated test condition. A mixed design ANOVA performed on subjective ratings with long-term context (related and unrelated) as within-subjects factor, and language (native and non-native) as between-subjects factor was performed in order to investigate whether the subjective ratings of listening effort varied across conditions and participants. Only a marginally significant main effect of language background was found [$F_{(1, 38)} = 3.27$, $p = 0.05$]. Overall, native participants rated the listening task as more effortful compared to non-native listeners. The effect of semantic context on subjective ratings of effort did not reach significance.

Table 4.9 Subjective ratings of listening effort from all participants, and from native and non-native listeners (0 = no effort, 7 = extreme effort).

Subjective ratings of listening effort						
Condition	**All**		**Native**		**Non-native**	
	M	***SD***	**M**	***SD***	**M**	***SD***
Related	3.8	*1.3*	4.1	*1.5*	3.7	*1.1*
Unrelated	4.5	*0.9*	4.8	*0.8*	4.5	*0.7*

A higher subjective listening effort in native compared to non-native listeners was unexpected, and it might be due to a less favourable SNR level for native listeners. To test this hypothesis, I first verified whether subjective ratings were consistent within participants (i.e. that participants who tended to give higher ratings for the unrelated condition also tended to give higher ratings for the related condition). A Pearson correlation coefficient was computed to assess the relationship between subjective ratings for the related and unrelated condition. Results indicated a significant positive correlation between the two variables [r = 0.496, n = 39, p = 0.001]. The mean value for related and unrelated subjective ratings was then computed for each participant. A correlation analysis did not reveal any correlation between SNR levels and subjective ratings of listening effort [r = - 0.160, n = 39, p = 0.33].

Lastly, an exploratory regression analysis was conducted to investigate whether the subjective ratings of listening effort for the related and unrelated conditions were able to predict the peak and mean pupil dilation for the corresponding experimental condition. Results indicated that the subjective measures of listening effort did not significantly contribute to estimating the pupil outcomes considered.

4.4 Discussion

Study 3 aimed at exploring how the presence of a consistent semantic context within a set of sentences contributed to modulate comprehension and listening effort in native and non-native listeners.

The analysis of intelligibility scores for the two groups of listeners showed that both native and non-native listeners were able to overall improve their comprehension when presented with related compared to unrelated sentences. However, as discussed in detail in section 4.3.1, the intelligibility level for related sentences was already higher than for unrelated sentences from the beginning section of blocks. A beneficial effect of long-term semantic context was instead expected to lead to a greater increase in intelligibility across sections (end compared to beginning) in the related compared to the unrelated condition. Two factors have been identified as potential explanation for this. First, the multiple repetition of keywords within related blocks may have contributed to enhance intelligibility scores already from the first section and onwards. Simultaneously, a beneficial effect of semantic context across utterances may have come into play already within the first section of related blocks. This would suggest that both native and non-native listeners may be capable to quickly pick up a conversation topic, and exploit it to create sensible predictions that can aid their speech understanding. If that is the case, this would expand our understanding of context benefit mechanisms, by suggesting that not only a sentence-level context is beneficial to listeners (Drager & Reichle, 2001; Wingfield & Tun, 2007; Zekveld, Rudner, et al., 2011), but similarly even the retention of a consistent semantic context across sentences may enable listeners to enhance their comprehension. Due to the widespread use of single-sentence type of stimuli in most experimental tasks, there are few investigations of how semantic benefit builds up within a longer communication unit, which is likely to be made of a series of connected utterances. Although the current experiment was designed to advance the understanding of how listeners are able to benefit from semantic context in real life communicative situations, the division of experimental blocks into three parts of 4 sentences each, and the sentence material used, might not have been ideal in capturing the benefit deriving from a consistent semantic context across multiple sentences. Overall, I argue that the combined effect of keywords repetition within related blocks, and an early onset of a long-term context semantic benefit might be responsible for an increased intelligibility level for related compared to unrelated sentences already present in the first section of blocks.

Additionally, it should be noted that native and non-native listeners improved their speech comprehension across sections to a comparable extent (approximately 3-5%) for the related and unrelated condition. This suggests that the improvement is likely to be due to habituation to the task and to the background noise, rather than to the experimental manipulation of semantic context. A comparable advantage was found for native and non-native listeners, although participants were able to tolerate a less background noise when attending to their second compared to first language.

Pupillometric measures in this study suggested that the long-term benefit in speech perception performance corresponded to a reduction in listening effort for native listeners. Reduced cognitive effort due to semantic benefit was previously found when considering context at a sentence-level (Winn, 2016). On the contrary, non-native listeners did not show reduced cognitive effort when attending to related compared to unrelated sentences, despite their performance results mirroring the intelligibility levels obtained by native speakers. However, caution should be adopted when evaluating this result. Indeed a significant effect of long-term context for native listeners was only found on the peak pupil dilation, and not reflected in the mean pattern of dilation. The mean pupil dilation is believed to be a more reliable and stable index of cognitive resource allocation (Ahern & Beatty, 1979; Verney et al., 2001), therefore it is desirable for future studies to replicate the present findings to ensure their reliability.

It should also be noted that the reduction in listening effort found for native listeners when attending to related compared to unrelated sentences might be partially explained by the increased intelligibility level achieved for the related compared to the unrelated condition. This is because we know that the pupil response is smaller for higher levels of intelligibility, when extremely low levels of intelligibility are not considered (Zekveld & Kramer, 2014). However, this is unlikely to explain the entire impact of long-term context on listening effort for native listeners. Indeed, although a similar difference in intelligibility levels was found for non-native listeners, they did not show the same reduction in listening effort as found for native listeners.

Taken together, the results indicate that the ability to create context-based predictions is not exclusive to native listeners. However, it is reasonable to speculate that the process of exploiting semantic cues does not entail a benefit in terms of a reduced cognitive cost for non-native listeners. Pupillometric results from this study are aligned with evidence from EEG studies showing a delayed N400 effect to semantic incongruity for L2 listeners compared to listeners attending to their first language (for a review see Moreno, Rodríguez-Fornells, & Laine, 2008). A possible functional interpretation of this delay suggests that the semantic integration of word meaning is more difficult for listeners attending to a non-native language (Hahne, 2001). The current findings suggest that behavioural results such as SNR levels and intelligibility scores are not always able to capture an exhaustive picture of the mechanisms involved during speech comprehension. Specifically, an improved performance can correspond either to a reduced or increased listening effort, which can potentially lead to a different ability to cope with complex multi-tasking scenarios in real life.

Additional thoughts on the ecological validity of the present study concern the type of predictions that listeners were likely to formulate based on the long-term context provided. Each of the sentences presented carried some sort of information about a given topic, but without including the construction of a factual narrative around it. When attending to related sentences, participants were able to predict the subject of the following sentence, and they were then more likely to expect to hear words relevant to the given semantic field rather than any other word. Therefore, semantic expectations were mainly built around lexicon, and given the nature of the stimuli presented it is reasonable to assume that participants' predictions did not include any expectation on the development of the narrative. On the contrary, during real life communication exchanges, the logic in the succession of narrated events is also likely to play an important role in anticipating the content of what we are about to hear. Importantly, while the ability to formulate lexical predictions is inherently linked to linguistic knowledge, it is not necessarily the same for expectations about the narrative content of speech. Therefore a more comprehensive understanding of how linguistic/lexical expectations and the communicative context (e.g. where and when the communication is happening, who is involved, etc.) can modulate listening effort and performance would benefit from the investigation of how predictions are formulated also on the basis of the wider communicative context. This could include the use of stories with a familiar pattern in the course of events (e.g. well-known stories or stories previously presented by means of images or animations), or the consideration of the pragmatic context. This approach has been previously taken in studies investigating speech comprehension in hearing impaired listeners, and labelled as “connected discourse tracking”, which typically used passages from simple novels aimed at low-proficiency readers. Its implementation is thought to be more representative of everyday speech communication compared to measures based on single sentences (De Filippo & Scott, 1978). The connected discourse tacking procedure was successfully used to investigate speech intelligibility in individuals with cochlear implants, and it was also adopted as a production training for hearing impaired children (Faulkner, Rosen, & Wilkinson, 2001; Levitt, Waltzman, Shapiro, & Cohen, 1986; Osberger, Johnson, & Miller, 1987). The same approach could be therefore fruitfully applied to expand our understanding of semantic context contribution to non-native speech comprehension.

Unlike in Study 1 and 2, for the present experiment stimuli presented were recorded from a native American speaker living in the state of Washington. In addition to the planned contrasts, the use of a different English accent is likely to have had an impact on speech perception during the experimental task for both native and non-native listeners. By looking at SNR levels for the high intelligibility condition in Study 1 (where 80% of intelligibility level was also targeted), results seems to suggest that both native and non-native listeners were able to tolerate a higher background noise in Study 1 compared to Study 3. This may suggest an increased difficulty in Study 3 compared to Study 1 potentially due to accent unfamiliarity. However, this difference may also be attributable to a variable complexity of the stimuli used. Indeed BEL sentences (presented in Study 1) have been specifically developed for testing non-native listeners, therefore might have been easier to understand for both groups of listeners. Previous research indicated that the cost of language processing under adverse listening conditions is influenced by variations in accent familiarity, with faster adaptation to familiar L1 accent compared to unfamiliar L1 or L2 accents (Adank, Evans, Stuart-Smith, & Scott, 2009). Pinet, Iverson, and Huckvale (2011) showed that listeners were more accurate at recognising accents that acoustically matched their own. Specifically, Southern British English (SE) listeners were more accurate in understanding SE speech compared to Irish English and to L2 accented English, proving that native listeners are better at recognising speech produced with their same L1 accent. As far as the study here presented, it is therefore possible that native listeners had to face an increased level of difficulty due to accent unfamiliarity compared to participants in Study 1 and 2. This may have led to an overall diminished language effect due to the increased processing cost acting against native listeners attending to an unfamiliar L1 accent. At the same time, Pinet et al. (2011) also found that non-native listeners became selectively more tuned to the accent of foreign language they were more exposed to (e.g. SE) as their experience and exposure to second language increased. Since participants recruited for the present experiment were living in London at the time of testing, it is likely that SE was the English accent they were more tuned to. If that is the case, non-native listeners too may have been challenged with additional difficulties arising from attending to an unfamiliar L2 accent. However this variable was not controlled for, and assumptions about its potential impact are difficult to confirm.

As previously mentioned, the selection of a shorter time window to be regarded as a baseline (500 ms) helped in minimising the impact of the experimenter using too fast a pace when initiating trials. Unfortunately however, this was not sufficient to exclude any effect of previous trial's dilation on pupil measures, as can be seen from figures 4.3 and 4.4. This should not change the overall pattern of results, but it may have interfered with the actual measurement of pupil dilation. This is because the pupil dilation was calculated over a baseline that was still partially dilated from the previous trial.

Subjective ratings of listening effort did not correlate with pupil results, and in fact revealed an inverse pattern with native listeners overall perceiving the speech comprehension task as more effortful compared to non-native listeners. This effect may be explained by the fact that native listeners were performing the task at a less favourable SNR level, therefore at a more difficult level if other variables (i.e. language background and intelligibility) are ignored. However, subjective ratings of effort did not correlate with SNR levels, and it is therefore difficult to interpret this effect. Surprisingly, there was no difference in subjective ratings between the related and unrelated condition, meaning that the small difference in intelligibility between the two conditions was not mirrored by a difference in the perceived effort. This may be due to the small magnitude of intelligibility difference between the related and unrelated condition (less than 10%). Previous research reporting a higher intelligibility level to be related to a lower subjective rating of effort indeed considered a greater range of intelligibility levels (Holube et al., 2016; Zekveld, Kramer, et al., 2011).

Lastly, an attempt was made to investigate whether L2 proficiency modulated listening effort for non-native listeners. Nevertheless, the analyses performed did not reveal any effect although there seems to be a difference when observing the pupillometric results plotted by proficiency. This could be due to the limited number of participants included in each proficiency group, and to the large individual variance typically seen in pupil measures. The easier difficulty level compared to Study 2 (80% vs 50% intelligibility level) may also have contributed to the lack of difference between the two proficiency groups. However, given the visual trend and the effect of proficiency found in Study 2 for the most difficult listening condition, it could be worth investigating further the presence of a proficiency effect by designing an ad hoc experiment.

To conclude, the results of the current study suggest that long-term semantic context only leads to a reduction in listening effort for native listeners, but not for non-native. This was found despite a similar pattern of intelligibility levels across related and unrelated sentences for both groups of listeners. However, caution should be used when considering this result, since the effect of long-term semantic context was only found to impact the peak pupil dilation, while no significant effect was found on the mean measures.

Chapter 5

5. General Discussion and Conclusions

The present dissertation set out to investigate differences in listening effort between native and non-native listeners during a speech perception task in noise, when intelligibility levels are equated. Despite an increasing attention towards the concept of listening effort, and the use of pupillometry as a means to investigate it, very little research to date has applied pupillometry to the investigation of differences in listening effort between native and non-native listeners during speech perception. Given the constantly growing number of people living and working in a country where a language foreign to them is spoken, it is increasingly relevant to understand the mechanisms leading to increased cognitive effort for non-native listeners.

The following main research questions were formulated:

1. Do native and non-native listeners performing at the same accuracy level differ in terms of cognitive effort required?
2. How does the presence of acoustic enhancements modulate listening effort in native and non-native listeners?
3. How does the availability of semantic cues within a sentence affect listening effort in native and non-native listeners?
4. Is the presence of semantic context spanning several sentences beneficial for native and non-native listeners in terms of listening effort reduction?
5. How does proficiency level affect listening effort in non-native listeners?

With the aim of answering the above questions, three studies were performed. From Study 1, presented in Chapter 2, I concluded that overall non-native listeners experienced a greater listening effort compared to native listeners, even when achieving the same intelligibility level. The second study, described in detail in Chapter 3, showed that attending to speech pronounced using a clear speaking style (therefore benefiting from acoustic enhancements) is an effective strategy to improve comprehension and to reduce listening effort for native and non-native listeners. Moreover, results showed that the presence of sentence-level semantic cues helped listeners in enhancing intelligibility, but did not lead to a reduced processing cost compared to meaningless sentences. Lastly, Study 3, presented in Chapter 4, revealed that a consistent semantic context across sentences led to better comprehension overall for both native and non-native listeners, but it was only for native listeners that the improvement also corresponded to a reduced listening effort as reflected in the peak pupil dilation. Effects of proficiency on the modulation of listening effort for non-native listeners remain not fully understood, even if a trend emerged suggesting that effort reduction after speech offset takes longer for less proficient non-native listeners. However, a consistent link was found in Study 2 and 3 between proficiency as measured via the Listening section of IELTS and the SNR thresholds in noise.

In this final chapter, findings from the three studies presented will be discussed in light of the ELU model (Rönnberg et al., 2019). Additionally, I will suggest how the model discussed might be expanded to better account for challenges that are specific to non-native speech understanding. Afterwards, I will address methodological concerns related to the potential confounding effect of motivation when adopting a between-subject design, with a focus on the comparisons between native and non-native listener groups. I will then suggest possible solutions to overcome these limitations. Lastly, I will share some final remarks on what I believe is the best perspective to interpret and value the evaluation of listening effort as measured by the pupil response.

5.1 Evaluation of the Ease of Language Understanding model

Various models aiming at providing a comprehensive account of the role of cognitive load among the mechanisms contributing to language understanding have been proposed and discussed in the literature (Pichora-Fuller et al., 2016; Rönnberg et al., 2019; Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008). One of the challenges for the construction of such models is being able to account for all factors that can potentially affect the ease of language understanding. This allows the generation empirical predictions based on the model components and their interaction. Typically, the most accepted and debated models have been formulated with a focus on explaining everyday challenges in language understanding stemming from impairments at all level of the auditory system. Models such as these are needed to explain difficulties experienced by hearing impaired listeners beyond their performance in pure-tone and speech audiometry. An increased awareness of the importance of auditory-cognitive interactions for the understanding of speech comprehension mechanisms has recently emerged. This is justified by a willingness to help individuals with hearing impairments, who frequently report the experience of listening being effortful, tiring or stressful despite having a good speech understanding. Therefore, the focus of the models proposed has primarily been on the cognitive mechanisms that come into play when listeners are attending to an impoverished signal (due to perceptual deficits or hearing aids), to speech masked by noise, or a combination of the two (Pichora-Fuller et al., 2016). Very limited attention is given to factors such as language proficiency or accent familiarity, and to the specific challenges encountered by listeners attending to a non-native language or to an unfamiliar accent. However, some challenges are certainly shared between hearing-impaired listeners, and those who are attending to a non-native language. It follows that some models not specifically formulated to account for challenges faced by non-native speech perception can still be useful for interpreting empirical data for the non-native population, and to make predictions about speech understanding and related listening effort. Here, I will discuss how the research findings presented in this dissertation can be interpreted in light of the most recent version of the ELU model (Rönnberg et al., 2019), and how such a model might be expanded to directly account for challenges arising during non-native speech perception.

To begin with, the ELU model (see Figure 1.2 in the opening Chapter) predicts that the degradation of the language signal increases the probability of a mismatch between the input stream of language and the stored mental representations of phonological and lexical units. As a result, the success of language understanding will become more dependent on WM and other executive functions such as inhibition and information updating, which will come into play to solve the mismatch. According to the ELU model, this explicit and deliberate involvement of WM is key to a successful language comprehension in challenging conditions, but it is also considered to be linked to an increase in cognitive effort. In Study 1, within each listener group (native and non-native), the low intelligibility condition corresponded to a less favourable SNR (therefore to a more degraded input signal) compared to the high intelligibility condition. Pupillometric results indeed confirmed that lower intelligibility levels led to greater cognitive load, as indicated by a greater pupil dilation. It is therefore reasonable to speculate that when overall intelligibility is considered (without accounting for the effects of semantic context), empirical predictions based on the ELU are verified for both native and non-native listeners. Further to that, speech clarity was directly manipulated in Study 2 and sentences pronounced either with a plain or clear speaking style were presented. By its very nature, clear speech is meant to facilitate listeners in the process of phoneme recognition, and therefore to reduce the occurrence of mismatches between input signal and stored phoneme representations. By applying the same rationale as above, clear speech is supposed to reduce the explicit reliance of speech understanding on WM and cognitive functions, and therefore a reduced cognitive load would be predicted by the ELU model when listeners are attending to clear compared to plain speech. Results from Study 2 showing a reduced listening effort when native and non-native listeners are attending to clear speech again confirms the predictions of the ELU model for both listener groups. Importantly, the confirmation of the accuracy of these predictions beyond the consideration of native normal hearing and hearing impaired listeners, expands the context within which the ELU model is able to make verified predictions of cognitive load. I believe this constitutes further evidence demonstrating the theoretical reliability of the model.

Nevertheless, there are factors that are specific to second language speech perception, and for which predictions based on the ELU model may not be as accurate. One of these concerns listeners' ability to take advantage of semantic cues. Rönnberg et al. (2019) reason that high lexical predictability and the availability of a coherent semantic context leads to a reduced reliance on cognitive "repair" functions such as WM, because the probability of mismatches is reduced by constraining the number of candidate words/phonemes. The hypothesis based on the ELU model would therefore predict a reduced listening effort when participants are attending to semantically predictable sentences. However, results from Study 2 and 3 did not systematically indicate a reduction in listening effort in the presence of an enhanced semantic context. Indeed, Study 2 did not show any difference in listening effort elicited by semantically predictable and anomalous sentences, while Study 3 suggested that the availability of a consistent semantic context across sentences only reduced listening effort for native listeners, but not for participants attending to a second language (although it is desirable that these findings are further verified by additional studies manipulating the availability of semantic context using different approaches). Difficulties specific to the perception of a non-native language might come into play to explain this misalignment between ELU model's prediction and actual results. Rönnberg et al. (2019) hypothesised a dual role played by WM in speech understanding within the ELU framework. First, a post-dictive role refers to the cognitive mechanisms that a listener may need to reconstruct what was said when a mismatch happens. This function is thought to be slow and explicit, and primarily responsible for an increase in cognitive effort. Second, a predictive role of WM pertains to the ability to use phonological and semantic information for predictive purposes, that is to pre-tune and focus attention on phonemes or words that the talker is more likely to use based on what the listeners has already heard. According to the authors of the ELU model, the predictive function is mainly implicit and automatic, and it is not considered to be the main cause of cognitive load. Postdiction and prediction are thought to be dynamically related during on-line language processing. While evidence from native listeners seems to support this general view, data from non-native populations has shown that the process of making predictions and exploiting the phonological and semantic context to prime and pre-tune speech perception does not happen with the same speed and ease when attending to a second language (Bradlow & Alexander, 2007; Cutler et al., 2004; Hahne, 2001; Mattys et al., 2010). The causes hypothesised for this non-native disadvantage include a reduced knowledge of the L2 at a phonological, lexical, syntactic and pragmatic level, potentially leading to the erroneous activation of word candidates (Cutler et al., 2006), and interference from L1 vocabulary, which is likely to make the word selection process more effortful (Broersma & Cutler, 2011; Spivey & Marian, 1999). Taken together, results presented in this thesis suggested that native listeners are able to rely on semantic context gaining a benefit both in terms of understanding (improved comprehension for semantically predictable compared to anomalous sentences in Study 2) and listening effort (reduced pupil response for related compared to unrelated sentences in Study 3), therefore following ELU model's predictions. However, while non-native listeners showed a similar ability to fruitfully exploit semantic context for improving comprehension, they did not benefit from any reduction in listening effort. In its current structure the ELU model is therefore ineffective in accounting for the additional challenges faced by non-native listeners. I believe that a useful addition to expand the applicability of the ELU model would be the inclusion of a "Language-specific knowledge" component feeding as a support into the pre-diction function. This would imply that native-like language knowledge allows for fast and implicit prediction making, but would also account for any increase in the processing load due to a reduced proficiency in L2. Moreover, the consideration of potential L1 interference would further improve the effectiveness of the model in generating realistic predictions for a wider range of listeners. The proposed addition to the ELU model's structure is graphically displayed in Figure 5.1, in green representing a facilitator component to the process, and in yellow to indicate a detrimental effect.

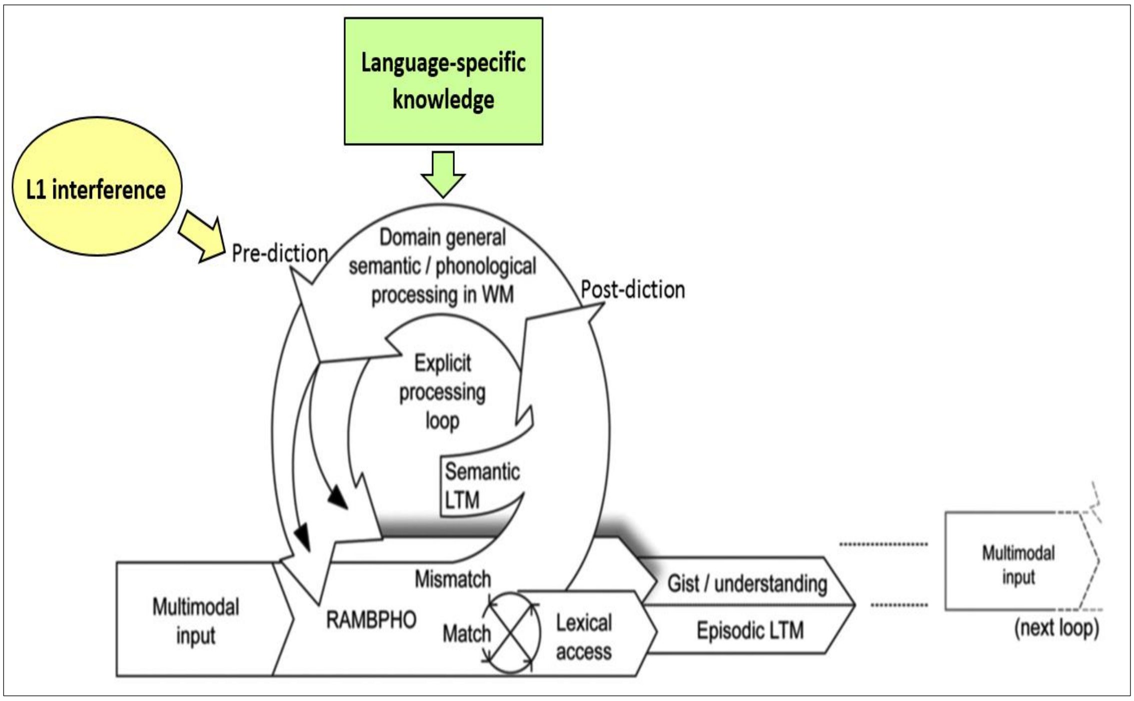

Figure 5.1 Visual representation of the proposed expansion of the ELU model, based on Rönnberg et al. (2019).

Specifically, the detrimental effect of L1 interference is intended to come into play both during phoneme recognition and word selection. The relative contribution of these two linguistic components (phonetic and lexical) could be tested by varying the expected interference of the speech material presented to non-native listeners. As an example, future studies might manipulate the degree of similarity between L1 and L2 phonemes, and investigate whether this would affect the cognitive effort elicited during their perception, since the level of difficulty in non-native phonemes discrimination has been associated to the degree of their similarity to L1 phonemes (Best et al., 2001), as extensively discussed in section 1.2.3. Similarly, the contribution of L1 lexical interference during non-native speech perception might be investigated by combining pupillometry data and information about the simultaneous consideration of lexical alternatives during spoken word perception, i.e. the visual world paradigm (Huettig et al., 2011).

On the other hand, a good starting point to test the proposed facilitating effect of language-specific knowledge might be to evaluate the effect of L2 vocabulary size in non-native listeners on listening effort. This approach is justified by previous research from Bundgaard-Nielsen, Best, and Tyler (2011) reporting a positive association between L2 vocabulary size and L2 vowel perception abilities for Japanese adult learner of English. This line of research could help in clarifying which processing steps are primarily involved in an increased effort when communicating in a second language.

Referring again to the ELU framework, research in non-native speech perception mainly focused on the ability of listeners to exploit the semantic context, however without incorporating in the experimental paradigm the distinction between predictive and postdictive abilities. It might be valuable in future work to consider the differential contribution of those two abilities. Specifically, the capacity to retrieve a lexical item that was either severely degraded or absent in the speech input could be evaluated by manipulating the amount of contextual, syntactic or morphological information available to the listener.

5.2 General remarks on the use of pupillometry in second language perception research

Pupillometry is becoming increasingly popular within the field of audiology as a valid tool to complement the assessment of listening effort, alongside with the measurement of performance and neural correlates. The investigation of the systematic changes in the pupil size has been fruitfully exploited to study the effects of speech signal manipulations, hearing impairment, age, attention and motivation on the resource allocation strategies during speech perception (for a review see Zekveld et al., 2018). As discussed in greater detail in the introductory chapter, this is justified since task-evoked pupillary responses have been proven to accomplish the three essential criteria for a reliable indicator of cognitive effort (Beatty, 1982; Kahneman, 1973). First, the pupil response is sensitive to within-task variations in the task demand (as for example different levels of intelligibility when performing the same type of task, as in Study 1). Second, it is sensitive to between task differences, such as auditory detection of sounds vs meaningful words identification, as in Kramer et al. (2012). Third, it is sensitive to between individual differences in cognitive effort due to group differences in the tested population, as in studies looking at pupil response in normal hearing vs hearing impaired listeners, such as in Wang et al. (2018). Therefore, the task-evoked pupillary response can be reasonably regarded as a reliable and sensitive measure of listening effort when experiments are thoughtfully designed.

Although only limited literature is currently available exploring pupil response during speech perception in a non-native language, evidence has consistently shown that an increased effort is required for non-native compared to native listeners. This consistency in results across studies is encouraging in validating pupillometry as a sensitive investigation technique to uncover listening effort differences between groups of participants with different linguistic backgrounds. As previously noted this opens the possibility of uncovering differences in listening effort even when listener groups are performing at optimal levels in terms of performance, but still experiencing differences in the ease of language processing.

However, one concern is the confounding effect that variations in the degree of motivation across listeners' groups could have on the pupil data. A greater willingness to understand speech or to perform better at the experimental task could in fact alter the allocation of resources while listening. Previous evidence has shown that a larger pupil response can be elicited by providing a high versus a low monetary compensation (Bijleveld, Custers, & Aarts, 2009; Knapen et al., 2016; Koelewijn, Zekveld, Lunner, & Kramer, 2018). Although no external manipulation of motivational factors was implemented in the research described in this dissertation, internally-driven variations might have occurred across participants. For example, non-native listeners volunteering to take part in the experiments might have been more motivated in performing well at the listening task compared to native listeners. They were recruited mainly from social media, and could have seen the experimental session as an opportunity to test or demonstrate their ability to understand English. On the contrary, native listeners were mainly recruited through the departmental subject pool, and were presumably more likely to have participated in previous experiments at UCL, therefore it is reasonable to assume that they approached the experimental session with a neutral attitude, rather than an occasion to prove the command of their mother tongue. This can indeed be considered a common downside shared by research aiming at comparing different populations, such as normal hearing and hearing impaired participants, where the latter group might be more motivated due a greater internal drive to overcome their hearing difficulties while performing the experimental task.

One potential strategy to encompass that in second language speech research, and to corroborate the accumulating evidence indicating an increased listening effort during non-native speech perception, would be to replicate this results in the context of a within-subject experiment. This would imply testing the same group of participants on a repeated measures design, presenting them with speech stimuli both in their first and second language. Although it is likely that listeners would still have a greater motivation in their L2 rather than their L1, a within-subject design would eliminate differences between participants' groups purely due to recruitment strategies. As an example, data about native and non-native speech perception would not come from listeners differing in terms of familiarity with the testing environment, as it is likely to happen with participants frequently taking part in experiments vs occasional participants. Additionally, such type of experimental design could allow a more systematic investigation of which individual cognitive abilities are linked with a greater increase in listening effort when switching from L1 to L2 speech perception. Nevertheless, when designing such experiments, the overall duration of the task should also be carefully considered, in order to prevent the confounding effect of fatigue on the pupil response. One option would be dividing the experiment in two sessions to be performed on different days, with the L1 and L2 condition randomised across participants. The effects of L2 proficiency on listening effort could also be further explored by implementing longitudinal studies, allowing to test the same set of participants at different point in times (e.g. by recruiting university international students at the beginning and at end of the academic year). This could also potentially shed light on which individual characteristics are associated with a proficiency-related change in listening effort. On the other hand however, important challenges should be addressed to implement a within-subject study investigating second language speech perception. An ad hoc corpus of stimuli ought to be designed, including speech material equated for difficulty level across languages of interest. This goal could prove to be difficult to achieve, particularly if refined manipulations of semantic context are planned.

One important limitation in applying pupillometry to listening effort research is the poor reliability of pupil data at individual level, as also emerged in the three studies presented here. Because of the large individual variability of pupil traces, it is difficult to draw reliable correlations between pupil response and cognitive measures or individual levels of language proficiency. However, this large individual variability is also shared by other physiological measures used in the study of language processing, as is for example the case in EEG research for the investigation of the late positive complexes in response to semantic anomalies (Kos, Van den Brink, & Hagoort, 2012). Therefore, the use of EEG measures might not be significantly better in capturing the role of individual differences in listening effort modulation. Indeed the recent literature looking at N400 response and cortical entrainment as a measure of listening effort, does not currently focus on the investigation of the role of individual differences among listeners (Song & Iverson, 2018). Behavioural measures such as dual task paradigms might therefore be better candidates to accomplish this objective, even though they lack the fine temporal resolution of pupillometric and EEG measures.

One additional focus of the present thesis was the investigation of the effects of L2 proficiency on non-native listening effort and performance. In particular, I explored whether the effort release after speech onset differed between high and low proficiency non-native listeners. Results were not consistent across studies and experimental conditions. However findings from Study 2 suggested a slower rate of listening effort release after speech offset for listeners with a low compared to high proficiency level. Nevertheless, this trend was only present in the most difficult condition, that is when acoustic cues were more degraded and semantic cues were not available. Interestingly, a prolonged pupil dilation was observed by Bradshaw (1968) during a mathematical task when participants were not able to solve a problem they were presented with, suggesting that a higher and more prolonged level of uncertainty might be reflected by the pupil response. In the case of non-native speech perception, a lower level of L2 proficiency might be linked to a higher level of uncertainty in word recognition, and therefore to a sustained listening effort for a prolonged period of time. This result is of particular relevance if we think about its implications in real life situations. Indeed, while in an experimental setting the pace of sentence presentation can be adapted and paused according the listener's needs, this is often not feasible while real conversations are happening, or may be highly disruptive causing cascading implications during everyday communication. This should be taken into consideration in learning contexts where L2 instructors are interacting with beginner learners, and appropriate pauses between utterances should be included in order to give the interlocutor enough time to process the content. More importantly, professional staff (e.g. in a medical or social setting) should be aware of the opportunity of reducing the speaking rate when communicating with non-native listeners, particularly when delivering unfamiliar or complicated content.

On a final note, I will share some concluding remarks and views on pupil response interpretation within the study of listening effort. All in all, an increased listening effort as measured by the pupil response is commonly regarded as a "negative" effect deriving from challenging listening conditions. This is understandable considering that the ideal listening experience to which listeners aim is effective language comprehension with the least possible effort. However, it is worth highlighting that an increased listening effort should not always be considered as negative. In the case of more challenging listening tasks, or for tasks implying the integration of different types of cues (e.g. a sparse semantic context to be integrated with degraded acoustic information), increased listening effort may simply indicate a greater reliance on top-down mechanisms. Such mechanisms are likely to call for a more extensive brain network activation, and may therefore come with a greater associated processing cost (Winn et al., 2018). Nevertheless, they are also a sign of the listener's capability to channel his/her cognitive resources to accomplish a listening goal. The debate about the association between cognitive ability and effort fits into this line of argumentation. There is no consensus yet on whether greater cognitive capacity is linked with increased or reduced listening effort. However, a flexible interplay between cognitive abilities, linguistic knowledge, task difficulty and cue availability seems to be the most accredited hypothesis (Rönnberg et al., 2019; Zekveld et al., 2018). In this regard, Winn et al. (2018) suggested that pupil dilation could be better interpreted as the effort exerted by listeners, modulated by a cost/benefit evaluation of allocating a certain amount of cognitive resources to a given listening task. Such an evaluation would be dependent on the specific features of the communicative environment, and on the listener's characteristics in terms of cognitive abilities, hearing status, language knowledge and motivation. The proposed view therefore places the individual, with his/her set of abilities and motivation, at the core of the listening effort evaluation.

Appendix A

Linguistic background questionnaire

The following questionnaire was sent to each non-native participants for completion. The questionnaire was implemented on-line on Google Forms.

Bilingual Language profile

We would like to ask you to answer the following questions about your language history, use, attitudes, and proficiency. This will help us to understand your profile and background as bilingual speaker in diverse settings. The survey consists of 27 questions and it will take approximately 10 minutes to complete. This is not a test, so there are no right or wrong answers. Please answer every question and give your answer sincerely.

1. Biographical information

Your name:

____

Your age:

____

Your gender:

- ◻ Male
- ◻ Female
- ◻ Other: ____________________

Current place of residence:

____

Highest level of formal education achieved:

- ◻ Primary or secondary school
- ◻ High school/College
- ◻ Bachelor degree
- ◻ Master degree
- ◻ PhD
- ◻ Other: ______________________

2. Language history

In this section, we would like you to answer some factual questions about your language history by placing a check in the appropriate box.

Which is the first language you started to speak?

____

Which is the first language of your parents?

____

Please, list the language you speak, in order of dominance as you perceive them:

____

At what age did you start learning English?

- ◻ Since birth
- ◻ 1
- ◻ (2, … 19)
- ◻ 20 or more

How many years have you formally studied English for?

- ◻ Less than one
- ◻ 1
- ◻ (2, … 14)
- ◻ 15 or more

How many years of your formal education (primary school through university) were taught in your FIRST LANGUAGE?

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

How many years of your formal education (primary school through university) were taught in ENGLISH?

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

How many years of your formal education (primary school through university) were taught in ANOTHER LANGUAGE? Please specify: ___________________

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

How many years have you spent in a COUNTRY where your FIRST LANGUAGE is spoken?

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

How many years have you spent in a COUNTRY where your ENGLISH is spoken?

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

How many years have you spent in a COUNTRY where your ANOTHER LANGUAGE is spoken? Please specify: ___________________

- □ Less than one
- □ 1
- □ (2, … 15)
- □ 16 or more

If there is any comment you wish to add regarding those questions, or anything else that you wish to add or clarify, please leave a comment here:

____

3. Language use

In this section, we would like you to answer some questions about your language use by placing a check in the appropriate box. Total use for all languages in a given question should equal 100%.

In an average week, what percentage of the time do you use your FIRST LANGUAGE with FRIENDS?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ENGLISH with FRIENDS?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ANOTHER LANGUAGE with FRIENDS?
Please specify: ___________________

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use your FIRST LANGUAGE with FAMILY members?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ENGLISH with FAMILY members?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ANOTHER LANGUAGE with FAMILY members?
Please specify: ___________________

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use your FIRST LANGUAGE at SCHOOL or WORK?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ENGLISH at SCHOOL or WORK?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

In an average week, what percentage of the time do you use ANOTHER LANGUAGE at SCHOOL or WORK?
Please specify: ___________________

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you talk to yourself, how often do you talk to yourself in your FIRST LANGUAGE?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you talk to yourself, how often do you talk to yourself in ENGLISH?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you talk to yourself, how often do you talk to yourself in ANOTHER LANGUAGE? Please specify: __________________

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you count, how often do you count in your FIRST LANGUAGE?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you count, how often do you count in ENGLISH?

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

When you count, how often do you count in ANOTHER LANGUAGE? Please specify: __________________

□ 0% □ 10% □ 20% □ 30% □ 40% □ 50% □ 60% □ 70% □ 80% □ 90% □ 100%

If there is any comment you wish to add regarding those questions, or anything else that you wish to add or clarify, please leave a comment here:

____

4. Language proficiency

In this section, we would like you to rate your language proficiency by giving marks from 0 to 6.

How well do you SPEAK each of the following languages? (0 = not well at all, 6 = very well)

Your first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

How well do you UNDERSTAND each of the following languages? (0 = not well at all, 6 = very well)

Your first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

How well do you READ each of the following languages? (0 = not well at all, 6 = very well)

Your first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

How well do you WRITE each of the following languages? (0 = not well at all, 6 = very well)

Your first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

If you have ever taken an English language proficiency test (e.g. IELTS, TOEFL), please tell us which test did you take, when, and which score you achieved in the test and in each subtest.

____

If there is any comment you wish to add regarding those questions, or anything else that you wish to add or clarify, please leave a comment here:

____

5. Language attitudes

In this section, we would like you to respond to statements about language attitudes by giving marks from 0 to 6.

I feel like myself when I speak…
(0 = strongly disagree, 6 = strongly agree)

My first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

Culturally, I identify myself with…
(0 = strongly disagree, 6 = strongly agree)

My first language culture: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

English culture: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

I want others to think I am a native speaker…
(0 = strongly disagree, 6 = strongly agree)

Of my first language: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

Of English: □ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6

If there is any comment you wish to add regarding those questions, or anything else that you wish to add or clarify, please leave a comment here:

____

Your answer has been recorded, thank you for completing the questionnaire!

Appendix B

Sentences from the Basic English Lexicon (BEL) sentence materials (Calandruccio & Smiljanic, 2012) were used as experimental stimuli in Study 1 and 2 (predictable stimuli).

List 1:

01x01 The PARK OPENS in ELEVEN MONTHS.

01x02 The THIRSTY KID DRINKS JUICE.

01x03 These BROWN MUSHROOMS TASTE AMAZING.

01x04 The TRAIN is FAST and VERY DANGEROUS.

01x05 The ANNOYING STUDENT ASKS many QUESTIONS.

01x06 The PERFORMER WEARS COLORFUL DRESSES.

01x07 A LAZY WORKER RESTS OFTEN.

01x08 My DOCTOR WORKS in that BUSY HOSPITAL.

01x09 He LOST his WHITE HAT TODAY.

01x10 The CITY SCHOOL is LARGE and CROWDED.

01x11 The WEAK PLANT is BARELY ALIVE.

01x 12 The KIDS ENJOYED the HOLIDAY PARADE.

01x 13 The RED VEGETABLES GROW in the GARDEN.

01x 14 The EGGS NEED MORE SALT.

01x 15 The OLD MEN MISSED HOME.

01x 16 The GIRL LOVES SWEET CANDY.

01x 17 The MILK and CHEESE SMELLED HORRIBLE.

01x 18 My GRANDMOTHER BAKED a CHOCOLATE CAKE.

01x 19 The BRIGHT SUN WARMS the GROUND

01x 20 His SISTER PLAYS with BEAUTIFUL TOYS.

01x 21 The PARTY GAME was REALLY EASY.

01x 22 Our CAT HATES TAKING a BATH.

01x 23 The GRAY MOUSE ATE the CHEESE.

01x 24 The BAR SELLS BEER on the WEEKEND.

01x 25 My STRONG FATHER CARRIED my BROTHER.

List 2:

02x01 The TWINS LIVE WITH their GRANDPARENTS.

02x02 Some PEOPLE DRINK BLACK COFFEE.

02x03 The WOMAN MET her FAVORITE ACTOR.

02x04 That BRIGHT LIGHT SEEMS FAR away.

02x05 The GLASS DISH BROKE in the KITCHEN.

02x06 My COUSIN OWNED a SILVER CAR.

02x07 The YOUNG PERFORMER LEARNED to SING.

02x08 The PROFESSORS WRITE SIMPLE PROBLEMS.

02x09 The BROWN BEARS EAT FRUIT.

02x10 The COUPLE KISSED AFTER DINNER.

02x11 The EXCITED TEENAGER was REALLY NOISY.

02x12 The PLAYERS FORGOT to bring LUNCH AGAIN.

02x13 The WEIRD NOISE UPSET the BABY.

02x14 The LADY WALKED DOWN the STREET.

02x15 The LOST DOG was HUNGRY and THIRSTY.

02x16 The SCARY MONKEY CHASED the CHILD.

02x17 Our SHY NEIGHBORS AVOID PEOPLE.

02x18 The FRUIT and SALAD TASTE FRESH.

02x19 Her LOUD COUGH SOUNDED HORRIBLE.

02x20 Her GRANDPARENTS are SERIOUS and SOMETIMES CRUEL.

02x21 The FIVE STUDENTS were LATE for CLASS.

02x22 The PRIVATE UNIVERSITY is not CHEAP.

02x23 The PLANE will LAND in TEN MINUTES.

02x24 Our FATHER WORKS in a LARGE OFFICE.

02x25 The TOURIST TRAVELED MANY PLACES.

List 3:

03x01 That STORE SELLS CHEAP CLOTHES.

03x02 My PUPPY CHASES CATS in the PARK.

03x03 The CHRISTMAS TREE LOOKS WONDERFUL.

03x04 The WAITER TOOK SHORT BREAKS.

03x05 A BIG BEAR SCARED the VISITORS.

03x06 The KING and QUEEN PLANNED a PARTY.

03x07 The ORANGE FIRE BURNED BRIGHTLY.

03x08 The HUNGRY TEENAGERS EAT SNACKS.

03x09 The RABBIT and MOUSE EXPLORED the FIELD.

03x10 The FOREIGN TOURIST was EXCITED and NERVOUS.

03x11 This NEW COMPUTER is QUITE USEFUL.

03x12 The POPULAR BAND PLAYED in a CONCERT.

03x13 The MOTHER MADE LUNCH for her CHILDREN.

03x14 The OPERA THEATER is FULL this EVENING.

03x15 The NURSE WORKS LATE NIGHTS.

03x16 The WILD HORSE JUMPED HIGH.

03x17 The BASEBALL BROKE a GLASS WINDOW.

03x18 A DARK SKY MEANS RAIN.

03x19 The FAMILY ATE in an EXPENSIVE RESTAURANT.

03x20 The STRONG ARMY WON the BATTLE.

03x21 My GRANDPA LOVES RED WINE.

03x22 The BIRD SINGS SONGS in the MORNING.

03x23 The BUSINESS CREATED MANY JOBS.

03x24 The BLUE BICYCLE went DOWN the PATH.

03x25 The CHEF ROASTS FRESH PORK.

List 4:

04x01 The ACTOR WORKED for SIX DAYS.

04x02 The FATHER HUGS his SAD DAUGHTER.

04x03 The KITCHEN GARBAGE SMELLED TERRIBLE.

04x04 That PRESIDENT LIVES in an EXPENSIVE APARTMENT.

04x05 A BASEBALL is HARD and PERFECTLY ROUND.

04x06 The RUDE JOKE UPSET my PARENTS.

04x07 The WILD ANIMALS SLEEP in the FOREST.

04x08 The DARK NIGHT BRINGS FEAR.

04x09 The CHEF COOKS WONDERFUL FOOD.

04x10 His FIRST GIRLFRIEND was ATTRACTIVE and SMART.

04x11 The HUSBAND and WIFE CUT the CAKE.

04x12 The HOMEMADE TEA was TOO HOT.

04x13 The BIRD FOUND a JUICY WORM.

04x14 The TOMATO PLANT GREW by the WINDOW.

04x15 The STUPID GUEST is NOT WELCOME.

04x16 My GRANDMOTHER DRINKS COLD BEER.

04x17 The WORRIED ADULT RAN HOME.

04x18 The NEWSPAPER COMES EVERY WEEKEND.

04x19 The TALENTED MUSICIAN KNOWS many SONGS.

04x20 The GRAY HORSE EATS GRASS.

04x21 The TREES GROW SWEET APPLES.

04x22 The POOR FAMILY ONLY eats RICE.

04x23 The CUSTOMER ENJOYED the MEAL and WINE.

04x24 The TWELVE CHILDREN were HIDING OUTSIDE.

04x25 The LAZY STUDENT FAILED the TEST.

List 5:

05x01 Their NEPHEW RAN AROUND the HOUSE.

05x02 The ENGLISH TEA SMELLED BAD.

05x03 My NEIGHBOR SINGS COUNTRY SONGS.

05x04 The FOOTBALL GAME ENDED EARLY.

05x05 The HUNGRY GIRL MADE a SANDWICH.

05x06 The SICK QUEEN RULED from her BED.

05x07 The ASSIGNMENT SEEMED SHORT and SIMPLE.

05x08 The TINY FLY BOTHERED EVERYONE.

05x09 A HAPPY MARRIAGE is VERY IMPORTANT.

05x10 The MAN WALKED in the FOREST ALONE.

05x11 The YELLOW CORN was SWEET and SALTY.

05x12 The TIRED BABY WENT to SLEEP.

05x13 The POPULAR CLUB is OFTEN FULL.

05x14 The CAT CHASED the DIRTY MOUSE.

05x15 Her BUSY DAUGHTER has MANY JOBS.

05x16 My MOTHER BAKES DELICIOUS COOKIES.

05x17 The NEW STUDENT ASKED QUESTIONS.

05x18 The FARMER WORKS a LONG DAY.

05x19 The SUN SETS in the LATE AFTERNOON.

05x20 The HOT COFFEE BURNED the BOY.

05x21 The FISH SWAM SLOWLY in the LAKE.

05x22 The CHEF PREPARES BREAKFAST in the KITCHEN.

05x23 Our COUSIN STARTS SCHOOL TOMORROW.

05x24 The DENTIST ENJOYS CANDY at NIGHT.

05x25 Our PROFESSOR ANSWERS EVERY QUESTION.

List 6:

06x01 The PLANTS and TREES LOOK BEAUTIFUL.

06x02 The TEAMS PLAY DIFFERENT SPORTS.

06x03 The WARM SUNSHINE FELT GREAT.

06x04 The CUTE PUPPY RAN HOME.

06x05 The STRANGE ANIMAL SCARED the BABY.

06x06 The WIND DAMAGED the TINY BOAT.

06x07 This OFFICE is COMFORTABLE and USUALLY QUIET.

06x08 The KIND GIRL HELPS STRANGERS.

06x09 The SPICY CARROTS were her FAVORITE DISH.

06x10 The TROPICAL BEACHES are HOT and CROWDED.

06x11 The RESTAURANT SELLS RED WINE.

06x12 The OLD AUNT was ALWAYS MEAN.

06x13 The FAT PIG SLEPT on the FLOOR.

06x14 She DROVE the BUS DOWN the STREET.

06x15 The MUSICIAN PLAYS MANY INSTRUMENTS.

06x16 The BLUEBERRY PIE BAKED in the OVEN.

06x17 The TEENAGER LIFTED a HEAVY BOX.

06x18 The MATH TEST was EASY to FINISH.

06x19 That MAN SWIMS in the COLD WATER.

06x20 The BASEBALL FLEW ACROSS the FIELD.

06x21 The TALENTED WRITER RECEIVED an AWARD.

06x22 The ADULTS LEARNED to DANCE in SCHOOL.

06x23 The BLACK CAT CLIMBED the TREE.

06x24 The STARS LIT the NIGHT SKY.

06x25 My GRANDPARENTS TOOK PICTURES on VACATION.

List 7:

07x01 The NEIGHBOR ran FROM the STRANGER.

07x02 Her THOUGHTFUL BOYFRIEND SENT FLOWERS.

07x03 Her BLACK SWEATER LOOKED FUNNY.

07x04 The MAN ATE a LARGE MEAL.

07x05 The STARVING DOG SMELLED the FOOD.

07x06 A FOREIGN COUNTRY is EXCITING to VISIT.

07x07 The FIGHTER is STRONG and VERY BRAVE.

07x08 The SAD PETS NEED LOVE.

07x09 His PARENTS TELL BORING STORIES.

07x10 The LAST YEAR was CALM and PEACEFUL.

07x11 That BOOK COST TEN DOLLARS.

07x12 A KIND WORD is ALWAYS APPRECIATED.

07x13 The PROUD FANS CHEERED for their TEAM.

07x14 The FRIED EGG was COOKED in BUTTER.

07x15 The NEWS was on TV EVERY MORNING.

07x16 The CHEF MADE FRESH NOODLES.

07x17 The CROWD ENJOYED the DANCING and DRUMS.

07x18 The CHURCH GROUP INSPIRED the COMMUNITY.
07x19 The PEOPLE SIT in the LIVING ROOM.
07x20 The CLEAN BEACHES have CLEAR WATER.
07x21 The SINGER and DANCER JOINED the ACTOR.
07x22 A LAZY CHILD SLEEPS OFTEN.
07x23 The MEAT TASTES GREAT with this SAUCE.
07x24 The PROFESSOR READ an INTERESTING NOVEL.
07x25 The DRIVER EARNED MONEY YESTERDAY.

List 8:

08x01 The COUPLES ENTERTAIN AROUND the HOLIDAYS.
08x02 The CLOSEST BEACH was REALLY CROWDED.
08x03 The STEAMED CHICKEN TASTED STRANGE.
08x04 The WORKERS NEED BETTER EDUCATION.
08x05 The HAPPY PET FOUND its TOY.
08x06 Her APARTMENT was NEAR the PRIVATE SCHOOL.
08x07 The YOUNGEST SISTER WATCHES TV.
08x08 The MAN LOST his HOUSE KEY.
08x09 The NEW GAME was my BIRTHDAY GIFT.
08x10 The DEEP OCEAN is DARK and DANGEROUS.
08x11 The WOODEN DOOR was HARD to CLOSE.
08x12 They BOUGHT THREE BLUE CARS.
08x13 The GRAPE JUICE SPILLED on the FLOOR.
08x14 The BEST EXPLANATION is OFTEN SIMPLE.
08x15 The TWO FRIENDS had a TERRIBLE FIGHT.
08x16 The RESTAURANT SERVES DUCK SOUP.
08x17 They ATE the ENTIRE CABBAGE QUICKLY.
08x18 Our TEAM PRACTICES EVERY NIGHT.
08x19 The SCARED DOG LOSES HOPE.
08x20 Our GRANDFATHER is SERIOUS and NEVER KIND.
08x21 The SUN SHINED BRIGHTLY in JUNE
08x22 The LADY SINGS a BEAUTIFUL SONG.
08x23 That PRETTY GIRL WON a PRIZE.
08x24 My BROTHER SLEEPS until LATE MORNING.
08x25 The OLD MAID LOOKS SAD.

List 9:

09x01 The LAWYER STUDIED for FIVE HOURS.

09x02 The CITY BUS is USUALLY EARLY.

09x03 His GIRLFRIEND LOVES CHINESE FOOD.

09x04 Their FAMOUS SON DANCED WELL.

09x05 The LONELY LADY CALLED her FRIEND.

09x06 She BOUGHT a PRETTY, COLORFUL HAT.

09x07 Her YOUNGEST CHILD HATES FRUIT.

09x08 The BEST WORKER WENT on the TRIP.

09x09 That TINY ANIMAL is CUTE but DANGEROUS.

09x10 Her UNCLE CRIED SOFTLY in his ROOM.

09x11 The CLASS LEARNED about EARTH SCIENCE.

09x12 The BOSS FIRED the LAZY WAITER.

09x13 The AUDIENCE ENJOYED the SONG and DANCE.

09x14 The CHEAP DRINKS ATTRACT CUSTOMERS.

09x15 The RICH ADULT OWNS many HOUSES.

09x16 The CROWD WATCHED the TALENTED PERFORMER.

09x17 The RAIN DESTROYED SOME PLANTS.

09x18 The SMALL BOY SEEMED SAD.

09x19 A SUMMER VACATION is ALWAYS RELAXING.

09x20 Some WRITERS TELL INTERESTING STORIES.

09x21 The BLACK BEAR was BIG and SCARY.

09x22 The PROFESSOR TRAVELED VERY FAR.

09x23 The CHILDREN PLAYED TOGETHER YESTERDAY.

09x24 The SPOILED POTATOES TASTED BAD.

09x25 She DROVE the CAR FAST and STRAIGHT.

List 10:

10x01 The CHRISTMAS SHOW INTERESTED my SON.

10x02 These COLLEGES OFFER MANY COURSES.

10x03 The BIG ROOM FELT EMPTY.

10x04 The PAINTER USED SOFT BRUSHES.

10x05 His GRANDMA and GRANDPA HELPED the KIDS.

10x06 The ANGRY BEAR SCARED the CAMPERS.

10x07 The STRANGER SEEMED TROUBLED and UNHAPPY.

10x08 The THIRSTY CAT DRANK MILK.

10x09 The YOUNG student TOOK the FINAL EXAM.

10x10 Her LAST BOYFRIEND was BORING and RUDE.

10x11 The WHITE ONION was CUT too THIN.

10x12 The MARKET SELLS DELICIOUS FOOD.

10x13 The TIRED CHILD CRIED for her MOTHER.

10x14 The CHEF COOKS PASTA EVERY day.

10x15 The BOILED FISH SMELLS BAD.

10x16 The NOISY GROUP ATTRACTS ATTENTION.

10x17 The TRAVELERS VISITED the ART MUSEUM.

10x18 The BIRD LAYS TINY EGGS.

10x19 The GIRL PLAYED with her BEST FRIEND.

10x20 The LONELY DUCK SWIMS in the LAKE.

10x21 My BAGS ARRIVED EARLY at the STATION.

10x22 The THREE SISTERS SHARED CLOTHES.

10x23 My GRANDMOTHER READ the NEWSPAPER QUICKLY.

10x24 The MEAN TEACHER is NEVER NICE.

10x25 The INSTRUCTOR GAVE BOOKS to her CLASS.

List 11:

11x01 The MEETING STARTS in TWENTY MINUTES.

11x02 The CUSTOMERS HATE BLACK TEA.

11x03 The SICK PERSON FEELS BETTER.

11x04 That BROWN BIRD is ALWAYS HERE.

11x05 The THREE COUSINS did their MATH HOMEWORK.

11x06 The DARK CLOUD COVERED the SKY.

11x07 The GROCERY STORE SELLS FOOD.

11x08 The MOVIE STARTED in the SMALL ROOM.

11x09 The CHICKEN SOUP was a TASTY MEAL.

11x10 The COOL NIGHT was COMFORTABLE and RELAXING.

11x11 The BIRTHDAY CARD was SENT LATE.

11x12 The SECRETARY LEARNED SPANISH EASILY.

11x13 The WHITE HORSE LIVES on a FARM.

11x14 Our APARTMENT NEEDS MORE WINDOWS.

11x15 Our MOTHER DRINKS ORANGE JUICE.

11x16 The DANGEROUS SNAKE bit the RABBIT.

11x17 The PROFESSOR GAVE an UNFAIR GRADE.

11x18 They PLAYED FAST MUSIC on the RADIO.

11x19 That ENGLISH TEST was REALLY DIFFICULT.

11x20 The BOYFRIEND and GIRLFRIEND PLANNED their WEDDING.

11x21 The KIDS SCREAMED LOUDLY in the PARK.

11x22 The TROUBLED SON STOLE MONEY.

11x23 His SISTER BUYS CAKE DAILY.

11x24 A LITTLE KITTEN CLIMBED over the FENCE.

11x25 The SNOWMAN had TWO GREEN GLOVES.

List 12:

12x01 The BALL ROLLED DOWN the HILL.

12x02 The BEDROOM RUG had a LARGE STAIN.

12x03 This EXPENSIVE DINNER TASTES GREAT.

12x04 The DOCTOR HELPED the SICK PATIENT.

12x05 Those CUTE ANIMALS CHEWED the PLANTS.

12x06 The GIRL PICKS PRETTY FLOWERS.

12x07 The BUSY FARMER GROWS POTATOES.

12x08 My BROTHER and SISTER RECEIVED GIFTS.

12x09 The INTERNATIONAL BUSINESS FAILED QUICKLY.

12x10 The STRONG WIND COOLED the AIR.

12x11 The TERRIBLE CUSTOMER was LOUD and ANGRY.

12x12 Her YOUNGEST SON was ALWAYS UPSET.

12x13 My YELLOW SHOES CAME in a BAG.

12x14 The RATS RAN through the DARK STREETS.

12x15 The GOAT EATS DRY LEAVES.

12x16 The FAT BUNNY LOVES CARROTS.

12x17 She WASHED and DRIED her CURLY HAIR.

12x18 The MAP SHOWS the CITY ROADS.

12x19 The HONEST MOTHER is LOVING and NICE.

12x20 The HAPPY CHILDREN LAUGH at the STORY.

12x21 Her UNCLE WAITS QUIETLY for the ANSWER.

12x22 They TOOK a SCHOOL PICTURE every YEAR.

12x23 The SHOPPER BOUGHT MANY THINGS.

12x24 The AIRPLANE FLEW in the BLUE SKY.

12x25 The CHEF BOILED CARROTS in the KITCHEN.

List 13:

13x01 The STUDENTS WALKED AROUND the CAMPUS.

13x02 Our CHILDREN LOVE GREEN VEGETABLES.

13x03 Her FAVORITE PANTS WERE RUINED.

13x04 The BOSS ASSIGNED a DIFFICULT PROJECT.

13x05 Their PETS LOOKED LOST and SAD.

13x06 The METAL KEY OPENED the DOOR.

13x07 Her BOYFRIEND WATCHED MOVIES with another GIRL.

13x08 The OLD GARBAGE ATTRACTS FLIES.

13x09 The WHITE and BROWN DOG was FRIENDLY.

13x10 Their FIRST DATE was FUN and ROMANTIC.

13x11 The KITTEN SAT on the WOODEN CHAIR.

13x12 A LONELY PERSON is USUALLY UNHAPPY.

13x13 The PARTY LASTED for THREE HOURS.

13x14 The TENNIS TEAM PRACTICED at NIGHT.

13x15 Those STORES SELL SPORTS CLOTHES.

13x16 Her LOUD VOICE SOUNDS BAD.

13x17 The MACHINE MADE a TERRIBLE NOISE.

13x18 The TRAIN STATION is UNDERGROUND and DIRTY.

13x19 The PASTA BOILED in the HOT WATER.

13x20 The FLAGS FLY HIGH and PROUD.

13x21 The SHOW ENDED EARLY TODAY.

13x22 A GOOD FRIEND TELLS the TRUTH.

13x23 The FAMILY LISTENS to CLASSICAL MUSIC.

13x24 This MEAL NEEDS SALT and PEPPER.

13x25 The ARTIST VISITED MANY MUSEUMS.

List 14:

14x01 The PICTURE HUNG ABOVE the DOOR.

14x02 They BOUGHT CLOTHING at the BUSY STORE.

14x03 The GREEN PLANTS LOOK HEALTHY.

14x04 My COUSIN BAKED me a BIRTHDAY CAKE.

14x05 The SLOW COMPUTER had MANY PROBLEMS.

14x06 The STUDENT STUDIES in the QUIET ROOM.

14x07 The LARGE FAMILY EXPECTED VISITORS.

14x08 The DOG CHASED FOUR RABBITS.

14x09 The SMALL FISH MARKET is CLEAN.

14x10 Their YOUNGEST DAUGHTER is TALL and THIN.

14x11 The MAN ATE FRIED CHICKEN.

14x12 That NEW BOOK is REALLY HELPFUL.

14x13 The RESTAURANT SERVES DINNER and DRINKS.

14x14 The FOREIGN LADY DREAMED of her HOME.

14x15 The BOSS TELLS HORRIBLE JOKES.

14x16 The LITTLE GIRL was ANGRY TODAY.

14x17 The COUPLE LIVES a PEACEFUL LIFE.

14x18 The HAPPY MONKEY SWINGS from the TREE.

14x19 The CLUB MEMBERS PLAYED SPORTS.

14x20 My PARENTS and GRANDPARENTS TRUST the POLICE.

14x21 The PERSON SOLD her HOUSE YESTERDAY.

14x22 The TWIN SISTERS WATCHED a MOVIE.

14x23 The AUTHOR WROTE a LONG NOVEL.

14x24 The SAD CHILD was TIRED and LONELY.

14x25 The DRIVER STOPPED SUDDENLY in the STREET.

List 15:

15x01 The WAITRESS SMOKED DURING her BREAK.

15x02 That FAST KITTEN CHASED a MOUSE.

15x03 The CHEAP CLOTHES LOOKED DIRTY.

15x04 They LOVED the FRENCH FOOD and DESSERT.

15x05 The ANGRY HUSBAND VISITED a LAWYER.

15x06 The FANS WATCHED FOOTBALL GAMES.

15x07 The UNHAPPY CHILD NEEDS HELP.

15x08 That BOY CARRIED SIX BAGS.

15x09 The VEGETABLE SOUP was a HEALTHY LUNCH.

15x10 The FARM TOWN was SMALL and NICE.

15x11 The SUN SHINES in the GREEN FOREST.

15x12 The ITALIAN RESTAURANT is REALLY POPULAR.

15x13 The STRESSFUL WEEK ENDED at the BAR.

15x14 The LONG PROJECT was COMPLETED on TIME.

15x15 The OLD SHIRT was warm and soft.

15x16 Their AUNT BUYS BIRTHDAY GIFTS.

15x17 The FAMILY CELEBRATED their FAVORITE HOLIDAY.

15x18 The BROWN BALL FELL off the TABLE.

15x19 The ACTOR SANG LOUDLY in the THEATER.

15x20 The four friends WENT on VACATION.

15x21 The TEACHER GAVE HOMEWORK DAILY.

15x22 The PREGNANT WOMAN LIKES CHOCOLATE.

15x23 They ROASTED SALTY MEAT in the PAN.

15x24 Those LITTLE KIDS are TIRED TODAY.

15x25 The DOGS DIG GIANT HOLES.

List 16:

16x01 The PERFORMER WORKED for LITTLE MONEY.

16x02 Her SON LOVES TOY CARS.

16x03 The MATH CLASS SEEMED USEFUL.

16x04 The FOUR COUSINS TRAVELED around the WORLD.

16x05 The STORMY WEATHER DESTROYED the HOME.

16x06 The ADULTS ATE LUNCH QUICKLY.

16x07 The CATHOLIC PRIEST SANG SONGS.

16x08 The ARTIST TOOK a BEAUTIFUL PICTURE.

16x09 The CLEAN KITCHEN has GLASS PLATES.

16x10 He SCREAMED LOUDLY in the CROWDED ROOM.

16x11 The CHERRY PIE was WARM and SWEET.

16x12 The BIG DISH was VERY HOT.

16x13 The EXCITED CHILDREN CHEERED for their TEAM.

16x14 The WAITER BROKE TEN GLASSES.

16x15 The RELIGIOUS COUPLE BELIEVES in GOD.

16x16 The MONKEY WANTS YELLOW BANANAS.

16x17 The BAD NEWS CAME SUDDENLY.

16x18 The UGLY RATS were QUIET and STILL.

16x19 The GARDENER GREW COLORFUL PEPPERS.

16x20 The SICK PATIENT RECEIVED many FLOWERS.

16x21 My SISTERS are FRIENDLY and ALWAYS TALKING.

16x22 The OCEAN LOOKED PERFECTLY CALM.

16x23 The TROPICAL FOREST had MANY TREES.

16x24 Our AUNT BAKES FISH WEEKLY.

16x25 The SMALL BUSINESS NEEDS SUPPORT.

List 17:

17x01 The ACTOR WALKED ACROSS the STAGE.

17x02 The COMPANY BUYS FOREIGN CARS.

17x03 The BAKED BREAD TASTED SWEET.

17x04 The SCHOOLS NEED MORE TEACHERS.

17x05 The TALENTED ARTIST DREW a PICTURE.

17x06 The PINK PIG is FAT and LAZY.

17x07 The TWO FRIENDS HIKED up the MOUNTAIN.

17x08 The ENGLISH CLASS READS BOOKS.

17x09 The FUNNY TELEVISION SHOW is POPULAR.

17x10 The BLUE HAT was STRANGE and UGLY.

17x11 The PRESIDENT GAVE a SHORT SPEECH.

17x12 The YOUNG WOMAN is VERY SMART.

17x13 The DIVORCED COUPLE SAT at the TABLE.

17x14 That GOAT FELL in the DEEP HOLE.

17x15 The GROUP HEARD SLOW MUSIC.

17x16 The HUGE SUPERMARKET OPENS TOMORROW.

17x17 She SPEAKS MANY DIFFERENT LANGUAGES.

17x18 The MOUSE FOUND TASTY CHEESE.

17x19 The VEGETABLES GREW in the GREEN GARDEN.

17x20 The FATHER and DAUGHTER SAW the MOVIE.

17x21 The HOT SUN WARMED the POOL.

17x22 The CHICKEN SANDWICH CAME with SALAD.

17x23 The TEN GIFTS were COMPLETELY BROKEN.

17x24 The DEDICATED NURSES HELP PATIENTS.

17x25 The LESSON SEEMED TOO DIFFICULT.

List 18:

18x01 The ONLY HOTEL is FAR and EXPENSIVE.

18x02 The AUTHOR WROTE THIRTY BOOKS.

18x03 The SPICY MEAL TASTED GREAT.

18x04 Our BAND PRACTICES in my SMALL GARAGE.

18x05 Our BUSY DAUGHTER JOINED many CLUBS.

18x06 The BUTCHER SELLS ITALIAN MEAT.

18x07 A FAITHFUL HUSBAND is ALWAYS HAPPY.

18x08 The KIND LADY GIVES ADVICE.

18x09 Her RIGHT ARM and LEG were BROKEN.

18x10 The PHONE MADE a LOUD NOISE.

18x11 The ENGLISH PROFESSOR never ASSIGNS HOMEWORK.

18x12 The CURIOUS VISITOR TRAVELED to the CITY.

18x13 The SICK NEIGHBOR ASKS for HELP.

18x14 The CATS DRINK WARM MILK.

18x15 The AMAZING DOCTOR SAVES LIVES.

18x16 The LIGHTNING ENDED the SOCCER GAME.

18x17 Our BEST PLAYER WON the AWARD.

18x18 The KEYS DISAPPEARED in the ORANGE BOX.

18x19 The MARKET was CROWDED and TOO FAR.

18x20 The OLD PINK SHIRT had no BUTTONS.

18x21 The TWO WAITERS SERVED BREAKFAST.

18x22 The RAIN LASTED for MANY WEEKS.

18x23 The HARDWORKING FARMER CUT the CORN.

18x24 The APPLE FELL on the WOOD FLOOR.

18x25 The COFFEE CAKE was a PERFECT DESSERT.

List 19:

19x01 That NEW STUDENT is QUIET and SHY.

19x02 The FAMILY LOST THEIR PET.

19x03 The JAZZ SINGER SOUNDED GREAT.

19x04 The TWINS RECEIVED the SAME GIFT.

19x05 The OFFICE PHONE RANG OFTEN.

19x06 The GOLD RING FIT her FINGER.

19x07 The PUPPY SLEPT PEACEFULLY on the BED.

19x08 A HUNGRY RABBIT EATS CARROTS.

19x09 The PLASTIC PLATE and CUP were CHEAP.

19x10 The FIRST QUESTION was CONFUSING and DIFFICULT.

19x11 Their NEIGHBORS TELL INTERESTING STORIES.

19x12 The RED APPLE is TOO SMALL.

19x13 A PRETTY BOAT went DOWN the RIVER.

19x14 The BORED CHILDREN LISTENED to MUSIC.

19x15 My GRANDFATHER MADE WOODEN CHAIRS.

19x16 The LIGHT BLUE JEANS are not EXPENSIVE.

19x17 My BROTHER JOINED the FUN GAME.

19x18 The FRIENDLY BABY HUGS EVERYONE.

19x19 Our BUSINESS PAID for the DAILY NEWSPAPER.

19x20 The RICE and BEANS LOOKED DRY.

19x21 The NURSES WORK HARD every DAY.

19x22 The TALL WOMAN WALKED AWAY.

19x23 The BIRD FLEW OVER the SEA.

19x24 The COUPLE SANG the SONG WELL.

19x25 The ARTIST STUDIES ITALIAN and FRENCH.

List 20:

20x01 My UNCLE KNOWS ABOUT the WAR.

20x02 The PLAYER KICKED the SOCCER BALL.

20x03 The HELPFUL NANNY CLEANED the HOUSE.

20x04 The RED GRAPES are BIG and TASTY.

20x05 The THIRSTY PIG DRANK WATER.

20x06 Our FAMILIES CELEBRATED the HOLIDAY TOGETHER.

20x07 The RICH DENTIST BOUGHT new TOOLS.

20x08 The LAST DAY was CRAZY and FUN.

20x09 The TEACHER CHOOSES DIFFICULT QUESTIONS.

20x10 That DEEP LAKE is QUITE COLD.

20x11 A LITTLE RABBIT RUNS through the FOREST.

20x12 The WORKER HURT his LEFT HAND.

20x13 His SPEECH was BORING and too LONG.

20x14 The LEAVES CHANGE COLOR in the FALL.

20x15 The ARTIST DREW on the YELLOW PAPER.

20x16 The TEAM SCORED GOALS EASILY.

20x17 He CUT the STEAK WITH a KNIFE.

20x18 The SOFT MUSIC PLEASED EVERYONE.

20x19 The SHY GUEST SPEAKS QUIETLY.

20x20 My AUNT RAISED FIVE CHILDREN.

20x21 The HORRIBLE NEWS UPSET my GRANDPARENTS.

20x22 Their OLDEST DAUGHTER PLAYS with TOYS.

20x23 The FUNNY MOVIE will END SOON.

20x24 The SCARED MOUSE STAYED in the HOLE.

20x25 The MEAN PEOPLE BROKE the RULES.

Appendix C

The following 20 lists of 25 sentences each were constructed based on BEL sentences (Calandruccio & Smiljanic, 2012, see appendix A), and presented in Study 2 as anomalous stimuli. Further details on anomalous sentences' creation are provided in section 3.2.2.

List 1:

01x01 The GARDEN TASTES VERY AMAZING.

01x02 The HORRIBLE PARADE was TAKING the KIDS.

01x03 These STRONG GROUNDS ASK for COUNTRIES.

01x04 The EGG is CROWDED and OFTEN FAST.

01x05 The ANNOYING DRESSES NEED many HOSPITALS.

01x06 The BEER WARMS the BROWN PARTY.

01x07 An OLD HOLIDAY OPENS OFTEN.

01x08 My CAT BAKED that BEAUTIFUL HAT.

01x09 He BARELY WORKS his GREY TRAIN.

01x10 The SALT MONTH is WHITE and EASY.

01x11 The PRIVATE VEGETABLE is ALWAYS LARGE.

01x12 The SCHOOL GROWS the EXPENSIVE BATH.

01x13 The THIRSTY PERFORMER MISSED the JUICE.

01x14 The MUSHROOMS WEAR WEAK CHEESE.

01x15 The LAZY SUN SMELLED the GIRL.

01x16 The PLANT LOVES BRIGHT CANDIES.

01x17 The PARK and QUESTIONS SELL FAMILIES.

01x18 My MEN LOST a BUSY FATHER.

01x19 The SWEET GRANDMOTHER PLAYS the CHEESE.

01x20 His STUDENT HATES the CHOCOLATE MOUSE.

01x21 The DOCTOR MILK was COLOURFUL TODAY.

01x22 Our BAR ENJOYED DRINKING a GAME.

01x23 The TASTY WEEKEND ATE the TOYS.

01x24 The CAKE RESTED the WORKER on the CITY.

01x25 My SOFT SISTER CARRIED my HOME.

List 2:

02x01 The MONKEY DRINKS after THEIR PROBLEMS.

02x02 Some SALADS TASTE HELPFUL FRUIT.

02x03 The LADY LIVES with her FAR CAR.

02x04 That SHY PEOPLE KISSED away SOON.

02x05 The THIRSTY BEAR SOUNDED in the COFFEE.

02x06 My WOMAN LANDED TEN PROFESSORS.

02x07 The BRIGHT GRANDPARENTS SEEMED to WALK.

02x08 The BABY EATS a LARGE NOISE.

02x09 The BROWN CHILD MET the TWINS.

02x10 The STREET WRITES DOWN the KITCHEN.

02x11 The EXCITED ACTOR was SOMETIMES SERIOUS.

02x12 The DOG AVOIDED to buy the QUESTIONS AGAIN.

02x13 The MANY PLAYERS SING the COUSIN.

02x14 The TOURIST BROKE WITH the UNIVERSITY.

02x15 The LONG NEIGHBOUR was CHEAP and FRESH.

02x16 The FIVE LUNCHES TRAVELLED the PLACES.

02x17 Our HUNGRY MINUTES OWNED the STUDENTS.

02x18 The PEOPLE and the COUGH WORK HORRIBLY.

02x19 Her SILVER OFFICE LEARNED the GRANDPARENTS.

02x20 Her DINNERS are LATE and NOT CRUEL.

02x21 The WEIRD TEACHER was LOUD for the MOUSE.

02x22 The GLASS COUPLE is REALLY LOST.

02x23 The PLANE will CHASE in a BLACK PERFORMER.

02x24 Our CLASS FORGOT a PRIVATE DISH.

02x25 The MONTHS LOVED SIMPLE MILK.

List 3:

03x01 That CAT LOVES SHORT RESTAURANTS.

03x02 My BASEBALL MADE CHILDREN in the KING.

03x03 The OPERA SKY EXPLORED EASILY.

03x04 The QUEEN PLANNED a NEW PORK.

03x05 A BIG ARMY JUMPED the BAND.

03x06 The HORSE and TEENAGER BROKE a FIELD.

03x07 The CHRISTMAS BATTLE PLEASED the COLOUR.

03x08 The GLASS BEAR SCARED the PARTY.

03x09 The TOURIST and RABBIT ATE the WAITER.

03x10 The HUNGRY LUNCH was BLUE and LATE.

03x11 This DARK BICYCLE is QUITE ORANGE.

03x12 The STRONG MOUSE ROASTS in a BUSINESS.

03x13 The NIGHT CHASES SONGS for its WINE.

03x14 The RED CHEF is FULL of CLOTHES.

03x15 The MOTHER LOOKS MANY FAMILIES.

03x16 The USEFUL PUPPY EATS the DENTIST.

03x17 The PATH CREATED a WONDERFUL NURSE.

03x18 A CHEAP CONCERT BURNED the BREAKS.

03x19 The VISITORS SINGS in a HIGH GRANDPA.

03x20 The FRESH PARK WORKS the BIRD.

03x21 My JOBS WON the FOREIGN STORE.

03x22 The WINDOW PLAYED THEATRE in the SNACKS.

03x23 The MORNING MEANS WILD TREES.

03x24 The EXCITED COMPUTER went BRIGHTLY this EVENING.

03x25 The RAIN SELLS POPULAR FIRE.

List 4:

04x01 The NIGHT GREW for TWELVE ADULTS.

04x02 The FOREST HUGS his LAZY DAUGHTER.

04x03 The TOMATO GARBAGE LIVES WORRIED.

04x04 That DAY GROWS in a RUDE CAKE.

04x05 The GRASS is STUPID and TOO HOT.

04x06 The ONLY ACTOR UPSET my HOME.

04x07 The KITCHEN PLANT CUTS in the APARTMENT.

04x08 The ATTRACTIVE CHEF WORKED the TREES.

04x09 The TEST COMES to the TALENTED HUSBAND.

04x10 His WONDERFUL PARENTS were POOR and HARD.

04x11 The TEA and FEAR FAILED the WORM.

04x12 The TERRIBLE SONG was NOT ROUND.

04x13 The GUEST RAN a SMART MUSICIAN.

04x14 The HOMEMADE WIFE was FOUND by the FAMILY.

04x15 The SWEET HORSE is PERFECTLY COLD.

04x16 My WEEKEND KNOWS SIX GRANDMOTHERS.

04x17 The SAD WINE EATS the WINDOW.

04x18 The FATHER SCARED EVERY STEAK.

04x19 The GRAY BIRD ENJOYED many BEERS.

04x20 The WELCOME APPLES DRINK RICE.

04x21 The MEAL COOKS the FIRST NEWSPAPER.

04x22 The DARK PRESIDENT only SMELLED FOOD.

04x23 The CUSTOMER BRINGS the BASEBALL and ANIMALS.

04x24 The JUICY STUDENT was SLEEPING OUTSIDE.

04x25 The WILD CHILDREN LOVED the JOKE.

List 5:

05x01 Their SANDWICH BAKES the FUN HOUSE.

05x02 The DELICIOUS SCHOOL SLEEPS BUSY.

05x03 My ASSIGNMENT CHASED POPULAR GAMES.

05x04 The COUNTRY CANDY ASKED ALONE.

05x05 The DIRTY FARMER ENDED a CORN.

05x06 The FOOTBALL BED ENJOYS its COFFEE.

05x07 The AFTERNOON WENT TINY and SHORT.

05x08 The NEW QUESTION SEEMED YELLOW.

05x09 A FULL NEPHEW is OFTEN IMPORTANT.

05x10 The QUEEN RULES in the GIRL TOMORROW.

05x11 The SICK BABY was LATE and LONG.

05x12 The FIVE PROFESSORS PREPARED the LEAVES.

05x13 The SWEET JOBS are SLOWLY TIRED.

05x14 EVERYONE MADE the BAD LAKE.

05x15 The OLDEST MOTHER has a DEEP BREAKFAST.

05x16 My MAN SWAM in SALTY QUESTIONS.

05x17 The HUNGRY SUN ANSWERS the NEIGHBOURS.

05x18 The DENTIST WALKED an ENGLISH BOY.

05x19 The COUSIN SMELLED the PRETTY GAME.

05x20 The HAPPY SONGS START the KITCHEN.

05x21 The FLY WORKS AROUND the FISH.

05x22 The STUDENT RUNS the NIGHT in the MOUSE.

05x23 Our MARRIAGE SINGS the DAUGHTER OUTSIDE.

05x24 The CAT BOTHERED the CHEF in the FOREST.

05x25 Our COOKIES SET EVERY DAY.

List 6:

06x01 The TREES and OFFICES LOOK BLACK.

06x02 The OVEN SCARED the TINY BABIES.

06x03 The GOLD PUPPY CELEBRATED KINDLY.

06x04 The RED RESTAURANT CLIMBED the SUNSHINE.

06x05 The DIFFERENT GRANDPA RAN the PLANTS.

06x06 The VACATION LIFTED the WARM STRANGERS.

06x07 This CARROT is STRANGE and ALWAYS COLD.

06x08 The COMFORTABLE STARS SELL the MUSICIAN.

06x09 The OLD DISHES were her EASY BOX.

06x10 The NIGHT TREES are FAT and TALENTED.

06x11 The SPORTS RECEIVED MANY TESTS.

06x12 The TROPICAL ANIMAL was USUALLY CROWDED.

06x13 The MATH GIRL FLEW on the INSTRUMENT.

06x14 She TOOK the SCHOOL ACROSS the FIELD.

06x15 The WIND HELPS the BLUEBERRY PICTURES.

06x16 The CUTE WATER SLEPT in the BOAT.

06x17 The TEENAGER DANCES a MEAN STREET.

06x18 The HEAVY FLOOR was HOT to DAMAGE.

06x19 That TEAM FELT in the SPICY PIE.

06x20 The BUS PLAYS DOWN the MAN.

06x21 The HORRIBLE ADULTS FINISHED the TRAIN.

06x22 The SKY LEARNED to DRIVE at HOME.

06x23 The QUIET BASEBALL BAKED the AUNT.

06x24 The WRITER LIT the FAVOURITE BEACHES.

06x25 My PIG CHASED the AWARD on the CAT.

List 7:

07x01 The WORD SITS OFTEN in the EGG.

07x02 Her APPRECIATED GARDEN MADE the NEWS.

07x03 Her FRIED STORIES were SENT YESTERDAY.

07x04 The COUNTRY SMELLED the LAST MONEY.

07x05 The CLEAR COMMUNITY ENJOYED the MAN.

07x06 A PROUD CHILD is SAD to COOK.

07x07 The BOOK is CALM and ALWAYS PEACEFUL.

07x08 The LIVING LOVE INSPIRED the BUTTER.

07x09 His SAUCE ATE LAZY FLOWERS.

07x10 The STARVING CROWD was KIND and FOREIGN.

07x11 That YEAR EARNED a LARGE CHEF.

07x12 The CHURCH FOOD is VERY FRESH.

07x13 The STRONG DOG READS for their STRANGER.

07x14 The THOUGHTFUL PETS were VISITED in the GROUP.

07x15 The BOYFRIEND was on the SWEATER FROM the BEACH.

07x16 The GROUND CHEERED the CROWDED ACTOR.

07x17 The TEAM RAN the DANCER and the TV.

07x18 The GREAT MORNING TASTES the DANCING.

07x19 The WATER SLEEPS in the BLACK DRIVER.

07x20 The CLEAN DOLLARS have a BRAVE NEIGHBOUR.

07x21 The FIGHTER and the ROOM TELL the MEAT.

07x22 The INTERESTING PEOPLE COST FUNNY.

07x23 The DRUMS LOOKED BORING with this SINGER.

07x24 The PARENTS NEED a DANCING FAN.

07x25 The NOODLES TOOK the NOVEL DOWN.

List 8:

08x01 The BROTHER PRACTICES NEAR the MAN.

08x02 The GRAPE COUPLE was NEVER STRANGE.

08x03 The CROWDED TOY LOOKS STEAMED.

08x04 The GIRL NEEDS a BIRTHDAY JUICE.

08x05 The SERIOUS SISTER FOUND her NIGHT.

08x06 Her GAME was AROUND the ENTIRE CABBAGE.

08x07 The DUCK TEAM LOST the DOG.

08x08 The SUN LOSES its BLUE GRANDFATHER.

08x09 The SCARED HOLIDAY was my OLD PET.

08x10 The PRIVATE APARTMENT is DEEP and SIMPLE.

08x11 The DARK FIGHT was HAPPY to SLEEP.

08x12 They ATE the CLOSEST HARD GIFT.

08x13 The LADY'S HOPES SHINED on the WORKERS.

08x14 The PRETTY CAR is BRIGHTLY KIND.

08x15 The BEST PRIZE had a NEW SONG.

08x16 The CHICKEN CLOSES a BETTER SCHOOL.

08x17 They WON the WOODEN MORNING QUICKLY.

08x18 Our FLOOR SINGS EVERY EDUCATION.

08x19 The BEAUTIFUL BEACH BOUGHT a RESTAURANT.

08x20 Our SONG is TERRIBLE and REALLY LATE.

08x21 The SUN TASTED SOFTLY in JUNE.

08x22 The LADY SINGS a BEAUTIFUL DOOR.

08x23 That HOUSE EXPLANATION SPILLED the TV.

08x24 My FRIENDS ENTERTAIN until THREE OCEANS.

08x25 The DANGEROUS JUNE TASTED SAD.

List 9:

09x01 The LAWYER TRAVELLED for DANGEROUS FRUITS.

09x02 The SUMMER EGG was EARLY YESTERDAY.

09x03 His CLASS ATTRACTS LONELY HATS.

09x04 Their RICH SCIENCE LOVES TOGETHER.

09x05 The RELAXING HOUSES HATE her LADY.

09x06 She PLAYED an INTERESTING and FAST ROOM.

09x07 Her BEST FOOD OWNS STORIES.

09x08 The SILVER SON ENJOYED the SALT.

09x09 That STRAIGHT DRINK is SCARY but BAD.

09x10 The RAIN DROVE ALWAYS in his CHILD.

09x11 The AUDIENCE CRIED about the CITY GIRLFRIEND.

09x12 The UNCLE TASTED the BLACK PARADE.

09x13 The PERFORMER FIRED the BOY and the VACATION.

09x14 The CHINESE CHILDREN BOUGHT WRITERS.

09x15 The PRETTY WORKER TELLS many ANIMALS.

09x16 The ADULT WATCHED the COLOURFUL CROWD.

09x17 The SONG DESTROYED the YOUNGEST BOSS.

09x18 The TALENTED BUS SEEMED CUTE.

09x19 An EARTH STATION is VERY ANGRY.

09x20 Some MUSEUMS CALLED the TINY ORANGE.

09x21 SOME RABBITS were RUDE and FAMOUS.

09x22 The PLANTS STUDIED the BEST BRUSHES.

09x23 The POTATOES WENT WELL SOFTLY.

09x24 The FIVE HOURS LEARNED CHEAP.

09x25 She FINISHED the CAR SMALL and FAR.

List 10:

10x01 The ART LAKE FELT my MOTHER.

10x02 These CHEFS LAY SOFT GROUPS.

10x03 The BOILED COURSES SEEMED ANGRY.

10x04 The PASTA GAVE a CHRISTMAS STATION.

10x05 His EXAM and ATTENTION TROUBLED the STRANGER.

10x06 The MEAN CAMPERS HELPED the KIDS.

10x07 The DUCK ARRIVED NICE and BORING.

10x08 MANY CATS READ the ONION.

10x09 The bad STUDENT PLAYED the FINAL FRIEND.

10x10 Her ORANGE BRUSHES were UNHAPPY and YOUNG.

10x11 The NOISY FISH was VISITED too DELICIOUS.

10x12 The BIRD DRANK a THIRSTY CLASS.

10x13 The THREE COLLEGES COOKED for her GRANDMOTHER.

10x14 The FOOD CUTS the GIRL every DAY.

10x15 The THIN ROOM SELLS LONELY.

10x16 The LAST MUSEUM ATTRACTS the MARKET.

10x17 The MILK INTERESTED the EMPTY EGGS.

10x18 The NEWSPAPER TOOK TINY SONS.

10x19 The BAG SWIMS with its RUDE TEACHER.

10x20 The BEST BOOKS CRIED in the BEAR.

10x21 My GRANDMA SCARED the TRAVELLERS QUICKLY.

10x22 The TIRED GRANDPA SHARED the INSTRUCTOR.

10x23 My PAINTER NEVER OFFERS the BOYFRIEND.

10x24 The BIG CHILD is SLOWLY STRANGE.

10x25 The CLOTHES SMELLED the SISTERS to her SHOW.

List 11:

11x01 The KITTEN LEARNED in a TROUBLED CARD.

11x02 The WINDOW NEEDS a CHICKEN JUICE.

11x03 The BROWN RABBIT HATES EASILY.

11x04 That UNFAIR SPANISH is ALWAYS BETTER.

11x05 The THREE ROOMS did their GROCERY BIRDS.

11x06 The DARK PERSON SELLS the WEDDING.

11x07 The ORANGE HOMEWORK GAVE the APARTMENT.

11x08 The MINUTES were SENT in the RELAXING MOVIE.

11x09 The LITTLE GLOVES were a BLACK CAKE.

11x10 The STRONG MILK was LOUD and SMALL.

11x11 The TASTY STORE was PLANNED GREEN.

11x12 The APPLES BUY a DARK GRANDMOTHER.

11x13 The COOL PARK STARTS on a SON.

11x14 Our MOTHER FEELS MORE CUSTOMERS.

11x15 Our SOUP STOLE the SICK FENCE.

11x16 The MATH RADIO PRACTICED the BEDROOM.

11x17 The FOOD COVERED the FAST BOYFRIEND.

11x18 They CLIMBED LATE TESTS on the CLOUD.

11x19 That COMFORTABLE GIRLFRIEND was LOUDLY ENGLISH.

11x20 The FARM and MEAL DRINK their SKY.

11x21 The SNAKE LIVES HERE in the GRADE.

11x22 The WHITE WIND OPENED the DINNER.

11x23 His SNOWMAN STARTED a TEA DAILY.

11x24 TWENTY PROFESSORS SCREAMED over the COUSINS.

11x25 The NIGHT had a DIFFICULT BIG SISTER.

List 12:

12x01 The DOCTOR RAN the CARROTS QUICKLY.

12x02 The CITY PATIENT had a CURLY MOTHER.

12x03 This DRY PICTURE BOILED STRONGLY.

12x04 The KITCHEN CHEWED the LOVING AIR.

12x05 Those HONEST YEARS TASTE the PLANTS.

12x06 The BUNNY LAUGHS for DARK ROADS.

12x07 The YOUNGEST STREET FLEW the CHILDREN.

12x08 My DINNER and SHOPPER FAILED the BUSINESS.

12x09 The BEDROOM CUSTOMER ALWAYS GROWS.

12x10 The UPSET HILL DRIED the SON.

12x11 The FAT LEAVES were HOT and BUSY.

12x12 Her SCHOOL STORY was QUIETLY HAPPY.

12x13 My BUSY HAIR CAME in the SKY.

12x14 The STAIN HELPED through the ANGRY BALL.

12x15 The ANSWER TOOK the INTERNATIONAL RATS.

12x16 The TERRIBLE MAP SHOWS GIFTS.

12x17 She FINISHED and RECEIVED UNHAPPY KEYS.

12x18 The WIND BOUGHT the BLUE CHEF.

12x19 The EXPENSIVE BROTHER is NICE and YELLOW.

12x20 The LOUD ANIMALS PICKED the UNCLE.

12x21 Her FLOWERS ROLLED DOWN for the GIRL.

12x22 They COOLED a CUTE BAG for every SHOE.

12x23 The POTATOES WASHED the SICK RUG.

12x24 The AIRPLANE EATS in the GREAT FARMER.

12x25 The GOAT LOVED THINGS in the CARD.

List 13:

13x01 The FLIES OPENED EARLY the PETS.

13x02 Our MOVIES PRACTICED UNHAPPY HOURS.

13x03 Her CROWDED NOISE WAS EXPENSIVE.

13x04 The CHILDREN MADE a TRAIN KEY.

13x05 Their TEAM LISTENED BAD and GREEN.

13x06 The TERRIBLE CAMPUS SELLS the SHOW.

13x07 Her PARTY BOILED MUSEUMS with another STATION.

13x08 The ROMANTIC PASTA VISITED the PROJECT.

13x09 The LONELY and BROWN TRUTH was RUINED.

13x10 Their YOUNGEST PROBLEM was GOOD and FRIENDLY.

13x11 The ARTIST LOOKED on the OLD WATER.

13x12 The HOT VOICE is AROUND SPORTS.

13x13 The STORES ASSIGNED for MANY BOYFRIENDS.

13x14 The UNDERGROUND DOG SITS at the CLOTHES.

13x15 Those FLAGS ATTRACT the METAL PERSON.

13x16 Her DIRTY CHAIR SOUNDS DIFFICULT.

13x17 The KITTEN TELLS a FIRST SALT.

13x18 The PROUD PANTS are FUN and HIGH.

13x19 The VEGETABLES LOVE in the FAVOURITE MACHINE.

13x20 The DOOR WALKED LOST and SAD.

13x21 The FAMILY ENDED ALWAYS TODAY.

13x22 A TENNIS FRIEND FLIES the PEPPER.

13x23 The GIRL WATCHED the WOODEN BOSS.

13x24 This STUDENT NEEDS MEAL and GARBAGE.

13x25 The MUSIC LASTED THREE COLLEGES.

List 14:

14x01 The MONKEY BOUGHT RABBITS TODAY.

14x02 They WATCHED PARENTS at the THIN POLICE.

14x03 The DRY SISTERS PICKED the MAP.

14x04 My STORE EXPECTED a FOREIGN DRINK.

14x05 The FRIED CLOTHING had MANY PICTURES.

14x06 The GIRL SOLD in the LITTLE AUTHOR.

14x07 The GREEN FAMILY PLAYED the CHICKEN.

14x08 The JOKES STUDIED FOUR DAUGHTERS.

14x09 The TIRED TWIN NOVELS are CLEAN.

14x10 Their YOUNGEST STREET is SLOW and HEALTHY.

14x11 The TREE WROTE a NEW COUSIN.

14x12 That DIRTY HILL was LONELY YESTERDAY.

14x13 The MARKET SWINGS COUPLES and SPORTS.

14x14 The FISH MOVIE LIVES of her DINNER.

14x15 The LIFE SERVES BIRTHDAY BOOKS.

14x16 The SMALL CAKE was SUDDENLY PEACEFUL.

14x17 The CHILD DREAMED a LONG HOME.

14x18 The LARGE MAN STOPPED from the GRANDPARENTS.

14x19 The HAPPY PERSON LOOKED the MEMBERS.

14x20 My PROBLEMS and VISITORS TRUST the DRIVER.

14x21 The BIRTHDAY ATE her DOG EARLY.

14x22 The HORRIBLE COMPUTER CHASED the HOUSE.

14x23 The PLANTS BAKED a TALL STUDENT.

14x24 The QUIET DOOR was SAD and HELPFUL.

14x25 The RESTAURANT TELLS QUICKLY in the ROOM.

List 15:

15x01 The TEACHER GAVE her SUN TODAY.

15x02 That LONG HELP ROASTED the MEAT.

15x03 The FOOTBALL HOUSE ENDED SOFTLY.

15x04 They CHASED the FOUR HOLIDAYS and SHIRTS.

15x05 The FAST PARK LIKES the LAWYER.

15x06 The DOGS VISITED the GREEN CLOTHES.

15x07 The VEGETABLE HUSBAND SMOKED the THEATRE.

15x08 That MOUSE SANG a STRESSFUL VACATION.

15x09 The FARM DESERT was a TIRED AUNT.

15x10 The BIRTHDAY TABLE was FRENCH and OLD.

15x11 The RESTAURANT WATCHED in the DIRTY CHOCOLATE.

15x12 The HEALTHY SUN is ALWAYS LITTLE.

15x13 The NICE TIME WENT at the LUNCH.

15x14 The PRIVATE ACTOR was CARRIED in the SNAKE.

15x15 The GIANT HOLES were SALTY and BROWN.

15x16 Their FRIENDS CELEBRATED the SMALL SOUP.

15x17 The HOMEWORK BUYS their WARM WEEK.

15x18 The UNHAPPY TOWN SHINES off the WAITRESS.

15x19 The KIDS LOOKED LOUDLY in the CHILD.

15x20 The ITALIAN QUEEN NEEDS the OFFICE.

15x21 The BOY COMPLETED the KITTEN EARLY.

15x22 The PREGNANT BAGS DIG the GAMES.

15x23 They COMPLETED the ANGRY FAMILY in the BAR.
15x24 Those SIX PANS are REALLY QUIET.
15x25 The FOOD LOVED the CHEAP GIFTS.

List 16:

16x01 The BANANAS SANG for TROPICAL MONEY.
16x02 Her COUSINS CELEBRATED LAZY DISHES.
16x03 The USEFUL VEGETABLE SEEMED EXCITED.
16x04 The TALKING WEATHER LOVED around the PICTURE.
16x05 The SMALL AUNT BELIEVES the MONKEY.
16x06 The FOREST LOOKED the FLOWERS WEEKLY.
16x07 The GLASS WAITER CHEERED the RATS.
16x08 The SISTERS GREW a STILL TEAM.
16x09 The TOY PLATES have CROWDED HOMES.
16x10 He WORKED SUDDENLY in the CHERRY PRIEST.
16x11 The WARM CARS were LITTLE and SWEET.
16x12 The TOY LUNCH was ALWAYS QUIET.
16x13 The CLEAN ROOM SCREAMED for their SONGS.
16x14 The BUSINESS DESTROYED the HOT SON.
16x15 The MATH GARDENER BAKES in the OCEAN.
16x16 The KITCHEN WANTS the FRIENDLY TREE.
16x17 The MANY PIES BROKE LOUDLY.
16x18 The COLOURFUL COUPLE was BAD and BEAUTIFUL.
16x19 The WEATHER NEEDS CALM CHILDREN.
16x20 The CATHOLIC WORLD took a STORMY CLASS.
16x21 My NEWS are SICK and VERY YELLOW.
16x22 The SUPPORT HIKED QUICKLY the SALAD.
16x23 The TEN PERFORMERS had RELIGIOUS GLASSES.
16x24 OUR GOD ate PATIENTS PERFECTLY.
16x25 The BIG FISH TRAVELLED the PEPPERS.

List 17:

17x01 The SHOW GREW ACROSS the PIG.
17x02 The GROUP HEARD HOT COMPANIES.
17x03 The DIFFERENT MUSIC CAME LAZY.
17x04 The PICTURE FOUND a TASTY CLASS.
17x05 The CHICKEN PATIENT TASTED the SUN.

17x06 The TEN LESSONS are DIVORCED and TALENTED.

17x07 The TWO WOMEN BROKE the SOUP.

17x08 The TELEVISION GARDEN READS the HOLE.

17x09 The DEEP FAT TEACHERS are STRANGE.

17x10 The SHORT MOUSE was FOREIGN and BROKEN.

17x11 The COUPLE HIKED MANY SCHOOLS.

17x12 The YOUNG NURSES are MORE POPULAR.

17x13 MANY DAUGHTERS WALKED at the STAGE.

17x14 That SUPERMARKET HELPS the FUNNY ACTOR.

17x15 The SALAD SEEMED a HUGE GOAT.

17x16 The BLUE ARTIST DREW COMPLETELY.

17x17 She WRITES PINK ENGLISH HOTELS.

17x18 The VEGETABLES OPEN a DIFFICULT HAT.

17x19 The PRESIDENT WARMED in the DEDICATED GIFTS.

17x20 The CHEESE and POOL GAVE the BREAD.

17x21 The UGLY BOOKS SAW the FATHER.

17x22 The SMART MOUNTAIN FEELS with the MOVIE.

17x23 The BAKED CARS were TOO GREEN.

17x24 The SLOW LANGUAGES BUY the TABLE.

17x25 The SANDWICH SPEAKS VERY SWEET.

List 18:

18x01 The FAR BUTTONS are FOREIGN and RAINY.

18x02 The AWARD ENDED the BEST FARMER.

18x03 The HAPPY LADY TASTED CROWDED.

18x04 Our BREAKFAST ASSIGNS in my FAR SHIRT.

18x05 Our ONLY WAITER made THIRTY TEACHERS.

18x06 The CAKE CUTS the ENGLISH DOCTOR.

18x07 A PINK MILK is TOO HARDWORKING.

18x08 The WARM MEAL DRINKS the CLUB.

18x09 Her LOUD PROFESSOR and LIGHTNING were BUSY.

18x10 The HOTEL SAVES an ORANGE CORN.

18x11 The TWO AUTHORS never LASTED the RAIN.

18x12 The CURIOUS HUSBAND WROTE the VISITOR.

18x13 The EXPENSIVE WEEKS JOINED for the MARKET.

18x14 The PHONE TRAVELLED the FAITHFUL MEAT.

18x15 The OLD HELP PRACTICES the ADVICE.

18x16 The APPLE WON the SICK CITY.

18x17 Our MANY ARMS GIVE the DESERT.

18x18 The NEIGHBOUR FELL in the SPICY PLAYER.

18x19 The FLOOR was RIGHT and ALWAYS GREAT.

18x20 The KIND SOCCER BAND had no CATS.

18x21 The ITALIAN GARAGE ASKS for the DAUGHTER.

18x22 The LIVES SELL PERFECT BOOKS.

18x23 The WOOD KEYS SERVED the BUTCHER.

18x24 The COFFEE OPENED on her AMAZING NOISE.

18x25 The BOX GAME was a DIRTY LEG.

List 19:

19x01 That CHEAP STUDENT is HARD and CONFUSING.

19x02 The RIVER EATS THEIR QUESTIONS.

19x03 The EXPENSIVE MUSIC WALKED FIRST.

19x04 The GAME LOST the SAME STORY.

19x05 The RICE BED JOINED OVER.

19x06 The OFFICE GIFT SANG her JEANS.

19x07 The PUPPY FITS OFTEN on the BROTHER.

19x08 A RED ARTIST TELLS the SONGS.

19x09 The BORED WOMAN and the APPLE were DRY.

19x10 The SHY CHAIRS were NEW and INTERESTING.

19x11 Their PLATES MADE the SOFT GRANDFATHER.

19x12 The TALL NEWSPAPER is OFTEN HUNGRY.

19x13 A WOODEN CUP went PEACEFULLY on the carrots.

19x14 The PRETTY PHONE PAID the BOAT.

19x15 My SEA RANG the QUIET BIRDS.

19x16 The OFFICE PLASTIC FAMILY is not FUN.

19x17 My SINGER LOOKED the CROWDED QUESTIONS.

19x18 The FRIENDLY RING LISTENED EVERYONE.

19x19 Our ITALIAN HUGS for the DAILY COUPLE.

19x20 The FRENCH and FINGER SOUNDED GOLD.

19x21 The BUSINESS STUDIES only EVERY NEIGHBOUR.

19x22 The JAZZ BABY SLEPT WELL.

19x23 The CHILDREN RECEIVED DOWN the RABBIT.

19x24 The NURSES FLEW the TWINS AWAY.

19x25 The LAKE GREW MOTHERS and PASTA.

List 20:

20x01 My KNIFE EASILY UPSETS the MILK.

20x02 The AUNT CLEANED the SOCCER RULES.

20x03 The RICH NANNY RUNS the MOVIE.

20x04 The COLD WORKER is DEEP and HORRIBLE.

20x05 The YELLOW PEOPLE RAISED the FOREST.

20x06 Our DAUGHTER PLAYS the PIG EASILY.

20x07 The OLDEST RABBIT DRANK new WARS.

20x08 The LAST TOOLS were FUNNY and LITTLE.

20x09 The DAY KNOWS the BORING UNCLE.

20x10 That TEAM GRAPE is TOO SHY.

20x11 A LEFT TOY CUTS through the WATER.

20x12 The MUSIC SPEAKS its RED HOLIDAY.

20x13 His LAKE was DIFFICULT and QUITE CRAZY.

20x14 The SPEECH CHOOSES the LEAVES in the ARTIST.

20x15 The HOUSE DREW on the BIG GOALS.

20x16 The GUESTS CHANGE the HOLE TOGETHER.

20x17 He ENDS the FALL ABOUT the STEAK.

20x18 The FUN BALL KICKED the COLOUR.

20x19 The SCARED PLAYER STAYED QUIETLY.

20x20 My CHILDREN SCORED the LONG DENTIST.

20x21 The FIVE TEACHERS HURT my MOUSE.

20x22 Their MEAN HAND CELEBRATED with QUESTIONS.

20x23 The HELPFUL FAMILIES will BREAK SOON.

20x24 The SOFT NEWS PLEASED the GRANDPARENTS.

20x25 The TASTY CLOTHES BOUGHT the BIRD.

Appendix D

The following sentences were presented as experimental stimuli in Study 3. They were adapted from the Connected Speech Test developed by Cox et al. (1987).

CODE	TOPIC	SENTENCE
01x01	windows	WINDOWS PROVIDE LIGHT and air to ROOMS.
01x02	windows	WINDOWS were once COVERED with CRUDE SHUTTERS.
01x03	windows	later, OILED PAPER was USED for WINDOWPANES.
01x04	windows	GLASS WINDOWS first APPEARED in ancient ROME.
01x05	windows	COLORED GLASS was used in EUROPEAN WINDOWS.
01x06	windows	the DESIGN of a WINDOW may be SIMPLE or DECORATIVE.
01x07	windows	some CHURCHES were FAMOUS for their BEAUTIFUL WINDOWS.
01x08	windows	these WINDOWS DISPLAYED PICTURES from the BIBLE.
01x09	windows	PIECES of GLASS were HELD together by LEAD.
01x10	windows	SUCH WINDOWS may be seen in FRENCH CATHEDRALS.
01x11	windows	ENGLISH CHURCHES also have STAINED glass WINDOWS.
01x12	windows	the LIGHT through the WINDOW creates BEAUTIFUL REFLECTIONS.
02x01	gloves	GLOVES are CLOTHING WORN on the HANDS.
02x02	gloves	the WORD GLOVE means PALM of the HAND.
02x03	gloves	CRUDE GLOVES were worn by PRIMITIVE MAN.
02x04	gloves	GARDENERS wear WORKING GLOVES to PROTECT their hands.
02x05	gloves	the ROMANS used GLOVES as a SIGN of RANK.
02x06	gloves	KNIGHTS used to FASTEN GLOVES to their HELMETS.
02x07	gloves	the GLOVES SHOWED their DEVOTION to their LADIES.
02x08	gloves	a GLOVE thrown on the GROUND SIGNALED a CHALLENGE.
02x09	gloves	KNIGHTS THREW them at their ENEMY's FEET.
02x10	gloves	FIGHTING STARTED when the ENEMY picked up the GLOVE.
02x11	gloves	DISPOSABLE GLOVES are used to PREVENT INFECTION.
02x12	gloves	BASEBALL players may wear GLOVES to CATCH the BALL.
03x01	umbrellas	PEOPLE use UMBRELLAS when it RAINS OUTSIDE.

03x02	umbrellas	UMBRELLAS were first USED in ANCIENT EGYPT.
03x03	umbrellas	they GAVE PROTECTION from the FIERCE SUNSHINE.
03x04	umbrellas	the WORD UMBRELLA actually means SMALL shadow.
03x05	umbrellas	SLAVES held UMBRELLAS OVER their MASTERS.
03x06	umbrellas	in early ROME, ONLY WOMEN used UMBRELLAS.
03x07	umbrellas	if a MAN did, HE was CONSIDERED WEAK.
03x08	umbrellas	UMBRELLAS were USED by both SEXES in ENGLAND.
03x09	umbrellas	UMBRELLAS used as SUNSHADES are CALLED PARASOLS.
03x10	umbrellas	the UMBRELLA is used on the WEATHER CHANNEL to indicate RAIN.
03x11	umbrellas	UMBRELLAS are used as LIGHT REFLECTOR in PHOTOGRAPHY.
03x12	umbrellas	most UMBRELLAS WORLDWIDE are MADE in CHINA.
04x01	giraffes	the GIRAFFE is the TALLEST WILD ANIMAL.
04x02	giraffes	it is THREE TIMES TALLER than a MAN.
04x03	giraffes	an ADULT GIRAFFE is EIGHTEEN feet HIGH.
04x04	giraffes	the GIRAFFE has an EXTREMELY LONG NECK.
04x05	giraffes	GIRAFFE have just SEVEN NECK BONES.
04x06	giraffes	the GIRAFFE 's BODY is about the SIZE of a HORSE 's.
04x07	giraffes	the BODY is SHAPED LIKE a TRIANGLE.
04x08	giraffes	AFRICA is the only COUNTRY where GIRAFFES live WILD.
04x09	giraffes	LARGE GROUPS of them are FOUND ON the PLAINS.
04x10	giraffes	THEY LIVE there with LIONS and ELEPHANTS.
04x11	giraffes	their FOOD source is LEAVES, FRUITS and FLOWERS.
04x12	giraffes	there are SEVEN EXTINCT SPECIES of GIRAFFE.
05x01	doves	a DOVE is a SMALL, TRIM BIRD.
05x02	doves	the BEST KNOWN is the MOURNING DOVE.
05x03	doves	the MOURNING DOVE lives in NORTH AMERICA.
05x04	doves	its NAME comes from its SAD MATING CALL.
05x05	doves	it is SOMETIMES INCORRECTLY CALLED TURTLEDOVE.
05x06	doves	the MOURNING DOVE is about a FOOT LONG.
05x07	doves	its BODY is BROWN with GRAY WINGS.
05x08	doves	it FEEDS on GRAINS, GRASSES and WEEDS.
05x09	doves	the MOURNING DOVE is a CARELESS HOUSEKEEPER.
05x10	doves	its NEST is just some STICKS TOSSED TOGETHER.

05x11	doves	they LAY ONE or TWO EGGS at a time.
05x12	doves	DOVES ADAPTED to most of the HABITATS on the PLANET.
06x01	carrot	a CARROT is a VEGETABLE RELATED to PARSLEY.
06x02	carrot	the PLANT of a CARROT probably ORIGINATED in PERSIA.
06x03	carrot	the long STEM of a CARROT GROWS UNDERGROUND.
06x04	carrot	it is this STEM that MOST PEOPLE EAT.
06x05	carrot	the LEAVES of the CARROT are ALSO EATEN.
06x06	carrot	they are OFTEN USED to FLAVOR FOODS.
06x07	carrot	the ROOTS CONTAIN high AMOUNTS of VITAMINS.
06x08	carrot	SPRING CROPS are GROWN in the western STATES.
06x09	carrot	the CROP is HARVESTED in one HUNDRED DAYS.
06x10	carrot	fall CROPS are GROWN in the NORTHERN STATES.
06x11	carrot	WINTER HARVESTS USUALLY come from CALIFORNIA.
06x12	carrot	CARROTS can be STORED for SEVERAL MONTHS.
07x01	grass	GRASS can GROW in ALL CLIMATES.
07x02	grass	GRASSES is not FOUND on the CONTINENT of ANTARCTICA.
07x03	grass	THERE are MANY FORMS of GRASSES.
07x04	grass	GRASSES may be ANNUAL or PERENNIAL HERBS.
07x05	grass	many GRASSES are IMPORTANT FOOD SOURCES.
07x06	grass	some GRASSES GROW higher than a MAN'S HEAD.
07x07	grass	among THESE are BAMBOO and SUGAR CANE.
07x08	grass	other TYPES are ONLY a few INCHES TALL.
07x09	grass	SOME GRASSES are as SLENDER as THREADS.
07x10	grass	OTHERS are STIFF enough to WITHSTAND a heavy SNOW.
07x11	grass	MOST GRASSES are FLOWERING PLANTS.
07x12	grass	these FLOWERS BLOOM MAINLY in the SPRING.
08x01	nails	NAILS are used to FASTEN WOOD TOGETHER.
08x02	nails	PIONEERS used WOODEN PEGS instead of NAILS.
08x03	nails	NAILS are typically DRIVEN into the WOOD by a HAMMER.
08x04	nails	one END of a NAIL is VERY SHARP.
08x05	nails	the POINT CREATES an OPENING for the NAIL.
08x06	nails	it also HELPS KEEP the WOOD from SPLITTING.
08x07	nails	at the NAIL's OTHER END is a HEAD.
08x08	nails	it PROVIDES a STRIKING SURFACE for the HAMMER.
08x09	nails	it also COVERS the NAIL HOLE in the WOOD.

08x10	nails	NAILS are made in a GREAT VARIETY of FORMS.
08x11	nails	there is a SPECIAL NAIL for EVERY PURPOSE.
08x12	nails	for MOST PURPOSES a ROUND NAIL will do.
09x01	woodpeckers	the WOODPECKER is a BIRD with a STRONG BEAK.
09x02	woodpeckers	it BORES HOLES in TREES looking for INSECTS.
09x03	woodpeckers	other BIRD and MAMMAL SPECIES also use those CAVITIES.
09x04	woodpeckers	WOODPECKERS LIVE in all PARTS of the WORLD.
09x05	woodpeckers	the MAJORITY of WOODPECKERS live SOLITARY LIVES.
09x06	woodpeckers	the TOES of WOODPECKERs are VERY UNUSUAL.
09x07	woodpeckers	two POINT FORWARD and two FACE BACKWARD.
09x08	woodpeckers	this ALLOWS the BIRD to CLING to TREES.
09x09	woodpeckers	the TAIL FEATHERS of a WOODPECKER are STIFF.
09x10	woodpeckers	THEY can USE their TAILS as a SUPPORT.
09x11	woodpeckers	they also USE their TAILS to GRASP TREES.
09x12	woodpeckers	WOODPECKERS HAVE long TONGUES with pointed TIPS.
10x01	owls	OWLS HUNT at NIGHT for their FOOD.
10x02	owls	they are BIRDS of PREY, LIKE EAGLES.
10x03	owls	these BIRDS KILL and EAT small ANIMALS.
10x04	owls	OWLS DEFEND our GARDENS by eating MICE.
10x05	owls	their VISION is GOOD even at FAR DISTANCES.
10x06	owls	different SPECIES of OWLS PRODUCE different SOUNDS.
10x07	owls	they are CLOSELY RELATED to NIGHT HAWKS.
10x08	owls	there are FIVE HUNDRED different KINDS of OWLS.
10x09	owls	they LIVE in both COLD and TROPICAL CLIMATES.
10x10	owls	OWLS usually LIVE ALONE in the FOREST.
10x11	owls	sometimes they LIVE on REMOTE SEA ISLANDS.
10x12	owls	OWLS are KNOWN for their SOLEMN EXPRESSION.
11x01	vegetables	VEGETABLES come from the LEAVES and FLOWERS of PLANTS.
11x02	vegetables	some VEGETABLES come FROM a PLANT's ROOTS.
11x03	vegetables	VEGETABLES play an important ROLE in HUMAN NUTRITION.
11x04	vegetables	the WORD VEGETABLE has SEVERAL MEANINGS.
11x05	vegetables	it is USED in the PHRASE VEGETABLE KINGDOM.
11x06	vegetables	this REFERS to the ENTIRE PLANT WORLD.

11x07	vegetables	the WORD VEGETABLE derives from the LATIN WORD for GROWING.
11x08	vegetables	SOME WILD VEGETABLES can be EATEN.
11x09	vegetables	VEGETABLES can be EATEN RAW or COOKED.
11x10	vegetables	the BEST way to COOK VEGETABLES is by STEAMING.
11x11	vegetables	they are usually CHOPPED or MASHED BEFORE EATEN.
11x12	vegetables	VEGETABLES are VERY DIFFERENT from FRUITS.
12x01	lemons	a LEMON is a YELLOW CITRUS FRUIT.
12x02	lemons	it GROWS in SOUTHERN CALIFORNIA and FLORIDA.
12x03	lemons	LEMON TREES have SPREADING BRANCHES.
12x04	lemons	they have GREEN LEAVES and LARGE FLOWERS.
12x05	lemons	the FLOWERS are WHITE, with PURPLE UNDERNEATH.
12x06	lemons	the LEMON FLOWER SMELLS SWEET.
12x07	lemons	some TYPES of LEMONS have NO SEEDS.
12x08	lemons	OTHER TYPES have MANY SEEDS.
12x09	lemons	their FRUIT is a SPECIAL TYPE of CITRUS.
12x10	lemons	its JUICE it's USED for CLEANING and COOKING.
12x11	lemons	it USUALLY has a VERY SOUR TASTE.
12x12	lemons	the PULP and ZEST are used in COOKING and BAKING.
13x01	violins	the VIOLIN is a POPULAR STRINGED INSTRUMENT.
13x02	violins	EARLY VIOLINS did not PRODUCE clear TONES.
13x03	violins	these VIOLINS SOUNDED VERY ROUGH.
13x04	violins	later VIOLIN MAKERS IMPROVED their CRAFT.
13x05	violins	their VIOLINS were EXTREMELY WELL MADE.
13x06	violins	the VIOLIN BECAME an INSTRUMENT for beautiful MUSIC.
13x07	violins	only SMALL CHANGES have occurred in VIOLIN DESIGN.
13x08	violins	VIOLINS must be MADE with GREAT CARE.
13x09	violins	a PERSON who MAKES VIOLINS is called a LUTHIER.
13x10	violins	PARTS of a VIOLIN are made from DIFFERENT types of WOOD.
13x11	violins	the WOOD USED greatly INFLUENCES the TONE.
13x12	violins	the PARTS must be GLUED TOGETHER by HAND.
14x01	wheat	WHEAT is a MAJOR SOURCE of FOOD.
14x02	wheat	it's a GRASS WIDELY CULTIVATED for its SEED.
14x03	wheat	MILLIONS of PEOPLE depend on WHEAT PRODUCTS.
14x04	wheat	it is the MOST WIDELY used HUMAN FOOD.

14x05	wheat	WHEAT is GROWN on the PLAINS of the united STATES.
14x06	wheat	MORE WHEAT is PRODUCED there than RICE.
14x07	wheat	HOWEVER, RICE is CHEAPER to PRODUCE.
14x08	wheat	it CAN be PLANTED and HARVESTED by HAND.
14x09	wheat	RAW WHEAT can be GROUND into FLOUR.
14x10	wheat	WHEAT is a SOURCE of ESSENTIAL NUTRIENTS.
14x11	wheat	the most COMMON forms of WHEAT are WHITE and RED.
14x12	wheat	IRELAND was the FIRST PRODUCER of WHEAT.
15x01	ice	ICE forms when WATER reaches the FREEZING POINT.
15x02	ice	lower TEMPERATURES are needed to FREEZE IMPURE WATER.
15x03	ice	on the EARTH 's SURFACE ice is abundant in the POLAR REGION.
15x04	ice	SNOWFLAKES and FROST are FORMS of ICE.
15x05	ice	LARGE BODIES of water FREEZE very SLOWLY.
15x06	ice	MOVING WATER takes even LONGER to FREEZE.
15x07	ice	it takes DAYS for ICE to FORM on a LAKE.
15x08	ice	it TAKES WEEKS for RIVERS to FREEZE.
15x09	ice	ICE can also FORM on ROADS and SIDEWALKS.
15x10	ice	this can MAKE TRAVELING VERY DANGEROUS.
15x11	ice	ICE was once USED to COOL REFRIGERATORS.
15x12	ice	ICE has an IMPORTANT role in the earth's WATER CYCLE.
16x01	donkeys	DONKEYS are SMALLER, STRONGER relatives of HORSES.
16x02	donkeys	it is FOUR FEET HIGH at the SHOULDERS.
16x03	donkeys	the DONKEY 's COAT is GRAY and BLACK.
16x04	donkeys	it has a DARK LINE ALONG its BACK.
16x05	donkeys	THIS ANIMAL is EXTREMELY INTELLIGENT.
16x06	donkeys	SURPRISINGLY, it is ALSO a SWIFT RUNNER.
16x07	donkeys	the WILD DONKEY is SHAPED like a ZEBRA.
16x08	donkeys	MAN has TAMED DONKEYS for his personal USE.
16x09	donkeys	DONKEYS are often USED for HARD LABOUR.
16x10	donkeys	MALE DONKEYS are often used to PRODUCE MULES.
16x11	donkeys	the PREGNANCY of a DONKEY lasts for TWELVE MONTHS.
16x12	donkeys	all DONKEYS are NOTED for their HUGE EARS.
17x01	guitars	the GUITAR is a STRINGED MUSICAL INSTRUMENT.

17x02	guitars	it's MADE of WOOD and has SIX STRINGS.
17x03	guitars	GUITARS are USED to ACCOMPANY SINGING.
17x04	guitars	THEY are PLAYED in GROUPS with other INSTRUMENTS.
17x05	guitars	a POPULAR STYLE of GUITAR has a FLAT top.
17x06	guitars	you TUNE a GUITAR by TWISTING the PEGS.
17x07	guitars	the NECK is HELD with the LEFT HAND.
17x08	guitars	the PLAYERS 's RIGHT HAND pulls the STRINGS.
17x09	guitars	he plays BASS NOTES with his RIGHT THUMB.
17x10	guitars	other NOTES are PLAYED with the first THREE FINGERS.
17x11	guitars	ELECTRIC GUITARS use an AMPLIFIER and a LOUDSPEAKER.
17x12	guitars	the ELECTRIC GUITAR has had a great INFLUENCE on popular CULTURE.
18x01	envelopes	an ENVELOPE is a POUCH CONTAINING a LETTER.
18x02	envelopes	the ADDRESS is WRITTEN ON the OUTSIDE.
18x03	envelopes	ENVELOPES may be used to PROTECT IMPORTANT DOCUMENTS.
18x04	envelopes	each ENVELOPE is a FOLDED SHEET of PAPER.
18x05	envelopes	ONE FLAP is COVERED with GLUE.
18x06	envelopes	the ENVELOPE is GLUED SHUT before MAILING.
18x07	envelopes	self-SEALING ENVELOPES use a SPECIAL GUM.
18x08	envelopes	PAPER ENVELOPES were DEVELOPED in CHINA.
18x09	envelopes	they were FIRST made in EIGHTEEN THIRTY NINE.
18x10	envelopes	BEFORE that time, LETTERS were SIMPLY FOLDED.
18x11	envelopes	ENVELOPS are AVAILABLE for full SIZE DOCUMENTS.
18x12	envelopes	RED ENVELOPS are used for MONETARY GIFTS.
19x01	grasshoppers	GRASSHOPPER refers to TWO TYPES of BUGS.
19x02	grasshoppers	they have LONG, THIN BACK LEGS.
19x03	grasshoppers	GRASSHOPPERS leap THROUGH FIELDS and MEADOWS.
19x04	grasshoppers	they can JUMP many TIMES their own LENGTH.
19x05	grasshoppers	a MAN could NEVER JUMP that FAR.
19x06	grasshoppers	LOCUSTS are a SPECIAL KIND of GRASSHOPPER.
19x07	grasshoppers	the DIFFERENCE BETWEEN the two is THEIR FEELERS.
19x08	grasshoppers	LOCUSTS have much SHORTER FEELERS than GRASSHOPPERS.
19x09	grasshoppers	GRASSHOPPERS are more GREEN in COLOR than LOCUSTS.

19x10	grasshoppers	LOCUSTS are USUALLY BROWN COLORED.
19x11	grasshoppers	they PROTECT THEMSELVES from PREDATORS by CAMOUFLAGE.
19x12	grasshoppers	GRASSHOPPERS have SPECIAL RECEPTORS for PRESSURE.
20x01	lettuce	LETTUCE is a green VEGETABLE with CRISP LEAVES.
20x02	lettuce	it is an ANNUAL PLANT of the DAISY FAMILY.
20x03	lettuce	it is USED to MAKE HEALTHY SALADS.
20x04	lettuce	LETTUCE is also USED for SOUPS and SANDWICHES.
20x05	lettuce	it GROWS in the NORTHERN HALF of the WORLD.
20x06	lettuce	there are LOOSE LEAF and HEAD LETTUCES.
20x07	lettuce	loose leaf LETTUCE is POPULAR in HOME GARDENS.
20x08	lettuce	this VARIETY is found MORE OFTEN in EUROPE.
20x09	lettuce	its LEAVES CURL LOOSELY INSIDE one another.
20x10	lettuce	most LETTUCE GROWN in AMERICA is HEAD lettuce.
20x11	lettuce	its LEAVES FOLD TIGHTLY OVER one another.
20x12	lettuce	the LEAVES FORM a BALL called a HEAD.
21x01	lawn	a LAWN is an AREA PLANTED with GRASS.
21x02	lawn	GREEN, TRIMMED LAWNS are a beautiful SIGHT.
21x03	lawn	people LIKE to PLANT LAWNS around their HOMES.
21x04	lawn	LAWNS are also used for SPORTS and OUTDOOR ACTIVITIES.
21x05	lawn	HOSPITALS OFTEN have LAWNS AROUND them.
21x06	lawn	MOST PUBLIC BUILDINGS have LAWNS.
21x07	lawn	LAWNS HELP to keep SOIL from ERODING.
21x08	lawn	a GOOD LAWN is very THICKLY PLANTED.
21x09	lawn	GRASSES GROW rapidly in the SPRING and AUTUMN.
21x10	lawn	there are FOUR HUNDRED PLANTS per square FOOT.
21x11	lawn	EACH PLANT has several BLADES of GRASS.
21x12	lawn	there are SEVERAL DIFFERENT KINDS of GRASSES.
22x01	cactus	the CACTUS is a PLANT with SHARP SPINES.
22x02	cactus	MANY SPECIES are used as ORNAMENTAL PLANTS.
22x03	cactus	five hundred DIFFERENT KINDS GROW in MEXICO.
22x04	cactus	nearly ALL CACTUS plants LIVE in AMERICA.
22x05	cactus	CACTUS LIVE best where there is LITTLE RAINFALL.
22x06	cactus	MOST CACTI is found in the SOUTHWEST DESERT.
22x07	cactus	PLANTS USUALLY make FOOD in their LEAVES.

22x08	cactus	the CACTUS does NOT have ANY LEAVES.
22x09	cactus	they have DISAPPEARED so the CACTUS can STAY MOIST.
22x10	cactus	the CACTUS STORES the WATER in its STEM.
22x11	cactus	the CACTUS can have ROOTS that SPREAD out WIDELY.
22x12	cactus	DESERT cactus FLOWERS BLOOM in the SPRING.
23x01	cabbage	CABBAGE is the most COMMON GARDEN VEGETABLE.
23x02	cabbage	it has THICK LEAVES that CURL INWARD.
23x03	cabbage	they FORM an EIGHT inch ROUND HEAD.
23x04	cabbage	the WORD CABBAGE is LATIN for "HEAD".
23x05	cabbage	the CABBAGE PLANT can live through SEVERAL FREEZES.
23x06	cabbage	it ALSO GROWS in the HEAT of SUMMER.
23x07	cabbage	early SPRING CABBAGE is PLANTED in GREENHOUSES.
23x08	cabbage	this PROTECTS the YOUNG PLANTS from FROST.
23x09	cabbage	AFTER six WEEKS they are MOVED OUTDOORS.
23x10	cabbage	TRANSLPANTING is done BEFORE the END of SPRING.
23x11	cabbage	CABBAGE heads can be GREEN, PURPLE and WHITE
23x12	cabbage	CABBAGE is a rich SOURCE of VITAMINS C and K.
24x01	gold	GOLD was one of the FIRST KNOWN METALS.
24x02	gold	for many YEARS GOLD has SYMBOLIZED WEALTH.
24x03	gold	even the EARLY CAVE MAN knew about GOLD.
24x04	gold	ancient EGYPTIANS HAMMERED GOLD into LEAVES.
24x05	gold	they USED these LEAVES to DECORATE their TOMBS.
24x06	gold	in the PAST GOLD was used as a MONETARY CURRENCY.
24x07	gold	a SCIENCE GREW up around EFFORTS to make GOLD.
24x08	gold	it STARTED DURING the MIDDLE AGES.
24x09	gold	the ancient SCIENTISTS NEVER ACHIEVED their GOAL.
24x10	gold	MODERN SCIENTISTS have made these DREAMS come TRUE.
24x11	gold	they now MAKE GOLD by a CHEMICAL PROCESS.
24x12	gold	GOLD is RESISTANT to MOST ACIDS.
25x01	weeds	WEEDS are CONSIDERED WORTHLESS PLANTS.
25x02	weeds	the DIFFERENCE between WEEDS and useful PLANTS is UNCLEAR.
25x03	weeds	WHERE a WEED grows DETERMINES its USEFULNESS.
25x04	weeds	OATS GROWING in a CORNFIELD are considered WEEDS.
25x05	weeds	OATS growing in an OATFIELD are USEFUL PLANTS.

25x06	weeds	much CROP DAMAGE is CAUSED by WEEDS.
25x07	weeds	WEED CONTROL is IMPORTANT in AGRICOLTURE.
25x08	weeds	FARMERS SPEND thousands of dollars for WEED SPRAYS.
25x09	weeds	CHEMICALS used to KILL WEEDS can be HARMFUL.
25x10	weeds	these CHEMICALS are sometimes FOUND in DRINKING WATER.
25x11	weeds	many PLANTS known as WEEDS have BENEFICIAL PROPERTIES.
25x12	weeds	WEEDS can be PESTS in a HOME GARDEN.
26x01	chimney	a CHIMNEY CARRIES SMOKE from a FIREPLACE.
26x02	chimney	it ALSO SUPPLIES the FIRE with OXYGEN.
26x03	chimney	WARM AIR is LIGHTER than COLD air.
26x04	chimney	WARM AIR above the FIRE tends to RISE.
26x05	chimney	as the WARM air RISES, COLD air RUSHES in.
26x06	chimney	a DRAFT is CREATED IN the CHIMNEY.
26x07	chimney	the DRAFT PROVIDES the OXYGEN needed for the FIRE.
26x08	chimney	CHIMNEYS must STAND HIGHER than the BUILDING.
26x09	chimney	OTHERWISE, the CHIMNEY will not DRAW PROPERLY.
26x10	chimney	CHIMNEYS can IMPROVE the APPEARANCE of a HOME.
26x11	chimney	CHIMNEYS APPEARED in EUROPE in the TWELFTH CENTURY.
26x12	chimney	ROMANS used TUBES inside the WALLS to draw SMOKE.
27x01	lead	LEAD is a SOFT, HEAVY, METAL.
27x02	lead	it is OFTEN COMBINED with OTHER METALS.
27x03	lead	many USEFUL OBJECTS contain some LEAD MIXTURE.
27x04	lead	the ROMANS used LEAD for WATER PIPES.
27x05	lead	their PUBLIC BATHS were LINED with LEAD.
27x06	lead	the WORD "PLUMBER" means a WORKER of LEAD.
27x07	lead	a LEAD ATOM has EIGHTY -two ELECTRONS.
27x08	lead	LEAD is one of the HEAVIEST KNOWN METALS.
27x09	lead	it is ELEVEN TIMES HEAVIER than WATER.
27x10	lead	the EXPRESSION "as HEAVY as LEAD" is COMMON.
27x11	lead	it DESCRIBES an OBJECT of GREAT WEIGHT.
27x12	lead	LEAD can be DANGEROUS for the NERVOUS SYSTEM.
28x01	lion	the LION is a WILD member of the CAT FAMILY.

28x02	lion	it is RELATED to the TIGER AND the BOBCAT.
28x03	lion	the LION and TIGER are the LARGEST CATS.
28x04	lion	the TIGER is the FIERCEST of ALL CATS.
28x05	lion	the LION is a STRONG, WILD CREATURE.
28x06	lion	it has a LARGE, HEAVY and POWERFUL BODY.
28x07	lion	its LONG MANE gives it a PROUD APPEARANCE.
28x08	lion	the LION is KNOWN as the "KING of JUNGLE".
28x09	lion	HOWEVER, they are SELDOM FOUND in the JUNGLE.
28x10	lion	they TYPICALLY INHABIT SAVANNA and GRASSLAND.
28x11	lion	LIONS are usually SOCIAL COMPARED to other CATS.
28x12	lion	LIONS SPEND much of their TIME RESTING.
29x01	zebra	a ZEBRA is an ANIMAL that LIVES in AFRICA.
29x02	zebra	it is a WILD ANIMAL that EATS GRASS.
29x03	zebra	it LOOKS VERY much LIKE a HORSE.
29x04	zebra	most ZEBRAS STAND four to five FEET HIGH.
29x05	zebra	the ZEBRA has a SURPRISINGLY different COLOR PATTERN.
29x06	zebra	ZEBRAS have parallel BLACK and WHITE STRIPES.
29x07	zebra	the STRIPES are ARRANGED in DISTINCTIVE PATTERNS.
29x08	zebra	these STRIPES RUN all OVER their BODIES.
29x09	zebra	they even RUN UP and DOWN their FACES.
29x10	zebra	the STRIPES also APPEAR on the ZEBRA 's EARS.
29x11	zebra	ZEBRAS can TURN their EARS in almost any DIRECTION.
29x12	zebra	the ZEBRA's EYES are on the SIDES of its HEAD.
30x01	wolf	the WOLF is a MEMBER of the DOG FAMILY.
30x02	wolf	a WOLF LOOKS like a SKINNY wild DOG.
30x03	wolf	it has a WIDE HEAD and POINTED NOSE.
30x04	wolf	WOLVES live in North AMERICA, EUROPE, and ASIA.
30x05	wolf	WOLVES USED to LIVE all over the united STATES.
30x06	wolf	GRAY WOLVES are SELDOM SEEN nowadays.
30x07	wolf	they LIVE in the ROCKIES and NORTHERN STATES.
30x08	wolf	wolves HUNT in PACKS and MATE for LIFE.
30x09	wolf	the AVERAGE WOLF PACK consists of TEN wolves.
30x10	wolf	a FEMALE WOLF gives BIRTH every other YEAR.
30x11	wolf	the WOLF 's STRONGEST SENSE is SMELL.
30x12	wolf	SMELL PLAYS a role in their COMMUNICATION.

31x01	orange	the ORANGE is the most IMPORTANT CITRUS FRUIT.
31x02	orange	ORANGE's trees BLOSSOM between APRIL and MAY.
31x03	orange	it is a GOOD SOURCE of VITAMIN C.
31x04	orange	it can be EATEN or MADE INTO JUICE.
31x05	orange	EATING ORANGES may PREVENT the common COLD.
31x06	orange	there are TWO DIFFERENT KINDS of ORANGES.
31x07	orange	the SWEET ORANGE is EATEN in the united STATES.
31x08	orange	it is THOUGHT to have COME FROM CHINA.
31x09	orange	the OTHER KIND of ORANGE is more BITTER.
31x10	orange	ORANGES are OFTEN USED in COOKING.
31x11	orange	the TANGERINE is often INCORRECTLY CALLED an ORANGE.
31x12	orange	the first ORANGE PRODUCER in the WORLD is BRAZIL.
32x01	oysters	OYSTERS are ANIMALS that live in SEA SHELLS.
32x02	oysters	the OYSTER LIVES in many PARTS of the WORLD.
32x03	oysters	it LIVES mostly in QUIET, SHALLOW WATERS.
32x04	oysters	some KINDS of OYSTERS are CONSIDERED a DELICACY.
32x05	oysters	it is MAN's MOST VALUABLE SEAFOOD.
32x06	oysters	the OYSTER's SHELL FORMS a SHELTER.
32x07	oysters	the SHELL is DIVIDED into TWO HALVES.
32x08	oysters	they are FASTENED TOGETHER at ONE END.
32x09	oysters	the LEFT HALF is LARGER and THICKER.
32x10	oysters	a MUSCLE ATTACHES the soft BODY to the SHELL.
32x11	oysters	a MUSCLE helps the OYSTER OPEN the SHELL.
32x12	oysters	their SHELLS are also USED to make DECORATIVE OBJECTS.
33x01	dice	DICE are CUBES used in GAMES of CHANCE.
33x02	dice	DICE are used to GENERATE RANDOM NUMBERS.
33x03	dice	it is UNCERTAIN WHERE DICE ORIGINATED.
33x04	dice	they may be MADE of IVORY, WOOD, or PLASTIC.
33x05	dice	a SINGLE such CUBE is CALLED a DIE.
33x06	dice	each SIDE of a DIE has ONE to SIX dots.
33x07	dice	DOTS on OPPOSITE SIDES add up to SEVEN.
33x08	dice	players TOSS the DICE on a FLAT SURFACE.
33x09	dice	the NUMBERS that COME up DECIDE the GAME.
33x10	dice	the COMBINATION of NUMBERS DEPENDS on CHANCE.
33x11	dice	a GAME of DICE is THEREFORE a GAMBLE.

33x12	dice	GAMBLING is not LEGAL in MOST STATES.
34x01	eagle	the EAGLE is a LARGE BIRD of PREY.
34x02	eagle	it has POWERFUL WINGS and SHARP EYES.
34x03	eagle	the EAGLE is a SYMBOL of COURAGE and FREEDOM.
34x04	eagle	the BALD EAGLE is AMERICA's national BIRD.
34x05	eagle	there are SEVERAL DIFFERENT KINDS of EAGLES.
34x06	eagle	each TYPE IS very DIFFERENT in SIZE and COLOR.
34x07	eagle	EAGLES have strong BEAKS and POWERFUL CLAWS.
34x08	eagle	the EAGLE's BEAK is as LONG as its HEAD.
34x09	eagle	there is a HOOK on the UPPER HALF of the BEAK.
34x10	eagle	the EAGLE uses its powerful BEAK to CATCH its PREY.
34x11	eagle	EAGLES normally BUILD their NESTS in tall TREES.
34x12	eagle	many CULTURES have DEPICTED EAGLES in their ART.
35x01	ear	the EAR is an IMPORTANT SENSORY ORGAN.
35x02	ear	the EAR has TWO MAIN PURPOSES.
35x03	ear	it allows PEOPLE to HEAR and MAINTAIN BALANCE.
35x04	ear	the EAR is usually DESCRIBED as having THREE PARTS.
35x05	ear	good HEARING allows PEOPLE to UNDERSTAND SPEECH.
35x06	ear	through SPEECH, we EXCHANGE IDEAS and OPINIONS.
35x07	ear	HEARING also MAKES man AWARE of DANGER.
35x08	ear	our EARS are PLACED on either SIDE of the HEAD.
35x09	ear	the EAR's BALANCE mechanism helps us WALK UPRIGHT.
35x10	ear	DAMAGE to this SECTION CAUSES STAGGERING.
35x11	ear	the PERSON also GETS DISORIENTED and DIZZY.
35x12	ear	this KIND of DIZZINESS is CALLED VERTIGO.
36x01	liver	the LIVER is a very IMPORTANT INTERNAL ORGAN.
36x02	liver	its MAIN FUNCTION is to FILTER the BLOOD.
36x03	liver	the LIVER PRODUCES BILE necessary for DIGESTION.
36x04	liver	the LIVER is the LARGEST ORGAN in MAN.
36x05	liver	it can WEIGH THREE to FOUR POUNDS.
36x06	liver	the LIVER is DARK RED or CHOCOLATE colored.
36x07	liver	it is LOCATED in the MIDDLE SECTION of the BODY.
36x08	liver	it SITS CLOSELY to the INTESTINES and KIDNEYS.
36x09	liver	a high DOSAGE of some MEDICINES can DAMAGE the LIVER.
36x10	liver	it is POSSIBLE to TRANSPLANT a LIVER from a DONOR.

36x11	liver	this ADVANCED OPERATION is VERY EXPENSIVE.
36x12	liver	HOWEVER, it is RESPONSIBLE for SAVING many LIVES.
37x01	leopard	the LEOPARD is a MEMBER of the CAT FAMILY.
37x02	leopard	it is the THIRD LARGEST CAT in the WORLD.
37x03	leopard	ONLY the LION and TIGER are LARGER.
37x04	leopard	its FUR is MARKED with SMALL SPOTS.
37x05	leopard	LEOPARDS LIVE in the JUNGLES of AFRICA.
37x06	leopard	they are EXCELLENT NIGHT TIME HUNTERS.
37x07	leopard	LEOPARDS are known for their ABILITY to CLIMB TREES.
37x08	leopard	LEOPARDS STAND almost two feet HIGH at the SHOULDERS.
37x09	leopard	a big MALE may MEASURE NINE FEET long.
37x10	leopard	it can WEIGH one HUNDRED and SIXTY POUNDS.
37x11	leopard	a large FEMALE will WEIGH only SEVENTY POUNDS.
37x12	leopard	LEOPARDS have only a FEW CUBS in a LITTER.
38x01	eye	the EYE is the most IMPORTANT SENSE ORGAN.
38x02	eye	we USE it to VIEW the WORLD AROUND us.
38x03	eye	the HUMAN EYE can DIFFERENTIATE ten million COLORS.
38x04	eye	the PUPIL REGULATES the LIGHT ENTERING the eye.
38x05	eye	the EYES are USED EVERYDAY in most ACTIVITIES.
38x06	eye	EYES are OUR WINDOWS to the WORLD.
38x07	eye	the LENS of the EYE COLLECTS LIGHT.
38x08	eye	the LIGHT is FOCUSED INSIDE the EYE.
38x09	eye	THIS INFORMATION is SENT to the BRAIN.
38x10	eye	the BRAIN then BEGINS to PROCESS the IMAGE.
38x11	eye	EYES help us to ENJOY BOOKS and PAINTINGS.
38x12	eye	EYESIGHT helps us CAPTURE BEAUTIFUL SUNSETS.
39x01	zipper	a ZIPPER is any KIND of SLIDE FASTENER.
39x02	zipper	the word ZIPPER REFLECTS the SOUND the DEVICE makes.
39x03	zipper	ALL ZIPPERS have two ROWS of TEETH.
39x04	zipper	the two EDGES of the ZIPPER FASTEN TOGETHER.
39x05	zipper	the TEETH HOLD the ZIPPER TOGETHER.
39x06	zipper	the EDGES stay FASTENED UNTILL they are RELEASED.
39x07	zipper	they are RELEASED by DRAWING the SLIDE BACK.
39x08	zipper	SLIDE ZIPPERS are often used to FASTEN CLOTHING.
39x09	zipper	THEY are USED on LUGGAGE and BRIEFCASES.

39x10	zipper	a ZIPPER is RELATIVELY CHEAP to PRODUCE.
39x11	zipper	the FIRST ZIPPER was INVENTED by an AMERICAN.
39x12	zipper	it was MADE of CONNECTED HOOKS and EYES.
40x01	egg	many kinds of ANIMALS and BIRDS PRODUCE EGGS.
40x02	egg	the main PURPOSE of EGGS is to BREED YOUNG.
40x03	egg	most YOUNG ANIMALS BEGIN as an EGG.
40x04	egg	PEOPLE usually THINK of the EGG as a FOOD.
40x05	egg	actually, FEW KINDS of EGGS are EATEN.
40x06	egg	BIRD's EGGS are LARGER than MAMMAL'S.
40x07	egg	most BIRD EGGS have an OVAL SHAPE.
40x08	egg	their EGGS CONTAIN FOOD for the young BIRD.
40x09	egg	young birds DEVELOP OUTSIDE the MOTHER's BODY.
40x10	egg	the OSTRICH EGG is the LARGEST TYPE.
40x11	egg	the HUMAN EGG is ONE of the SMALLEST.
40x12	egg	DECORATED EGGS are an EASTER TRADITION.
41x01	clocks	CLOCKS are INSTRUMENTS that can MEASURE TIME.
41x02	clocks	the CLOCK is one of the OLDEST HUMAN INVENTION.
41x03	clocks	they DIVIDE DAYS into REGULAR INTERVALS.
41x04	clocks	ORIGINALLY SHADOWS were used to MARK TIME.
41x05	clocks	the SHORTEST SHADOWS OCCUR around MIDDAY.
41x06	clocks	LONGER SHADOWS occur in MORNING and late AFTERNOON.
41x07	clocks	the FIRST CLOCK INVENTED was the SUNDIAL.
41x08	clocks	later, the WATER CLOCK was DEVELOPED in CHINA.
41x09	clocks	it could MEASURE TIME on CLOUDY DAYS.
41x10	clocks	WATER CLOCKS were used for several THOUSAND YEARS.
41x11	clocks	early GREEKS and ROMANS also USED CLOCKS.
41x12	clocks	the WORD CLOCK is DERIVED from CELTIC.
42x01	kangaroo	the KANGAROO CARRIES its YOUNG in a POUCH.
42x02	kangaroo	the POUCH is LOCATED OUTSIDE of the ABDOMEN.
42x03	kangaroo	ANIMALS with POUCHES are NOT found in AMERICA.
42x04	kangaroo	the KANGAROO's NATIVE COUNTRY is AUSTRALIA.
42x05	kangaroo	there are MANY DIFFERENT KINDS of KANGAROOS.
42x06	kangaroo	the SMALLEST are the SAME SIZE as a RABBIT.
42x07	kangaroo	the LARGEST are nearly SEVEN FEET TALL.
42x08	kangaroo	their BACK LEGS are LARGER than their FRONT legs.

42x09	kangaroo	all SPECIES of KANGAROO are STRICTLY HERBIVORES.
42x10	kangaroo	KANGAROOS LIVE in SMALL GROUPS.
42x11	kangaroo	KANGAROO FOSSILS have RECENTLY been FOUND.
42x12	kangaroo	PREHISTORIC KANGAROOS GREW to be very LARGE.
43x01	camel	CAMELS are MAMMALS living in ARID COUNTRIES.
43x02	camel	THEY are STRONG and RESLIANT ANIMALS.
43x03	camel	most CAMELS LIVING TODAY are DOMESTICATED.
43x04	camel	CAMELS were ONCE a SYMBOL of WEALTH.
43x05	camel	in the BIBLE, GOD gave CAMELS to ABRAHAM.
43x06	camel	CAMELS were also USED for MILITARY PURPOSES.
43x07	camel	their HUMPS are RESERVOIRS of FATTY TISSUE.
43x08	camel	they can ENDURE long, HARD DESERT JOURNEYS.
43x09	camel	CAMELS can TRAVEL many miles WITHOUT needing WATER.
43x10	camel	without the CAMEL, MAN couldn't TRAVEL the DESERTS.
43x11	camel	CAMELS can LIVE up to about FIFTY YEARS.
43x12	camel	CAMEL CARAVANS are still SEEN in the SAHARA.
44x01	goose	the GOOSE is a WEB FOOTED BIRD.
44x02	goose	the GOOSE is CLOSELY RELATED it the DUCK.
44x03	goose	a GOOSE is LARGER THAN a DUCK.
44x04	goose	its NECK is SLIGHTLY LONGER than a DUCK's.
44x05	goose	there are FORTY DIFFERENT VARIETIES of GEESE.
44x06	goose	SEVENTEEN kinds of WILD GEESE live in AMERICA.
44x07	goose	GEESE live in PERMANENT PAIRS throughout the YEAR.
44x08	goose	they are TERRITORIAL DURING the NESTING SEASON.
44x09	goose	GEESE are KNOWN to MOVE with the SEASONS.
44x10	goose	THEY FLY north in SUMMER and south in WINTER.
44x11	goose	some FLY as FAR NORTH as the ARCTIC.
44x12	goose	others FLY as FAR SOUTH as MEXICO.
45x01	dictionary	a DICTIONARY LISTS the MEANING of WORDS.
45x02	dictionary	WORDS in a DICTIONARY are ARRANGED ALPHABETICALLY.
45x03	dictionary	it allows a PERSON DEFINE a WORD QUICKLY.
45x04	dictionary	BILINGUAL DICTIONARIES are used to TRANSLATE WORDS.
45x05	dictionary	a DICTIONARY CONTAINS over six HUNDRED thousand WORDS.

45x06	dictionary	a FIFTH grade CHILD knows two THOUSAND WORDS.
45x07	dictionary	ADULTS and CHILDREN NEED to use DICTIONARIES.
45x08	dictionary	many WEBSITES OPERATE as ONLINE DICTIONARIES.
45x09	dictionary	DICTIONARIES TELL us many USEFUL things about WORDS.
45x10	dictionary	every DICTIONARY shows the CORRECT SPELLING of a WORD.
45x11	dictionary	it also SHOWS HOW a WORD is PRONOUNCED.
45x12	dictionary	DICTIONARIES are very HELPFUL for CREATIVE WRITING.

References

Abercombie, D. (1967). *Elements of general phonetics*: Aldine Pub. Company.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of experimental psychology: Human perception and performance, 35*(2), 520.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289-1292.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*(9), 839-843.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods, 43*(3), 761-770.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin, 91*(2), 276.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science, 5*(10), 371-372.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology, 2*, 142-162.

Beatty, J., & Wagoner, B. L. (1977). *Pupillometric signs of brain activation vary with level of cognitive processing*. Retrieved from

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language, 28*(1), 1-11.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: influence of age and hearing impairment. *Brain research bulletin, 91*, 21-30.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2014). *Objective assessment of listening effort in the oscillatory eeg: Comparison of different hearing aid configurations.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends in amplification, 17*(2), 75-93.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science, 20*(11), 1313-1315.

Bilger, R. C., Nuetzel, J., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research, 27*(1), 32-48.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: a pupillometry study. *Frontiers in neuroscience, 12*, 152.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339-2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272-284.

Bradshaw, J. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology, 20*(2), 116-122.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology, 64*(1), 74-95.

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America, 131*(2), 1449-1464.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527-2538.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*(3), 433-461.

Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences, 114*(34), 9032-9037.

Calandruccio, L., & Smiljanic, R. (2012). New sentence recognition materials developed using a basic non-native English lexicon. *Journal of Speech, Language, and Hearing Research, 55*(5), 1342-1355.

Chapman, L. R., & Hallowell, B. (2015). A novel pupillometric method for indexing word difficulty in individuals with and without aphasia. *Journal of Speech, Language, and Hearing Research, 58*(5), 1508-1520.

Clopper, C. G., & Bradlow, A. R. (2006). Effects of dialect variation on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *119*(5), 3424-3424.

Cohen, J. (1988). Statistical power analysis for the behavior science. *Lawrance Eribaum Association*.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562-1573.

Cooke, M., Lecumberri, M. G., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America, 123*(1), 414-427.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8*(5 Suppl), 119S-126S.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269-284.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668-3678.

Cvijanović, N., Kechichian, P., Janse, K., & Kohlrausch, A. (2017). Effects of noise on arousal in a speech communication setting. *Speech Communication, 88*, 127-136.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior, 19*(4), 450-466.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top–down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23*(12), 3914-3932.

De Filippo, C. L., & Scott, B. L. (1978). A method for training and evaluating the reception of ongoing speech. *The Journal of the Acoustical Society of America, 63*(4), 1186-1192.

Denby, T. N. (2013). The Filtering Listener: Dispersion in Exemplar Theory.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences, 109*(29), 11854-11859.

Docherty, G. J. (2003). Speaker, community, identity: Empirical and theoretical perspectives on sociophonetic variation. Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences.

Drager, K. D., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *Journal of Speech, Language, and Hearing Research, 44*(5), 1052-1057.

Drullman, R. (1995). Speech intelligibility in noise: relative contribution of speech elements above and below the noise level. *The Journal of the Acoustical Society of America, 98*(3), 1796-1798.

Duffy, J. R., & Giolas, T. G. (1974). Sentence intelligibility as a function of key word selection. *Journal of Speech, Language, and Hearing Research, 17*(4), 631-637.

Ellis, R. (1994). *The study of second language acquisition*: Oxford University.

Elshtain, E. L., & Schaefer, T. (1968). Effects of storage load and word frequency on pupillary responses during short-term memory. *Psychonomic Science*, *12*(4), 143-144.

Elston-Güttler, K. E., & Gunter, T. C. (2009). Fine-tuned: Phonology and semantics affect first- to second-language zooming in. *Journal of Cognitive Neuroscience*, *21*(1), 180-196.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(4), 639-645.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report, 1*, 22-30.

Faulkner, A., Rosen, S., & Wilkinson, L. (2001). Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear and Hearing, 22*(5), 431-438.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America, 116*(4), 2365-2373.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of child language, 16*(3), 477-501.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E. (2012). The role of input in second language speech learning. . *VIth International Conference on Native and Non-native accent of English. Lodz, Poland 6-8 December 2012.*

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*(2), 169-186.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(04), 527-552.

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

Francis, A. L., Tigchelaar, L. J., Zhang, R., & Zekveld, A. A. (2018). Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech. *Journal of Speech, Language, and Hearing Research, 61*(7), 1815-1830.

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology, 66*(12), 2289-2294.

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research, 53*(1), 18-33.

Garland, S. (2007). The bilingual spectrum. *Orlando, FL: Guirnalda*.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International journal of audiology, 43*(2), 85-99.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology, 24*(1), 63-68.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*(2), 103-127.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America, 63*(1), 223-230.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 252-269.

Goldinger, S. D. (2007). *A complementary-systems approach to abstract and episodic speech perception.* Paper presented at the Proceedings of the 16th International Congress of Phonetic Sciences.

Gosselin, P. A., & Gagne, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research, 54*(3), 944-958.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*(4), 457-461.

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*(3), 1197-1208.

Grynpas, J., Baker, R., & Hazan, V. (2011). *Clear speech strategies and speech perception in adverse listening conditions.* Paper presented at the Proceedings of the 17th International Congress of Phonetic Science.

Gustafsson, H. Å., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America, 95*(1), 518-529.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of psycholinguistic research, 30*(3), 251-266.

Hazan, V., Grynpas, J., & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions? *The Journal of the Acoustical Society of America, 132*(5), EL371-EL377.

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication, 52*(11), 996-1009.

Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in systems neuroscience, 8*, 35.

Heinrich, W. (1896). der Sinnesorgane. *Zeitschrift fu r Psychologie und Physiologie der Sinnesorgane, 9*, 342.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349-350.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics, 72*(5), 1218-1227.

Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends in hearing, 20*, 2331216516667734.

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International journal of audiology, 52*(11), 753-761.

Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International journal of audiology, 49*(12), 928-932.

Hsia, H. (1977). Redundancy: Is it the lost key to better communication? *AV communication review, 25*(1), 63-85.

https://www.ielts.org/. (2017).

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica, 137*(2), 151-171.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48*(3), 598-612.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*(1), 553-562.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(01), 145-160.

Johnson, K. (2008). 15 Speaker Normalization in Speech Perception. *The handbook of speech perception*, 363.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47*(2), 310.

Kahneman, D. (1973). *Attention and effort*: Citeseer.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337-1351.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 336.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963-973.

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with “holes''in the spectrum. *The Journal of the Acoustical Society of America, 112*(3), 1102-1111.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5*(2), 115-133.

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in psychology, 5*, 651.

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2015). Native and non-native speech perception by hearing-impaired listeners in noise-and speech maskers. *Trends in hearing, 19*, 2331216515579127.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS One, 11*(5), e0155574.

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America, 139*(4), 1618-1636.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing, 33*(2), 291-300.

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research, 367*, 106-112.

Kos, M., Van den Brink, D., & Hagoort, P. (2012). Individual variation in the late positive complex to semantic anomalies. *Frontiers in psychology, 3*, 318.

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2012). Processing load during listening: The influence of task characteristics on the pupil response. *Language and cognitive processes, 28*(4), 426-442.

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

Kuchinsky, S. E., Ahlstrom, J. B., Vaden Jr, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology, 50*(1), 23-34.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & psychophysics, 50*(2), 93-107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274): Springer.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850-11857.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience, 5*(11), 831-843.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684-686.

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*, 61.

Laeng, B., Sirois, S., & Gredebäck, G. (2012a). Pupillometry a window to the preconscious? *Perspectives on psychological science*, *7*(1), 18-27.

Laeng, B., Sirois, S., & Gredebäck, G. (2012b). Pupillometry: A window to the preconscious? *Perspectives on psychological science*, *7*(1), 18-27.

Landgraf, S., Van der Meer, E., & Krueger, F. (2010). Cognitive resource allocation for neural activity underlying mathematical cognition: a multi-method study. *Zdm, 42*(6), 579-590.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects Desempeño cognitivo y percepción del esfuerzo en tareas de procesamiento del lenguaje: Efectos de las diferentes condiciones de fondo en sujetos normales e hipoacúsicos. *International journal of audiology*, *44*(3), 131-143.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature reviews neuroscience, 9*(12), 920.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11), 864-886.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2B), 467-477.

Levitt, H., Waltzman, S., Shapiro, W., & Cohen, N. (1986). Evaluation of a cochlear prosthesis using connected discourse tracking. *Journal of rehabilitation research and development, 23*(1), 147-154.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673-680.

Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic bulletin & review, 23*(2), 412-425.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74*(6), 431.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. i., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96*(4), 2076-2087.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input1. *Applied linguistics, 4*(2), 126-141.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001-1010.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., . . . Froehlich, M. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America, 127*(3), 1491-1505.

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology, 22*(2), 113-122.

MacMahon, M. K. (1991). The woman behind 'Arthur'. *Journal of the International Phonetic Association, 21*(01), 29-31.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*: Routledge.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods, 50*(1), 94-106.

Mattys, S. L., Carroll, L. M., Li, C. K., & Chan, S. L. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication, 52*(11), 887-899.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and cognitive processes, 27*(7-8), 953-978.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3), 686-693.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology, 18*(1), 1-86.

McGarrigle, R., Gustafson, S. J., Hornsby, B. W., & Bess, F. H. (2018). Behavioral Measures of Listening Effort in School-Age Children: Examining the Effects of Signal-to-Noise Ratio, Hearing Loss, and Amplification. *Ear and Hearing*.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International journal of audiology*.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of experimental psychology, 41*(5), 329.

Mirman, D. (2014). *Growth curve analysis and visualization using R*: CRC Press Boca Raton, FL.

Moore, T. M., & Picou, E. M. (2018). A Potential Bias in Subjective Ratings of Mental Effort. *Journal of Speech, Language, and Hearing Research, 61*(9), 2405-2421.

Moreno, E. M., Rodríguez-Fornells, A., & Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics, 21*(6), 477-508.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews, 33*(7), 1004-1023.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(3), 299-325.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42-46.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage, 55*(2), 713-723.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *The Journal of Neuroscience, 32*(36), 12376-12383.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., . . . Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research, 351*, 68-79.

Osberger, M. J., Johnson, D. L., & Miller, J. D. (1987). Use of connected discourse tracking to train functional speech skills. *Ear and Hearing, 8*(1), 31-36.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83*(1), 56-64.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America, 95*(3), 1581-1592.

Peelle, J. E., Gross, J., & Davis, M. H. (2012). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378-1387.

Pfenninger, S. E., & Singleton, D. (2018). Starting Age Overshadowed: The Primacy of Differential Environmental and Family Support Effects on Second Language Attainment in an Instructional Context. *Language Learning*. doi:https://doi.org/10.1111/lang.12318

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., . . . Mackersie, C. L. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S.

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653-1662.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*(3), 560-569.

Pollack, I., Rubenstein, H., & Decker, L. (1959). Intelligibility of known and unknown message sets. *The Journal of the Acoustical Society of America, 31*(3), 273-279.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., McNamara, J., & Wiliams, S. (2004). Vision: the eye *Neuroscience* (3rd ed., Vol. 1, pp. 179-198).

R Core Team. (2017). R: A language and environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2018). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, 1-14.

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature communications*, *7*, 13289.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology, 22*(2), 204-207.

Richter, M., Gendolla, G. H., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know *Advances in motivation science* (Vol. 3, pp. 149-186): Elsevier.

Ringbom, H. (1980). On the distinction between second-language acquisition and foreign-language learning. *AFinLAn vuosikirja*, 37-44.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(03), 465-485.

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International journal of audiology, 58*(5), 247-261.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Pichora-Fuller, M. K. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in systems neuroscience, 7*.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International journal of audiology, 47*(sup2), S99-S105.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*(5), 1147-1153.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & psychophysics, 31*(4), 307-314.

Sauseng, P., & Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews, 32*(5), 1001-1013.

Savin, H. B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America, 35*(2), 200-206.

Scharenborg, O., Coumans, J. M., & van Hout, R. (2017). The Effect of Background Noise on the Word Activation Process in Nonnative Spoken-Word Recognition.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17*(1), 133-145.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in psychology, 5*.

Siegle, G. J., Steinhauer, S. R., Stenger, V. A., Konecky, R., & Carter, C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *Neuroimage, 20*(1), 114-124.

Simantiraki, O., Cooke, M., & King, S. (2018). Impact of Different Speech Types on Listening Effort. *Proc. Interspeech 2018*, 2267-2271.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(6), 679-692.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass, 3*(1), 236-264.

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America, 130*(6), 4020-4031.

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition, 179*, 163-170.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5): Pearson Boston, MA.

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and cognitive processes, 1*(4), 297-308.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America, 88*(2), 663-666.

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, *283*(5401), 549-554.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication, 49*(1), 2-7.

Van der Vlugt, M., & Nooteboom, S. (1986). Auditory word recognition is not more sensitive to word-initial than to word-final stimulus information. *IPO Annual Progress Report*.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and speech*, *53*(4), 510-540.

Van Engen, K. J., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research, 57*(5), 1908-1918.

Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35*(1-2), 103-113.

Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology, 38*(01), 76-83.

Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology, 52*(1), 23-36.

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear and Hearing, 39*(3), 573-582.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of memory and language, 50*(1), 1-25.

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). *WAIS-IV: Wechsler adult intelligence scale*: Pearson San Antonio, TX.

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in psychology, 7*, 345.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *The Journal of Neuroscience, 32*(40), 14010-14021.

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology, 18*(7), 548-558.

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in hearing, 20,* 2331216516669723.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing, 36*(4), e153.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in hearing, 22*, 2331216518800869.

Wode, H., Burmeister, H., & Ufert, D. (1980). *Phonology in L2 acquisition*: Universität Kiel.

Wood, K. V. (1995). Marker proteins for gene expression. *Current opinion in biotechnology, 6*(1), 50-58.

Zekveld, A. A., George, E. L., Kramer, S. E., Goverts, S. T., & Houtgast, T. (2007). The development of the text reception threshold test: A visual analogue of the speech reception threshold test. *Journal of Speech, Language, and Hearing Research, 50*(3), 576-584.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage, 101*, 76-86.

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in hearing, 22*, 2331216518777174.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology, 51*(3), 277-284.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31*(4), 480-490.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing, 32*(4), 498-510.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., & Rönnberg, J. (2011). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and Hearing, 32*(6), e16-e25.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language, 122*(2), 103-113.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in neuroscience, 8*, 88.