Spatial listening and semantic interference in a dual-language context

Emily Allen-Rice

PhD

University of York

Psychology

December 2025

Abstract

Understanding target speech amid competing speech (i.e., a masker) is an everyday challenge, and for bilinguals it may be compounded when the masker is another known language (i.e., a dual-language context). This thesis examines how forms of masking affect bilingual listening. Chapter 2 examines energetic masking (EM), a type of acoustic disruption that can be alleviated through spatial separation in a process known as spatial release from masking (SRM). For bilinguals, prior research suggests EM hinders L2 processing more than L1, suggesting larger SRM benefits for L2 than L1 listening. Across selective and divided listening tasks, Chapter 2 shows SRM benefited listening, regardless of whether the target language was L1 or L2. However, this benefit was reduced when listeners tracked both talkers simultaneously, likely due to cognitive demands associated with ear-switching. Chapter 3 investigates informational masking (IM), examining whether semantic overlap between target and masker speech disrupts listening. Previous research demonstrates that masker meaningfulness *per se* impacts listening, and that lexical-semantic links are stronger for L1 than L2. Therefore, maskers semantically related to a target might impact listening more than semantically unrelated maskers, particularly when the masker is L1 compared to L2 (for an L1 target). Across experiments, semantic overlap did not affect listening, regardless of masker language. Yet in single-language (L1-L1) contexts with minimal acoustic cues for stream segregation, listeners appeared to use semantic coherence when deciding if a heard word belonged to the target or masker. Therefore, semantic overlap between speech streams did not affect target transcription but did impact error types. Overall, while semantic cues from the masker do not impact listening, bilinguals nevertheless rely on low-level acoustic cues to stream speech in adverse listening conditions.

Table of Contents

List of Figures

Acknowledgements

Most importantly, I would like to thank my three excellent supervisors: Sven Mattys, Angela de Bruin, and Sarah Knight. I could not have completed this thesis without your expert guidance, thorough feedback, and constant support. You have all been incredibly patient with me and have taught me so much. Completing this thesis has been a huge challenge, but your kindness and support have encouraged me immensely.

I would also like to thank the other former and current members of Speech Lab throughout my PhD: Alex Mepham, Lotte Eijk, Sophie Meekings, Georgie Maher, Rachel Yue Zheng, Ethan Brims, and Emily Shepherdson. I have loved getting to know you all, and my research has been inspired by your ideas and passion for your own fascinating projects. I am extremely grateful to María Muradás-Taylor for taking the time to record all the Chapter 3 stimuli, and Georgie Maher and Miriam Tucker for helping to score the transcription results in Chapter 3. Thank you also to my thesis advisory panel: Dan Baker, Gareth Gaskell, and Silvia Gennari, for the conversations that have challenged me, motivated me, and made me a better researcher. I am also grateful to my examiners, Silvia Gennari and Emma Holmes, for giving their time and expertise to review this thesis. I especially appreciate Emma's willingness to travel up to York for the viva.

Finally, I want to thank my family for always encouraging me to achieve my goals and supporting me throughout my whole life. I would particularly like to thank Louis. You have been beside me through every high and low of this PhD, celebrating successes and helping me through the setbacks. You have listened to countless practice talks, asked all the right questions, and given me the confidence I needed to keep going. Thank you for everything.

Author's Declaration

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

All experiments were designed by me, with input from my supervisors, Professor Sven Mattys, Dr Angela de Bruin, and Dr Sarah Knight, and with additional guidance from my thesis advisory panel, Professor Dan Baker, Professor Gareth Gaskell, and Dr Silvia Gennari. María Muradás-Taylor contributed to Chapter 3 by checking the Spanish accuracy of the stimuli and recording them, and Georgie Maher and Miriam Tucker contributed to Chapter 3 by assisting with scoring the listening task results. All data analyses and all written components of this thesis were carried out by me, with the assistance of my supervisors.

A version of Chapter 2 has been accepted for publication as: Allen-Rice, E., Knight, S., de Bruin, A., & Mattys, S. (in press). Selective and divided listening in a dual-language context. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218251386821

Chapter 1: Introduction

1.1 Overview

1.1.1 Energetic and informational masking

Listening to a talker while another person is talking nearby can be challenging. This is in part because the competing talker exerts energetic masking (EM), which occurs when the background and target talkers share frequency components at the same time (spectrotemporal overlap; Brungart, 2001; Pollack, 1975). Furthermore, various factors beyond EM can interfere with intelligibility. These include modulation masking, the interference of competing amplitude modulations between target and masker even when there is little or no spectral overlap between them (Stone et al., 2011). Additionally, the competing talker may exert higher-level masking upon the target beyond EM. For example, the competing talker may use words that capture the listener's attention (i.e., their name), speak in a familiar (versus unfamiliar) language or accent, or speak about a topic that is semantically similar to the target speech. This can interfere with the target beyond acoustic overlap and is known as informational masking (IM; Pollack, 1975).

One way to reduce EM is by spatially separating the two talkers, resulting in spatial release from EM (SRM), which improves listening performance (e.g., Ihlefeld & Shinn-Cunningham, 2008a). However, spatial separation can be detrimental to listening if the listener attempts to attend to both talkers at the same time. Spatial separation in this case can be cognitively demanding due to the need to shift attention between locations (e.g., Treisman, 1971). Despite this increased cognitive demand, listening to both talkers at once is still easier when the talkers are spatially separated compared to collocated (Knight et al., 2023). This demonstrates that the beneficial effects of SRM outweigh the cognitive costs of spatial separation.

This reduction of EM can increase listening performance, but listening performance may also be increased by reducing IM. Research shows that some participants notice when their own name appears in unattended masker speech while focusing on a target (e.g., Moray, 1959; Röer & Cowan, 2021; Wood & Cowan, 1995). This reflects a distracting effect of the masker that is not due to spectral-temporal overlap (EM), but rather to IM. Similarly, when EM is controlled, a masker babble spoken in a more familiar language disrupts target listening more than babble in a less familiar language (e.g., Van Engen, 2010), again illustrating the impact of IM. A less explored form of IM concerns the degree of semantic overlap between target and masker speech. Some research has explored whether the semantic meaningfulness of the masker impacts listening to a meaningful target, but findings have been inconsistent (e.g., Brouwer et al., 2012; Calandruccio et al., 2018). However, no research has directly investigated how varying semantic similarity between two meaningful, complex speech streams influences listening to a target (i.e., varying the semantic similarity between the two speech streams). It may be easier to ignore a competing talker when their speech is semantically unrelated (neutral) to the target than when it is semantically related. This method of increasing listening performance relies on the assumption that listeners unintentionally process semantic content of masker speech.

These findings highlight the potential interactions between acoustic, semantic and attentional factors during speech perception in adverse conditions. However, existing research mostly focuses on monolingual listeners and single-language contexts (i.e., when only one language is being spoken), leaving open the question of how bilingual listeners manage challenges such as EM and IM, particularly when target and masker speech occur in different, but known, languages. To address this, it is necessary to consider how bilinguals process and manage lexical information across their two languages.

1.1.2. Theoretical models of bilingual lexical access

Several cognitive models have been developed to explain how bilinguals process words and manage competition between their two languages. These models of bilingual processing aim to explain patterns such as cross-language interference, competition between words from different languages, and differences in processing speed depending on language proficiency. A notable model is the Bilingual Interactive Activation Plus model (BIA+; Dijkstra & van Heuven, 2002). This model evolved from the earlier Bilingual Interactive Activation model (BIA; Dijkstra et al., 1998) and proposes that bilinguals recognise visually presented words in either language through bottom-up processes. Low-level features of the input (e.g., letters) are activated first, and these activate other higher-level representations, including semantics (word meaning).

A key assumption of the BIA+ is parallel, non-selective activation, meaning that words from both languages are activated simultaneously, even in monolingual contexts (i.e., when only one language is being used). Empirical evidence supports the principle of non-selective activation in both the visual and auditory domains. For example, Dijkstra et al. (1999) asked Dutch-English bilinguals to identify visually presented English (second language, L2) words which were visually masked. Responses were slower when the English word was phonologically similar to a Dutch (first language, L1) word, indicating cross-language activation. Furthermore, Spivey and Marian (1999) found that when Russian-English bilinguals were instructed to pick up an object in either English or Russian during a monolingual session, they briefly looked at a distractor object whose name in the irrelevant language was phonologically similar to the spoken word more often than at a control distractor object.

Within the BIA+, activation is based on the low-level features of the input, at which point the language is not always identifiable yet. This parallel activation of words in both languages creates competition, since multiple words are simultaneously activated. The BIA+ accounts for this competition through several mechanisms. At the orthographic or phonological level, the word form that best matches the input gradually accumulates the most activation, allowing it to be selected. At the lexical level, words that share overlapping features inhibit one another laterally, reducing activation for less likely words. The BIA+ model also includes a higher-level task schema, which (depending on, for example, the linguistic and non-linguistic context) can modify the response decision criteria. While the BIA+ argues this task schema does not influence the word recognition system directly, other versions of the model do also include top-down control over the lexicon. For instance, the BIA includes higher-level language nodes that can modulate activation of words in the target language and suppression of words in the non-target language.

Crucially, the BIA+ accounts for the influence of word frequency on processing speed: words with higher frequencies (i.e., which are encountered more often by the individual) have higher resting activation levels and are therefore activated and recognised more quickly. Monolingual research can illustrate this principle of lexical activation. For example, in a visual-world task, Dahan et al. (2001) found that participants asked to move a picture on a screen (e.g., bench) were more likely to fixate on a more frequently used competitor (e.g., bed) than on a less frequently used competitor (e.g., bell), demonstrating that higher-frequency words are activated more rapidly. In bilinguals, words from a lower proficiency and less used L2 are typically used less frequently than words from their L1, which means they generally have lower resting activation levels. As a result, L2 words tend to be activated, identified, and processed more slowly – this is referred to as L2 delay (Dijkstra et al., 1998; Dijkstra & van Heuven, 2002).

While the BIA+ was primarily developed to explain visual word recognition, the model also incorporates phonological representations and acknowledges its potential applications to spoken word recognition. Building on this idea, the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS; Shook & Marian, 2012) applies similar principles to the auditory domain. Like the BIA+, the BLINCS is a connectionist, bottom-up model; however, it focuses on phonological rather than orthographic input. In the BLINCS, phonological features of spoken words (e.g., phonemes) are first activated, and these activate lexical and semantic representations. The BLINCS also maintains the principle of non-selective access, which is supported by research showing that bilinguals activate words from both languages when processing speech (e.g., Spivey & Marian, 1999; Weber & Cutler, 2004). Furthermore, L2 delay is also a feature of the BLINCS model. It proposes that L2 words often have weaker phonological representations and lower resting activation levels, leading to slower activation during auditory processing, much like in the BIA+. Therefore, when bilinguals process speech, L2 words are generally recognised more slowly than L1 words due to both their weaker phonological representations and their lower frequency of use (Shook & Marian, 2012).

1.1.3. Bilingual listening in dual-language contexts

Dual-language contexts occur when a listener is presented with two languages simultaneously, for example when the target speech is in one language and the masker speech is in another. Within dual-language contexts, bilingual listeners (who may also understand both languages) may experience unique challenges, such as cross-language activation, competition between lexical representations in the two languages, and differential effects of spatial and semantic masking depending on the language of the target and masker. Investigating bilingual speech perception in a dual-language context can therefore reveal how mechanisms of speech perception are affected when two languages compete and must be controlled.

This thesis investigates how bilingual speech perception is affected by two factors: (1) the trade-off between SRM and the cognitive costs of spatial separation (Chapter 2), and (2) the impact of the semantic content of masker speech on the perception of target speech (Chapter 3). The trade-off between SRM and the cognitive costs of spatial separation has not yet been explored in dual-language contexts. While monolingual research has shown that spatial separation improves target perception despite the additional cognitive demands of spatial listening, this balance may shift in bilingual listeners. Previous findings suggest that L1 and L2 processing are differently affected by EM, with L2 processing more negatively impacted by its presence (e.g., Cooke et al., 2008), potentially resulting in the L2 benefitting *more* than the L1 from SRM. Models of bilingual lexical access (e.g., Dijkstra & van Heuven, 2002; Shook & Marian, 2012) propose that L2 words are generally slower to activate due to their lower resting activation levels (e.g., Dijkgraaf et al., 2019), meaning that it could be more cognitively demanding to listen to the L2 than L1, so it could be more negatively impacted by the cognitive costs of spatial separation.

Investigating the impact of masker semantic content on target perception is novel even within monolingual contexts: no known study has examined how the degree of semantic relatedness between target and masker speech influences target transcription using naturalistic (i.e., semantically meaningful) and complex sentences in a speech-on-speech task. Additionally, this question is especially relevant in a dual-language context (i.e., when the two talkers are speaking different languages known to the listener), such as when the target speech is in the listener's L1 and the masker speech is in the listener's L2. In such cases, it is possible that semantic representations in the L2 are activated more slowly than those in the L1, as predicted by the L2 delay described in the BIA+ (Dijkstra & van Heuven, 2002) and BLINCS (Shook & Marian, 2012) models. Additionally, the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) posits that L2 words have weaker connections to their semantic representations than L1 words. This would mean that the semantic content of L2 speech might interfere with listening less than the semantic content of L1 speech. Therefore, investigating how spatial and semantic release from masking interact with bilingual listeners' need to control two languages simultaneously can advance our understanding of bilingual speech perception in adverse listening conditions.

1.2. Spatial release from energetic masking

1.2.1. The effect of spatial release from masking on L1 listening

Many researchers have provided descriptions of EM, with Pollack (1975) stating that it arises from the physical interference of background noise with the to-be-attended speech signal due to overlapping acoustic properties. This definition is expanded on by Brungart (2001), who states that EM occurs when both the target speech and competing noise contain energy in the same critical frequency bands at the same time. In other words, EM occurs when background noise interacts with the target signal at the lowest levels of speech perception (i.e., within the cochlea). In their review, Mattys et al. (2012) describe the extrinsic degradation of target speech due to EM as an adverse listening condition which constitutes an impediment to speech perception.

Culling and Stone (2017) explain that listeners can improve their listening performance by taking advantage of differences between the target speech and masker. One such difference is the presence of short-term dips in the masker’s intensity, which can reveal brief “glimpses” of the target speech (e.g., Cooke, 2003). A further method of reducing EM is by spatially separating the two stimuli, which creates interaural timing and intensity differences between the two sounds (e.g., Best et al., 2006; Ihlefeld & Shinn-Cunningham, 2008a). For example, if two talkers are collocated, their speech is equally intense at each ear, whereas when the two talkers are spatially separated, their speech will have different levels of intensity at each ear, due to each talker being physically closer to one ear than the other. Listeners use these differences to improve their listening performance, and this effect is known as spatial release from EM (SRM). This was investigated in Ihlefeld and Shinn-Cunningham’s (2008a) selective listening experiment. They found that listeners are more accurate at reporting target sentences when the masker is spatially separated from the target by 90° azimuth (e.g., the target is presented directly in front of the listener, and the masker is presented to the listener’s left ear) compared to when the target and masker are collocated. Similar results have been found by Marrone et al. (2008) and Pinto et al. (2020). Additionally, Domingo et al. (2019) found that separating the target from the masker by just 15° azimuth was beneficial to listening to a target when the target sentence was spoken in a familiar voice.

When the stimuli are spatially separated to the extent that each stream is presented solely to one ear (dichotic), EM is removed entirely; listeners can focus attention on one ear and ignore the other (Cherry, 1953; Dalton & Fraenkel, 2012). Therefore, the listener can engage their cognitive resources at the target ear to the point of sometimes experiencing a complete lack of awareness for the masker, a phenomenon referred to as inattentional deafness (Dalton & Fraenkel, 2012) in the masker ear. Cherry (1953) conducted an experiment in which a participant was presented with two passages of dichotic speech and was tasked with focussing on one and ignoring the other. The participant experienced no difficulty in listening to either speech input while rejecting the other. Furthermore, the participant could not recall anything that was presented to the unattended ear and did not notice when the unattended message switched languages from English to German.

Listeners' ability to direct their attention towards one sound source and away from another in a different location has also been demonstrated by Dalton and Fraenkel (2012). They presented participants with a binaural recording in which it sounded as though two women were conversing to one side of the participant and two men were conversing to the other side. Experimenters asked half the participants to listen to the women's conversation, and the other half to listen to the men's conversation. During the recording, another talker "entered" from one side of the auditory scene, repeating the phrase "I am a gorilla". When asked if they noticed anything unusual about the scene, most participants attending the women's conversation (far from the gorilla) did not report hearing the gorilla, whereas most attending the men's conversation (close to the gorilla) did.

These studies demonstrate that a listener can focus on target stimuli and ignore maskers more successfully when the targets and maskers are separated in space. This is particularly true when the sounds are fully dichotic, and thus when there is no EM, as they can attend to one ear and completely ignore the other. Listeners are unable to do this when stimuli are collocated, due to the requirement of parsing the sounds into separate streams.

1.2.2. The effect of energetic masking on L2 listening

Evidence suggests that L2 processing is more cognitively demanding than L1 processing (e.g., Kilman et al., 2014). When an L2 is acquired later in life, it often has an overall lower proficiency level and/or is used less often than the L1. According to the BIA+ (Dijkstra & van Heuven, 2002) and BLINCS (Shook & Marian, 2012) models, this lower frequency of use of L2 means that individual L2 word representations have lower resting activation levels than individual L1 word representations. Consequently, L2 words often take longer to process than L1 words. Eye-tracking studies by Weber and Cutler (2004) and Marian and Spivey (2003) demonstrate that, while both L1 and L2 words compete for activation during spoken word recognition, L1 words are more readily activated. This means that, as well as being processed more quickly per se, the L1 can also interfere more strongly when processing the L2, making it more cognitively demanding for a bilingual to focus on L2 when L1 is the masker than vice versa. This is supported by evidence that L1-based phonological competitors are activated during L2 word recognition, as reflected in increased fixations in eye-tracking studies, whereas comparable effects are not observed during L1 processing (e.g., Weber & Cutler, 2004).

This L2 delay (Dijkstra & van Heuven, 2002) may be exacerbated under adverse listening conditions, such as when EM is present. This was first discussed by Mayo et al. (1997), who conducted a speech in noise test with English-speaking monolinguals, early bilinguals (their L1 was Spanish, and they learnt fluent English before the age of six years), and late bilinguals (their L1 was Spanish, and they learnt fluent English after the age of 14 years). Bilinguals listening to their L2 were more disadvantaged by the addition of a babble masker compared to monolingual listeners, and this effect was magnified for late bilinguals.

This finding that EM negatively impacts listening to L2 more than L1 has been widely replicated (e. g., Black & Hast, 1962; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Gat & Keith, 1978; Zinszer et al., 2019). Researchers have suggested that EM negatively impacts L2 more than L1 because L2 listeners are having to cope with imperfect signals caused by masking as well as their reduced knowledge of the language (Garcia Lecumberri et al., 2010). Indeed, a higher L2 proficiency has been associated with improved L2 speech perception in noise (Kilman et al., 2014). Furthermore, L1 listeners experience less long-lasting disruption from EM and can more effectively overcome its interfering effects, demonstrating that the negative effect of EM is weaker for L1 compared to L2 (Cutler et al., 2008).

As research has demonstrated that L2 is more negatively impacted by EM (e.g., Cooke et al., 2008), it is possible that spatial release from EM (SRM, i.e., reducing EM through spatial separation) may benefit L2 more than L1. This was investigated by Hui et al. (2021) in a selective listening study. Their L1 participants were English speakers and their L2 participants had a variety of L1s and had moved to an English-speaking country before the age of seven, so were assumed to have a high L2-English proficiency. The target speech consisted of nonsense English sentences and was always played from a loudspeaker at 0° azimuth. A pink noise masker was presented at varying locations from speakers around the participant at a signal-to-noise ratio (SNR) of -3 dB (i.e., the masker sound level is 3 dB higher than the target sound level). The results indicated that SRM benefited L1 listeners more than L2 listeners.

These results run contrary to the hypothesis outlined previously, namely that SRM would benefit L2 more than L1. Hui et al. (2021) argued that a listener more familiar with a language is better able to take advantage of spatial cues. However, it may also have been the case that the masker’s sound level was too high relative to the target (low SNR) for the L2 listeners to overcome given L2’s vulnerability to EM. In other words, the relative SNR at the better ear may have still been sufficiently challenging for L2 listeners that no benefit for spatial separation was observed.

Ezzatian et al. (2010) conducted a similar selective listening experiment using a speech masker. Their L1 participants were native English speakers, and their L2 participants had arrived in an English-speaking country between the ages of seven and 14 and judged themselves to be fluent in English. Participants listened to nonsense English target sentences which were perceived as being either in the same location or 180° apart from a masker (also nonsense English sentences). Contrasting with Hui et al. (2021), they found that SRM did not impact L1 and L2 to different extents. While previous research often demonstrates that EM affects L1 and L2 differently (e.g., Cooke et al., 2008), the absence of such a difference in the Ezzatian study could suggest spatial separation may operate differently to simply reducing EM through masker intensity. One possible reason for the discrepancy between Hui et al. and Ezzatian et al. could be the type of masker used: Ezzatian et al. used a competing talker, which tends to produce strong interference in speech perception across L1 and L2 listeners (Cooke et al., 2008). In contrast, Hui et al. used an unintelligible speech masker, which L1 listeners might be better equipped to ignore due to greater familiarity with the target language, potentially leading to different patterns of spatial release between L1 and L2 listeners.

Furthermore, these studies compared different groups of L1 and L2 listeners. Studies often use just one language for both language conditions (L1 and L2) and therefore compare bilinguals using their L2 with monolinguals using their L1 (e.g. Cooke et al., 2008; Ezzatian et al., 2010; Hui et al., 2021; Jin & Liu, 2012; Scharenborg et al., 2018). As a result, observed differences between groups may reflect differences between monolinguals and bilinguals rather than differences between L1 and L2 processing. Supporting this issue, Blumenfeld and Marian (2011) found that while monolinguals and bilinguals performed similarly in a word recognition task (selecting the correct image of an auditorily-presented word from a four-image grid), they performed differently in a subsequent priming probe task, in which the participants were asked to select a grey (rather than black) asterisk. Monolinguals showed slower responses when the probe appeared in the location of a competitor image (a word phonologically similar to the target) than in the location of the target, whereas bilinguals did not show this effect. This suggests that, even when both groups are tested in the same language, bilinguals inhibit previously activated competitor words less strongly than monolinguals, demonstrating a difference in how monolinguals and bilinguals process language, even within the same language. Therefore, this reduces the validity of conclusions based on differences (or a lack thereof) between L1 and L2 processing in studies that use monolinguals as the L1 group and bilinguals as the L2 group. Stronger conclusions about bilingual listening could be drawn from a within-subjects design, with bilingual participants completing a speech in noise task in their L1 and L2, reducing the likelihood of results being found due to extraneous variables (such as differences between monolinguals and bilinguals) rather than due to differences between L1 and L2 processing as such.

Furthermore, the effects of SRM may differ when listening in a dual-language context when both L1 and L2 are present. This has not previously been investigated but is critical to explore as it considers a bilingual's unique need to activate one language while avoiding interference from the other language. Chapter 2 of this thesis will therefore investigate how SRM affects L1 versus L2 processing while these languages are competing for activation. Specifically, the first empirical experiment within this thesis (Chapter 2, Experiment 1) investigated the effect of SRM in L1 and L2 using a selective listening task with a within-subjects design. This addressed the research question: Is the extent of spatial release from EM affected by whether the listener attends to L1 or L2 in a dual-language context? Unbalanced Spanish-English bilingual participants (Spanish as L1) were presented with pairs of sentences, one Spanish and one English, which were dichotic or collocated. Participants attended to their L1 (Spanish) in one condition, and to their L2 (English) in another condition.

With respect to implementing SRM in our experiments in Chapter 2, dichotic listening was used as a way of simulating maximal spatial separation. Although spatial hearing is best implemented using a combination of interaural time differences (ITDs) and interaural level differences (ILDs), ILDs have been shown to be a dominant spatial cue at high frequencies and to partly contribute to the localisation of low-frequency sounds (Middlebrooks & Green, 1991; Middlebrooks & Onsan, 2012). As such, dichotic listening meets our goals of reducing EM and improving streaming through increased perceived left and right (L-R) separation. A similar ILD-only method was used successfully by Knight et al. (2023) to simulate a four-step change in spatial location using L-R mixtures with graded intensity differences, and by Calcus et al. (2015), who compared various dichotic listening conditions with a diotic baseline to evaluate spatial masking release.

Presenting participants with two known languages simultaneously makes this experiment unique compared to previous research. Not only are participants focusing on one speaker and ignoring the other (the standard requirement for speech-in-noise experiments), but they must also control their languages so that they activate one while managing interference from the other (Dijkstra & van Heuven, 2002; Green, 1998; Shook &

Marian, 2012). This demonstrated the effect of SRM on each language in a dual-language context.

This research is important because it advances our understanding of bilingual speech processing mechanisms. While prior studies have shown that SRM benefits target transcription in monolinguals and bilinguals in a single-language context, this study explores whether spatial cues continue to aid listening when the speech streams are already more distinct due to language differences. Moreover, the findings will clarify whether differences in L1 and L2 processing, as proposed by the bilingual word processing models (e.g., BIA+ and BLINCS), result in different effects of spatial separation on the two languages.

1.3. The costs and benefits of spatial separation under divided attention

1.3.1. Divided listening and the attentional spotlight

While selective attention refers to the allocation of attention to a single stimulus while ignoring competing information (e.g., Bater & Jordan, 2020), divided attention involves distributing attention across multiple concurrent stimuli or tasks (e.g., Kahneman, 1973). SRM improves listening ability when selectively attending to one stimulus and ignoring another. However, if the listener is attending to *two* simultaneous stimuli, then the beneficial effect of SRM may be partially offset by increasing cognitive demand in spatial attention control. Mattys et al. (2012) class cognitive load as an adverse listening condition, specifically, a receiver limitation. This means that cognitive load can impact listening performance due to the extra demands placed on a limited pool of cognitive resources that are required for successful listening. Therefore, this introduces the question: when spatially separating two stimuli, does the benefit of SRM outweigh the cognitive costs of spatial separation?

The increased cognitive demand when attending to two spatially separated stimuli can be interpreted within the Attentional Spotlight Theory, posited by Posner et al. (1980). When paying attention to a stimulus, individuals direct their focus of attention – their "attentional spotlight" – towards the area in which the stimulus occurs and become unaware of stimuli occurring in locations outside of that area. This was demonstrated in the Cherry (1953) and Dalton and Fraenkel (2012) studies mentioned previously - participants were able to fully focus their spotlight on one region of space and become attentionally deaf to other regions. However, this becomes problematic when there are two to-be-attended signals in two different locations, as the attentional spotlight must be divided, broadened or flickered between the two stimuli.

1.3.1.1. The visual attentional spotlight

Researchers have debated how we manipulate our attentional spotlights to attend to two spatially separated visual stimuli. Fernandes et al. (2011) presented participants with a central image with digits on the periphery, their task was to determine whether the digits were even or odd. They found that the attentional spotlight broadened and narrowed depending on the emotional content of the central image: negative images led to less accurate decisions made about the digits, suggesting the attentional spotlight was narrower in these circumstances. This indicates that the attentional spotlight can be broadened to encompass spatially separated stimuli, but the ease of this depends on the content of the stimuli.

Other research contradicts this, instead stating that the spotlight must be moved between stimuli. Gabbay et al. (2019) conducted an experiment in which participants searched for targets in varying locations based on colour. They concluded that attention must be redirected to a new location for 300 ms before it is entirely disengaged from its previous location, meaning that attention is temporarily split between two locations before the attentional spotlight focuses completely on the new location. This conclusion has been corroborated (e.g., Müller et al., 2003). If the conclusions from the visual domain can be generalised to the auditory domain, then, for 300 ms, listeners may effortlessly pay attention to spatially separated stimuli simultaneously. However, this would only be useful for listening to individual phonemes, benefiting the listener minimally if they are attempting to listen to longer stimuli, such as sentences.

The idea of a moving spotlight is supported by Van Rullen et al. (2007). In their cued visual target detection task, they found that participants serially attended to different cued locations whilst awaiting the target, rather than focussing on multiple locations at once. Therefore, they suggest that the spotlight is not broadened, but that it flickers between targets, and only one spatial location can be attended to at a time. Furthermore, research indicates that the ability to shift attention between spatial locations is associated with cognitive skills (i.e., working memory capacity), as those with high working memory capacities were better able to shift their attention between spatially separated stimuli (e.g., Kane et al., 2001). If conclusions about visual attention can be generalised to auditory attention, attending to two simultaneous spatially separated talkers would rely on the listener’s ability to rapidly shift their attention between the two talkers, which may require cognitive control (e.g., Kane et al., 2001).

1.3.1.2. The auditory attentional spotlight

Although there is less research investigating the notion of an attentional spotlight in the auditory domain than the visual domain, there is a long history of research on attention ear-switching. This refers to a situation in which participants must switch their focus between the two ears – effectively forcing the participant to flicker their attentional spotlight between spatial locations. Treisman (1971) presented participants with an auditory list of digits, which were either presented from the same location, or alternated between ears. The listener was tasked with recalling the list of digits, and recall was found to be less accurate when the digits were alternating between ears compared to when they were presented from the same location, with maximal costs when the switching rate was high. Similar studies corroborate this finding (e.g., Axelrod et al., 1968; Huggins, 1964; Samuel, 1991), and neurological evidence shows that children with normal hearing (but not those with hearing loss, even when using hearing aids) exhibit preparatory brain activity prior to spatially shifting their attention (Holmes et al., 2017). The observation that performance declines when attention must be redirected across spatial locations, combined with the presence of anticipatory neural activity, suggests that shifting spatial attention imposes a cognitive load.

Colflesh and Conway (2007) found that increased working memory capacity was associated with improved ability to attend to two dichotic auditory stimuli. Participants were required to attend to a message in one ear, but asked to respond when their name was spoken in their other ear. Those with a high working memory capacity detected their name 66.7% of the time, compared to those with low working memory capacity, who detected their name 34.5% of the time. This result suggests that individuals with more cognitive resources in the form of higher working memory capacity are better able to overcome the increased cognitive demand associated with listening to dichotic stimuli.

Overall, research demonstrates that listening to two simultaneous talkers who are spatially separated is cognitively demanding (e.g., Treisman, 1971) due to the need to shift attention between two locations, and this can reduce listening performance. However, working memory capacity can mitigate these cognitive costs of spatial separation. Those with a higher working memory capacity are better able to flexibly control their attention between two spatial locations, allowing them to listen to dichotic stimuli more effectively.

1.3.2. Spatially manipulating energetic masking and cognitive demand

1.3.2.1. Divided auditory attention in L1 listening

Divided listening tasks allow us to investigate the trade-off between two aspects of spatial separation: SRM and increased cognitive demand. If listeners are better able to attend to two talkers who are collocated compared to dichotic, this indicates that the cognitive costs of spatial separation outweigh SRM. However, if listeners are better able to attend to two talkers who are dichotic compared to collocated, then the benefits of SRM outweigh the cognitive costs of spatial separation.

In their divided listening study, Pinto et al. (2020) found that listening to spatially separated stimuli (presented at 40° and -40° azimuth) resulted in more accurate transcriptions of the target than in a collocated condition. Ihlefeld and Shinn-Cunningham (2008b) report a similar finding. These results suggest that SRM outweighs the cognitive costs of spatial separation. However, neither study used *dichotic* stimuli for their spatial separation conditions: Pinto et al.’s (2020) stimuli were 80° separated and Ihlefeld and Shinn-Cunningham’s (2008b) stimuli were 90° separated. According to Posner’s (1975) Attentional Spotlight Theory, the more the stimuli are separated, the more cognitively challenging it becomes to attend to both stimuli. This is because greater separation implies a larger attentional shift between locations, which may increase the time and effort required to reorient attention between stimuli. Therefore, it is possible that when cognitive load is further increased and EM is further decreased with increased spatial separation, different results may be found. If the stimuli are presented at 180° separation (e.g.., one sound in each ear, dichotic presentation), the levels of EM are minimised (simultaneously, cognitive costs may increase), potentially resulting in increased transcription accuracy.

Best et al. (2006) conducted a divided listening study with targets 0°, 30°, 60°, 90°, 120°, 150°, or 180° azimuth separated from each other. Participants were best able to attend to both stimuli simultaneously at 90° to 120° separation, with performance dropping as the targets became closer together (increased EM) or further apart (increased cognitive load). This indicates that listening performance is optimal at an intermediate level of spatial separation, when EM and cognitive demand levels are both moderate, rather than one absent and the other extreme. However, Best et al.’s stimuli were edited to be spectrally interleaved, which means that EM was lower than is usual in natural multi-talker listening conditions. Thus, EM was low even in the collocated condition. It is possible that EM no longer had an effect at 90° to 120° separation, and from that point, increasing spatial separation only increased cognitive costs, but no longer reduced EM. Therefore, performance may have dropped beyond that point because participants were no longer further benefiting from SRM but were further disadvantaged by the increased cognitive demand.

A similar divided listening study addressing this question was conducted by Knight et al. (2023). In one experiment, they used unfiltered stimuli which exerted natural EM, thus ensuring that spatially separating the stimuli would reduce EM. They tested listening performance under four levels of spatial separation: collocated, near (presented at -30° and 30° azimuth), far (presented at -60° and 60° azimuth), and dichotic (described by the authors as -90° and 90° azimuth, with one sentence presented to each ear). They found that as spatial separation increased, performance also increased, with maximal performance in the dichotic condition, demonstrating that the positive effect of SRM outweighed the cognitive costs of spatial separation.

In a second experiment, Knight et al. (2023) filtered the stimuli into non-overlapping frequency bands, neutralising EM in all conditions. This time, after completing the same divided listening task with the new filtered stimuli, there were no differences in performance between the collocated, near and far conditions, but there was a significant drop in performance in the dichotic condition. This suggests that controlling attention between two ears is more cognitively demanding than focusing on two stimuli which originate from similar spatial locations. This second experiment therefore suggests that the results from the first experiment do not indicate a total lack of cognitive costs in spatially separated conditions, but, rather, that those costs are outweighed by the benefits of SRM. Furthermore, Knight et al. (2023) found that working memory capacity was positively correlated with listening performance only in their second experiment, when EM was completely removed in all spatial conditions. This indicates that the listener can make use of cognitive resources to improve listening performance when more acoustic information is available (i.e., EM is removed).

1.3.2.2. Divided auditory attention in L2 listening

While there is some evidence that the beneficial effects of SRM outweigh the cognitive costs of spatial separation during divided-attention listening in monolingual participants in a single-language context, there is no research investigating this in bilingual listeners in a dual-language context. The theoretical importance of this becomes apparent when considering the trade-off between EM and cognition. As discussed earlier in this review, L2 is more negatively impacted by EM than is L1 (e.g., Cooke et al., 2008). This suggests that reducing EM through spatial separation would benefit L2 to a greater extent than L1, although this prediction does not align with previous research investigating SRM in L1 compared to L2 (Ezzatian et al., 2010; Hui et al., 2021). However, those studies do not consider the cognitive costs associated with spatial separation when attending to two simultaneous stimuli. As L2 is more cognitively demanding to process than L1 (Dijkstra & van Heuven, 2002), it is possible that increasing cognitive demand further by spatially separating the stimuli may negatively impact L2 to a larger extent than L1. As L2 could benefit from SRM more than L1, but could also be more negatively impacted by the cognitive costs of spatial separation, it is unclear how spatial separation might affect the two languages differently. Therefore, the second empirical experiment in this thesis (Chapter 2 Experiment 2) addressed the following research question: How does the reduction of EM and increased cognitive demand caused by spatial separation impact L1 and L2 in a dual-language context?

This research question was investigated using a divided listening task, in which participants were presented with two simultaneous sentences, one L1 (Spanish) and the other L2 (English). Each pair of sentences were presented as collocated or dichotic.

Participants were told which talker to transcribe after hearing each sentence pair, therefore had to attend to both talkers simultaneously to perform successfully.

One key question this study explored is how SRM interacts with cognitive load in a dual-language context. Prior research with monolinguals suggests that the benefits of SRM outweigh the cognitive costs of spatial separation (e.g., Knight et al., 2023). However, the dual-language aspect of this study could alter that balance. Because of the acoustic-phonetic and prosodic differences between languages (e.g., Van Engen & Bradlow, 2007), using a dual-language design reduces EM across both spatial conditions, but cognitive demand is increased overall across both spatial conditions due to the need to process two languages rather than one. This is indicated by interpreting studies showing that performance at processing two languages simultaneously when interpreting increases with working memory capacity (e.g., Christoffels et al., 2006). This could increase the likelihood of observing a reversal of the typical trade off (i.e., rather than the reduction of EM outweighing the cognitive costs of spatial separation, the opposite pattern may be found). This experiment provides insight into bilinguals’ ability to comprehend two languages simultaneously, as opposed to attending selectively to one while ignoring the other, and clarifies how increased cognitive demands influence how effectively listeners use spatial cues when listening.

1.4. Semantic relatedness between the masker and the target

1.4.1. Introducing informational masking

Thus far, this thesis has focused on EM, and how it can be reduced through spatial separation. Chapter 3 focuses on informational masking (IM) and returns to selective (rather than divided) attention. Having considered the effect of spatial separation on EM in L1 versus L2, this section considers the effect of semantic release from masking in L1 versus L2. IM is extremely broad, so this section focuses specifically on the semantic content of the masker - how semantic relatedness between a masker and target can impact the ability to accurately attend to that target. We define semantic release from masking as the hypothetical performance improvement when the semantic overlap or relatedness between the target and masker speech is reduced. Our main question is whether semantic release from masking affects listening performance differently when the masker is L1 compared to when the masker is L2. To define IM, Shinn-Cunningham (2013) states that, while EM is disruption of the target speech at the auditory periphery, IM represents any effects of the masker which have not been accounted for by EM. For example, competing speech may capture the attention of the listener, cause semantic interference, or increase cognitive load, which are higher-level post-periphery consequences of the masking (Kidd et al., 2008).

A masker can have minimal EM, but still exert IM (Shinn-Cunningham, 2008), for example, when the acoustic information of the target is completely available to the listener, but another sound (which does not necessarily overlap with the target spectro-temporally) distracts the listener and prevents them from attending to the target. There may also be situations in which the listener can parse the two sounds (i.e., separate them into separate streams), but fail to allocate attention to the target stream, resulting in incorrectly perceiving portions of the masker as the target speech (Cooke et al., 2008).

1.4.2. The impact of semantic content of the masker

1.4.2.1. Semantic processing of unattended speech in selective attention

When selectively listening to a target, the masker may interfere based on how semantically related it is to the target. For example, the semantic content of the masker might attract the listener’s attention, shifting their focus from the target, or it might cause semantic confusion, making it difficult to distinguish which words belong to the target versus the masker. Both cases represent instances of IM, where interference arises from the semantic content of the competing speech. Reducing the semantic relatedness between the target and masker may therefore lead to a release from semantic interference, making it easier for the listener to separate the two speech streams. However, this depends on the assumption that listeners involuntarily process the semantic content of background speech while selectively attending to a target.

An early theory of attention relating to the semantic content of background speech is Broadbent’s (1958) Filter Theory. This posits that an early-filter mechanism prevents unattended stimuli from being processed beyond their basic physical properties. Therefore, in the case of selective listening, if a target is being attended to, the semantic content of the background speech should not be processed. Filter Theory has been supported by so-called “cocktail party” experiments demonstrating that when attention is focused on a target speaker, many high-level properties of background speech remain unnoticed by the listener (e.g., Cherry, 1953; Dalton & Fraenkel, 2012). However, further experiments indicate that some participants do notice if their name appears in the background speech (e.g., Moray, 1959; Röer & Cowan, 2021; Wood & Cowan, 1995). Although this effect is only found for some participants (around 33%), and primarily in participants with a lower working memory score (Conway et al., 2001), it does not align with Filter Theory, because, for the listeners to have noticed their name, they must have processed more than simply the physical properties of the background speech.

As Broadbent's (1958) Filter Theory cannot account for this, Treisman's (1964) Attenuation Theory was developed. This stipulates that rather than an early filter which completely filters out the semantic content of the irrelevant speech, there is a late-filter mechanism which merely attenuates the irrelevant speech. This means that the background speech is co-activated along with the target speech and stored until it is deemed irrelevant and rejected. This would enable some of the background speech to be processed at higher, semantic levels, which means that the semantic content of the background speech could potentially exert IM. This claim is supported by the previously mentioned research showing that some listeners notice their name in the unattended stream (e.g., Conway et al., 2001; Moray, 1959; Röer & Cowan, 2021; Wood & Cowan, 1995). Additionally, Attenuation Theory is supported by research demonstrating that, when asked to tap in response to certain words (i.e., a prime word) in a target stream, participants occasionally also respond to these prime words when they appear in the unattended message (e.g., Treisman & Geffen, 1967; Treisman & Riley, 1969). This indicates that listeners unintentionally processed the unattended stream to recognise the prime words from that task-irrelevant stream.

Lavie’s (1995) Perceptual Load Theory (or Load Theory) offers an intermediate account, proposing that the extent to which distractors are processed depends on the availability of perceptual resources. Under low-load conditions, spare attentional capacity leads to increased processing of task-irrelevant stimuli, resulting in greater distractor interference. In contrast, under high-load conditions, attentional resources are fully engaged by the task, leaving fewer resources available for processing irrelevant information and reducing distractor processing. Lavie (1995) demonstrated this pattern in visual attention, where distractors impaired performance under low-load conditions but had a smaller effect under high-load conditions. Load Theory was developed for visual attention, and Murphey et al.’s (2017) review of its application to the auditory domain found mixed results. Some studies supported load-depended distractor processing during listening (e.g., Francis, 2010), whereas others report interference from irrelevant sounds regardless of load (e.g., Murphey et al., 2013).

While it remains unclear whether co-activation of the masker depends on task load, some research (e.g., Conway et al., 2001; Moray, 1959; Röer & Cowan, 2021; Treisman & Geffen, 1967; Treisman & Riley, 1969; Wood & Cowan, 1995) does indicate that when selectively attending to a target talker, listeners may also semantically process to-be-ignored masker speech. However, these studies used maskers consisting of individual words (such as the listener’s name or a prime word). In speech-on-speech situations, semantic interference might be stronger because naturalistic sentences offer more opportunities for semantic confusion. Alternatively, individual words that listeners are primed to detect might more effectively capture attention (even in the masker stream) and disrupt listening to the target more than full, naturalistic sentences.

1.4.2.2. Semantically meaningful versus anomalous maskers

Brouwer et al. (2012) investigated how semantic content of the masker impacts listening performance in a speech-on-speech experiment, and thus whether or not the semantic content of the background speech is processed. They presented participants with a semantically meaningful target sentence (e.g., *He played with his train*) with simultaneous two-talker babble, which either contained semantically meaningful (e.g., *Rice is often served in round bowls*) or semantically anomalous (e.g., *The great car met the milk*) sentences. They found that the targets were recalled more accurately when they were presented with a semantically anomalous masker compared to a semantically meaningful masker. These results suggest that to-be-ignored speech is processed at the semantic level, as the semantic content of the masker speech impacted the listener's ability to understand the target. However, other studies have only found this effect in older adults (Rossi-Katz & Arehart, 2009; Tun et al., 2002), although it is noted that Tun et al.'s (2002) younger participants did recognise more items from the meaningful distractors than the anomalous distractors in a subsequent surprise recognition test. This suggests that the younger participants did process the to-be-ignored maskers at the semantic level but could prevent any "in-the-moment" semantic interference from impacting overall listening performance.

Additionally, Calandruccio et al. (2018) conducted a similar experiment to Brouwer et al. (2012) and found that the semantically meaningful babble was no more distracting than the semantically anomalous babble. Calandruccio et al. (2018) claim that the discrepancy between their findings and those of Brouwer et al. may be due to differences in the control and design of the masker stimuli. Specifically, in Brouwer et al.'s study, all semantically anomalous masker sentences shared the same syntactic structure (they were structurally identical to each other) and were composed exclusively of monosyllabic words. In contrast, the semantically meaningful maskers varied in both syntactic structure and word length, including a mix of monosyllabic and disyllabic words. Calandruccio et al. (2018) argue that these differences between the meaningful and anomalous maskers may have been an extraneous variable impacting Brouwer et al.'s results. The anomalous maskers, by being more internally consistent and predictable in structure, may have been easier to ignore, whereas the meaningful maskers, being more variable, may have attracted the listener's attention, not because of their semantic content, but because of their lack of rhythmic or syntactic predictability. This raises the possibility that Brouwer et al.'s observed difference between meaningful and anomalous maskers may be due to these structural differences, creating a potential confound.

1.4.2.3. Semantic relatedness between target and masker

The studies mentioned in the previous section focused on semantic meaningfulness, comparing semantically meaningful maskers (e.g., *He played with his train*) to semantically anomalous maskers (e.g., *The great car met the milk*). In contrast, the present investigation focuses on semantic similarity between the target and the masker. For example, for the target *The cat casually scratches the chair*, a semantically similar masker might be *The dog anxiously gnaws the table*, while a semantically dissimilar (neutral) masker might be *The lily slowly opens its petals*.

Research has shown that anomalous sentences can disrupt short-term memory by demanding additional time and cognitive resources to resolve the unexpected content (Röer et al., 2019). As a result, comparing meaningful and anomalous masker speech introduces a confound: meaningful sentences may cause semantic confusion, whereas anomalous sentences may be disruptive because they are unexpected. This overlap could obscure genuine differences between conditions. To avoid this confound and to better reflect real-world listening scenarios, in which anomalous speech is rare, we focused on meaningful masker sentences that are either semantically similar or semantically dissimilar to the target. This approach provides increased validity in testing whether semantic content of background speech interferes with speech perception.

So far, there are few studies comparing how semantic relatedness between target and masker affects target processing. The most relevant study is by Villard et al. (2024). Participants were presented with a target sentence, always in the structure *Betsy sees a [noun]*, simultaneously with a masker sentence, either in the structure *Donna gives a [noun]* or *Molly wants a [noun]*. Participants were then asked to select the image that depicted the noun Betsy saw. Within each trial, the masker noun was either semantically related to the target noun (e.g., both nouns were birds) or semantically different (e.g., the target noun was a bird, and the masker noun was an item of clothing). This study showed no significant difference in accuracy between the two conditions, suggesting that the semantic content of the irrelevant noun was not sufficiently processed for it to exert IM, supporting the idea of an early-filter mechanism (e.g., Broadbent et al., 1958) preventing the unnecessary processing of irrelevant stimuli.

Before accepting their conclusion, however, there are aspects of Villard et al.'s (2024) experiment to consider. Their focus was to test whether aphasic participants and aged-matched controls differed in linguistic-semantic processing. Their sample consisted of eight individuals suffering with aphasia and eight age-matched controls - the lack of effect of masker semantic relatedness was found for both groups. The small sample size may have resulted in a lack of statistical power, potentially resulting in a type II error. Additionally, the process of selectively listening to a target and ignoring a semantically related (or unrelated) masker may operate differently in aphasic and older participants compared to young, healthy individuals. Furthermore, performance was measured on only one keyword at the end of very short sentences, whereas a task involving more complex and naturalistic sentential semantics may produce a different result. For example, longer sentences with more keywords provide the listener with more chances to be distracted by the masker, allowing for more variation in listening performance.

1.4.3. The semantic relatedness effect in a dual-language context

Research investigating the effect of masker-target semantic overlap on listening performance is limited and inconsistent. However, some research has demonstrated that semantic meaningfulness of a masker impacts listening performance (e.g., Brouwer et al., 2012), so it is possible that a to-be-ignored masker is processed at the semantic level. Therefore, a masker which semantically overlaps with the target might cause more confusion and interference compared to a neutral masker when the listener recalls the target.

It is also important to consider whether unattended background speech is semantically processed when it is in the listener's L2 versus L1. Bilingual language processing models, such as the BIA+ (Dijkstra & van Heuven, 2002) and BLINCS (Shook & Marian, 2012) suggest that both languages are automatically activated in parallel during speech processing. However, these models also assume that L2 words typically have lower resting activation levels than L1 words, resulting in slower and weaker activation of L2 words – this slower recognition is known as L2 delay.

Hierarchical models of bilingualism posit that there is cross-language semantic priming (e.g., French & Jacquet, 2004), with the Revised Hierarchical Model (Kroll & Stewart, 1994) suggesting that semantic concepts in the L2 are processed through the L1, particularly for bilinguals with low L2 proficiency. This suggests that the semantic content of a masker may to some extent be processed similarly in an L1 and L2, but more slowly in the L2 given its route through the L1. Further research shows that semantic information is indeed processed more slowly in the L2 than in the L1 (Dijkgraaf et al., 2019), and that semantic priming in the L2 takes longer to develop (Wen & van Heuven, 2017). The BIA+ and BLINCS models also posit that semantic representations are slower to activate in the L2 than L1 (Dijkstra & van Heuven, 2002; Shook & Marian, 2012). As a result, the semantic content of an L2 masker may interfere less with target speech than the semantic content of an L1 masker, since its semantic content may not be activated quickly enough to cause semantic confusion between masker and target.

These differences in semantic processing between the L1 and L2 may depend on the proficiency levels in each language. Cai et al.’s (2021) Chinese-English bilingual listeners were asked to name images in their L2 (English) which were presented simultaneously with semantically related or neutral L2 auditorily presented distractor words. Only the bilinguals highly proficient in their L2 performed significantly more slowly when the distractor was semantically related compared to neutral. This suggests that as language proficiency increases, listeners are more likely to semantically process irrelevant L2 speech, increasing the potential for semantic interference from the L2. However, it should be noted that this study focused on language production rather than processing, and there is no research comparing how much semantic processing an L2 compared to an L1 masker receives when processing an L1 target (i.e., comparing a dual- and single-language context).

Beyond influences of semantic content in the L1 versus L2 masker, in general, it is well established that, in the speech-perception domain, L2 maskers can interfere with L1 target processing, although typically to a lesser extent than L1 maskers. This reduced interference may stem from (at least) two distinct sources: differences in acoustic similarity between target and masker speech, and differences in resting activation levels between the L1 and L2.

At the acoustic level, the Target-Masker Similarity Hypothesis (e.g., Brouwer et al., 2012; Mepham et al., 2022) posits that listeners find it more difficult to segregate two streams of speech when they are acoustically similar. More specifically, they claim that a masker of the same language as the target may cause greater disruption than a masker in a different language to the target, due to the higher phonological similarity (Pigoli et al., 2018). This prediction is supported by Smith et al. (2024), who found that both English monolingual and Mandarin-English bilingual individuals were more negatively impacted by a masker which was the same language as the target than a masker that differed from the target language. This effect was found both when the target and masker were L1 (Mandarin) and when they were L2 (English), highlighting that the similarity between target and masker languages is more important than language dominance to determine degree of masking.

Another explanation of an L1 masker being more disruptive to an L1 target than an L2 masker involves the relative resting activation levels of words in L1 and L2, as proposed by the BIA+ and BLINCS models. These models assume that words in both languages are processed in parallel and compete for activation, but L1 words typically have higher resting activation levels due to greater frequency of use and proficiency. This makes L1 words faster to activate and more likely to interfere, potentially including when they are not the focus of attention. This is supported by research demonstrating that the L1 interferes with the L2 more than vice versa (e.g., Marian & Spivey, 2003). This leads to the next research question within the empirical chapters of this thesis: Does the semantic content of L1 and L2 maskers impact listening to an L1 target?

In Chapter 3, bilingual participants completed a selective listening experiment in which they were presented with an L1 target simultaneously with an L1 or L2 masker. The masker was either semantically competing with the target (e.g., *The cat casually scratches the chair* and *The dog anxiously gnaws the table*) or semantically unrelated (neutral) to the target (e.g., *The cat casually scratches the chair* and *The lily slowly opens its petals*).

This study revealed whether listeners inadvertently process the semantic content of masker sentences, even when selectively attending to a target sentence. Such findings help distinguish between Treisman's (1964) Attenuation Theory, which proposes that unattended information is still partially processed and can activate semantic representations, and Broadbent's (1958) Filter Theory, which posits that unattended input is filtered out early and does not undergo semantic processing. Furthermore, if semantic interference from the masker is observed, the study tests the bottom-up activation assumptions of bilingual lexical access models, such as BIA+ and BLINCS. Specifically, it assessed whether L2 maskers produce less semantic interference than L1 maskers, consistent with the idea that L2 words have lower resting activation levels and slower access to semantic representations compared to L1 words.

As previously described, Tun et al. (2002) found that while the semantic content of the masker speech did not impact listening performance (except in older participants), the semantically meaningful maskers were better remembered than the semantically anomalous maskers (by the younger participants). This indicates that the semantic content of the maskers was indeed processed; however, listeners used a mechanism (potentially increased attentional control) to prevent this semantic content from interfering with their ability to listen to the target. Therefore, the final empirical experiment in this thesis (Chapter 3 Experiment 2) made use of a surprise recognition task to investigate whether semantically competing maskers are better recognised than neutral maskers. If the semantically competing maskers are better recognised than the neutral maskers, this will provide further support for the hypothesis that (a) the semantic content of background speech is co-activated with the target and semantically processed and (b) that this can potentially lead to semantic interference.

1.5. Summary

The aim of this thesis is to explore how bilingual listeners use spatial and semantic cues to enhance listening performance in adverse listening conditions. Chapter 2 examines how the trade-off between spatial release from energetic masking (EM) and the cognitive costs of spatial separation affects bilinguals listening in a dual-language context – which has not previously been investigated. This demonstrates how bilinguals exploit spatial cues to mitigate the negative effects of EM, and whether this is still effective under increased cognitive demands, such as when the listener attends to two simultaneous talkers. Chapter 3 addresses the impact of informational masking (IM). It investigates whether the semantic content of background speech is unintentionally processed and exerts IM, and whether this effect varies depending on the language of the masker. This has not previously been investigated in a speech-on-speech experiment using complex and naturalistic stimuli, so provides insights into the mechanisms that listeners use when selectively listening. Together, both chapters demonstrate how bilinguals control two languages simultaneously in dual-language contexts and reveals the strategies that they use to parse simultaneous talkers, including making use of lower-level phonological and spatial differences between talkers and higher-level semantic differences.

Chapter 2[1]: Spatial listening in a dual-language context

2.1. Abstract

Spectro-temporal overlap between two simultaneous talkers (energetic masking, EM) reduces intelligibility. However, masking can be mitigated by spatially separating the two talkers. This effect is referred to as spatial release from masking (SRM). Because masking, and signal degradation in general, negatively impacts second language (L2) more than first language (L1) processing, we hypothesised that L2 might benefit more from SRM than L1 when the listener must selectively attend to the target and ignore the masker. Under divided attention, however, spatial separation might introduce cognitive costs due to the need for spatial attentional control. Such costs might be particularly pronounced for L2 given the already high cognitive demands associated with L2 processing and therefore mitigate the L2 SRM benefit. Using a dual-language context, Spanish (L1) - English (L2) bilinguals heard one English sentence and one Spanish sentence simultaneously, either collocated or dichotic (one sentence in each ear), with dichotic listening simulating maximal spatial separation. Participants completed a selective-attention task (track one talker – Experiment 1) and a divided-attention task (track both talkers – Experiment 2). In both experiments, performance was higher for L1 than L2. SRM (better performance in dichotic

---

[1] This chapter has been adapted from: Allen-Rice, E., Knight, S., de Bruin, A., & Mattys, S. (in press). Selective and divided listening in a dual-language context. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218251386821

than collocated listening) was similar for L1 and L2. SRM was smaller under divided than selective attention, suggesting that increased cognitive costs can reduce SRM. The results show that bilinguals benefit from SRM equally in L1 and L2, even in dual-language contexts and under high cognitive load.

2.2. Introduction

In a busy room, it can be challenging to listen to one talker while ignoring competing talkers nearby. This challenge is due in part to energetic masking (EM), the spectrotemporal overlap between a masker talker and a target at the cochlear level (Brungart, 2001; Pollack, 1975). However, EM is reduced when sound sources are spatially separated: at a given ear, there is an increase in the number and duration of spectro-temporal regions in which the energy in the target exceeds that in the masker (Edmonds & Culling, 2006); listeners can also use interaural cues (such as level and timing differences) to distinguish different streams. Spatial separation can also reduce informational masking (IM) and higher-level interference (e.g., Pollack, 1975) due to, for example, misallocation of informational content across the two speech streams (Oh et al., 2021)[2]. Improved intelligibility resulting from spatial separation is referred to as spatial release from masking (SRM, e.g., Litovsky, 2012).

SRM is largest when two sound sources are maximally separated, with the target presented to one ear and a masker to the other (dichotic presentation, Calcus et al., 2015). Listeners can benefit from SRM both when they know in advance which speaker to focus on (selective listening, Ihlefeld & Shinn-Cunningham, 2008a) and when they must listen to the two talkers simultaneously (divided listening, Best et al., 2006; Ihlefeld & Shinn-Cunningham, 2008b; Knight et al., 2023; Pinto et al., 2020). Such effects are well established for monolinguals, but less well understood for bilinguals. In bilinguals, masking appears to negatively impact second language (L2) processing more than first language (L1) processing

---

[2] As this study focuses on SRM in a dual-language context (i.e., two speech streams in different languages), IM through misallocation of information between the speech streams is expected to be low. Therefore, this paper focuses primarily on SRM as a reduction in EM.

(e.g., Black & Hast, 1962; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Gat & Keith, 1978; Mayo et al., 1997; Zinszer et al., 2019). While some research indicates that this effect is reduced when the task is restricted to simple tasks such as phoneme processing (e.g., Cutler et al., 2004; for comparisons of different types of bilinguals, see Flege & Liu, 2002; MacKay et al., 2001; but see Cutler et al., 2008 for opposite findings), it has been observed for sentence comprehension tasks across various studies (e.g., Black & Hast, 1962; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Gat & Keith, 1978; Mayo et al., 1997; Zinszer et al., 2019). Therefore, when processing sentences, SRM could be expected to benefit L2 more than L1. Only two studies have examined this possibility. Both studies compared native (L1) and non-native (L2) listeners in a single-language context (where only one language is used) rather than a single group of bilinguals tested in L1 versus L2 conditions in a dual-language context.

Hui et al. (2021) recruited L1 and L2 speakers of English, of whom the L2 speakers had various L1s and were either born in an English-speaking country, or moved to an English-speaking country before the age of seven. These listeners were presented with meaningless English target sentences (e.g., *A spider will drain a fork*) perceived as coming from straight ahead (0° azimuth) while pink noise was played at 0°, 45°, 90°, 135°, or 180° azimuth. Participants were asked to transcribe the target sentences. The results showed that, while L1 and L2 listeners performed similarly when the target sentences and the noise were collocated (0°; running counter to research demonstrating that EM impacts L2 more than L1), L1 listeners benefited more than L2 listeners from spatial separation. The authors interpreted their results as showing that L1 listeners were better able to make use of spatial cues to improve their listening performance, whereas L2 listeners failed to make use of these cues. These results were supported by a later study which used babble masker speech

(Hui et al., 2022); however, this study involved a comparison of early and late-immersed English speakers, rather than a comparison of L1 and L2 processing.

However, Ezzatian et al. (2010) did not find SRM differences between L1 and L2 speakers. Their sample consisted of L1 and L2 speakers of English, of whom the L2 speakers had various L1s and arrived in Canada from a non-English speaking country after the age of seven years. These listeners were presented with English target sentences in a two-talker masker (with both masker talkers speaking English) or a speech-shaped noise masker. The target and masker were either collocated or perceived as originating from different locations. Participants were asked to repeat back the target sentences. Both listener groups benefited equally from spatial separation, regardless of masker type. The authors suggested that L1 and L2 listeners can benefit equally from stream segregation cues that do not rely on linguistic or semantic knowledge.

The incompatible results of these two experiments could be due to methodological differences such as the masker used (pink noise in Hui et al., 2021; two-talker masker and speech-shaped noise in Ezzatian et al., 2010) and the task (transcription in Hui et al. 2021; immediate verbal recall in Ezzatian et al. 2010). Additionally, both studies used a between-subjects design, with (likely) monolingual participants in the L1 condition and bilingual participants in the L2 condition. Even within their L1 (English), bilinguals have been shown to use different speech processing mechanisms to monolinguals (Blumenfeld & Marian, 2011). Furthermore, between-subjects designs have the disadvantage of allowing variables other than language (L1/L2) to influence the results (e.g., immigration status, number of languages known, country of origin, etc.). To better understand the effects of SRM on L1 versus L2 processing, a comparison of language effects within bilinguals is needed.

So far, the focus has been on how release from EM through spatial separation facilitates listening. However, listening to multiple talkers also requires attentional control. Spivey and Marian (1999) showed that when Russian-English bilinguals were asked to pick up an object in a single-language context, they briefly glanced at a distractor object whose name initially shared phonetic similarities with the target word in the non-relevant language. This suggests that both languages remain active, creating cross-linguistic interference. As a result, bilinguals experience a higher cognitive load associated with attentional control, because they must manage competition and interference between the co-activated languages. This cognitive load might be highest when having to process L2, as listeners need more time to access L2 words than L1 words (e.g., McLaughlin et al., 2004). Furthermore, the L2 can experience more interference from L1 than vice versa (e.g., Green, 1998). Therefore, more cognitive resources might be needed to control interference from L1 while processing L2. Language competition and its associated cognitive load are likely to be even greater in a dual-language context, where bilinguals need to manage the multiple languages they hear (cf. Green & Abutalebi, 2013, for a discussion of language control in single- and dual-language contexts).

While SRM can be beneficial during both selective and divided listening, cognitive load and the need for attentional control are higher during divided listening, and particularly when the talkers are in different spatial locations (Treisman, 1971). If a listener is attempting to listen to two talkers who are spatially separated, shifting attention between the two locations may be cognitively demanding (e.g., Axelrod et al., 1968; Axelrod & Powazek, 1972; Treisman, 1971). This was demonstrated in Knight et al.'s (2023) study using a divided-listening paradigm in which listeners had to attend to two simultaneous talkers. When naturalistic stimuli were presented, listening performance was higher as spatial separation increased, reflecting SRM. However, when EM was eliminated by playing the two voices in non-overlapping frequency bands, listening performance was higher in the collocated compared to dichotic condition. Thus, when spatial separation is not associated with SRM, processing costs are apparent. Knight et al. also used the letter number sequencing task (LNS, Wechsler, 1997) to measure working memory (WM) capacity and found that LNS performance was positively correlated with listening performance in the condition where EM was eliminated. This provides further evidence that spatial control of attention during divided listening is cognitively costly.

Divided listening has been studied primarily in monolingual listeners. However, the potential interplay between SRM and attentional costs might differ between L1 and L2 processing. SRM may benefit L2 more than L1 due to the L2 being more negatively affected by masking than L1 (e.g., Cooke et al., 2008). At the same time (and especially in divided listening situations), the increased need for attentional control created by spatial separation may negatively impact L2 processing more than L1 processing due to the general increase in cognitive demands during L2 processing. To investigate these questions, a selective attention experiment was conducted to assess the effects on L1 and L2 processing of EM alone (and its release through spatial separation; Experiment 1) before introducing the cognitive load component via a divided-attention condition (Experiment 2).

2.3. Experiment 1: Selective Attention

2.3.1. Pre-registration and ethics

The pre-registrations for Experiment 1 (https://osf.io/fv52q) and Experiment 2 (https://osf.io/gqh7r) are available on the Open Science Framework (OSF). Ethical approval for both experiments was granted by the ethics committee at the Department of Psychology of the University of York (ref. 2230). All participants provided informed consent before taking part, and all procedures were performed in compliance with relevant laws and institutional guidelines. Additionally, all participants were paid at a rate of £6 an hour for their completion of each experiment and associated screening.

2.3.2. Hypotheses

The goal of Experiment 1 was to test whether the magnitude of SRM differs during L1 and L2 processing when the two languages are heard in a dual-language context. Spanish (L1) – English (L2) bilinguals heard a Spanish sentence and an English sentence simultaneously and were asked to selectively attend to and transcribe either the Spanish sentence or the English sentence. The two sentences were played either diotically (perceived as being collocated) or dichotically (one sentence in each ear). WM capacity and language background were also measured. Our hypotheses were as follows:

- H1: Performance (i.e., transcription accuracy) will be higher when attending to L1 than L2 (i.e., main effect of language). This is in line with previous work showing poorer L2 than L1 listening (e.g., Spivey & Marian, 1999).
- H2: Performance will be higher when the stimuli are dichotic compared to collocated (i.e., main effect of spatial separation, SRM). This is based on previous research demonstrating SRM in monolingual listeners (e.g., Marrone et al., 2008).
- H3: SRM will be larger for L2 than L1 listening (i.e., interaction between language and spatial separation). This hypothesis aligns with a substantial body of research showing EM is more detrimental to L2 than L1 processing when the task involves sentence processing (Black & Hast, 1962; Cooke et al., 2008; Garcia Lecumberri &

Cooke, 2006; Gat & Keith, 1978; Mayo et al., 1997; Zinszer et al., 2019). This suggests that L2 should benefit more from spatial separation (and thus the reduction of EM) than L1.

2.3.3. Methods

2.3.3.1. Participants

Participants were recruited through Prolific (www.prolific.co; Prolific, 2014) and tested online using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Data from 100 participants were collected, although two participants were removed from analysis due to low performance (performance below 20% correct in any listening condition) in the main listening task. The sample size was chosen based on Brysbaert and Stevens' (2018) rule of thumb of having a minimum of 1,600 observations per combination of conditions in mixed-effects analyses. With 20 trials per combination of conditions (L1-collocated, L1-dichotic, L2-collocated, L2-dichotic), a minimum of 80 participants was required. We aimed to invite the same participants to take part in Experiment 2, so recruited 100 participants in Experiment 1 to allow for attrition.

The 98 participants included in the Experiment 1 analyses were Spanish-English bilinguals (mean age = 25.20 years, SD = 3.56 years; 49 female, 45 male, 3 other, 1 did not disclose their gender identity). All participants started learning Spanish from birth (L1) and English from childhood (L2), and all were highly proficient in both languages, but more so in Spanish (see Table 1 for further details).

A screening procedure was used to ensure participants met the following, pre-registered inclusion criteria: aged between 18 and 35 years inclusive; had no self-reported neurological condition or language or reading difficulties (including dyslexia); had normal hearing and normal (or corrected-to-normal) vision; were living in the UK or Spain at the time of testing; started acquiring Spanish before the age of three years and English aged three or older; had a self-reported English understanding proficiency of at least 5/10 and self-reported English reading, speaking, or writing proficiency of at least 4/10; had lived in Spain for more than five years and lived in an English-speaking country for no more than ten years; scored at least 60% on the LexTALEs (see section 2.3.3.2. Materials, LexTALE and Lextale-Esp); and reported they did not look up words during the tasks. These criteria ensured that Spanish was the dominant language in terms of age of acquisition and proficiency, whilst English was still sufficiently proficient to understand the stimuli in the selective attention task. We also pre-registered the criterion that participants had to have lived in an English-speaking country for at least six months, but ultimately this criterion had to be removed to facilitate recruitment.

Table 1 - Language background of the Experiment 1 sample (N = 98)

Language background measure	**L1-Spanish**				**L2-English**			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Age of acquisition (years)	0.00	0.00	0.00	0.00	5.79	2.40	3.00	14.00
LexTALE proficiency (%)[a, b]	92.30	5.73	70.83	100.00	80.46	8.55	62.50	96.25
Self-reported understanding (/10)	9.97	0.30	7.00	10.00	8.30	1.04	5.00	10.00
Self-reported speaking (/10)	9.99	0.10	9.00	10.00	7.29	1.24	4.00	10.00
Self-reported reading (/10)	9.97	0.30	7.00	10.00	8.89	0.96	5.00	10.00
Self-reported writing (/10)	9.98	0.14	9.00	10.00	7.67	1.25	5.00	10.00
Language use (/4)[c]	Mean = 2.97, SD = 0.47, Min = 1.21, Max = 3.76							

[a] The LexTALE is a standardised test of language proficiency (see section 2.3.3.2. Materials, LexTALE and Lextale-Esp for details).

[b] The L1-Spanish values exclude one participant who scored 5% in the Lextale-Esp but demonstrated high Spanish proficiency throughout the study. Because we assume that they misunderstood the instructions for the Lextale-Esp, we have included their other scores in the analyses.

[c] This value was calculated from the Language and Social Background Questionnaire (see section 2.3.3.2. Materials, Language background questionnaire for details) in which participants were asked to state which language they used in particular situations. 0 = All English, 1 = Mostly English, 2 = Half Each Language, 3 = Mostly Spanish, 4 = All Spanish.

2.3.3.2. Materials

Language background questionnaire. A language frequency of use score was calculated based on a set of questions taken from the Language and Social Background Questionnaire (Anderson et al., 2018). These questions asked participants how often they used each language with different people, in different contexts, and for different activities. Their scores were coded as follows: All English = 0, Mostly English = 1, Half Each Language = 2, Mostly Spanish = 3, and All Spanish = 4. A mean score was calculated for each participant across the 20 included items. In the same questionnaire, we also asked participants to self-rate their language proficiency and age of language acquisition.

LexTALE and Lextale-Esp. The LexTALE (Lemhöfer & Broersma, 2012), a short vocabulary test, was used to measure participants' English proficiency. Spanish proficiency was measured using the Lextale-Esp (Izura et al., 2014). The order in which participants completed the LexTALE and Lextale-Esp tests was counterbalanced. On each trial (60 experimental LexTALE trials, 90 experimental Lextale-Esp trials), participants saw a string of letters, which formed either a word or a non-word in the relevant language. Participants indicated via a button press whether the word was real or not. In each test, participants' score was calculated as:

$$\frac{\left(\frac{number\ of\ words\ correct}{number\ of\ words\ in\ total} \times 100\right) + \left(\frac{number\ of\ nonwords\ correct}{number\ of\ nonwords\ in\ total} \times 100\right)}{2}$$

Letter number sequencing task. The letter number sequencing task (LNS) was used as a measure of WM (Wechsler, 1997). The LNS was designed for verbal administration. However, a visually administered version was used in this study to avoid specifying a particular language for task completion and reducing language biases in bilinguals (cf. Mielicki et al., 2018).

The LNS consisted of seven sets, with three trials within each set. In the first set, trials consisted of two-item alphanumeric strings, and string length increased by one letter or number each set up to a maximum of eight items. Stimuli were created using Microsoft PowerPoint (Microsoft Corporation, 2018). Letters and numbers were presented sequentially for 2,000 ms each, with no inter-item interval. Participants were asked to pay attention to each alphanumeric string, then reorder the letters and numbers, with the numbers in ascending order followed by the letters in alphabetical order (e.g., R4D → 4DR; V5J1 → 15JV). They were asked not to write down the letters and numbers during presentation or reorder the letters and numbers once they were typed on the response page. The task ended either when the seventh set was complete, or after a participant had incorrectly responded to an entire set of three trials. Participants scored one point for each correct trial (i.e., a maximum score of 21).

Selective attention task. On each trial of the selective attention task, participants heard a pair of simultaneous sentences played as either collocated or dichotic. One sentence in each pair was L1-Spanish, the other was L2-English, and one talker was female and the other was male. Each sentence contained five keywords which were used for scoring (see section 2.3.4. Analyses).

The 168 English sentences were taken from the IEEE corpus (Rothauser, 1969). An example IEEE sentence is *The <u>jacket</u> <u>hung</u> on the <u>back</u> of the <u>wide</u> <u>chair</u>* (keywords underlined). Each sentence was spoken by a male and a female talker, both native English speakers. The Spanish sentences were taken from the Sharvard corpus (Aubanel et al., 2014). This corpus was created based on the IEEE corpus, where each Sharvard sentence is a translation of an IEEE sentence, to ensure the corpora are semantically comparable. Each Sharvard sentence was spoken by a male and female talker, both native Spanish speakers with a northern Spanish accent, which is considered a standard Spanish accent. The selected Spanish sentences were different from the 168 selected English sentences (i.e., not translation). Sentences with salient cross-linguistic cognates or false friends were excluded. IEEE sentences were excluded if they contained keywords with a low frequency in English (Zipf-Frequency < 3, SUBTLEX-UK, van Heuven et al., 2014). Finally, Sharvard sentences were excluded if they contained a keyword with a diacritic that could change the word's meaning if omitted. For example, *caminó* means *he walked,* but when the diacritic is omitted, *camino*, which means *path*.

Each of the 168 English sentences were paired with one of the 168 Spanish sentences (with the female/male English versions of each sentence paired with the male/female Spanish versions of the chosen competing sentence). Each sentence pair was chosen manually to make sentence durations within a pair as similar as possible. There were nevertheless small duration discrepancies: the average difference in sentence duration was 243 ms. We therefore calculated the average duration of each pair of sentences and compressed or expanded the longer and shorter sentences in the pair to the average duration of the pair using Audacity (Audacity Team, 2014).

Every sentence was resampled using Praat (Boersma & Weenink, 2023) to a sampling rate of 44,100 Hz and 16-bit resolution. The English sentences and the Spanish female sentences were then filtered to match the long-term average spectrum (LTAS) of the Spanish male sentences using Praat. Every sentence was root-mean-square (RMS)-equalised to an intended presentation level of 60 dB prior to playback using Praat. Finally, each sentence was combined with its partner sentence using MATLAB (The MathWorks Inc., 2022). A mono file and a stereo file (with one voice in each channel to simulate spatial separation) were generated for each sentence pair to create the collocated and dichotic conditions respectively.

The sentence pairs were randomly assigned to one of two lists, ensuring that participants completing both Experiment 1 and Experiment 2 would hear different stimuli in each experiment. Each sentence pair was randomly assigned to one of four conditions, which determined attended language (L1 (Spanish) or L2 (English)) and spatial separation (collocated or dichotic).

2.3.3.3. Procedure

Participants first completed a screening session, which consisted of the language background questionnaire and LexTALEs. Some participants also completed the LNS in the screening session (N = 19), while others completed it after the selective attention task (N = 79)[3]. A few days later (mean = 4.17 days, SD = 5.51 days), participants who met the language requirements for the experiment were invited to complete the selective attention task.

Before starting the selective attention task, participants were instructed to adjust their volume to a comfortable listening level using a brief segment of white noise that had been RMS-equalised to the same level as the stimuli in the selective attention task (60 dB). Next, a headphone check was completed to ensure participants were wearing functioning stereo headphones (Woods et al., 2017). In this task, the use of antiphase audio for some of the tones meant that the task could only be successfully completed with stereo headphones.

Each participant was randomly assigned to List 1 stimuli or List 2 stimuli, and to one of four groups determining which talker (female or male) would speak which language (Spanish or English) and in which ear (left or right; dichotic blocks only) throughout the entire task. Instructions were presented in both Spanish and English throughout the selective attention task. Participants were told at the beginning of the task which talker would be speaking which language, and, for the dichotic blocks, which language would be in which ear. Participants then completed four blocks, with twenty trials per block plus one practice trial per block. Blocks were presented in a random order, and each block

---

[3] This change was made for cost-efficiency reasons, i.e., to reduce the time to complete the screening so we could minimise compensation for participants who ultimately did not meet the eligibility criteria.

represented one condition of the selective attention task (L1-collocated, L2-collocated, L1-dichotic, and L2-dichotic). Participants were told at the beginning of each block whether the sentences would be collocated or dichotic, and which voice to attend to throughout the block. On each trial, participants were asked to transcribe the sentence spoken by the talker they were attending to.

Participants also completed an English vocabulary check after the selective attention task to confirm they were familiar with the words in the L2 stimuli. In this task, participants saw 40 written English keywords from the selective attention task with three possible Spanish translation options. Participants were asked to click on the correct Spanish translation of the English keyword that was presented. All participants passed the English vocabulary check, with an average score of 99.77% (SD = 0.72%).

2.3.4. Analyses

Participants were excluded from analysis if they: scored fewer than three points on the LNS or reported that they cheated (i.e., wrote down the letters/numbers); failed the headphone test twice, indicating that they were not wearing functioning stereo headphones; reported that they cheated in the selective attention task (i.e., wrote down the sentences before the response page); or scored less than 85% in the English vocabulary check. For these reasons, 11 participants were unable to complete the main listening task or excluded from analyses. Any participant who achieved an average score of less than 0.2 (20% correct) in any of the four conditions in the selective attention task (N = 2), which is equivalent to less than one keyword correct per sentence, was also excluded from analysis. This eligibility criterion was not pre-registered but was added to ensure that all participants could later attempt the more challenging divided attention task.

Each target sentence in the selective attention task contained five keywords. For each keyword, participants received a score of 1 or 0 depending on whether the keyword was correctly transcribed, and an average score was computed for each trial (e.g., two out of five keywords correctly identified resulted in a score of 0.4 for that trial). The following deviations from the target were still scored as correct:

- Any typed word phonologically identical to the keyword, even if the typed word was the same as another existing word (e.g., the keyword was *bear* and the participant typed *bare*).
- Any typed word phonologically dissimilar to the keyword (and any other real word) but spelt correctly except for either one omitted letter (e.g., *young* typed as *yong*), one added letter (e.g., *Christmas* typed as *Christmast*), or one pair of consecutive letters switched over (e.g., *light* typed as *lihgt*).
- Two correctly spelt words merged into one word (e.g., *paper bag* typed as *paperbag*).
- Added diacritics (e.g., *niños* typed as *ñiños*) or omitted diacritics (e.g., *ratón* typed as *raton*).

The data were analysed using generalised linear mixed effects models (GLMMs) in R (v. 4.3.1), using RStudio (v. 2023.06.0+421) with the lme4 package (v. 1.1.34, Bates et al., 2015). A binomial distribution with a logit link was used to model the by-trial scores. The model converged with a full random-effect structure including participant and item (sentence pairs) intercepts, and random slopes by participant for language, spatial separation, and the interaction between them. There were no by-item random slopes as none of the predictors were manipulated within items. Language (L1-Spanish and L2-English), spatial separation (dichotic and collocated), and their interaction were included as fixed effects. Both effects were sum coded (0.5 versus -0.5 for dichotic versus collocated and L1-Spanish versus L2-English). The model used the BOBYQA optimiser (Powell, 2009) and a maximum of 10^9 iterations. The full Hypothesis-Testing Model was as follows:

```
glmer(TranscriptionAccuracy + Language*SpatialSeparation +
(1 + Language*SpatialSeparation|Participant) + (1|Item),
family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”,
optCtrol = list(maxfun = 1e9)))
```

The data were also analysed using exploratory models to investigate individual differences in transcription accuracy. All continuous independent variables were standardised by centring data points on the mean and z-transforming to facilitate interpretation and comparison of effect sizes. The exploratory models used the same random effects structure as the Hypothesis-Testing Model.

The Language History Model used only L2 performance as the outcome variable, and included L2-English age of acquisition, L2-English proficiency (LexTALE score), and frequency of language use (LSBQ score) as fixed effects to investigate the potential role of individual differences in L2 background. These variables were allowed to interact with spatial separation in the model. As this model included many predictors, we also conducted model comparisons to determine which (if any) of these variables explained transcription performance in the selective attention task. The Reduced Language History Model removed any fixed effects and interactions that were non-significant according to the fixed effects table. The Language History Model and Reduced Language History Model were compared using a likelihood ratio test. If a model was found to fit the data better than the other (demonstrated by a smaller AIC value), only that model was retained. If there was no significant difference between the models’ AIC values, the simpler model (i.e., with fewer fixed effects and interactions) was retained. A second exploratory model, the LNS Model, examined the potential role of WM in selective attention listening. This included LNS score as a fixed effect.

2.3.5. Results

The results from the selective attention task are displayed in Figure 1. In the Hypothesis-Testing Model, a significant main effect of language was found (β = 2.66, SE = 0.17, z = 15.36, *p* < .001), with performance for L2-English (mean = 0.66, SD = 0.16) lower than for L1-Spanish (mean = 0.95, SD = 0.04). A significant main effect of spatial separation was also found (β = 0.93, SE = 0.15, z = 6.20, *p* < .001), with performance for collocated stimuli (mean = 0.76, SD = 0.09) lower than for dichotic stimuli (mean = 0.85, SD = 0.09). However, there was no significant interaction between language and spatial separation (β = 0.28, SE = 0.30, z = 0.92, *p* = .357). This indicates that neither language benefited from spatial separation significantly more than the other (average SRM scores (i.e., dichotic score minus collocated score): 0.04 (SD = 0.05) for L1-Spanish; 0.14 (SD = 0.10) for L2-English).

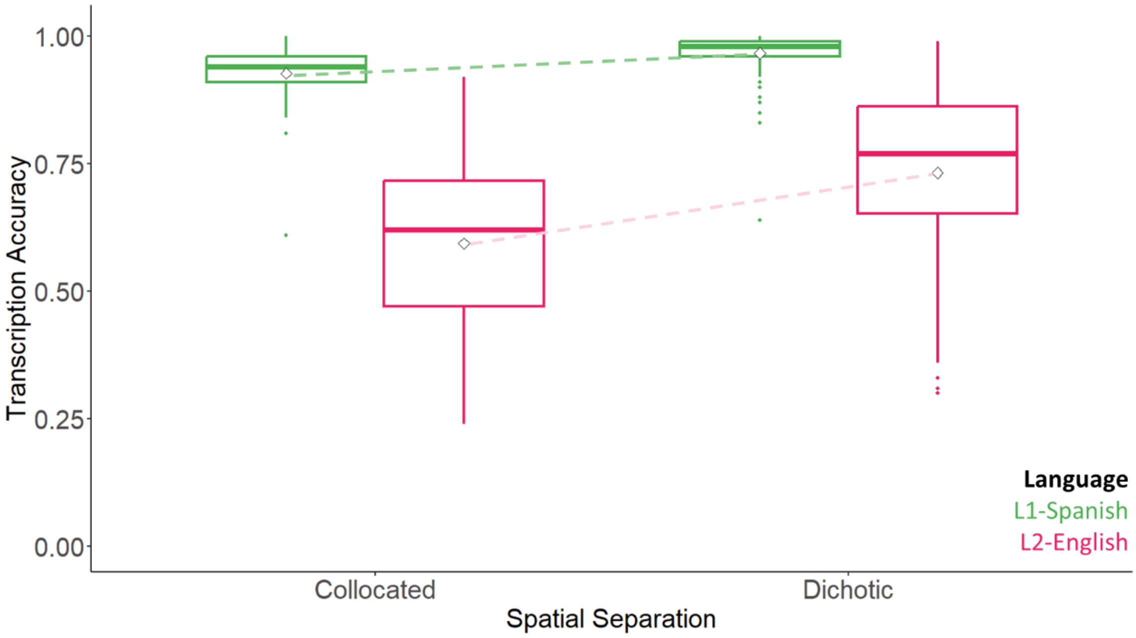


Figure 1 - Average transcription accuracy for each language and spatial separation level (Experiment 1)

Note. Transcription accuracy is the average proportion of keywords correctly transcribed within one condition (L1-collocated, L1-dichotic, L2-collocated, L2-dichotic). Open diamonds depict the mean score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average scores beyond 3 SDs from the mean.

The Language History Model and Reduced Language History Model comparisons showed no significant difference between the models with and without the non-significant fixed effects and interactions (*p* = .217). Therefore, the simpler version of the model (Reduced Language History Model) was used, without the non-significant fixed effects and interactions. The Reduced Language History Model, run on the L2-English results only, included spatial separation, L2-English proficiency, and frequency of language use as fixed effects, and the interaction between spatial separation and L2-English proficiency.

This model showed that, in addition to a main effect of spatial separation (β = 0.80, SE = 0.20, z = 3.97, *p* < .001), there were significant main effects of L2-English proficiency (β = 0.39, SE = 0.08, z = 4.92, *p* < .001) and frequency of language use (β = -0.25, SE = 0.08, z = -3.17, *p* = .002). Those who demonstrated higher L2-English proficiency in the LexTALE had higher L2-English transcription accuracy. A more frequent use of L2-English was associated with higher L2-English transcription accuracy. There was also a significant interaction between spatial separation and L2-English proficiency (β = 0.11, SE = 0.06, z = 2.01, *p* = .045). Follow-up pairwise comparisons of the estimated slopes indicated that the effect of L2-English proficiency on performance (restricted to the L2-English condition) was significantly positive in both spatial conditions. However, the effect was weaker in the collocated (β = 0.33, SE = 0.08, z = 4.24, *p* < .001) than the dichotic (β = 0.44, SE = 0.09, z = 4.98, *p* < .001) condition. Therefore, increased L2-English proficiency benefited L2-English transcription accuracy more when the L2-English target was spatially separated from the masker. Therefore, those with a higher L2-English proficiency demonstrated a larger effect of SRM when attending to L2-English.

The LNS Model, in line with the Hypothesis-Testing Model, showed main effects of language (β = 2.66, SE = 0.17, z = 15.49, *p* < .001) and spatial separation (β = 0.93, SE = 0.15, z = 6.20, *p* < .001) on transcription accuracy, but no interaction between language and spatial separation (β = 0.27, SE = 0.30, z = 0.92, *p* = .359). There was no main effect of LNS score (β = -0.03, SE = 0.07, z = -0.38, *p* = .703) and no interaction between LNS score and spatial separation (β = -0.05, SE = 0.07, z = -0.69, *p* = .491). There was a significant interaction between LNS score and language (β = 0.22, SE = 0.12, z = 2.10, *p* = .036). However, follow-up pairwise comparisons of the estimated slopes indicated that the LNS score did not significantly impact transcription accuracy when the target was either L1-Spanish (β = 0.09, SE = 0.09, z = 1.01, *p* = .312) or L2-English (β = -0.14, SE = 0.09, z = -1.56, *p* = .120). There was no significant three-way interaction between LNS score, language, and spatial separation (β = -0.06, SE = 0.05, z = -1.32, *p* = .188).

2.3.6. Discussion

Previous research has demonstrated that SRM is beneficial for monolinguals listening in a single-language context (e.g., Marrone et al., 2008). Experiment 1 examined whether the magnitude of SRM differed during L1 and L2 processing when the two languages are heard simultaneously.

As research has demonstrated that L2 is more negatively impacted by EM than L1 (e.g., Cooke et al., 2008), we expected that SRM would benefit L2 processing more than L1 processing. Contrary to this hypothesis, there was no statistically significant interaction between language and spatial separation. This finding is in line with Ezzatian et al. (2010), who also showed no difference in SRM magnitude for L1 and L2 processing when comparing separate groups of L1 and L2 speakers in single-language two-talker contexts. This suggests that reducing EM through spatial separation might be qualitatively different from reducing EM through other methods (e.g., reducing the signal-to-noise ratio of a collocated pair of talkers). This finding is also in line with research indicating that the L2 disadvantage in noise is not observed when the task consists of identifying phonemes, rather than engaging in word or sentence processing (e.g., Cutler et al., 2004). However, our task involved sentence identification, and participants were required to identify full words, which should have optimised the chance of seeing an L2 disadvantage in noise. We nevertheless did not find any different impact of spatial separation on L2 than L1 processing, thereby demonstrating another situation in which the L2 disadvantage in noise is not apparent.

Listeners with higher L2 proficiency were overall better able to transcribe the L2. As L2 proficiency was measured using the LexTALE, which tests vocabulary knowledge, this suggests that participants with a larger L2 vocabulary understood more of the keywords presented in the task, thus improving their overall L2 transcription accuracy. Similarly, using L2 more frequently was associated with improved overall L2 transcription accuracy. This may be because individuals who are exposed to their L2 in a larger variety of situations (a high score on the frequency of language use questions in the LSBQ indicates that the listener encounters the L2 more often and potentially also in more varied situations) are more practised at listening to and streaming the L2 in a range of challenging listening environments.

However, those with higher L2 proficiency demonstrated a larger effect of SRM (with no such interaction observed for L2 use). If a less proficient language is affected more by EM and therefore benefits more from SRM, the opposite pattern should have been expected (bilinguals with a lower L2 proficiency should have benefitted more). However, the observed proficiency relationship indirectly aligns with Hui et al.'s (2021) findings that L1 listeners (those with a higher language proficiency) were better able to make use of spatial cues to improve their transcription accuracy than L2 listeners (with a lower proficiency). Having stronger lexical representations may allow listeners to recognise words more quickly, enabling them to better benefit from cues arising from spatial separation (e.g., interaural level differences). Our experiment did not include listeners with poor English vocabulary (as we required a minimum of 60% performance in the LexTALE). However, it is possible that more unbalanced bilinguals with a low L2 proficiency may show a larger SRM effect for their L1 than L2 – as was found by Hui et al. (2021).

We additionally explored how WM capacity is associated with listening performance and found that WM capacity (as measured by the LNS) did not predict the magnitude of SRM. It should be noted that the LNS scores were higher than expected, with a mean score of 15.95 (SD = 2.89) (compared to a mean score of 12.60 (SD = 3.70) in Knight et al., 2023). It is possible that the task, as implemented in the present study, did not reliably measure WM capacity. We therefore modified the LNS in Experiment 2 to make it more in line with the version used by Knight et al. (2023) and will further discuss the WM findings across the two experiments in the General Discussion.

In sum, Experiment 1 shows that both L1 and L2 processing benefited to a similar extent from spatial release from EM. As explained in the Introduction, a selective attention design was used in Experiment 1 to investigate SRM in L1 and L2 with a single target voice. However, when listeners must attend to two spatially separated talkers (during divided rather than selective listening), they must use cognitive control processes to shift their attention between the two locations. Thus, the benefit of spatial separation may be reduced by these additional cognitive costs, and these cognitive costs might affect L2 processing in particular. Experiment 2 tested these predictions by implementing a divided attention version of Experiment 1.

2.4. Experiment 2: Divided Attention

2.4.1. Hypotheses

The goal of Experiment 2 was to investigate whether spatial separation has a different impact on L1 and L2 processing during divided-attention listening, when spatial separation is associated with cognitive costs. The methodology of Experiment 2 was identical to that of Experiment 1, except that participants were required to attend to both talkers, rather than just one. Our hypotheses were as follows:

- H1: As in Experiment 1, performance will be higher when reporting L1 compared to L2 (i.e., a main effect of language).

- H2: As in Experiment 1, performance will be higher when the stimuli are dichotic compared to collocated (i.e., a main effect of spatial separation).
- H3 provides two competing hypotheses regarding the nature of a possible interaction between language and spatial separation:
    - H3 (i): SRM will be larger for L2 than L1 listening, because EM negatively impacts L2 listening more than L1 listening. However, as this pattern was not found in Experiment 1, a possible alternative for this interaction is described in H3(ii).
    - H3 (ii): The cognitive costs associated with tracking two talkers in the dichotic condition will be more detrimental for L2 than L1 processing, because processing L2 is more cognitively demanding than processing L1 (e.g., Bsharat-Maalouf et al., 2023). Therefore, any SRM benefits should be reduced in the L2 compared to L1 condition.

In addition to these main hypotheses, we considered the association of WM with listening performance. WM is particularly relevant for this experiment because of the requirement in the dichotic condition to flexibly shift attention between two locations – a process which requires increased spatial attentional control (Treisman, 1971) and may draw on WM resources (Knight et al., 2023). We therefore hypothesised that higher transcription performance in the dichotic condition would be associated with higher LNS scores, more so than in the collocated condition (i.e., an interaction between spatial separation and LNS score). Furthermore, we expected this relationship to be stronger when processing L2, as it is more cognitively demanding to listen to than L1 (e.g., Bsharat-Maalouf et al., 2023) (i.e., a three-way interaction between language, spatial separation, and LNS score).

Comparing Experiments 1 and 2, we hypothesised that transcription accuracy would be lower in Experiment 2 than Experiment 1 due to the additional demands of tracking both talkers rather than one (i.e., a main effect of task). We expected L2 processing to be more negatively impacted than L1 processing by the requirement to listen to two talkers (i.e., an interaction between language and task) as the increased interference of L1 on L2 was expected to be stronger when both languages are attended to compared to one. Finally, SRM was expected to be smaller in Experiment 2 than Experiment 1 because of the larger cognitive cost for dichotic as opposed to collocated listening when tracking two talkers rather than one (i.e., an interaction of spatial separation and task).

2.4.2. Methods

2.4.2.1. Participants

All 98 participants from Experiment 1 were invited to participate in Experiment 2. Only 66 of those invited completed Experiment 2. Therefore, we recruited a further 14 participants to complete Experiment 2 (referred to as “new participants''), using the same screening criteria as in Experiment 1. In total, data from 80 participants were collected and analysed in Experiment 2 (mean age = 26.00 years, SD = 4.00 years; 38 female, 39 male, 3 other). See Appendix A.1 for further participant information.

2.4.2.2. Procedure

The new participants completed the screening and were invited to complete the divided attention task if they met the same criteria as described in Experiment 1. Returning participants who had been presented with List 1 in the selective attention task were presented with List 2 in the divided attention task, and vice versa. Rather than each sentence pair being presented as either collocated *or* dichotic across all participants (as in Experiment 1), each sentence pair was presented as collocated for half of the participants and as dichotic for the other half.

The divided attention task procedure was almost identical to that in the selective attention task. Participants were told they would hear pairs of sentences played simultaneously, presented as either collocated or dichotic. They were told which talker would be speaking which language and in which ear (in the dichotic condition) throughout the task. However, participants were asked to attend to both talkers at the same time and were only told which voice to report *after* the presentation of each sentence pair. Participants completed two blocks (collocated and dichotic), each containing two conditions (report-L1 and report-L2, with the L1/L2 order randomised). Each block included 40 trials (20 report-L1 trials and 20 report-L2 trials), and participants were given a break between the two blocks and halfway through each block. Block order was counterbalanced.

After completing the divided attention task, all participants completed a new version of the LNS. This was modified to address concerns about validity at measuring WM (cf. section 2.3.6. Discussion). The modifications included limiting the presentation of each letter and number to 500 ms and adding a blank screen between each letter and number for 500 ms, more accurately replicating the oral version of the task. The same instructions were given to participants as in Experiment 1, but the numbers and letters were rearranged to remove potential practice effects for returning participants. Finally, participants completed the English vocabulary check for the list they were presented with in the divided attention task. All participants passed the English vocabulary check, with an average score of 99.78% (SD = 0.81%).

2.4.3. Analyses

The divided attention task was scored in the same way as the selective attention task. The pre-registration indicated that any participant who scored an average of less than 0.2 in any condition (N = 14) would be excluded from analysis. This eligibility criterion was removed due to the challenging nature of the divided attention task.

Experiment 2 was analysed similarly to Experiment 1, using the same analysis tools and packages. Unlike in the Experiment 1 analyses, the Experiment 2 Hypothesis-Testing Model included spatial separation as a by-item random slope (since each item was presented as dichotic for half of the participants and collocated for the other half). LNS scores were standardised by centring all data points on the mean and z-transforming, then added as a fixed effect. The full Hypothesis-Testing Model was as follows:

```
glmer(TranscriptionAccuracy ~ Language*SpatialSeparation*LNS +
(1 + Language*SpatialSeparation|Participant) +
(1 + SpatialSeparation|Item), family = binomial(link = “logit”),
glmerControl(optimizer = “bobyqa”, optCtrol = list(maxfun = 1e9)))
```

We also compared the two experiments to assess how attending to one talker in a dual-language context differs from attending to two talkers. The 66 participants who completed both experiments were included in the Comparison Model (mean age = 25.44 years, SD = 3.55 years; 32 female, 33 male, 1 other; see Appendix A.2 for further Comparison Model participant details). The only additional fixed effect was task (selective attention sum coded as -0.5, divided attention sum coded as 0.5). Task was also included as a by-participant and a by-item random slope. Due to a lack of convergence, correlations between slopes and intercepts were removed but otherwise the maximal model structure was used. The model used the BOBYQA optimiser (Powell, 2009) and a maximum of 10^9 interactions. The Comparison Model was as follows:

```
glmer(TranscriptionAccuracy ~ Language*SpatialSeparation*Task +
(1 + Language*SpatialSeparation*Task||Participant) +
(1 + SpatialSeparation + Task||Item), family = binomial(link = “logit”),
glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 1e9)))
```

2.4.4. Results

In the Hypothesis-Testing Model (Figure 2), a significant main effect of language was found (β = 1.25, SE = 0.13, z = 9.78, *p* < .001), with performance for L2-English (mean = 0.33, SD = 0.12) lower than for L1-Spanish (mean = 0.59, SD = 0.12). A significant main effect of spatial separation was also found (β = 0.35, SE = 0.06, z = 5.37, *p* < .001), with performance for collocated stimuli (mean = 0.42, SD = 0.09) lower than for dichotic stimuli (mean = 0.50, SD = 0.10). There was no significant interaction between language and spatial separation (β = -0.05, SE = 0.14, z = -0.32, *p* = .747). This indicates that spatial separation is beneficial for listening in both languages, but not one significantly more than the other (average SRM scores = 0.07 (SD = 0.15) for L1-Spanish; 0.08 (SD = 0.10) for L2-English).

There was a significant main effect of LNS score (β = 0.12, SE = 0.05, z = 2.55, *p* = .011), with higher LNS scores associated with higher overall transcription accuracy. However, there was no significant interaction between language and LNS score (β = -0.02, SE = 0.10, z = -0.23, *p* = .815), spatial separation and LNS score (β = 0.34, SE = 0.04, z = 0.80, *p* = .424), nor a three-way interaction between LNS score, spatial separation, and language

($\beta = -0.07$, SE = 0.11, z = -0.68, $p = .495$)[4]. Figure 3 depicts the relationship between LNS score and transcription accuracy.

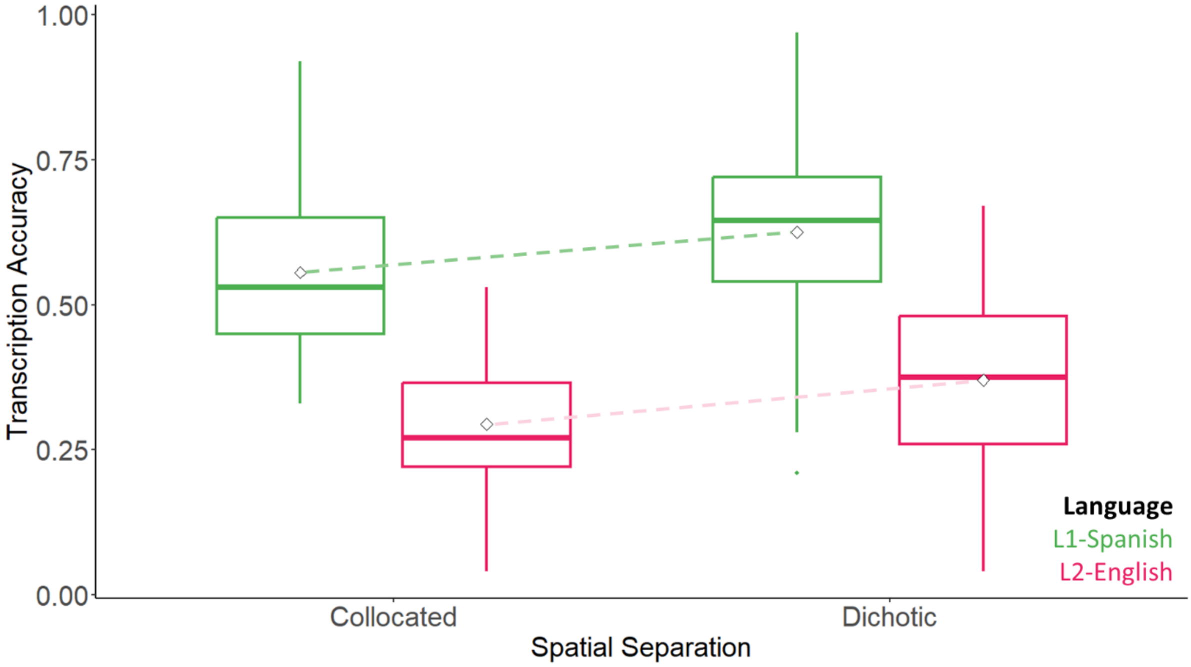


Figure 2 - Average transcription accuracy for each language and spatial separation level (Experiment 2)

Note. Transcription accuracy is the average proportion of keywords correctly transcribed within one condition (L1-collocated, L1-dichotic, L2-collocated, L2-dichotic). Open diamonds depict the mean score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average scores beyond 3 SDs from the mean.

---

[4] When the LNS Model for Experiment 1 was rerun using the Experiment 2 LNS scores (N = 66), LNS score was no longer associated with transcription accuracy (See Appendix A.3 for full GLMM output).

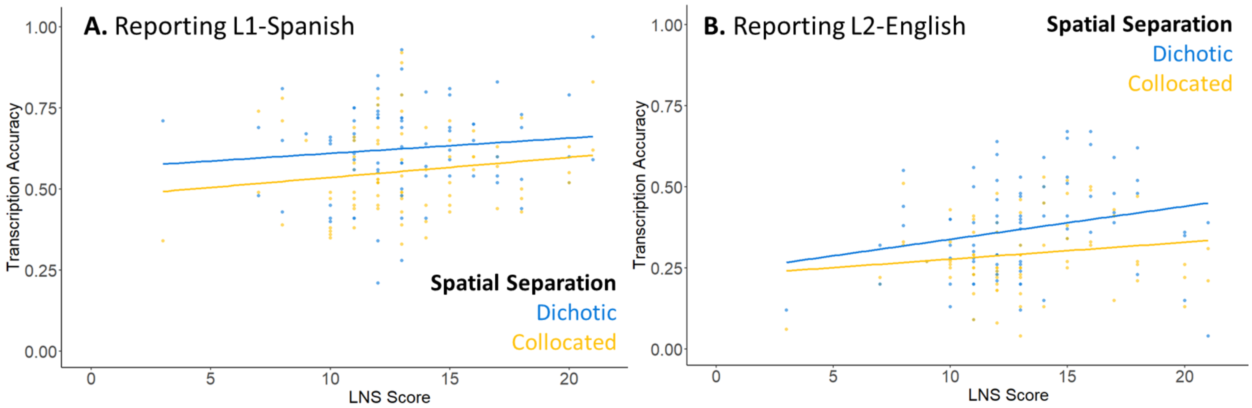


Figure 3 - Relationship between WM (measured by the LNS) and transcription accuracy in Experiment 2

Note. Transcription accuracy and LNS score (using the LNS described in Experiment 2) in the divided attention task. Panel A is transcription accuracy for L1-Spanish target, panel B is transcription accuracy for L2-English target. Each dot is the average accuracy of a participant. Linear trends are shown as regression lines.

In the Comparison Model (which included task), the same main effects were observed as reported above (main effect of language: β = 1.97, SE = 0.15, z = 13.08, *p* < .001; main effect of spatial separation: β = 0.64, SE = 0.08, z = 8.18, *p* < .001; no interaction between language and spatial separation: β = 0.16, SD = 0.16, z = 0.96, *p* = .335). In terms of task effects, a significant main effect of task was found (β = -2.32, SE = 0.08, z = -27.82, *p* < .001), with lower performance for the divided attention task (mean = 0.47, SD = 0.08) than for the selective attention task (mean = 0.80, SD = 0.08).

There was no significant three-way interaction between language, spatial separation, and task (β = -0.47, SD = 0.28, z = -1.69, *p* = .091), suggesting that, for both tasks, SRM benefits are similar when listening to L1 and L2. However, there was a significant interaction between language and task (β = -1.31, SD = 0.15, z = -8.52, *p* < .001). Follow-up pairwise comparisons using estimated marginal means revealed that performance in the divided attention task was significantly lower than in the selective attention task, although the magnitude of this effect differed between language conditions. When the target was L2, the odds of a correct response in the selective attention task were 5.26 times greater than in the divided attention task (SE = 0.57, z = 15.30, *p* < .001). When the target was L1, the odds of a correct response in the selective attention task were 19.59 times greater than in the divided attention task (SE = 2.32, z = 25.14, *p* < .001). This indicates that L1 processing was more negatively impacted than L2 processing by the requirement to listen to two talkers (average difference between the selective attention and divided attention tasks = 0.35 (SD = 0.11) for L1-Spanish processing; 0.32 (SD = 0.10) for L2-English processing).

There was also a significant interaction between spatial separation and task (β = -0.62, SD = 0.13, z = -4.79, *p* < .001). Follow-up pairwise comparisons using estimated marginal means revealed that performance in the dichotic condition was significantly higher than in the collocated condition, although the magnitude of this effect differed between task conditions. In the selective attention task, the odds of a correct response for the dichotic stimuli were 2.57 times greater than for the collocated stimuli (SE = 0.32, z = 7.56, *p* < .001). In the divided attention task, the odds of a correct response for the dichotic stimuli were 1.39 times greater than for the collocated stimuli (SE = 0.10, z = 4.74, *p* < .001). This demonstrated that SRM was larger in the selective attention task than in the divided attention task (average SRM scores = 0.09 (SD = 0.06) for selective attention; 0.07 (SD = 0.08) for divided attention).

2.4.5. Discussion

Experiment 2 examined whether the magnitude of SRM differed between L1 and L2 processing when bilingual listeners attend to two simultaneous talkers in a dual-language context. The same pattern of results was found as in Experiment 1. Transcription accuracy was higher for L1 than L2, and when the sentences were dichotic compared to collocated.

However, contrary to our hypotheses, SRM did not impact processing L1 and L2 to different extents.

The presence of an SRM effect in a divided listening task with bilinguals is consistent with research with monolinguals (e.g., Knight et al., 2023), demonstrating that the overall benefit of spatial separation outweighs the cognitive costs associated with dividing attention between spatial locations. SRM benefits outweighing cognitive costs applied to both L1 and L2, with no significant differences between the two languages. It is possible that both hypothesised interactions were true: that is, that L2 benefited from SRM more than L1 while also being more negatively impacted by the increased cognitive demands associated with spatial separation. Therefore, these effects could have balanced each other out and manifested in an overall lack of a significant interaction. However, in combination with results from Experiment 1, which indicated that L1 and L2 were not impacted by spatial separation to different extents, it is more parsimonious to conclude that L1 and L2 did not fundamentally differ in how much they benefited from SRM in both selective and divided listening circumstances. Furthermore, these findings suggest that L1 and L2 are not differently impacted by the cognitive costs associated with dividing attention between dichotic stimuli.

We also found that that increased WM capacity was associated with increased transcription accuracy in the divided attention task. This suggests that listeners draw on their WM resources to process L1 and L2 simultaneously. Although WM capacity was associated with overall transcription accuracy, it was not associated with SRM magnitude. The implications of this result are discussed further in the General Discussion.

2.5. General Discussion

Listening in noisy environments can be challenging, partly due to energetic masking (EM; spectrotemporal overlap between sounds). Masking can be reduced by spatial separation of sound sources, with improved listening performance as a result (spatial release from masking, SRM, e.g., Ihlefeld & Shinn-Cunningham, 2008a; Pinto et al., 2020). Here, we showed across two experiments that when bilinguals listen to their two known languages at the same time, they benefit from SRM in both L1 and L2. However, when attending to two talkers simultaneously as opposed to just one, spatial separation is also associated with costs due to the need to shift attention between two distinct locations (Best et al., 2006; Knight et al., 2023). These costs notwithstanding, we found that bilinguals continue to benefit from SRM during divided listening in both L1 and L2, in line with previous research with monolinguals showing that SRM can outweigh costs associated with spatial separation (e.g., Ihlefeld & Shinn-Cunningham, 2008b; Knight et al., 2023; Pinto et al., 2020).

Although both experiments showed a benefit of spatial separation (SRM), the magnitude of SRM was smaller in Experiment 2 (divided attention) than in Experiment 1 (selective attention). This suggests that the cost of spatial separation was magnified when tracking two speakers, hence reducing the observable SRM benefit. This result is inconsistent with that of Pinto et al. (2020), who found SRM to be similar in their selective and distributed attention conditions. However, this discrepancy may be due to different levels of spatial separation. In Pinto et al.'s spatially separated conditions, the stimuli were separated by 160°, whereas ours were dichotic, allowing for a maximal impact of SRM. Additionally, Pinto et al.'s task involved target word detection, rather than entire sentence transcription; this may lead to SRM differences because listeners can use streaming cues more effectively when listening to full sentences than single words. Instead, our findings are consistent with Knight et al.'s (2023) finding that, once EM is controlled, spatial separation negatively impacts listening performance when attending to two stimuli simultaneously. Both Knight et al.'s study and the present one demonstrate that there are costs of spatial separation during divided attention, but these are outweighed by SRM.

One possible source of costs associated with spatial separation during divided-attention listening is the cognitive demand of shifting attention between two spatial locations. This aligns with Posner et al.'s (1980) Attentional Spotlight Theory, which likens attention to a movable spotlight, enhancing focus on specific areas for better processing. Although this claim originally applied to visual selective attention, it also aligns with the mechanisms of attending to spatially separated auditory stimuli, as illustrated by Best et al. (2006). In their study, stimuli presented closer together (falling within the hypothetical attentional spotlight) were easier to process than spatially separated stimuli, due to the need to shift the attentional spotlight between locations. Treisman (1971) demonstrated the cognitive costs associated with this effect using an ear-switching paradigm. Participants recalled lists of digits presented either to both ears at the same time, or alternately to each ear. Recall performance was worse when the digits were presented to alternating ears, indicating that shifting attention between two ears was cognitively demanding. Together with our study, these results provide evidence that spatial separation is associated with a processing cost due to shifts of spatial attention between two simultaneously presented stimuli.

It is possible that individual differences may affect listeners' ability to cope with such increased cognitive demands. Our study addressed this question by investigating the relationship between WM (as measured by the LNS) and transcription accuracy. WM capacity was associated with overall transcription accuracy when attending to two talkers (Experiment 2), but not when attending to one talker (Experiment 1). It is likely that the relationship between LNS score and transcription accuracy in the divided attention task is indicative of the increased task demands associated with processing and maintaining two, compared to one, sentences in memory simultaneously, and/or with processing two languages simultaneously.

Crucially, while previous studies have focused on monolingual listeners in a single-language context, our studies examined SRM in bilingual listeners in a dual-language context. We showed that both L1 and L2 benefit from SRM, and that they do so to similar extents. Research into the field of simultaneous interpretation (listening to one language while translating into another) shows an important role of WM in interpretation performance and speed (Macnamara & Conway, 2016; Tzou, 2008). This aligns with our conclusion that WM resources are used for processing two languages simultaneously. Some previous research on Portuguese-English bilinguals has indicated that WM capacity might be particularly associated with L2 processing (Fay & Buchweitz, 2014), which is inconsistent with our own lack of interaction between language and WM capacity. However, Kim et al. (2022) showed in their study with Korean-English bilinguals that WM is only associated with L2 listening when the stimuli are long passages, rather than short sentences as in this study. Indeed, Fay and Buchweitz's (2014) stimuli were also long passages followed by comprehension questions. Thus, cognitive demands (and the potential role of WM) might have a larger influence when bilinguals are asked to process longer and/or more complex passages in their L2 than when processing the relatively simple sentences used here. Other individual differences in cognitive abilities might also be relevant for dual-language contexts such as the one we used. For example, differences in inhibition abilities may impact selective more than divided attention performance, because inhibition of irrelevant stimuli is an important component of selective attention (Neill et al., 1995). Inhibition may be particularly relevant for bilinguals in a dual-language context, through inhibition of the non-target language (Declerck & Koch, 2023). Future studies on bilingual selective and divided listening should therefore measure additional cognitive variables such as inhibition.

Research has indicated that L2 is more negatively impacted by EM than L1 (e.g., Cooke et al., 2008) and is more cognitively demanding to process (e.g., Spivey & Marian, 1999). We therefore hypothesised that spatial separation would impact L1 and L2 differently. We expected that, during divided attention, L2 processing would be more negatively affected by cognitive costs associated with spatial separation. We found no evidence that L1 and L2 were differently impacted by SRM or cognitive costs, indicating that bilinguals use spatial separation to improve listening to both languages, even in cases where spatial separation should be associated with cognitive costs. Although there is a limited amount of literature investigating how SRM impacts L1 versus L2 processing, the results of this study are consistent with Ezzatian et al. (2010). Thus, across different types of maskers and when using a dual- rather than single-language context, L2 (and L1) continues to benefit from spatial separation, even when cognitive demands are increased during a more demanding divided-listening task.

When considering language differences (regardless of SRM), we observed more similar performance between L1 and L2 in Experiment 2 than in Experiment 1. This was unexpected as we predicted that L2 processing would be more negatively impacted by the added costs of divided attention, due to L2's presumed greater cognitive demands (Spivey & Marian, 1999). Thus, L1 processing, rather than L2 processing, might be more negatively impacted by the added cost of dividing attention. One potential explanation is that bilingual listeners may attempt to compensate for their lower L2 proficiency and activation levels in challenging conditions by purposefully focusing more on L2 than L1. (N.B. This would only apply to the divided listening task because listeners always knew in advance which voice to report during the selective listening task in Experiment 1.) However, listeners may overcompensate when using this tactic, focusing too much on their L2 and thus unknowingly limiting their L1 performance. This explanation, while hypothetical, is supported by anecdotal feedback from some participants indicating that they deliberately focused more on their L2 due to the challenging nature of the task. This raises questions for further research about subjective techniques used by bilinguals in dual-language contexts, and how successful these techniques are.

An additional consideration is that, in both experiments, changes in target language were always paired with changes in masker language. Therefore, the language effect could reflect either target language proficiency or the degree of interference from the masker language. This limitation could be addressed by reducing the linguistic interference from the masker by time-reversing it, preserving acoustic properties, but removing linguistic content. This manipulation has been shown to produce linguistic release from masking (e.g., Mepham et al., 2022; Mepham et al., 2025). If similar language effects are observed when the masker is time-reversed, this would suggest that target language proficiency underlies the effect. Alternatively, if the effect is diminished when the masker is time-reversed, this would imply that interference from the masker language contributes to the effect, potentially due to the higher resting activation levels of the L1 making it more interfering when reporting the L2 (e.g., Dijkstra & van Heuven, 2002). Linguistic release from masking can also clarify the role of SRM: if the spatial separation effect persists when the masker is time-reversed and therefore exerts lower levels of informational masking (e.g., Rhebergen et al., 2005), this would indicate that its benefit arises primarily through reducing energetic masking specifically.

In summary, our study demonstrates that spatial separation improves listening performance through SRM, even when attention is divided between two talkers. However, the benefits of SRM are smaller during divided attention listening due to the cognitive costs associated with shifting attention between spatial locations. Contrary to expectations, these costs do not differentially affect L1 and L2 performance, suggesting that bilinguals use spatial separation similarly to improve listening in both languages. Our findings align with previous studies highlighting the cognitive demands of divided listening but challenge assumptions about the disproportionate impact of these demands on L2 processing.

Chapter 3: Semantic interference from first- and second-language maskers during speech-on-speech listening

3.1. Abstract

When selectively listening to target speech, it is an open question whether the semantic content of background speech is blocked using a semantic filter mechanism, or whether it is unintentionally co-activated with the target. This study addresses this question using a speech-on-speech task, in which participants transcribe target sentences presented simultaneously with masker sentences that are either semantically related (semantically competing, SC) or unrelated (neutral, N) to the target. If SC maskers disrupt target transcription more than N maskers, it would suggest that lexical representations of unattended masker speech are co-activated with the target and undergo semantic processing. Across two experiments, we found that the semantic relationship between target and masker speech (i.e., SC or N) did not impact target transcription accuracy. This suggests that the masker speech was not semantically processed. However, SC maskers elicited more intrusion mistakes than N maskers (i.e., listeners incorrectly reported SC masker words more often than N masker words). A surprise recognition task (Experiment 2) revealed that this semantic intrusion effect likely resulted from listeners mis-streaming target and masker, rather than from target-masker co-activation. Therefore, these

experiments suggest listeners do not process the semantic content of masker speech, indicating reliance on an effective semantic filter mechanism.

3.2. Introduction

A challenge during speech-on-speech listening is the spectro-temporal overlap between target and masker speech, known as energetic masking (EM). However, various factors beyond EM can affect intelligibility. These include modulation masking, which is the interference of competing amplitude modulations between the target and the masker even when there is little or no spectral overlap between them (Stone et al., 2011). The focus of the present study, however, is on a higher-level type of interference, known as informational masking (IM, e.g., Pollack, 1975). This involves disruption to the target that arises from higher-level perceptual or cognitive factors, rather than from the physical overlap of sounds. For example, listening could be impaired by familiarity with the language of the competing speech. One under-investigated form of IM is semantic interference between the target and masker. For instance, a masker which semantically overlaps with the target (e.g., *The cat casually scratches the chair* and *The dog anxiously gnaws the table*) may impose a greater amount of IM than a masker with weaker semantic overlap with the target (e.g., *The cat casually scratches the chair* and *The lily slowly opens its petals*). Examining this question can help to understand the mechanisms of selective attention in listening, specifically, the extent to which unattended speech is activated and processed, and how listeners overcome potential semantic interference.

Two competing mechanisms are proposed to explain how listeners cope with a semantic overlap between simultaneous speech streams: the Semantic Filter account and the Target-Masker Co-Activation account. According to the Semantic Filter account, when selectively attending to a target stream, listeners apply an attentional filter that blocks semantic processing of the masker. In this view, the basic physical properties of the masker (e.g., pitch) may be processed, but its lexical-semantic content is not, as described by Brodbeck et al. (2018). This claim is reminiscent of Broadbent's (1958) Early Filter Theory, which proposes that unattended speech is rejected prior to semantic analysis. More recently, Marsh and Campbell (2016) proposed a neurological early filter mechanism consistent with Broadbent's theory – it states that systems within the frontal lobe constrain the processing of complex sounds at the level of the brainstem, limiting the distracting effects of background sound entering the higher auditory system. Furthermore, neurological evidence shows that when listening to target speech, the brain uses phonological differences between target and masker to enhance neural tracking of the target and block any activation of the masker stream (e.g., Lakatos et al., 2013; Szalárdy et al., 2020; Wang et al., 2019). This indicates that listeners may rely on low-level streaming mechanisms that prevent high-level, including semantic, processing of unattended speech. If the Semantic Filter account is correct, masker semantic content should exert little or no effect on the intelligibility of the target, regardless of the semantic overlap between target and masker.

In contrast, the Target-Masker Co-Activation account proposes that listeners semantically process both streams, activating semantic representations of both target and masker content. This account is described by Kiefer (2002), whose EEG study showed that semantic representations are activated automatically, not strategically, allowing for unattended stimuli to undergo semantic processing. As a result, when a target and a masker overlap semantically, the masker may interfere with comprehension of the target, as it becomes unclear which words belong to the target and which words belong to the masker. This interpretation aligns with Treisman's (1964) Attenuation Theory, which posits that unattended speech is not completely filtered out, but rather attenuated, undergoing partial semantic processing. According to this view, the unattended stream is ultimately discarded once it is deemed irrelevant, but not before semantic overlap can cause interference. Thus, according to the co-activation account, target comprehension should be poorer if the masker is semantically related to the target than if it is semantically neutral.

Some behavioural research supports the Target-Masker Co-Activation account by demonstrating that when participants are instructed to tap their finger in response to specific words in a target stream, they occasionally respond to those same words when presented in the unattended stream (e.g., Treisman & Geffen, 1967; Treisman & Riley, 1969). This suggests that, even when selectively attending to a target stream, listeners process some content from the masker stream. Similarly, participants have been shown to notice their own name when it appears in a masker stream (e.g., Röer & Cowan, 2021; Wood & Cowan, 1995). However, this is only observed in a third of participants and is more common among individuals with a lower working memory capacity (Conway et al., 2001). Together, these findings indicate that at least some listeners may co-activate the semantic representations of both target and masker streams during selective listening.

However, it is possible that in these studies, participants are not fully semantically processing the masker speech but are instead primed to detect specific expected words. That is, the presence of salient words (i.e., the participant’s name or words they are expecting in the target stream) may lead to initial lexical activation (an early stage of listening where sound patterns momentarily match stored word representations) without full semantic processing. Thus, these words can capture attention without implying comprehension of the masker content. Rather than using single, salient distractor words, some speech-on-speech studies have employed complete masker sentences alongside the target to assess whether continuous background speech is activated and processed. Results from this research are inconsistent. Brouwer et al. (2012) found that when listening to a meaningful target sentence, a meaningful masker (i.e., *Rice is often served in round bowls*) was more disruptive than a meaningless masker (i.e., *The great car met the milk*). This design primarily addresses whether the semantic content of the masker *per se* exerts IM, rather than the semantic relationship between target and masker. Nevertheless, Brouwer et al.'s results suggest that there is semantic processing of background speech, weakening support for a strict semantic filter blocking all high-level processing of irrelevant speech.

However, Calandruccio et al. (2018) found no effects of masker semantic meaningfulness on target recall using a similar design to Brouwer et al. (2012), but with more controlled stimuli (the meaningful and anomalous maskers used by Calandruccio et al. (2018) were more comparable to each other in terms of syntactic structure and the number of syllables per word). This diminishes support for a semantic interference effect, suggesting that masker content may not reach higher-level processing, potentially being filtered out early in the listening process. Furthermore, other similar studies have observed a masker meaningfulness effect, but only in older adults (e.g., Rossi-Katz & Arehart, 2009; Tun et al., 2002). This pattern aligns with research showing that IM generally becomes more disruptive in older adults (e.g., Ezzatian et al., 2015; Helfer & Freyman, 2008; Zobel et al., 2019) and could support the idea of a semantic filter mechanism that becomes less effective with age.

However, the research discussed so far (e.g., Brouwer et al., 2012; Calandruccio et al., 2018; Rossi-Katz & Arehart, 2009) has not directly examined the role of semantic overlap between target and masker. Instead, most behavioural speech-on-speech studies have focused on whether the semantic content of the masker *per se* exerts IM by comparing syntactically well-formed and meaningful sentences with anomalous sentences. While valuable in demonstrating whether the semantic meaningfulness of background speech exerts IM, this approach has a potential confound, which may have led to null effects in some studies (e.g., Calandruccio et al., 2018). Specifically, the anomalous sentences may have been disruptive because of the unexpected semantics. Anomalous background speech may briefly disrupt short-term memory due to the time and cognitive effort required to make sense of the unexpected and unusual stimuli (Röer et al., 2019). While other unexpected maskers, such as time-reversed speech, do not seem to impair performance (e.g., Mepham et al., 2022; Mepham et al., 2025), this may be because the anomaly is acoustic and prosodic, rather than semantic. Semantically anomalous maskers might interfere through the cognitive demands of processing illogical stimuli, while meaningful maskers could hinder performance via semantic or linguistic interference, resulting in no observed effect of masker meaningfulness on transcription accuracy even though semantic content is, in fact, being processed.

To address this limitation, it is necessary to investigate how the semantic content of masker speech impacts listening using more naturalistic stimuli. One approach is to use meaningful masker sentences but to manipulate the semantic relatedness between masker and target speech. For example, a target sentence (*The cat casually scratches the chair*) can be paired with a semantically competing masker (*The dog anxiously gnaws the table*) or a semantically unrelated (neutral) masker (*The lily slowly opens its petals*). This design allows us to examine how the semantic *relationship* between target and masker affects transcription accuracy, rather than merely the semantic *content* of the masker, and avoids disruptions due to anomalous sentence processing.

To my knowledge, the only speech-on-speech study that directly investigated how performance is influenced by target-masker semantic relatedness was conducted by Villard et al. (2024). This study manipulated semantic relatedness of the masker to the target by using target sentences that always followed the pattern *Betsy sees a [NOUN],* and masker sentences that followed the pattern *Donna gives a [NOUN]* or *Molly wants a [NOUN]*, with masker nouns either semantically related to the target (e.g., *broccoli* and *carrot*) or semantically unrelated (e.g., *broccoli* and *train*). Participants were asked to identify the target noun from a visually presented matrix. Villard et al. found that the degree of semantic relatedness between target and masker did not affect target noun identification, suggesting that a semantic filter mechanism blocked semantic processing of irrelevant speech. However, Villard et al., who were primarily interested in the effects of aphasia on masked speech perception, used a small sample of 16 middle-aged adults (mean age = 56.22 years, SD = 9.11), eight of whom had aphasia. Furthermore, since participants knew that *Betsy* and *sees* always signified a target, streaming was relatively easy, with only one further keyword to identify. Semantic interference effects might still emerge in varied and longer sentences that are closer to real-life listening conditions.

Therefore, the first aim of the present study is to investigate whether semantic information in masker speech is processed. This will be done by examining whether a semantic interference effect occurs when there is a semantic relationship between a meaningful target and masker. The sentences in the present study are longer and more varied than those used by Villard et al. (2024), providing more opportunities for semantic overlap and increasing the likelihood of detecting semantic interference effects. If a semantic filter is used to prevent semantic processing of the masker, the degree of semantic overlap between target and masker should not influence performance, and semantically related maskers should therefore not disrupt listening more than unrelated maskers. In contrast, if target-masker co-activation occurs and the background speech is semantically processed, maskers with greater semantic overlap with the target are expected to interfere more with target processing than maskers with less semantic overlap.

Our study included familiarity of the masker language as an additional variable of interest. While processing semantic information in one's first language (L1) occurs rapidly, it is thought to be delayed and potentially weaker when listening to a second language (L2). According to the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994), L1 lexical representations have stronger links to their semantic meanings than do L2 representations. The RHM also argues that semantic concepts in the L2 are processed through the L1, slowing activation of semantic concepts when processing the L2. Other bilingual language processing models such as the Bilingual Interactive Activation Plus (BIA+, Dijkstra & van Heuven, 2002) and the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS, Shook & Marian, 2012) models also state that bilinguals have slower access to L2 than L1 semantic representations. Empirical studies indeed show that semantic information is processed more slowly in the L2 than in the L1 (Dijkgraaf et al., 2019), with semantic priming in the L2 taking longer to build (Wen & van Heuven, 2017). Therefore, if the Target-Masker Co-Activation account is true, slower L2 semantic activation should result in reduced semantic interference when the masker is L2 compared to L1. We therefore compared semantic interference in a single-language context (i.e., when target and masker are in L1) and in a dual-language context (i.e., when the target is L1 and the masker is L2).

The Target-Masker Similarity Hypothesis (Brouwer et al., 2012) posits that a masker more similar to the target (e.g., phonologically similar) is likely to disrupt listening to a target to a larger extent. Consequently, listening should be easier in a dual-language context than in a single-language context, this finding is supported by previous empirical research (e.g., Calandruccio et al., 2013; Calandruccio & Zhou, 2014; Mepham et al., 2022, 2025).

Additionally, when the target is always L1, this comparison of a single- and dual-language context tests the BIA+ and BLINCS principle that the L2 is slower to process than the L1 when it is *in the unattended stream.* If this principle holds when considering masker language, then listening to the L1 should be easier when the masker is L2 than L1 because of the lower resting activation levels of L2 making it less interfering. This idea is consistent with the Target-Masker Similarity Hypothesis.

3.3. Experiment 1

3.3.1. Pre-registration and ethical approval

The pre-registrations for Experiment 1 (https://osf.io/aetbw) and Experiment 2 (https://osf.io/w8yh4) are available on the Open Science Framework. Ethical approval for both experiments was granted by the ethics committee at the Department of Psychology of the University of York (ID: 202440). All participants provided informed consent before taking part, and all procedures were performed in compliance with relevant laws and institutional guidelines. All participants were paid at a rate of £6 an hour for their completion of each experiment and/or pilot testing.

3.3.2. Hypotheses

Experiment 1 aimed to investigate whether semantic overlap between target and masker impacts target perception, and whether any semantic interference effect is larger when the masker is L1 compared to L2. We recruited Spanish (L1) - English (L2) bilingual participants to complete a speech-on-speech task in both a single-language context (target and masker in L1-Spanish) and a dual-language context (target in L1-Spanish and masker in L2-English). L1-Spanish speakers were chosen to maximise recruitment, as there are many L1-Spanish speakers with a high L2-English proficiency registered on our selected recruitment platform (Prolific, www.prolific.co). Targets and maskers were always semantically meaningful sentences, but the masker semantic content was either neutral relative to the target sentence (N; i.e., low semantic relatedness) or semantically competing (SC; i.e., high semantic relatedness). The semantic relatedness of the maskers to the targets was manipulated across four keywords. The hypotheses were as follows:

- H1: If the semantic content of a masker is processed (supporting the Target-Masker Co-Activation account), target transcription accuracy will be higher when the masker is N compared to SC (i.e., a main effect of masker semantic content). This would support the idea that target and masker are semantically co-activated and would be in line with Brouwer et al.'s (2012) findings, although counter to research by Calandruccio et al. (2013) and Villard et al. (2024). This experiment will increase the likelihood of finding a semantic relatedness effect by addressing stimulus and design issues in previous research, specifically by manipulating semantic relatedness (rather than meaningfulness) across several target words in the context of complex sentences.
- H2: Transcription accuracy for L1 targets will be higher when the masker is L2 compared to L1 (i.e., a main effect of masker language). This is in line with the Target-Masker Similarity Hypothesis, which posits that the more linguistically similar a masker is to the target, the more disruptive it is (e.g., Brouwer et al., 2012; Mepham et al., 2022). It is also in line with the BIA+ (Dijkstra & van Heuven, 2002) and BLINCS (Shook & Marian, 2012) models, which posit lower resting activation levels for L2 than L1, potentially making L2 less interfering as a masker.

- H3 provides two possible scenarios with respect to an interaction between masker language and masker semantic content, both made with the assumption that a semantic interference effect is found (i.e., H1 accepted).
    - H3(i): If an L2 masker can be more easily ignored due to its lower resting activation levels, it may be less likely to be semantically processed. Therefore, the semantic interference effect may be smaller when the masker is L2 than L1 (i.e., a significant interaction between masker semantic content and masker language).
    - H3(ii): If the L1 is easier to transcribe when the masker is L2 than L1 purely because of phonological differences between target and masker speech (i.e., as explained by the Target-Masker Similarity Hypothesis) rather than because of L1-L2 differences in activation levels, there will be no difference in semantic interference between L1 and L2 maskers (i.e., no significant interaction between masker language and masker semantic content).

We also investigated performance over the course of each test block. Research demonstrates that, with exposure, listeners may adapt to maskers and thus improve their ability to parse the speech streams and transcribe the target (e.g., Bent et al., 2009; Marrufo-Pérez & Lopez-Poveda, 2025; Mepham et al., 2022). Additionally, listeners may invest more effort when the masker is particularly challenging (e.g., Brown & Strand, 2018), for example when the masker is L1 compared to L2, or SC compared to N. Thus, it is possible that when the masker poses a greater challenge, performance will be very low at first but will increase at a faster rate than when the masker is less challenging. Therefore, we hypothesised that participants' transcription accuracy would increase throughout a block (i.e., a main effect of trial number) and, critically, that it would improve faster when the masker is SC compared to N (i.e., a significant interaction between trial number and masker semantic content), and when the masker is L1 compared to L2 (i.e., a significant interaction between trial number and masker language).

3.3.3. Methods

3.3.3.1. Participants

Participants were recruited through Prolific (www.prolific.co) and tested online using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Brysbaert and Stevens' (2018) rule of thumb of having 1,600 observations per condition for a mixed model suggests a minimum of 80 participants (20 trials per combination of conditions: L1-N, L1-SC, L2-N, L2-SC). However, we also included trial number as a continuous fixed effect. Since the continuous nature of trial number makes it difficult to predict power, we decided to use a larger sample size of 120.

The 120 participants included in the analyses were Spanish-English bilinguals (mean age = 26.73, SD = 3.82; 62 females, 57 males, and 1 other). A questionnaire and Lextale-Esp (see section 3.3.3.2. Materials for further details) were used to ensure that invited participants met the following, pre-registered inclusion criteria: they were aged between 18 and 35 years old; had no self-reported neurological conditions or language or reading difficulties (including dyslexia); had normal hearing and normal (or corrected to normal) vision; were living in the UK or Spain at the time of testing; started acquiring Spanish before the age of three years; had lived in Spain for at least 10 years; scored at least 60% on the Lextale-Esp (see section 3.3.3.2. Materials, LexTALE and Lextale-Esp); and scored at least 70% on the English vocabulary check (see section 3.3.3.2. Materials, English vocabulary check). These criteria ensured that Spanish was the participants' L1. As participants were never asked to transcribe English, there was no minimum LexTALE score to participate in the experiment. However, we did still include an English vocabulary check to ensure participants were familiar with the L2 words in the masker. On average, participants showed a very high L1-Spanish proficiency level and an intermediate to high L2-English level. See Table 2 for further details about the participants' language history. Based on the described exclusion criteria, 25 participants were excluded from analyses (therefore, a total of 145 participants were tested, but only 120 used in analyses).

Table 2 - Language background for Experiment 1 sample (N = 120)

Language background measure	**L1-Spanish**				**L2-English**			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Age of acquisition (years)	0.01	0.09	0.00	1.00	5.92	3.06	0.00	15.00
LexTale proficiency (%)[a]	90.75	7.37	63.33	100.00	80.03	10.44	50.00	97.50
Self-reported understanding (/10)	9.93	0.33	7.00	10.00	8.13	1.38	3.00	10.00
Self-reported speaking (/10)	9.92	0.46	6.00	10.00	7.17	1.61	3.00	10.00
Self-reported reading (/10)	9.91	0.39	7.00	10.00	8.67	1.16	5.00	10.00
Self-reported writing (/10)	9.85	0.64	5.00	10.00	7.63	1.46	3.00	10.00
Language frequency of use (/4)[b]	Mean = 2.94; SD = 0.53; Min = 0.50; Max = 3.65							

[a] The L1-Spanish values exclude the data from one participant who scored 5% in the Lextale-Esp (see section 3.3.3.2. Materials, LexTALE and Lextale-Esp). Due to that participant's normal performance in the listening task and high self-reported Spanish proficiency, it was assumed that their low Lextale-Esp score was due to confusing the response buttons.

[b] This value is from the Language Background Questionnaire (see section 3.3.3.2. Materials, Language background questionnaire) and determines how frequently participants use each language in their daily lives (0 = All English, 1 = Mostly English, 2 = Half each language, 3 = Mostly Spanish, 4 = All Spanish).

3.3.3.2. Materials

Language background questionnaire. Twenty questions from the Language and Social Background Questionnaire (Anderson et al., 2018) were used to establish a language frequency-of-use score. Participants indicated how often they used each language (Spanish and English) with different people, in different contexts, and for different activities. For each item, their responses were coded as: All English = 0, Mostly English = 1, Half Each Language

= 2, Mostly Spanish = 3, and All Spanish = 4. A mean score was calculated from these 20 items for each participant. This questionnaire also included questions about language proficiency and age of language acquisition.

LexTALE and Lextale-Esp. English proficiency was measured with the LexTALE (Lemhöfer & Broersma, 2012), a vocabulary test appropriate for online testing. Spanish proficiency was measured with the Lextale-Esp (Izura et al., 2014). On each trial, participants saw a string of letters (60 experimental LexTALE trials, 90 experimental Lextale-Esp trials), which formed either a word or a non-word in the assessed language. Participants were asked to identify whether or not each string of letters was a real word by pressing a keyboard button. The order in which the participants completed the two vocabulary tests was counterbalanced. For each test, a final score was calculated for each participant using the standardised calculation:

$$\frac{\left(\frac{number\ of\ words\ correct}{number\ of\ words\ in\ total} \times 100\right) + \left(\frac{number\ of\ nonwords\ correct}{number\ of\ nonwords\ in\ total} \times 100\right)}{2}$$

Listening task. As there was no known corpus of sentences that manipulated inter-sentence semantic relatedness, we created our own stimuli for this study. A list of 100 sentences was created - this was the *target list*. An English and a Spanish version of each of the 100 sentences in the target list were created. The sentences in the Spanish list were direct translations of the sentences in the English list (i.e., word order and structure were identical), and each sentence contained four keywords: two nouns, one adverb, and one verb. Each sentence was either in the structure: *noun adverb verb noun* (e.g., *The <u>daisy</u> <u>gradually</u> <u>wilts</u> in the <u>heat</u>/La <u>margarita</u> <u>gradualmente</u> se <u>marchita</u> con el <u>calor</u>*, keywords underlined) or *noun verb adverb noun* (e.g., *The <u>deer</u> <u>runs</u> <u>quickly</u> across the <u>field</u>/El <u>ciervo</u>*

corre rápidamente por el campo), whichever order was more natural in English (as the Spanish adverb-verb order is more flexible).

For each sentence in the target list, a new sentence was created which was designed to be semantically related to it as well as identical in structure and keyword position. This new list of 100 sentences (translated into Spanish and English) was the *semantically competing list.* Finally, every sentence in the semantically competing list was paired with a sentence that was semantically unrelated to both the target sentence and the semantically related sentence (but with the same sentence structure). This was called the *neutral list.* See Table 3 for two L2-English triplets of sentences illustrating this design. See Appendices B.1 and B.2 for all sentences used in this experiment.

Table 3 - Example sentence sets

Target sentence	**Semantically competing (SC) sentence**	**Neutral (N) sentence**
El gato casualmente rasca la silla. *The cat casually scratches the chair.*	El perro ansiosamente roe la mesa. *The dog anxiously gnaws the table.*	El lirio lentamente abre sus pétalos. *The lily slowly opens its petals.*
La planta gradualmente crece sus raíces. *The plant gradually grows its roots.*	El lirio lentamente abre sus pétalos. *The lily slowly opens its petals.*	El perro ansiosamente roe la mesa. *The dog anxiously gnaws the table.*

Note. Examples of sentence triplets: target, semantically competing (SC), and neutral (N) sentences. SC maskers were also N maskers for other target sentences. Within the experiment, target sentences were always presented in L1-Spanish, and masker sentences were presented in L1-Spanish or L2-English depending on the masker language condition.

Throughout the entire stimulus set, no keyword was repeated, except for adverbs due to the relatively small number of Spanish adverbs. A latent semantic analysis (LSA, scored /1; Dumais et al., 1988) was conducted on each full target sentence with its SC and N masker to measure how semantically related the masker sentences were to the target sentences. The semantically competing list was more semantically related to the target list (mean LSA score for the 80 experimental trials = 0.63, SD = 0.07) than the neutral list (0.37, SD = 0.07), and this difference was significant ($t = -24.04$, $p < .001$).

All sentences were recorded using Audacity (Audacity Team, 2014) by a female speaker of L1-Spanish and L2-English. The speaker acquired Spanish from birth, English during childhood, and self-rated their proficiency as being high in both languages. The recording took place in a sound-attenuating booth using a SHURE SM58 vocal microphone. These recordings were edited to remove silence before and after the utterance using Audacity, then root-mean-square (RMS)-equalised to an intended presentation level of 60 dB prior to playback using Praat (Boersma & Weenink, 2023). Background noise was reduced in Audacity by selecting sections of the audio where no speech was present, capturing the background noise, and applying noise reduction to the entire clip.

The Spanish and English sentences were then matched on their long-term average spectrum (LTAS) using Praat, thus increasing the comparability of the amount of EM exerted across all listening conditions. For every sentence pair (320 sentence pairs in total: 80 targets paired with four possible maskers [L1-N, L1-SC, L2-N, L2-SC]), the average duration was calculated, and the longer audio clip in each pair was time-compressed to the pair average while the shorter was expanded to the pair average. This was to ensure that each target was an identical length to each of its four paired sentences.

Stimulus piloting. We ran a pilot experiment (N = 20, no overlap with the main experiment) to find the correct perceived vocal-tract length (VTL) and fundamental frequency (F0) parameters to ensure the target and masker sentences were perceptually streamable from each other and reduce ceiling and floor effects. Participants completed a short version of the listening task (involving 20 trials). For every five participants, we used a new version of the manipulation described in Darwin et al. (2003) to adjust the F0 and VTL of the talkers (Gaudrain et al., 2009; Smith et al., 2007) until transcription performance was around 70% correct.

As a result of the piloting, the target F0 was manipulated to be 96% of its original F0, and the masker F0 was manipulated to be 104% of its original F0. The VTL of the speaker was lengthened to be 103% the length of the original for both the target and masker voices, to ensure the voices remained natural sounding. The pilot also ensured that the Spanish and English sentences were intelligible individually after this manipulation – participants heard five Spanish sentences and five English sentences in isolation and were asked to transcribe them (Spanish = 94.00% correct, SD = 13.19; English = 68.00% correct, SD = 17.64).

The pilot experiment also checked the semantic relationships between the targets and maskers at the keyword level (an additional check of semantic overlap to the LSA). Participants were shown 200 pairs of keywords (i.e., target keywords with their corresponding N or SC masker keyword) and asked to rate on a scale from 1 to 100 how semantically related the words within each pair were. The 80 target sentences which had the least overlap between the SC and target semantic ratings versus the N and target semantic ratings were chosen to be experimental trials. The average rating for the 80 SC pairs was 71.71 (SD = 10.52) and the average rating for the 80 N pairs was 13.33 (SD = 7.15), and this difference between SC and N pair semantic ratings was significant (t = -31.62, *p* < .001). The remaining 20 sentences were used as practice trials and target voice examples.

English vocabulary check. To ensure that participants had an adequate understanding of the L2-English maskers, they completed a brief vocabulary check consisting of 40 English words that had appeared within the masker sentences during the listening task. For each word, participants were presented with four possible Spanish translations and asked to select the correct one. They advanced to the next trial by clicking on one of the options. Their performance was used as an inclusion criterion, as described in section 3.3.3.1. Participants.

3.3.3.3. Procedure

Before starting the experiment, participants were provided with a brief segment of white noise that had been RMS-equalised to an intended presentation level of 60 dB prior to playback (the same level as the stimuli in the listening task). Participants used this sound to adjust their volume to a comfortable listening level. Next, participants completed a headphone check to ensure they were wearing functioning stereo headphones (Woods et al., 2017). The use of antiphase audio for some of the tones meant that the task could only be successfully completed with stereo headphones. If participants failed the headphone check twice (i.e., responded incorrectly to two or more of the six trials), they could not continue with the experiment.

Participants were then randomly assigned to one of four groups, which determined which masker (L1-N, L1-SC, L2-N, L2-SC) they would hear with each target sentence. Participants were presented with instructions for the listening task in Spanish and English, then completed four blocks (one for each masker type) of 20 trials each. The blocks were presented in a random order, and the trials within each block were presented in a random order. All trials within a block belonged to the same condition (e.g., L1-N) as we were interested in investigating whether listeners adapt to maskers over time (i.e., across trials throughout a block). Before starting each block, participants were provided with three target sentences to [re]familiarise themselves with the target voice, and they could play these as many times as needed. They then completed two practice trials, in which they were provided with the correct target transcription after each trial and could replay the sentence pair as many times as needed. For the 20 experimental trials in each block, participants were presented with a sentence pair, then were asked to type what they heard of the target sentence after the offset of the stimulus. After finishing the listening task, participants completed the English vocabulary check, the LexTALE and Lextale-Esp (in a counterbalanced order), then the language background questionnaire.

3.3.4. Analyses

Each target sentence in the listening task contained four keywords. For each keyword, participants received a score of 1 or 0 depending on whether the keyword was correctly transcribed, and an average score was computed for each trial (e.g., three out of four keywords correctly identified resulted in a score of 0.75 for that trial). The following deviations from the target were still scored as correct:

- Any typed word phonologically identical to the keyword, even if the typed word was the same as another existing word (e.g., the keyword was *bear* and the participant typed *bare*).
- Any typed word phonologically dissimilar to the keyword (and any other real word), but spelt correctly except for either one omitted letter (e.g. the keyword was *young* and the participant typed *yong*), one added letter (e.g. the keyword was *Christmas* and the participant typed *Christmast*), or one pair of consecutive letters switched over (e.g. the keyword was *light* and the participant typed *lihgt*).
- If two correctly spelt words were merged into one word (e.g. the two consecutive words were *paper bag* and the participant typed *paperbag*).

- Any diacritics were ignored - if a participant incorrectly added a diacritic (e.g. the keyword was *niños* and the participant typed *ñiños*) or excluded a diacritic (e.g. the keyword was *ratón* and the participant typed *raton*).
- Although it was not pre-registered, a further scoring rule was deemed appropriate. An answer was scored as correct if a participant incorrectly switched one letter for another, even if the typed word was not a homophone of the correct keyword, if the typed word was not another existing word (e.g., the keyword was *brain* and the participant typed *braun*).

The data were analysed using generalised linear mixed effect models (GLMMs) in R (v. 4.4.3 (2025-02-28 ucrt)), on RStudio (v. 2025.05.1 Build 513) with the lme4 package (v. 1.1.36, Bates et al., 2015). A binomial distribution with a logit link was used to model the by-trial scores. The model converged with a full random-effect structure including participant and item (target sentence) intercepts, and random slopes by participant and by item for masker semantic content, masker language, and the interaction between them. The fixed effects were: masker semantic content (N and SC), masker language (L1-Spanish and L2-English), trial number (continuous variable (1-20) rescaled and recentred (mean = 0, range = -1.65 – 1.65)) and the interactions between them. Categorical effects were sum coded (0.5 versus -0.5 for L1 versus L2 and SC versus N). The model used the BOBYQA optimiser (Powell, 2009) and a maximum of 10^9 iterations. The full Target Accuracy Model was as follows:

```
glmer(TranscriptionAccuracy ~
MaskerSemanticContent*MaskerLanguage*TrialNumber +
(1 + MaskerSemanticContent*MaskerLanguage|Participant) +
(1 + MaskerSemanticContent*MaskerLanguage|Item),
family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”,
optCtrl = list(maxfun = 1e9)))
```

The data were also analysed using exploratory models to investigate intrusions, which are instances where the participant incorrectly transcribed a keyword from the masker sentence instead of the target. Intrusions made up 34.69% of the errors in the listening task (across all conditions). These data were analysed because the rate of intrusions could constitute additional evidence of semantic interference, alongside overall transcription accuracy: if listeners process the semantic content of masker speech, they may be more likely to experience intrusions from an SC than N masker due to increased semantic interference from more closely related items. The intrusion rate was scored based on the masker keywords (rather than the target keywords) using the same rules as detailed above. The Intrusion Model used the same fixed and random effects structures as the Target Accuracy Model. Due to a lack of convergence, it was run without intercept/slope correlations. See Appendix B.3 for a description of the Intrusion Error Model, which considered intrusion rate as a proportion of target transcription errors (i.e., whenever a target word was not correctly transcribed, how frequently that error was the result of a masker intrusion). Other than intrusions, errors primarily consisted of phonologically similar substitutions to either the target or masker keyword (i.e., words sharing phonetic overlap, e.g., typing *destruye* instead of *instruye*), or omissions (e.g., participants entered an ellipsis in place of a keyword).

3.3.5. Results

3.3.5.1. Target Transcription Accuracy

There was no significant main effect of masker semantic content (β = -0.01, SE = 0.10, z = -0.09, *p* = .931) in the Target Accuracy Model, with transcription accuracy near identical for N maskers (mean = 0.65, SD = 0.12) and SC maskers (mean = 0.65, SD = 0.10). However, there was a significant main effect of masker language (β = -2.36, SE = 0.11, z = -21.11, *p* < .001), with poorer transcription accuracy when the masker was L1-Spanish (mean = 0.46, SD = 0.13) than L2-English (mean = 0.85, SD = 0.09; Figure 4). There was also a significant main effect of trial number (β = 0.07, SE = 0.01, z = 5.04, *p* < .001), with transcription accuracy increasing across trials.

There was no significant interaction between masker semantic content and masker language (β = 0.06, SE = 0.16, z = 0.38, *p* = .702), with neither the L1 nor L2 condition showing an effect of masker semantic content (average L1 semantic interference effect (i.e., N score minus SC score) = -0.001, SD = 0.18; average L2 semantic interference effect = -0.001, SD = 0.08). However, there was a significant interaction between trial number and masker semantic content (β = -0.06, SE = 0.03, z = -2.24, *p* = .025, Figure 5). Follow-up pairwise comparisons of the estimated slopes indicated that the effect of trial number was significantly positive in both masker semantic content conditions. However, the effect was weaker in the SC condition (β = 0.04, SE = 0.02, z = 1.99, *p* = .047) than in the N condition (β = 0.10, SE = 0.02, z = 5.14, *p* < .001). There was no significant interaction between trial number and masker language (β = 0.02, SE = 0.03, z = 0.85, *p* = .395), nor a three-way interaction between trial number, masker semantic content and masker language (β = -0.10, SE = 0.05, z = -1.85, *p* = .064).

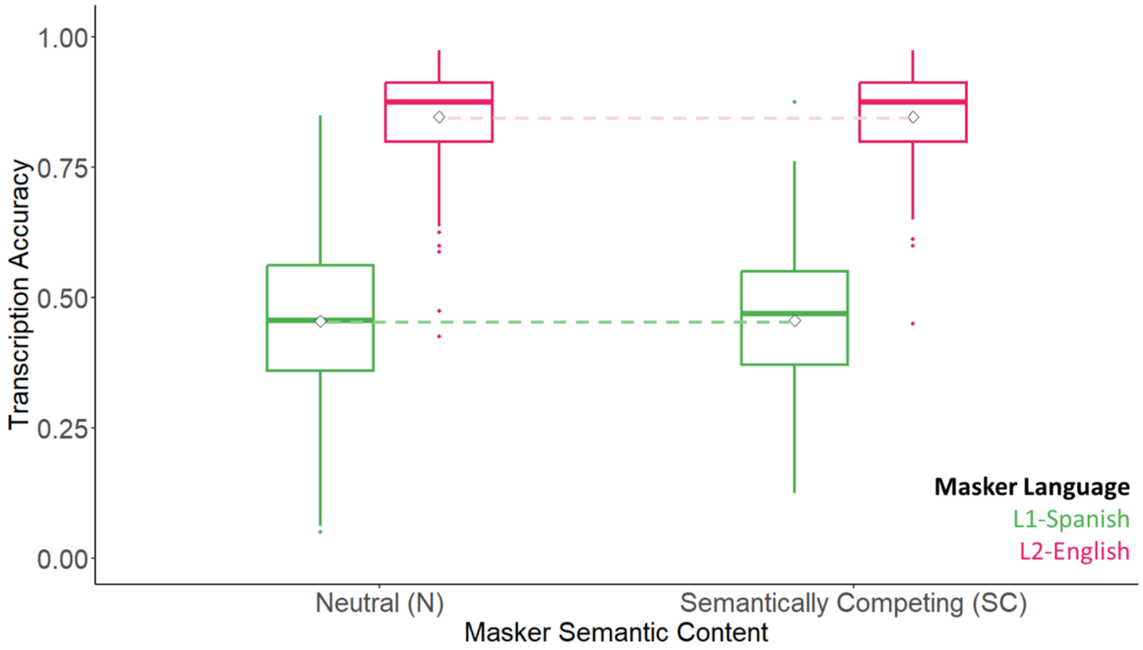


Figure 4 - Average transcription accuracy for each masker semantic content and masker language level

Note. Open diamonds depict the mean score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average scores beyond 3 SDs from the mean.

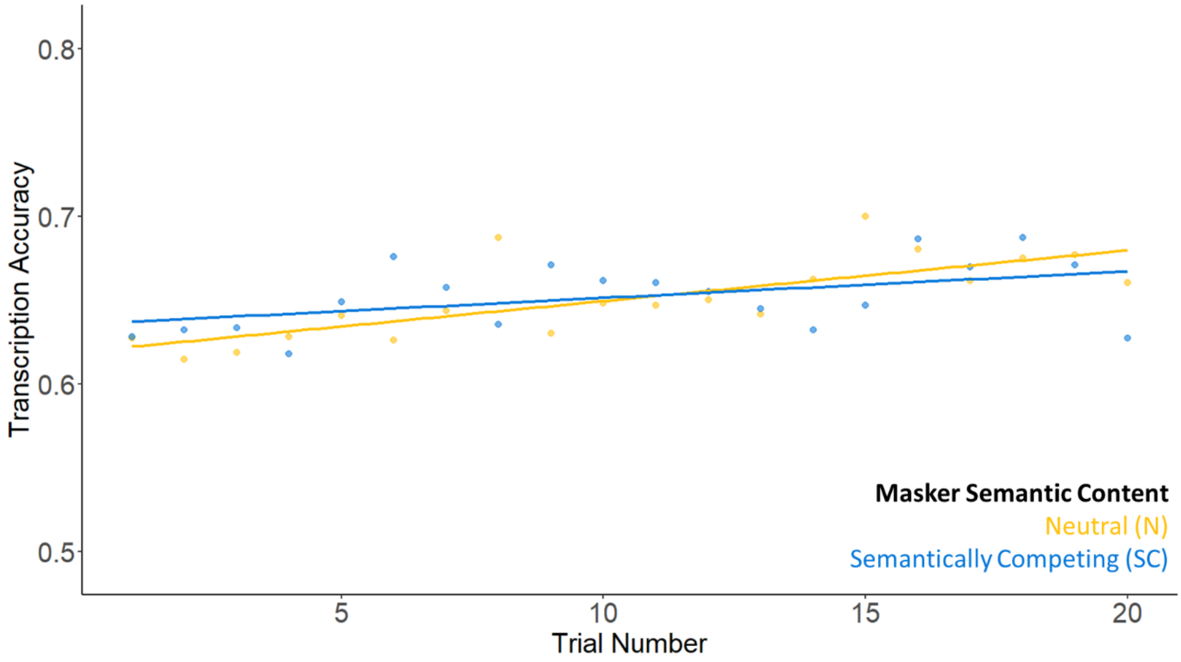


Figure 5 - Average transcription accuracy on each trial for N and SC maskers

Note. Average transcription accuracy (proportion of keywords correctly transcribed) for each trial (1-20) collapsed across masker language.

3.3.5.2. Intrusions

The exploratory analysis of the intrusion rates showed a significant main effect of masker semantic content (β = 1.08, SE = 0.20, z = 5.34, *p* < .001) in the Intrusion Model (Figure 6), with more intrusions from the SC maskers (mean = 0.15, SD = 0.06) than from the N maskers (mean = 0.10, SD = 0.06). There was also a significant main effect of masker language (β = 5.05, SE = 0.21, z = 23.95, *p* < .001), with more intrusions from L1-Spanish (mean = 0.24, SD = 0.10) than from L2-English maskers (mean = 0.004, SD = 0.01). However, there was no significant main effect of trial number (β = -0.09, SE = 0.08, z = -1.20, *p* = .230).

There was a significant interaction between masker semantic content and masker language (β = -0.79, SE = 0.38, z = -2.12, *p* = .034). Follow-up pairwise comparisons using estimated marginal means revealed that when the masker was L1-Spanish, the odds of an intrusion from an SC masker were 1.99 times greater than from an N masker (SE = 0.32, z = 4.31, *p* < .001). When the masker was L2-English, the odds of an intrusion from an SC masker were 4.40 times greater than from an N masker (SE = 1.57, z = 4.15, *p* < .001). Nonetheless, the extremely low intrusion rate when the masker was L2 suggests a floor effect, which means that the larger odds for this condition likely overstate the relevance of this effect. Intrusions from L2 maskers were virtually absent regardless of masker semantic content. None of the interactions including trial number were significant (all *p* > .300). The same pattern of results was found in the Intrusion Error Model (see Appendix B.4 for the full GLMM output).

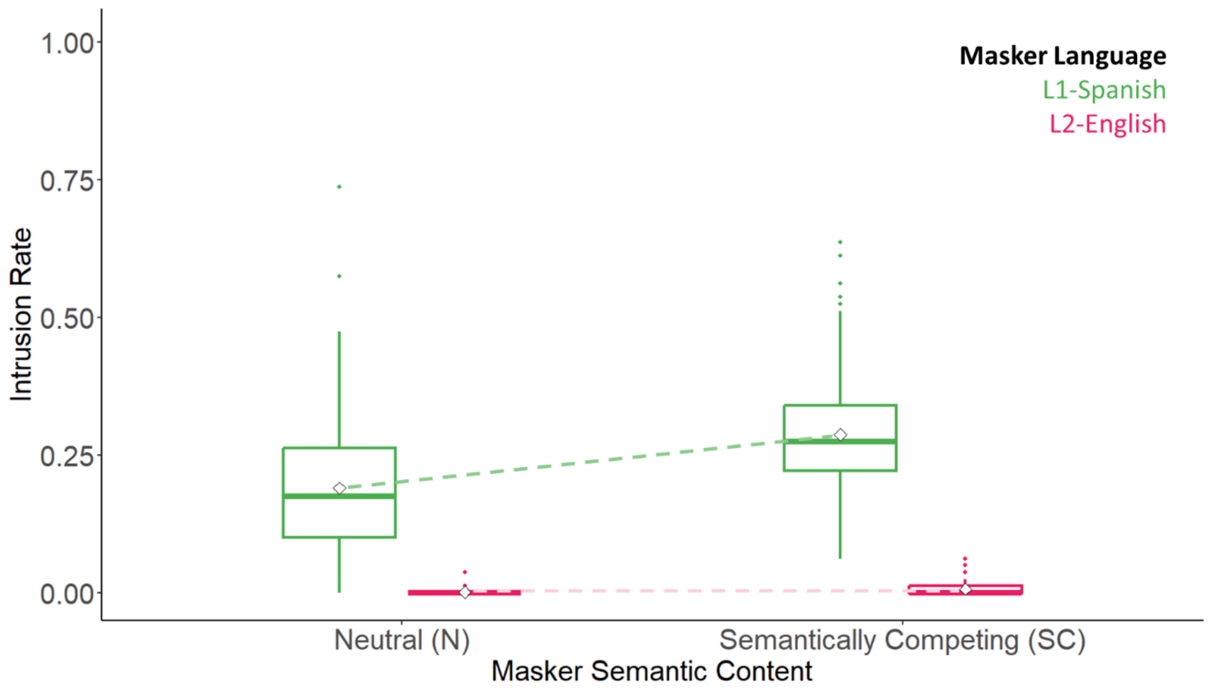


Figure 6 - Average intrusion rate for each masker language and masker semantic content level

Note. Open diamonds depict the mean intrusion score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average intrusion rates beyond 3 SDs from the mean.

3.3.6. Discussion

Experiment 1 aimed to investigate whether listeners process the semantic content of irrelevant background speech when they are listening to another talker, and thus, whether semantic overlap between target and masker can cause interference. Two possible mechanisms were considered: Semantic Filter or Target-Masker Co-Activation. Experiment 1 also investigated whether semantic interference is affected by whether the masker is in L1 or L2, given an L1 target.

Based on the findings from Brouwer et al.'s (2012) study and other research (e.g., Röer & Cowan, 2021; Treisman & Geffen, 1967; Treisman & Riley, 1969; Wood & Cowan, 1995) supporting Treisman's (1964) Attenuation Theory, we expected that semantically competing (SC) maskers would negatively impact transcription accuracy more than neutral (N) maskers. This would support the Target-Masker Co-Activation account – that listeners activate both target and masker when listening and process the masker along with the target, enabling any semantic overlap between target and masker to exert IM.

Our results indicate that, regardless of the language of the masker, semantic relatedness between target and masker does not impact target transcription accuracy. This was the case even though the pilot and LSA scores clearly demonstrated that the SC maskers were more semantically related to the target than were the N maskers. This result is inconsistent with our first hypothesis (H1) and runs counter to previous research showing semantic interference from an L1 masker (e.g., Brouwer et al., 2012). Instead, our findings are in line with previous studies showing a lack of semantic interference from the masker (e.g., Calandruccio et al., 2018; Villard et al., 2024) and suggest that background speech is not semantically processed, at least to the extent that it interferes with target identification. This is consistent with the Semantic Filter account, indicating that listeners can block the activation of semantic representations from the masker (e.g., Lakatos et al., 2013; Szalárdy et al., 2020; Wang et al., 2019).

Our results also showed that transcription of the L1 target was more accurate when the masker was L2 compared to L1. This supports previous research demonstrating that it is easier to understand target sentences when the masker speech is in another language (e.g., Brouwer et al., 2012; Calandruccio et al., 2017; Mepham et al., 2022), as postulated by the Target-Masker Similarity Hypothesis (Brouwer et al., 2012). This result can be interpreted as indicating that bilingual listeners make use of differences between the two languages (e.g., different phonological inventories) to facilitate streaming and segregation of the target and masker. However, the masker language effect can also be explained by differences in L1 and

L2 activation levels (e.g., Dijkstra & van Heuven, 2002). Lower L2 resting activation levels could mean that there is less interference with the (L1) target and/or make it easier for participants to inhibit the L2 masker. It is likely that both mechanisms underlie the observed higher performance in the L2 masker condition compared to the L1 masker condition when the target is presented in the L1.

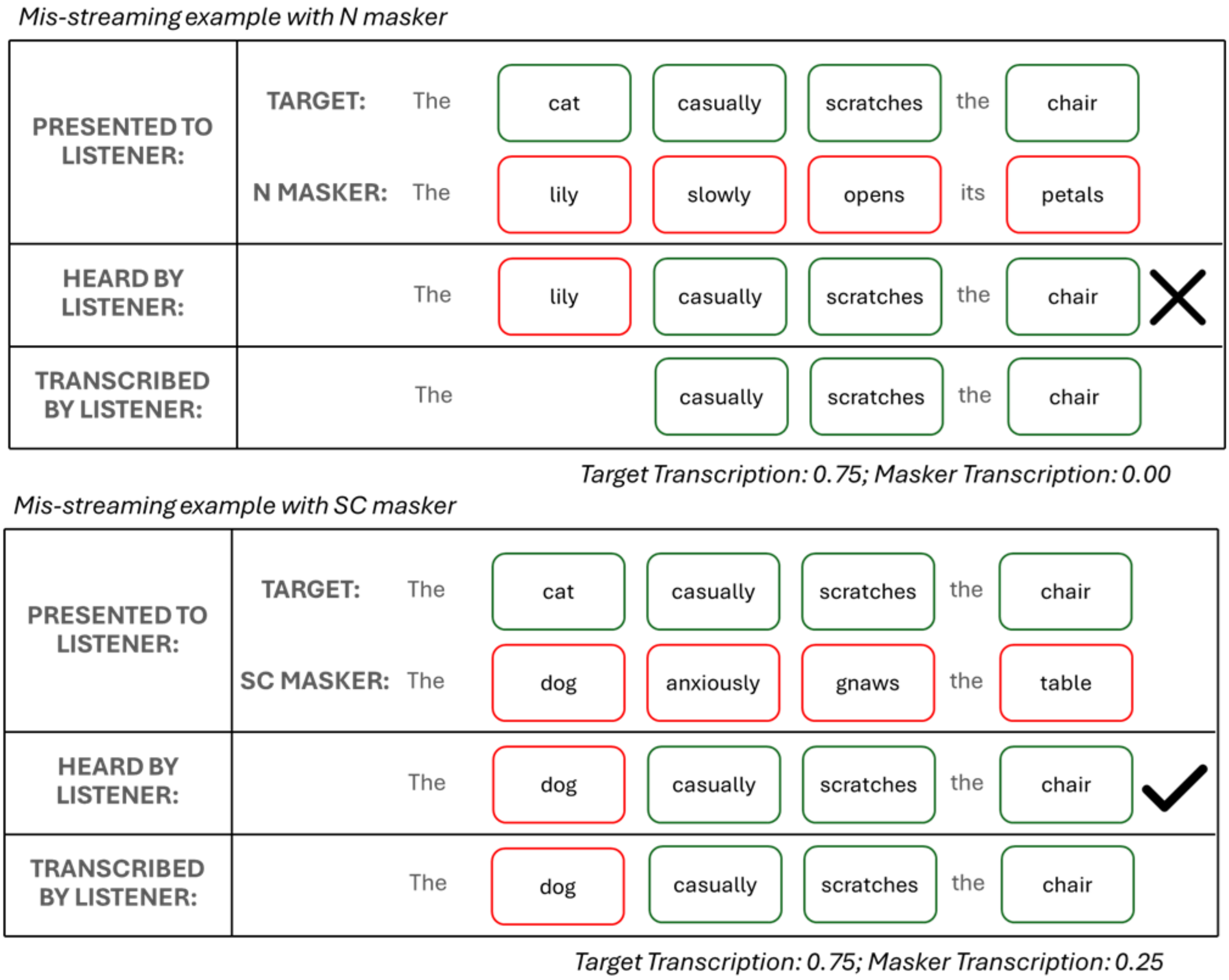
As predicted, transcription accuracy improved during the course of a block. This indicates that listeners can learn to adapt to a masker, improving their ability to suppress it, which is in line with previous research (e.g., Bent et al., 2009; Marrufo-Pérez & Lopez-Poveda, 2025; Mepham et al., 2022, 2025). However, the improvement was comparable for both masker languages, which is contrary to our hypothesis that participants would show faster adaptation to the more challenging L1 masker (e.g., Brown & Strand, 2018). Instead, the results demonstrate that proficient bilingual listeners can learn to direct their attention away from background speech regardless of the language of that speech.

However, transcription accuracy improved more rapidly when the masker was semantically neutral (N) than semantically competing (SC) with the target, indicating that, despite the lack of an overall semantic interference effect on transcription accuracy, SC maskers were more challenging to adapt to. This interaction weakens support for the Semantic Filter account. Indeed, it may be the case that suppressing N compared to SC maskers was less cognitively demanding. Therefore, in the N masker condition, participants had a larger opportunity to improve their performance over time. However, any differences observed were small.

To explore any potential influences of semantic relatedness further, we investigated whether SC maskers intruded on target transcription more than N maskers. Differences in intrusions between SC and N maskers could suggest that maskers are indeed semantically activated and processed and could cause confusion despite not affecting overall target transcription accuracy. As expected, since the task involved transcription in L1, there were almost no intrusions from an L2 masker. More interestingly, there were more intrusions from SC than N maskers, demonstrating a main effect of masker semantic content on intrusion rate. Thus, although the target transcription data suggest participants use a Semantic Filter mechanism to block out the masker content, the intrusions suggest that both target and masker were co-activated to some level, thus creating semantic interference - at least when the target and masker were both L1-Spanish. Therefore, listeners somehow managed to prevent semantic interference from impacting transcription accuracy. Although the semantic intrusion effect was stronger when the masker was L2-English compared to L1-Spanish according to the odds ratio, the floor effect found in the L2-English masker condition reduces the interpretability of this interaction and reduces the relevance of the semantic intrusion effect when the masker was L2-English.

A mechanism intermediate to Target-Masker Co-Activation and the Semantic Filter can be considered. It is possible that masker speech is semantically processed at an early stage of listening, but listeners can suppress interference once the masker is deemed irrelevant. Consistent with this possibility, Tun et al. (2002) conducted a speech-on-speech study in which participants attended to speech streams while ignoring a semantically meaningful or anomalous masker stream (but the semantic relationship of the masker to the target was not manipulated). Younger adults did not show an effect of masker meaningfulness on target listening performance. However, they later recalled the words from the meaningful masker better than those from the meaningless masker. This suggests that the masker semantic content was processed. However, this did not impair their ability to recall the target stream, suggesting that listeners can suppress semantic interference from the masker even if it is initially processed. In the current study, the larger number of SC than N masker intrusions could therefore indicate that the target and masker were co-activated, but with listeners being able to suppress this semantic interference effect from the masker enough to prevent it from impacting their overall target transcription accuracy.

However, increased SC masker intrusions with no effect on target transcription accuracy could also reflect the fact that listeners sometimes unintentionally focused on the masker speech stream (especially considering the acoustic similarity of the two sentence streams). Mis-streaming (but without any co-activation) could then result in more SC than N intrusions on the grounds of semantic coherence, as assessed by the listener at the transcription stage. Take for instance the target *The cat casually scratches the chair*, presented against the SC masker *The dog anxiously gnaws the table* versus the N masker *The lily slowly opens its petals*. If the listener unintentionally focused on the start of the masker sentence, then switched to the target, participants in the SC condition might hear *The dog casually scratches the chair*, which is semantically coherent, whereas participants in the N condition might hear *The lily casually scratches the chair*, which is semantically incoherent. Based on the coherence of what they heard, the listener may then accept or reject the mis-streamed masker words, with a higher likelihood of reporting the masker word in the SC condition than the N condition (see Figure 7 for a visualisation). This would lead to a higher intrusion rate in the SC than N condition despite no difference in terms of target transcription accuracy.



Mis-streaming example with N masker

PRESENTED TO LISTENER:	**TARGET:**	The	cat	casually	scratches	the	chair	
	N MASKER:	The	lily	slowly	opens	its	petals	
HEARD BY LISTENER:		The	lily	casually	scratches	the	chair	✗
TRANSCRIBED BY LISTENER:		The		casually	scratches	the	chair	

Target Transcription: 0.75; Masker Transcription: 0.00

Mis-streaming example with SC masker

PRESENTED TO LISTENER:	**TARGET:**	The	cat	casually	scratches	the	chair	
	SC MASKER:	The	dog	anxiously	gnaws	the	table	
HEARD BY LISTENER:		The	dog	casually	scratches	the	chair	✓
TRANSCRIBED BY LISTENER:		The	dog	casually	scratches	the	chair	

Target Transcription: 0.75; Masker Transcription: 0.25

Figure 7 - Visualisation of how a participant might accept or reject a mis-streamed masker keyword

Note. Within both examples, a participant has mis-streamed the sentences, hearing the first word of the masker before switching to the target stream. *Lily* is not retained for transcription because it is semantically incompatible with the target sentence, whereas *dog* is transcribed because it is semantically compatible with the target sentence. Therefore, only *dog* is scored as an intrusion. Thus, in the N masker condition, the participant receives an Intrusion Rate score of 0, whereas in the SC masker condition, the participant receives an Intrusion Rate score of 0.25 (for *dog*), despite the same listening process occurring in both examples. In both conditions, the participant would receive a Target Transcription score of 0.75 (for *casually*, *scratches*, and *chair*).

In sum, the semantic intrusion effect in the absence of a semantic effect on target transcription accuracy could be explained in two ways. It could be due to the genuine co-activation of target and masker, and subsequent successful suppression of the masker stream. Alternatively, the target and masker may not be co-activated, and the semantic intrusion effect could be explained by the listeners transiently mis-streaming the target and the masker. Listeners are then more likely to transcribe an SC masker word than an N masker word on the grounds of semantic coherence, resulting in an apparent semantic intrusion effect. Experiment 2 was designed to disambiguate these two explanations.

3.4. Experiment 2

3.4.1. Hypotheses

Experiment 2 had two aims: (1) to replicate the findings from Experiment 1, and (2) to explore potential explanations for the greater number of intrusions from a semantically competing (SC) compared to neutral (N) masker. Given that Experiment 1 showed only a very small number of intrusions in the L2 masker condition, Experiment 2 did not include the L2 masker condition. Instead, participants completed the listening task described in Experiment 1 but only including the L1-Spanish masker blocks. The outcome variables were target transcription accuracy and intrusion rate.

We expected to replicate the results of Experiment 1. Therefore, it was hypothesised that masker semantic content would not significantly impact target transcription accuracy. However, we expected significantly more intrusions when the masker was SC compared to N (i.e., a significant main effect of masker semantic content on intrusion rate). We also expected target transcription accuracy to increase across trials (i.e., a significant main effect of trial number on target transcription accuracy) and this effect to be larger in the N masker block compared to the SC masker block (i.e., a significant interaction between trial number and masker semantic content on target transcription accuracy). We expected no effect of trial number on intrusion rate, regardless of masker semantic content.

To address the second aim of Experiment 2, participants completed a surprise recognition task following the listening task. This was to investigate which masker words were processed, regardless of whether they were transcribed as a result of a semantic coherence judgement in the listening task. During this task, they were shown written versions of words that they had heard in the listening task (both target and masker keywords) and new words. They were asked to indicate whether each word had appeared in the listening task. There are three possible outcomes from this analysis.

Firstly, if both speech streams are genuinely semantically co-activated during the listening task (e.g., Brungart & Simpson, 2003; Conway et al., 2001; Rivenez et al., 2006; Röer & Cowan, 2021), and if SC maskers produce greater semantic interference than N maskers (in line with Brouwer et al., 2012), then SC keywords should be better recognised than N keywords in the recognition task.

If, instead, the semantic intrusion effect arises from mis-streaming and misallocation of words between the two competing sentences, without semantic co-activation, then SC and N keywords should be recognised to a similar extent in the recognition task. This is because mis-streaming is assumed to occur equally often for SC and N maskers.

Alternatively, if the mis-streaming account is true, but the process of rejecting a masker keyword as a target requires engagement of cognitive processes, then N keywords may be better recognised than SC keywords in the recognition task, due to their being overall more SC than N intrusions.

3.4.2. Methods

3.4.2.1. Participants

Participants were recruited through Prolific (www.prolific.co) and tested online using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). To ensure adequate power for replication of Experiment 1's listening task, we aimed to collect data from 128 participants, allowing for equal sample sizes across 16 counterbalancing groups. All participants completed both conditions of the listening task (SC and N). For the recognition task, each participant was assigned to one of two conditions (SC or N). Due to an experimenter error affecting the SC condition of the recognition task, additional participants had to be recruited. As a result, listening task analyses included data from 159 participants (mean age = 27.58 years, SD = 3.80; 85 female, 70 male, 3 other, and 1 preferred not to say; see Table 4 for more details), and recognition task analyses included data from 115 participants (mean age = 27.64; SD = 3.88; 56 female, 56 male, 2 other, 1 preferred not to say, see Appendix B.5 for more details), all of whom were also included in the listening task analyses.

All participants were L1-Spanish speakers, whose use of Prolific (an English platform) indicated proficient knowledge of English. A questionnaire was used to ensure participants met the following, pre-registered inclusion criteria: aged between 18 and 35 years, no self-reported neurological condition or language or reading difficulties (including dyslexia); normal hearing and normal (or corrected to normal) vision; acquired Spanish before the age of three years; resided in Spain for at least 10 years; and scored at least 60% on the Lextale-Esp (see section 3.3.3.2. Materials, LexTALE and Lextale-Esp above). A total of 19 participants were excluded from analyses based on these criteria (therefore, a total of 178 participants were tested, with 159 of those included in the listening task analyses, and 115 of those included in the recognition task analysis). To address problems with slow recruitment, we removed the requirement that participants had to live in the UK or Spain at the time of testing. Instead, the participants could reside in any country except those in North or South America.

Table 4 - L1-Spanish language background for Experiment 2 listening task sample (N = 159)

Language background measure	Mean	SD	Min	Max
Age of acquisition (years)	0.02	0.18	0.00	2.00
LexTale proficiency (%)[a]	91.74	7.95	60.00	100.00
Self-reported understanding (/10)	9.94	0.27	8.00	10.00
Self-reported speaking (/10)	9.94	0.26	8.00	10.00
Self-reported reading (/10)	9.98	0.14	9.00	10.00
Self-reported writing (/10)	9.88	0.44	8.00	10.00
Language frequency of use (/4)[b]	Mean = 2.89, SD = 0.71, Min = 0.94, Max = 4.00			

[a] The L1-Spanish values exclude the data from one participant who scored 9.17% in the Lextale-Esp (see section 3.3.3.2. Materials, LexTALE and Lextale-Esp above). After completing the experiment, they stated that they mixed up the response buttons.

[b] This value is from the language background questionnaire (see section 3.3.3.2. Materials, Language background questionnaire above) and determines how frequently participants use Spanish in their daily lives (0 = All another language, 1 = Mostly another language, 2 = Half each language, 3 = Mostly Spanish, 4 = All Spanish).

3.4.2.2. Materials

The materials used in Experiment 2 were largely identical to those used in Experiment 1, with a few exceptions. The language background questionnaire was modified to remove items regarding English language background and proficiency. The frequency of use items were reworded to refer to “another language” rather than “English”. The English LexTALE was omitted from this experiment; however, the Lextale-Esp was administered as in Experiment 1. The listening task was identical to that of Experiment 1, except that the two L2 masker blocks (i.e., when masker language was L2-English) were removed, leaving two blocks of 20 trials each – both with an L1-Spanish target and masker. The English vocabulary check used in Experiment 1 was not required in Experiment 2 because there were no English maskers.

The only additional task in Experiment 2 was the recognition task. In this task, participants saw 90 words, one at a time (one word per trial). For each word, they indicated whether they had heard it before during the listening task. There were three types of word, with 30 items of each type:

- Target-Heard (T-H): target keywords that participants were presented with during the listening task.
- Masker-Heard (M-H): masker keywords that participants were presented with during the listening task.
- Masker-Not Heard (M-NH): masker keywords that participants had not been presented with during the listening task.

The T-H and M-H words were drawn from each participant's final 10 trials of the final block of the listening task. This reduced the likelihood that memory would affect performance in the recognition task. Because the listening task used a block design, participants completed the recognition task for only one condition (either N or SC). As a result, analyses of the recognition task used a between-subjects design. The M-NH words were the alternative masker keywords corresponding to the T-H items (i.e., if the participant was in the SC condition of the recognition task, they might see *cat* as a T-H word, *dog* as a M-H word, and *lily* as a M-NH word). Therefore, if a participant's final block in the listening task was the SC condition, the M-NH words were the N masker keywords associated with the T-H words (see Figure 8 for a visualisation of this design).

Participants were shown every keyword from the final 10 trials of their last listening task block (plus the corresponding M-NH words), except for adverb keywords, which were excluded to reduce task duration and because they lacked uniqueness. Since the aim of Experiment 2 was to investigate the semantic intrusion effect, only the M-H trials were included in the main analysis of the recognition task.

Listening Task example trial:

Target: The cat casually scratches the chair.
SC Masker: The dog anxiously gnaws the table.

Recognition Task example stimuli:

T-H	*M-H (used in analyses)*	*M-NH*
Cat	Dog	Lily
Scratches	Gnaws	Opens
Chair	Table	Petals

Figure 8 - Example of the recognition task design

Note. If a participant was in the SC recognition task condition, then in the final block of the listening task they might have been presented with *The cat casually scratches the chair* (target) simultaneously with *The dog anxiously gnaws the table* (SC masker). The N masker alternative to this target (which the participant did not hear) is *The lily slowly opens its petals*. In the recognition task, the participant would then be presented with the following words: *cat* (T-H)*, scratches* (T-H)*, chair* (T-H)*, dog* (M-H)*, gnaws* (M-H)*, table* (M-H)*, lily* (M-NH)*, opens* (M-NH)*, petals* (M-NH). This example is in English for clarity, but all keywords were presented in Spanish in the recognition task.

3.4.2.3. Procedure

Participants first completed a headphone check. As in Experiment 1, if they failed this twice, they could not complete the rest of the experiment. This was followed by the listening task, which proceeded as in Experiment 1, with the exception that only the L1-Spanish masker language conditions were included. Each participant completed two blocks (N and SC) of 20 trials each.

The listening task was followed by the surprise recognition task, which the participants had not been told about to make sure they did not intentionally memorise the stimuli from the listening task. During the recognition task, participants were presented with words on a screen, displayed sequentially in a randomised order. They were asked whether they heard each word in the listening task and responded via a keyboard button press. Immediately after responding, the next trial began. The recognition task was followed by the Lextale-Esp and the language background questionnaire.

3.4.3. Analyses

The listening task data were scored in the same way as in Experiment 1, using the same scoring rules when misspellings were made. The data were analysed using GLMMs in R (v. 4.4.1 (2024-06-14 ucrt)), on RStudio (v. 2024.12.1.563) with the lme4 package (v. 1.1.35.5, (Bates et al., 2015). A binomial distribution with a logit link was used to model the by-trial scores. The model converged with a full random-effect structure including participant and item (target sentences) intercepts, and random slopes by participant and by item for masker semantic content. Fixed effects were masker semantic content (N and SC), trial number (continuous variable (1-20) rescaled and recentred (mean = 0, range = -1.65 – 1.65)) and the interaction between them. Categorical effects were sum coded (0.5 versus -0.5 for SC versus N). The model used the BOBYQA optimiser (Powell, 2009) and a maximum of 10^9 iterations. The full Transcription Accuracy Model was as follows:

```
glmer(TranscriptionAccuracy ~ MaskerSemanticContent*TrialNumber +
(1 + MaskerSemanticContent|Participant) + (1 + MaskerSemanticContent|Item),
family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”,
optCtrl = list(maxfun = 1e9)))
```

The Intrusion Model was identical, except that intrusion rate (proportion of masker words transcribed in each trial) was the outcome variable. An Intrusion Error Model (as described in Appendix B.3) was also run, using the same random and fixed effects structure. Both intrusion models converged with intercept/slope correlations.

As we were primarily interested in the semantic intrusion effect, only the M-H trials were included in the analysis of the recognition task data, with the outcome variable being whether or not the keyword was recognised (recognised = 1, not recognised = 0). The data were analysed using GLMMs and included random slopes by item for masker semantic content. The model converged with a full random-effect structure including participant and item (keyword) intercepts. The fixed effect was masker semantic content (N and SC), which was sum coded (0.5 versus -0.5 for SC versus N). The model used the BOBYQA optimiser and a maximum of 10^9 iterations. The full Recognition Model was as follows:

```
glmer(RecognitionRate ~ MaskerSemanticContent + (1|Participant) +
(1+ MaskerSemanticContent|Item), family = binomial(link = "logit"),
glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e9)))
```

To address the potential confound that memory for typed words may be stronger than for words that were not typed (e.g., masker words that were heard but rejected as targets and therefore not transcribed), another version of the model was run: Recognition Transcription Model. This included a fixed effect of transcription status (i.e., whether or not the word was transcribed by the participant in the listening task: -0.5 for transcribed, 0.5 for not transcribed), as well as the interaction between transcription status and masker semantic content. In the pre-registration, we proposed addressing this confound by rerunning the Recognition Model after having removed all trials in which the participant had transcribed the keyword in the listening task. However, removing data points would reduce statistical power, therefore, the Recognition Transcription Model was used instead to account for this confound while retaining all trials, preserving statistical power. In accordance with the pre-registration, the recognition models were also conducted excluding participants with a *d prime* score below 0.50 in the T-H and M-NH conditions (28 participants excluded), as these scores indicated less reliable responses.

3.4.4. Results

3.4.4.1. Listening Task Replication

Target transcription accuracy. There was no significant main effect of masker semantic content on target transcription accuracy (β = 0.10, SE = 0.13, z = 0.80, *p* = .427; Figure 9), with average transcription accuracy comparable when the masker was N (mean = 0.39, SD = 0.16) compared to SC (mean = 0.40, SD = 0.14). There was also no significant main effect of trial number (β = 0.03, SE = 0.01, z = 1.89, *p* = .059). However, there was a significant interaction between trial number and masker semantic content (β = -0.16, SE = 0.03, z = -5.62, *p* < .001; Figure 10). Follow-up pairwise comparisons of the estimated slopes indicated that the effect of trial number on target transcription accuracy was significantly positive in the N condition (β = 0.11, SE = 0.02, z = 5.27, *p* < .001) and significantly negative in the SC condition (β = -0.05, SE = 0.02, z = -2.66, *p* = .008).

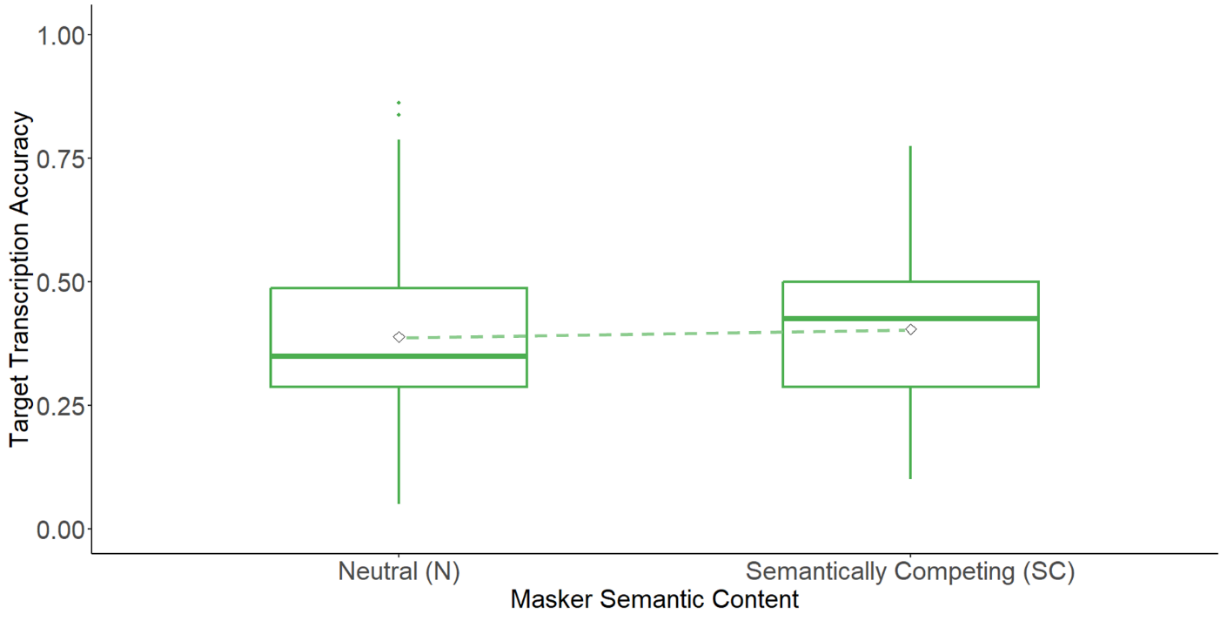


Figure 9 - Effect of masker semantic content on target transcription accuracy

Note. Open diamonds depict the mean target transcription accuracy score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average target transcription accuracies beyond 3 SDs from the mean.

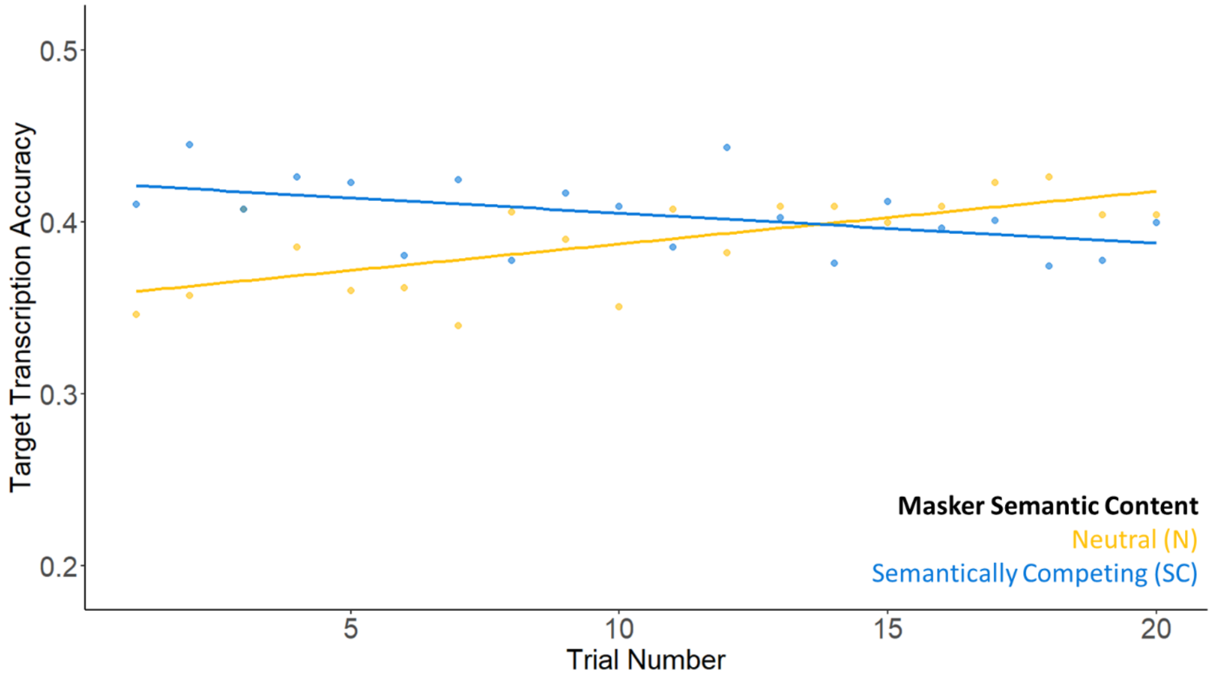


Figure 10 - Change in target transcription accuracy across trials

Note. Average target transcription accuracy (proportion of target keywords correctly transcribed) for each trial (1-20).

Intrusion rate. There was a significant main effect of masker semantic content on intrusion rate (β = 0.62, SE = 0.14, z = 4.29, *p* < .001) in the Intrusion Model (Figure 11), with more intrusions from the SC maskers (mean = 0.33, SD = 0.12) than from the N maskers (mean = 0.24, SD = 0.13). There was also a significant main effect of trial number (β = 0.06, SE = 0.02, z = 3.99, *p* < .001), with intrusions becoming more frequent across trials (Figure 12). There was no significant interaction between trial number and masker semantic content (β = 0.02, SE = 0.03, z = 0.54, *p* = .590). The same pattern of results was found in the Intrusion Errors Model, see Appendix B.6 for the full GLMM output.

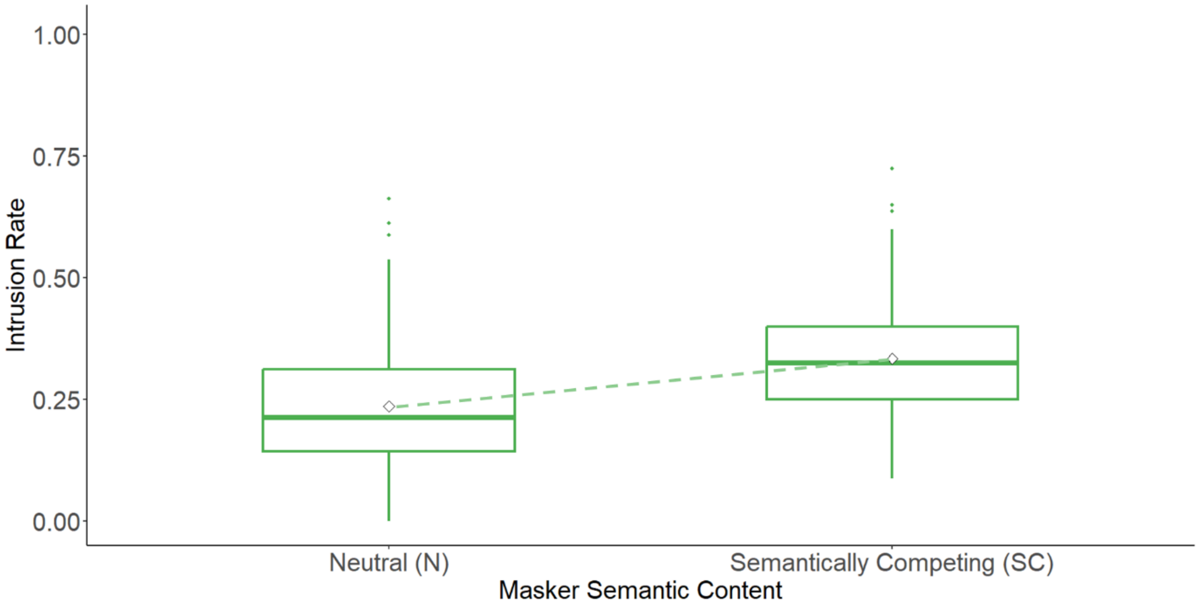


Figure 11 - Effect of masker semantic content on intrusion rate

Note. Open diamonds depict the mean intrusion score for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average intrusion rates beyond 3 SDs from the mean.

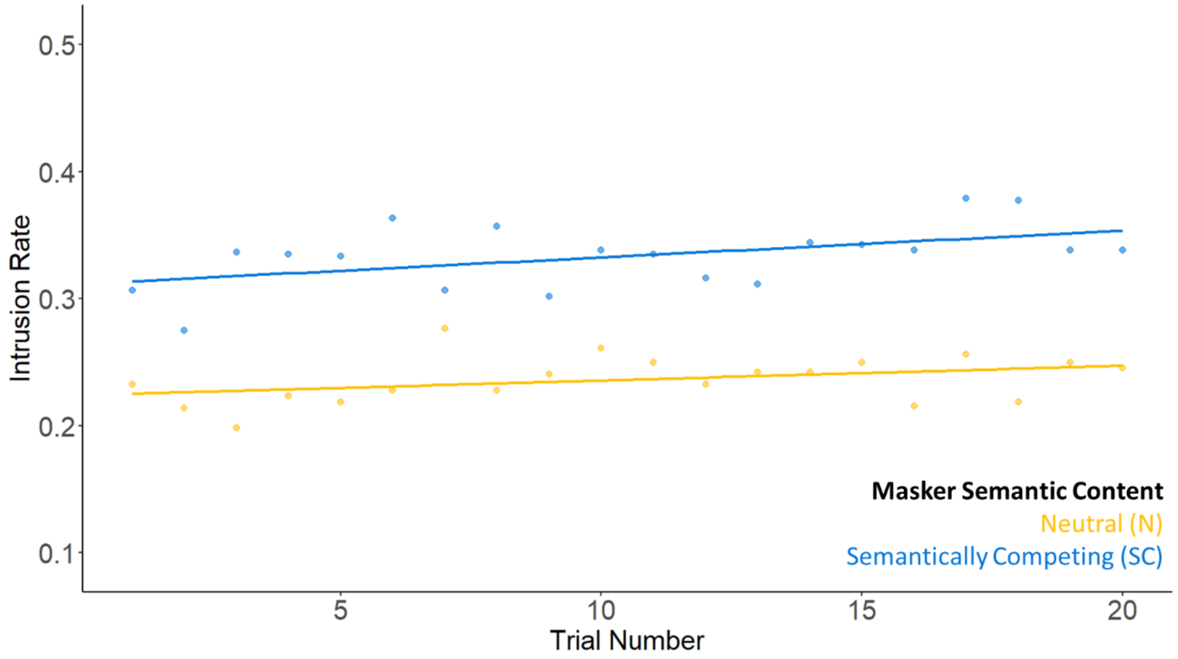


Figure 12 - Change in intrusion rate across trials

Note. Average intrusion rate (proportion of masker keywords transcribed) for each trial (1-20).

3.4.4.2. Recognition Task

There was a significant main effect of masker semantic content on recognition rate (β = 0.49, SE = 0.17, z = 2.93, *p* = .003), with SC maskers (mean = 0.52, SD = 0.16) better recognised than N maskers (mean = 0.42, SD = 0.17) in the Recognition Model. However, in the Recognition Transcription Model, the main effect of masker semantic content was no longer significant (β = 0.14, SE = 0.16, z = 0.89, *p* = .373), whereas there was a significant main effect of transcription status (β = 2.71, SE = 0.12, z = 22.73, *p* < .001), with the words that were transcribed in the listening task (mean = 0.84, SD = 0.17) better recognised than those that were not transcribed (mean = 0.31, SD = 0.17). There was no significant interaction between transcription status and masker semantic content (β = -0.22, SE = 0.23, z = -0.95, *p* = .345; Figure 13). When comparing the two recognition models with an ANOVA, the Recognition Transcription Model fitted the data significantly better than the Recognition Model (*p* < .001), as indicated by a smaller AIC value in the Recognition Transcription Model

(3745.70) than in the Recognition Model (4441.10). The same pattern of results (for both recognition models) was found when running the analyses without participants whose *d prime* was less than 0.50 (see Appendices B.7 and B.8).

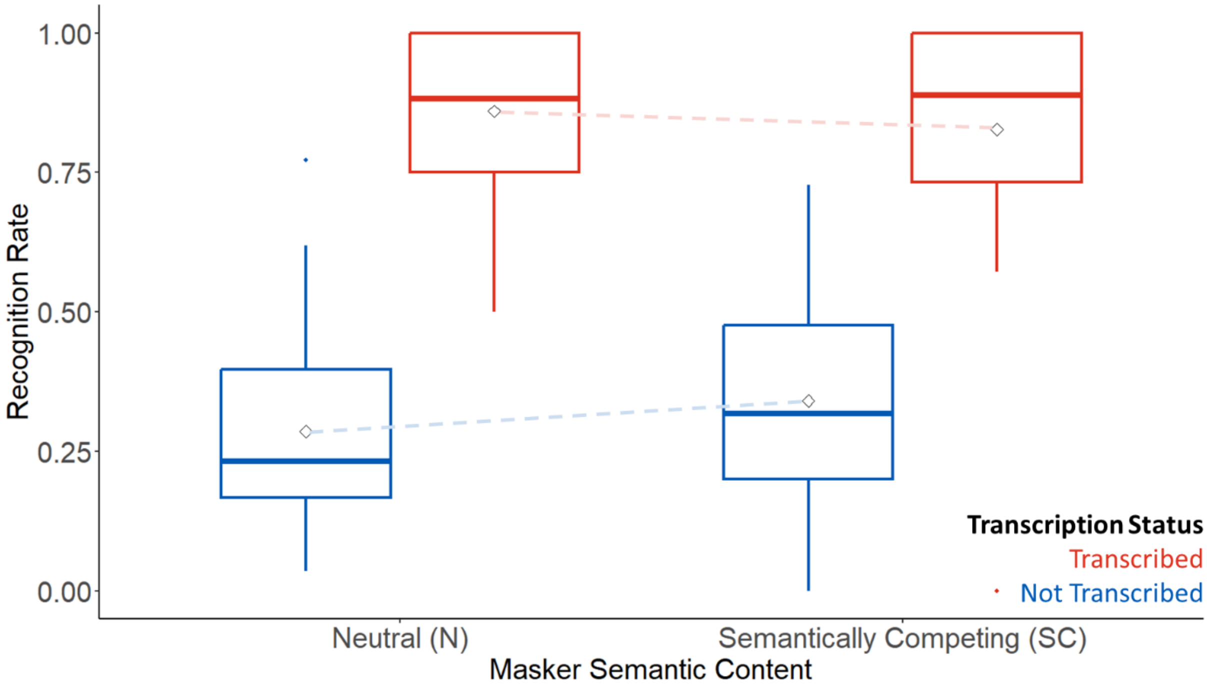


Figure 13 - Effect of masker semantic content and transcription status on recognition rate

Note. Open diamonds depict the mean recognition rate for each condition. Whiskers extend to 3 standard deviations; upper and lower horizontal lines on each box refer to Q1 and Q3; central horizontal line on each box refers to Q2 (median score). Dots depict participants with average recognition rates beyond 3 SDs from the mean.

These results confirm the main findings from Experiment 1, namely that a) listeners' target transcription is not influenced by the semantic content of the masker, but b) when mistakes are made, intrusions more often stem from SC than N maskers. Once the effect of whether a word was transcribed in the listening task was accounted for, the recognition task analysis showed that recall for masker words was not impacted by whether the word was semantically related or unrelated to the target in the listening task. This indicates that the semantic intrusion effect may be the result of listeners using semantic coherence to decide whether they have mis-streamed the sentences, as discussed further in section 3.5. General Discussion.

While the main findings regarding semantic intrusions were replicated from Experiment 1, many of the results involving trial number were not. Both experiments showed an increase in transcription accuracy throughout the block when the masker was neutral. This aligns with previous research indicating that listeners can adapt to speech maskers with practice over time (e.g., Bent et al., 2009; Marrufo-Pérez & Lopez-Poveda, 2025; Mepham et al., 2022, 2025). However, SC maskers (our main condition of interest) showed a positive relationship with trial number in Experiment 1 and a negative relationship in Experiment 2. Across the two experiments, there were therefore no stable changes in transcription accuracy for SC maskers within each block. Similarly, the intrusion analyses showed an increase of intrusions throughout the block (main effect of trial number) in Experiment 2 but not in Experiment 1. These inconsistent patterns across experiments could be due to the relatively small number of trials within each block compared to other studies investigating the effects of masker adaptation over time (e.g., Mepham et al., 2022, 2025), reducing the likelihood of finding a concrete effect of masker adaptation.

3.5. General Discussion

The overarching aim of this study was to investigate whether semantic overlap between target and background speech exerts informational masking (IM) and disrupts transcription accuracy during selective listening. Answering this question can help to determine which of two possible mechanisms are used to cope with semantic content from background speech when selectively listening: the Semantic Filter mechanism, which is suggested to block semantic representations of the masker from being activated; or the Target-Masker Co-Activation mechanism, which is suggested to result in the activation and processing of the semantic content of unattended background speech. To test whether semantic interference varied as a function of masker language, we investigated whether masker language modulated the hypothesised semantic interference effect, given the weaker connections for L2 words and semantic representations relative to L1 (e.g., Kroll & Stewart, 1994). However, Experiment 1 demonstrated that semantic interference effects did not depend on masker language, so this variable was removed for Experiment 2.

Findings from both experiments consistently demonstrated that the semantic overlap between target and masker did not affect target transcription accuracy. This result supports the idea of a Semantic Filter blocking the activation of masker semantic content, preventing it from causing confusion when listening to the target. This also aligns with Broadbent's (1958) Early-Filter Theory and previous speech-on-speech research (e.g., Calandruccio et al., 2018; Villard et al., 2024). This demonstrates that even when the masker semantically overlaps with the target (rather than is either semantically meaningful or not), it still does not interfere with target perception.

However, both experiments also consistently demonstrated that SC maskers intruded on target transcription more than N maskers: When transcription mistakes were made, they were more likely to be borrowed from the masker keywords in the semantically competing condition then in the neutral condition. This suggested an intermediate potential mechanism: that the target and masker are co-activated, as described by the Target-Masker Co-Activation account, but listeners can successfully suppress the resulting semantic confusion from impacting overall target perception. However, the Semantic Filter account could not be ruled out at this point, as the listener may only be processing the semantic content of one stream at a time, but may mis-stream the two sentences (i.e., switch between the two speech streams), and use semantic coherence judgements at the transcription stage to decide whether to accept or reject a heard word as part of the target sentence (as described in the Experiment 1 Discussion). Therefore, Experiment 2 aimed to investigate these two explanations using a surprise recognition task.

In Experiment 2 of the current study, the initial recognition analysis model indicated that SC masker words were better recognised than N masker words. This suggested that the masker words were activated and processed to an extent, with the SC masker words being better recognised because of the larger IM they exert due to semantic confusion. This supported the Target-Masker Co-Activation hypothesis. However, after controlling for whether a word had been transcribed during the listening task, this effect disappeared. This suggests that in the first initial model, the SC words were better remembered because they were transcribed more often (i.e., there were more SC than N intrusions). Once this was controlled for, neither masker type caused significantly more semantic confusion than the other. This suggests that target and masker are not simultaneously co-activated, but rather the listener occasionally mis-streams the two sentences and makes semantic coherence judgements to determine which words they heard are part of the target. This mechanism, rather than co-activation, could be responsible for the increased intrusions from SC than N maskers.

The Semantic Filter mechanism suggests that listeners rely on low level, bottom-up (phonological, spatial, etc.) cues to identify the target stream. When these cues are limited (e.g., when the voices of two talkers are similar, as in the current study), attention may unintentionally shift between competing talkers. In the current study, this could result in a listener hearing the beginning of one sentence (e.g., “*The dog…*”) and the continuation of another (e.g., “*… casually scratches the chair*”), while the listener is unaware that a stream shift has happened. When attention is allocated to a talker, whether or not it is the correct target stream, cognitive processes block the activation of semantic representations of any other speech. Consequently, for the sentences *The dog anxiously gnaws the table* and *The cat casually scratches the chair,* only *cat* or *dog* is processed, not both. This mechanism prevents the semantic content of competing speech, or the semantic overlap between the competing speech and target, from affecting target perception, explaining why masker semantic content did not impact target perception. However, masker semantic content affects intrusion rate because unintentional shifts between streams are more likely to be noticed when masker intrusions are not semantically coherent with the target, which is more likely when the masker is N compared to SC.

However, these findings do not directly oppose co-activation. It is possible that the high acoustical similarity between the two speech streams made this task particularly challenging, thus depleting listeners' cognitive resources (c.f. Lavie, 1995) and preventing them from simultaneously processing both the target and masker semantically. In an easier task (e.g., if the target and masker voices were easier to parse due to larger F0 differences between the voices), listeners might have spare cognitive resources allowing for co-activation, potentially impacting target transcription. Therefore, future research should investigate whether a semantic interference effect is found when the two simultaneous talkers have more acoustically different voices, perhaps using speech reception thresholds (SRT) rather than transcription accuracy to avoid ceiling effects.

While previous research supports the idea that listeners can semantically co-activate both target and masker speech (e.g., Brouwer et al., 2012; Brungart & Simpson, 2003; Conway et al., 2001; Rivenez et al., 2006; Röer & Cowan, 2021), the current study suggests that, at least under cognitively demanding conditions, the semantic overlap between target and masker does not exert IM, indicating that unattended maskers are not semantically processed. This is likely because auditory processes prioritise the identification of the target speech, so suppress the semantic processing of the masker. In the case of stream uncertainty (i.e., when the speech streams are challenging to parse, resulting in mis-streaming) and a lack of visual or spatial cues, listeners make use of semantic coherence to resolve this uncertainty, sometimes resulting in intrusions. Although this study does not provide evidence supporting Target-Masker Co-Activation, it does provide evidence of how listeners strategically attempt to manage selective listening in an adverse listening condition, potentially using a semantic filter mechanism.

Chapter 4: General Discussion

4.1. Summary of findings

The aim of this thesis was to examine speech processing in bilingual individuals, with a particular focus on listening in adverse conditions. A further goal was to investigate these processes in dual-language contexts (i.e., two known languages spoken simultaneously), which remain under-explored despite their relevance for understanding how bilinguals manage two languages concurrently. Specifically, this thesis investigated how bilinguals cope with different forms of adverse listening conditions. Chapter 2 examined low-level influences of energetic masking (EM) and the extent to which spatial separation between talkers mitigates EM and facilitates processing each language. Chapter 3 addressed higher-level influences on listening, focusing on how the semantic relationship between target and masker speech affects listening to a target – a novel research question even in monolinguals in single-language contexts.

Previous research on monolingual listeners demonstrates the beneficial effect of spatial release from masking (SRM; e.g., Ihlefeld & Shinn-Cunningham, 2008a; Knight et al., 2023; Marrone et al., 2008; Pinto et al., 2020), which is when EM is reduced through separating the two sound signals in space, increasing interaural differences and enhancing the streamability of concurrent speech signals. Even when attending to two voices simultaneously (i.e., divided attention) and spatial separation introduces additional cognitive demands, such as shifting attention across ears, the benefits of SRM outweigh these cognitive costs (e.g., Knight et al., 2023). While this is established in monolingual listeners, these effects may differ for bilinguals because EM disproportionately disrupts L2 processing compared to L1 (e.g., Black & Hast, 1962; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Gat & Keith, 1978; Mayo et al., 1997; Zinszer et al., 2019), and L2 listening is typically more cognitively demanding to process than L1 (e.g., Bsharat-Maalouf et al., 2023). This has not previously been investigated in a within-subjects design that avoids the confounds of comparing monolingual and bilingual groups. Therefore, Chapter 2 investigated whether the extent of SRM is affected by whether the listener attends to L1 or L2 in a dual-language context, which was examined in a selective listening experiment. Furthermore, Chapter 2 investigated whether the trade-off between SRM and increased cognitive costs caused by spatial separation operates differently for L1 and L2 in a dual-language context, which was examined in a divided listening experiment.

The findings from Chapter 2 show that bilinguals more effectively transcribed their L1 while ignoring their L2 than the reverse, consistent with previous literature (e.g., Spivey & Marian, 1999). Additionally, spatial separation improved listening accuracy even when it introduced cognitive costs, which aligns with previous research (e.g., Knight et al., 2023). Although L2 was expected to benefit more from SRM than L1, because L2 listening is more susceptible to EM when masking is associated with signal-to-noise ratio (SNR) or intensity changes (e.g., Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006), the magnitude of SRM did not differ across languages. This pattern is consistent with Ezzatian et al. (2010).

These results suggest that the often-reported L1 advantage under EM (e.g., Cooke et al., 2008) may reflect differences in the use of specific acoustic cues. L1 listeners exploit spectral-temporal and SNR-based cues more efficiently than L2 listeners (e.g., Garcia Lecumberri & Cooke, 2006), whereas binaural cues such as interaural level differences (ILDs) appear to support L1 and L2 processing to a similar extent. Thus, it may not be *masking* that disproportionately affects L2 listening, but rather the particular cues available to resolve it.

If L1 and L2 listeners can benefit from binaural cues (such as ILDs) to the same extent, then SRM should not vary across target languages. Consequently, the target language in Chapter 2 did not modulate the effect of spatial separation, nor was either language impacted disproportionately when attending to both talkers simultaneously. A comparison between the selective and divided listening experiments demonstrated that the magnitude of SRM was smaller when listening to two talkers compared to one, in line with Knight et al. (2023). Research on ear-switching (e.g., Axelrod et al., 1968; Huggins, 1964; Samuel, 1991; Treisman, 1971) suggests that the reduced SRM when attending to two talkers likely reflects the cognitive costs of shifting attention between ears.

While Chapter 2 focused on EM, Chapter 3 examined a form of informational masking (IM), which reflects higher level disruption to listening (e.g., Pollack et al., 1975). Specifically, Chapter 3 investigated how semantic overlap between target and masker speech influences listening to the target. This question has not previously been tested using naturalistic sentences containing multiple keywords, even in monolinguals. Addressing this question provides insight into whether background speech undergoes semantic processing when unattended. Competing theories predict different outcomes: a Semantic Filter mechanism posits that the listener blocks high-level processing of unattended speech (e.g., Broadbent, 1958; Brodbeck et al., 2018; Marsh & Campbell, 2016), whereas a Target-Masker Co-Activation account predicts temporary activation of both target and masker representations (e.g., Kiefer, 2002; Treisman, 1964). Prior research comparing semantically meaningful and meaningless maskers has resulted in inconsistent conclusions (e.g., Brouwer et al., 2012; Calandruccio et al., 2018). This issue is particularly relevant in a dual-language context, as L2 words have weaker connections to semantic representations than L1 words (Kroll & Stewart, 1994), potentially attenuating semantic interference. Chapter 3 of this thesis therefore investigated whether the semantic content of L1 and L2 maskers impacted listening to an L2 target.

Results from Chapter 3 showed that bilingual listeners understood the L1 target more accurately when the masker was L2 compared to L1, consistent with Chapter 2 and previous research (e.g., Brouwer et al., 2012). Contrary to predictions, however, semantic relatedness between target and masker did not affect transcription accuracy, suggesting that a semantic filter prevents high-level processing of unattended speech. This pattern was consistent regardless of masker language. Nevertheless, exploratory analyses revealed more intrusions from semantically competing than neutral maskers. Tun et al. (2002) suggested that listeners (specifically young adults) may process background speech semantically without allowing it to impair listening to the target.

Therefore, a follow-up experiment aimed to replicate the main findings: semantic overlap between target and masker does not impact transcription accuracy but does affect intrusion rate. In addition, a surprise recognition task was included to assess whether semantically competing keywords were better remembered than neutral keywords. If so, this would indicate that the masker was processed and the cognitive process of resolving semantic conflict between target words and competing masker words would lead to deeper encoding. This would corroborate Tun et al.’s (2002) finding that the masker is semantically processed, yet listeners prevent this interference from affecting target perception.

Unlike Tun et al. (2002), however, listeners did not recognise masker words more accurately when they were semantically competing. This pattern suggests that listeners process only one stream at a time (consistent with the Semantic Filter account), but occasionally mis-stream the sentences. When mis-streaming occurs, semantically competing masker words intrude more often because they form coherent sentences in combination with the rest of the target sentence. Thus, although the masker semantic content (regardless of masker language) does not impact accuracy at listening to an L1 target, at least in a single-language context with few phonological cues to stream segregation, the semantic content of the masker can influence the type of errors that listeners produce.

Taken together, Chapters 2 and 3 show that bilingual listeners rely on a range of cues to parse competing speech streams and support selective listening. They effectively use low-level binaural cues (e.g., ILDs) and phonological cues (e.g., phonological differences between languages), although the ability to exploit these cues diminishes as cognitive demands increase. Bilingual listeners also appear to prevent the semantic activation or processing of unattended speech in adverse listening conditions, which may facilitate selective listening. Additionally, in tasks with minimal cues for stream segregation, listeners appear to rely on the semantic content of the target when deciding which stream a given word belongs to in cases of mis-streaming target and masker sentences. This strategy, however, does not aid target perception and instead increases the likelihood of masker intrusions when the two streams are semantically similar.

4.2. Links to theoretical models

4.2.1. Brouwer et al.’s (2012) Target-Masker Similarity Hypothesis

The Target-Masker Similarity Hypothesis proposes that listeners exploit differences, particularly phonological differences, between speech streams to facilitate stream segregation and improve target perception (Brouwer et al., 2012). Previous research supports this, demonstrating that listening to a target is easier when the masker is in a different language to the target (e.g., Calandruccio et al., 2013; Mepham et al., 2022, 2025; Van Engen, 2010). These findings indicate that cross-language phonological differences help listeners to parse competing streams.

A similar pattern emerged in Experiment 1 of Chapter 3: listeners understood the L1-Spanish target more accurately when the masker was L2-English compared to L1-Spanish. However, this effect is conflated with the Language Familiarity Effect, illustrated by Mepham et al. (2022). Their study showed that Mandarin-English bilinguals listening to English were more disrupted by a Mandarin masker than English monolinguals were under the same conditions, presumably because the bilinguals were more familiar with the Mandarin masker. More generally, a masker that is familiar to the listener (e.g., L1) tends to be more disruptive than a less familiar masker (e.g., L2). Additionally, Lew et al. (2024) found that listening to an L2 target was always easier when the masker was L2 compared to L1 – disputing the Target-Masker Similarity Hypothesis in favour of the Language Familiarity Effect. Because the target language in Experiment 1 of Chapter 3 was always L1, the single-language context necessarily involved an L1 masker, making it inherently more familiar than the L2 masker used in the dual-language context. Previous research is inconsistent in terms of which effect is stronger: Mepham et al. (2022) suggest that the differences predicted by the Target-Masker Similarity Hypothesis are larger than those driven by familiarity; however, Lew et al. finds the opposite pattern of results. The experiments in this thesis cannot disentangle these influences without an additional experiment with an L2 target.

Brouwer et al. (2012) also discuss linguistic similarity, proposing that semantically meaningful maskers should disrupt listening to meaningful targets more than semantically anomalous maskers, because of the similarity in the linguistic content of the two streams (i.e., two meaningful sentences are more linguistically similar than when one is meaningful and one is anomalous). In Chapter 3, however, semantic similarity between target and masker did not affect listening accuracy, which might appear inconsistent with this aspect of the Target-Masker Similarity Hypothesis. Yet it is important to note a methodological distinction: Brouwer et al. manipulated the semantic *meaningfulness* of the masker itself, whereas Chapter 3 manipulated the semantic *relationship* between target and masker. This reflects a difference between linguistic content and semantic content. Brouwer et al.'s anomalous maskers were syntactically well-formed, but lacked coherent sentence-level meaning, and therefore may have contained less linguistic information than their meaningful maskers. Contrastingly, both the semantically competing and neutral maskers in Chapter 3 of this thesis were meaningful, and thus, contained the same amount of linguistic information as each other, only differing in the amount of semantic overlap with the target.

Consequently, the findings from Chapter 3 do not necessarily contradict the Target-Masker Similarity Hypothesis with respect to linguistic similarity; rather, they indicate that semantic similarity between streams does not impair listening accuracy in the same manner as linguistic meaningfulness of the masker. This aligns with the idea of a broad-but-shallow level of linguistic analysis: once the masker is detected as meaningful, deeper semantic processing does not occur, and the masker is recognised as meaningful without its specific semantic content being interpreted. This idea is supported by Brown et al. (2023) who found that semantic violations in task-irrelevant speech elicited neural response, but the listeners did not report noticing these semantic violations of the masker, suggesting that while the semantic meaningfulness of the speech was processed generally, it was not processed deeply.

Overall, the evidence suggests a nuanced pattern. Listeners appear unable to filter out masker familiarity (e.g., Lew et al., 2024; Mepham et al., 2022), and Brouwer et al.'s findings indicate that they also cannot filter out linguistic similarity in the form of meaningfulness. This conclusion is supported by neurological research demonstrating that selective attention does not fully eliminate linguistic processing of task-irrelevant speech (Har-shai Yahav & Zion Golumbic, 2021). However, the experiments in Chapter 3 indicate that listeners can filter out the semantic content of the masker.

The Language Familiarity Effect could reflect low-level processes: listeners are more attuned to the phonological attributes of their L1 than their L2 (e.g., rhythm of the speech), consistent with evidence from speaker discrimination tasks showing that such effects arise from phonological familiarity rather than from higher-level attributes of the target speech (e.g., Fleming et al., 2014; Johnson et al., 2017). Alternatively, it could also reflect ease of lexical access: higher-frequency words (i.e., L1) are activated more quickly than lower-frequency words (i.e., L2), as indicated by models of bilingual lexical access (e.g., Dijkstra & van Heuven, 2002; Shook & Marian, 2012), making familiar language maskers more disruptive.

4.2.2. Theories of Attention

Posner et al.'s (1980) Attentional Spotlight Theory, originally developed for vision, proposes that attention can be directed to only one spatial location at a time. Although direct evidence in the auditory domain is limited, applying this theory to listening suggests that attending to two dichotic speech streams would require rapid shifts of attention between the ears. This ear-switching process appears to disrupt listening efficiency (e.g., Axelrod et al., 1968; Huggins, 1964; Samuel, 1991; Treisman, 1971). Experiment 2 of Chapter 2 (the divided listening task) provides support for the relevance of the attentional spotlight in the auditory domain: many listeners' qualitative responses indicated that they purposefully alternated their attention between ears to capture as much of both sentences as possible. When this became tiring, several participants reported choosing one sentence stream to focus on.

The comparison between the two experiments in Chapter 2 further suggests that SRM was reduced in Experiment 2 relative to Experiment 1. This reduction was interpreted as evidence that spatial separation imposed additional cognitive demands under dichotic listening, likely due to the need for rapid attentional shifts between ears. This increased demand affected only the dichotic condition, as in the collocated condition, both speech streams were perceived as originating from the same spatial location and therefore fell within an attentional spotlight, eliminating the need for spatial shifting. Similar results were found by Knight et al. (2023): When the effects of EM were controlled for (i.e., no EM even in the collocated condition), listening performance was poorer in the dichotic condition, likely due to the cognitively demanding task of shifting attention between ears.

The experiments in Chapter 3 shifted focus from spatial attention to the processing of unattended speech, linking more directly to Broadbent’s (1958) Early-Filter Theory and Treisman’s (1964) Attenuation Theory. The results aligned more closely with an early semantic filter mechanism akin to Broadbent’s proposal: semantic overlap between target and masker did not impair transcription accuracy, indicating that semantic content of the masker was not activated to a level that interfered with target processing. This pattern suggests that listeners effectively prevent semantic activation of irrelevant speech, allowing them to allocate cognitive resources to the target stream in adverse listening conditions.

A further attentional theory relevant to this thesis is Lavie’s (1995) Perceptual Load Theory. This argues that when a task is highly demanding, it exhausts available perceptual resources, leaving little capacity to process distractors, which can explain phenomena such as inattentional deafness (e.g., Cherry, 1958; Dalton & Fraenkel, 2012). When a task is less demanding, spare perceptual resources may be involuntarily allocated to task-irrelevant stimuli, potentially decreasing selective attention efficiency. In this respect, Perceptual Load Theory represents an intermediary model of attention relative to Broadbent's (1958) Early Filter Theory and Treisman's (1964) Attenuation Theory: distractors are neither fully blocked nor always partially processed; instead, their processing depends on task demands.

Like Posner et al.'s (1980) Attentional Spotlight Theory, Perceptual Load Theory (or Load Theory) was originally developed with reference to the visual domain. However, the results of Chapter 3 align closely with Load Theory. In these experiments, the simultaneous voices were highly similar, only differing slightly in terms of fundamental frequency (F0), creating a demanding perceptual task. Zekveld et al. (2014) demonstrated with pupillometry that voice similarity between target and masker speech increases perceptual load because listeners rely on acoustic cues to segregate speech streams. Under such high-load conditions, listeners may allocate nearly all available perceptual resources to the target, effectively preventing activation or processing of the masker speech. Conversely, in a task with lower perceptual load (e.g., more distinct voices), spare resources could be available for processing distractors, increasing the likelihood of semantic interference. These findings suggest that a semantic filter operates robustly under high-load conditions to facilitate selective attention, whereas in lower-load situations, the filter may be attenuated, allowing some semantic processing of unattended speech. However, as the talkers become more phonologically distinct, masker semantic interference might not be observed due to it being easier to parse the masker and talker voices, even if the masker is processed.

Previous research leaves the applicability of Load Theory to the auditory domain uncertain. While studies such as Cherry (1958) and Dalton and Fraenkel (2012) demonstrate inattentional deafness, these manipulations involved spatial separation rather than load per se. Francis (2010) found that increasing perceptual load when responding to a nonspeech cue led to less interference from an irrelevant talker, supporting Load Theory's application to audition, although not directly to speech-on-speech listening. Additionally, Dekerle et al. (2014) showed reduced high-level processing of irrelevant speech under higher load conditions. However, other findings challenge Load Theory in auditory contexts: Lynch (2021) found no evidence for load effects in diotic or dichotic listening, and Murphy et al. (2013) observed no inattentional deafness regardless of task demand. Murphy et al. (2017) highlight the inconsistency of empirical findings, with some studies reporting inattentional deafness under high load and others finding none. Thus, although Perceptual Load Theory is a useful framework for interpreting some aspects of the Chapter 3 findings, its general relevance to speech-on-speech tasks remains inconclusive.

4.2.3. Bilingual language processing models

Models of bilingual language control have been central to interpreting the findings of this thesis. In particular, the Bilingual Interactive Activation Plus (BIA+, Dijkstra & van Heuven, 2002) and the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS, Shook & Marian, 2012) models propose that L1 words have higher resting activation levels than L2 words, making L1 processing faster and more efficient. This difference is partly attributed to the general lower frequency of L2 word use compared to L1. This principle of bilingual language processing models is supported by previous research demonstrating asymmetric translation priming: L1 has a larger priming effect on L2 than vice versa, meaning that L1 words have higher resting activation levels, making them more effective primes (Wen & van Heuven, 2017). The results from Chapters 2 and 3 are also broadly consistent with these models. For example, Chapter 2 showed higher transcription accuracy for L1 sentences than L2 sentences, consistent with the prediction that L1's higher resting activation facilitates processing. However, because the sentence pairs in Chapter 2 were always presented as L1-L2 or L2-L1, it remains unclear whether the L1 advantage reflects easier processing when it is the target, or whether the L1 also acts as a more disruptive masker due to its higher activation. Despite this uncertainty, however, the overall pattern of results aligns with principles of the BIA+ and BLINCS models.

The experiments in this thesis allowed Spanish-English bilinguals who lived in the UK at the time of experimentation to participate. While participants generally reported more frequent current use of Spanish than English, some reported the opposite, potentially altering the relative resting activation levels of L1 and L2 lexicons. Both the BIA+ and BLINCS suggest that resting activation levels are influenced by the frequency that the individual uses each word: more frequently used words, regardless of language, are activated more rapidly, a principle that is supported by empirical research demonstrating that more frequently used words activate the other language translation equivalents in lexical decision tasks more strongly (e.g., Chaouch-Orozco et al., 2020). Therefore, some participants may have exhibited a reduced L1 processing advantage, or in some cases, faster activation of L2-English words compared to L1-Spanish words. Indeed, language use scores collected through the language background questionnaires in Chapter 2 showed that participants who reported more frequent L2-English use achieved higher scores on the L2-English trials, consistent with the BIA+ and BLINCS predictions. Importantly, however, frequency of L2-English use did not modulate the magnitude of SRM.

Chapter 3 examined semantic processing of unattended speech, drawing on the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994), which posits stronger connections between L1 words and their semantic representations than between L2 words and their semantic representations. Previous research supports the idea that semantic information is processed more slowly in the L2 than L1 (Dijkgraaf et al., 2019), and that semantic priming in the L2 takes longer to develop (Wen & van Heuven, 2016). Therefore, it was hypothesised that semantic interference from a masker would be greater if the masker was L1 than L2. This hypothesis was not supported, although neither was it contradicted, as no semantic interference was observed for either L1 or L2 maskers. As discussed in Chapter 3, these findings suggest that listeners do not semantically process unattended masker speech under high cognitive load, regardless of masker language. Therefore, no definitive conclusions can be drawn regarding differential semantic processing of L1 versus L2, and the results neither support nor contradict the RHM.

4.3. Methodological strengths

4.3.1. Within-subjects design

Much research comparing L1 and L2 processing relies on between-group designs in which monolinguals represent the L1 condition and bilinguals represent the L2 condition. Although this approach carries practical benefits, such as requiring only one target language and avoiding confounds caused by differences between languages, it introduces an interpretive limitation. Specifically, a between-subjects design may conflate differences between monolinguals and bilinguals with differences between L1 and L2 processing. This distinction is important given that bilingual language processing models (e.g., BIA+; Dijkstra & van Heuven, 2002; BLINCS; Shook & Marian, 2012) posit that both languages remain active during comprehension. As a result, effects attributed to processing in L2 versus L1 may instead arise from differences in cross-language activation between bilinguals and monolinguals.

Empirical evidence supports this concern. Blumenfeld and Marian (2011) reported that monolinguals and bilinguals showed comparable performance in an initial word recognition task yet differed in a subsequent priming probe task. Monolinguals exhibited slower responses when the probe appeared at the location of a phonological competitor image than in the location of the target, whereas bilinguals showed no such effect. This suggests that bilinguals inhibit previously activated competitor words less strongly than monolinguals, even when both groups operate in the same language. Such differences likely stem from bilinguals' continual management of two linguistic systems (Bialystok et al., 2012; Green, 1998), which shapes the underlying processing mechanisms.

These findings demonstrate a limitation of studies that compare monolinguals in an L1 condition with bilinguals in an L2 condition: any observed difference may reflect group-level processing differences rather than genuine differences between L1 and L2 comprehension. The experiments in this thesis instead used a within-subjects design, in which each participant completed both the L1 and L2 conditions. This approach removes confounds associated with comparing monolinguals with bilinguals. This design also increases statistical power by reducing between-subject variability and allows for direct comparisons of L1 and L2 processing within the same listener. Consequently, it provides clearer evidence about how language experience shapes processing, rather than conflating these effects with differences in participant groups.

4.3.2. Dataset of semantically matched sentence stimuli

In speech-on-speech research, several established corpora exist; however, no existing corpus met the specific requirements of Chapter 3: namely, a set of sentence stimuli in which targets and maskers were systematically constructed to be either semantically related or semantically unrelated. As a result, a new sentence dataset was developed to address the research question posed in Chapter 3.

The resulting sentence dataset comprises 100 triplets, each containing a target sentence, a semantically related masker, and a semantically unrelated masker. Each masker sentence was designed to serve as a semantically competing masker for one target and a neutral masker for another, resulting in a total of 100 targets and 100 maskers. All sentences were produced in both Spanish and English, with grammar and meaning verified by native speakers of each relevant language, and all keywords, except for adverbs, were unique across the entire sentence dataset. The two language versions were matched for syntactic structure (each triplet uses either a *noun–adverb–verb–noun* or a *noun–verb-adverb-noun* pattern). This allowed the sentence dataset to be used to compare L1 and L2 directly because the stimuli in each language are directly comparable.

Quantitative validation further supports the sentence dataset. Latent semantic analysis (Dumais et al., 1988) confirmed that, at the sentence level, targets are more semantically related to their semantically competing than their neutral maskers. Participant ratings provided further evidence of this at the keyword level. These checks strengthen the construct validity of the semantic relatedness manipulation in Chapter 3, ensuring that any observed effects can be attributed to the semantic relationship between target and masker.

This careful construction of a validated new sentence dataset allowed for a novel set of experiments: the semantic relationship between target and masker is manipulated rather than the meaningfulness of the masker (e.g., Brouwer et al., 2012; Calandruccio et al., 2018). Additionally, in contrast to Villard et al. (2024), semantic relatedness varies across multiple keywords embedded within naturalistic sentences, maximising sensitivity to a potential semantic interference effect if one was present.

4.4. Methodological and talker/listener considerations

4.4.1 Simulated spatial separation

A methodological consideration concerns the use of the term “spatial separation” in Chapter 2. Strictly speaking, this manipulation only constituted a *simulation* of spatial separation, as participants listened via headphones and only interaural level differences (ILDs) were varied, with the two talkers in the separated conditions always being presented dichotically (i.e., one in each stereo channel); interaural timing differences (ITDs) were not incorporated. Nonetheless, ILDs are a dominant spatial cue at high frequencies and contribute to localisation of low-frequency sounds (Middlebrooks & Green, 1991; Middlebrooks & Onsan, 2012). Even so, including ITDs and monaural cues, which can be simulated using head-related transfer functions (HRTFs), would have strengthened the ecological validity of the spatial manipulation.

A fully spatial listening environment would require a free-field setup with multiple loudspeakers positioned around the listener. This was not feasible for the studies reported in Chapter 2 due to the highly specific participant population required. Recruiting approximately 100 in-person participants who shared the same L1 and L2 and were proficient in their L2 to the correct level would have been impractical. Conducting the experiments online therefore necessitated a headphone-based approach. In this context, the simplified dichotic method of spatial separation provided a practical means of manipulating spatial cues while allowing for adequate sample sizes for statistical power. Additionally, focusing on a single dominant spatial cue offered an additional methodological benefit. By maximising ILDs and keeping ITDs constant, the experiments isolated one component of spatial listening, enabling a clearer assessment of its contribution to performance without conflating effects of multiple spatial cues.

Comparable dichotic approaches have been used effectively in previous work. Knight et al. (2023) simulated a four-step change in spatial location using left-right mixtures with graded intensity differences, and Calcus et al. (2015), employed ILD-based dichotic configurations to examine SRM relative to a diotic baseline. In light of this, the present dichotic manipulation provides a reasonable basis for interpreting the findings as reflecting an extreme version of SRM rather than simply a dichotic listening advantage.

4.4.2. Bilingual considerations

4.4.2.1. Proficiency levels

Individual differences among bilingual participants likely influenced performance in the experiments reported in this thesis. All bilingual participants had high L2-English proficiency, which was a practical necessity because recruitment was conducted via Prolific, an English-language platform. Additionally, participants required an appropriate level of L2-English proficiency for the listener to sufficiently attend to it in Chapter 2, and to investigate the potential influence of semantic content of the masker in Chapter 3. In Chapters 2 and 3, all participants (except those in Experiment 2 of Chapter 3) achieved at least 60% on the LexTALE, and many lived in the UK, resulting in some participants using L2-English more frequently than L1-Spanish. Nevertheless, on average, participants reported greater use of L1-Spanish than L2-English.

In Chapter 2, it was predicted that L2 listening would benefit more from SRM than L1, as research indicates that EM disproportionately disrupts L2 more than L1 processing (e.g., Cooke et al., 2008). On the one hand, if participants had been more unbalanced bilinguals, with much higher L1 than L2 proficiency, this might have maximised the chance that the expected interaction was found, as indicated by previous research demonstrating that late bilinguals show a larger disadvantage at listening to their L2 in noise compared to early bilinguals (Regalado et al., 2019). On the other hand, Experiment 1 of Chapter 2 revealed that SRM for L2 was actually larger among participants with higher L2 proficiency, suggesting that lower-proficiency listeners are less likely to exhibit a stronger SRM effect (see also Hui et al., 2021). However, this was not replicated in Experiment 2 when attending to both talkers simultaneously. Including participants with a very low L2 proficiency could also have confounded investigation of L2 masker effects, as the masker may have functioned as an unfamiliar language.

4.4.2.2. Language combinations

In both Chapters 2 and 3, all participants understood both Spanish and English; however, different language combinations could result in different outcomes. Calandruccio et al. (2013) showed that maskers are more disruptive when their language is phonologically similar to the target, consistent with Brouwer et al.’s (2012) Target-Masker Similarity Hypothesis: for an English target, English maskers were most disruptive, followed by

Spanish, then Mandarin maskers. Therefore, bilinguals whose languages are more phonologically distinct (e.g., English and Mandarin) would likely demonstrate higher overall transcription accuracy in both Chapter 2 and the dual-language conditions of Chapter 3. Conversely, with greater phonological differences between languages, listeners might rely less on spatial cues for stream segregation, instead leveraging phonological differences to distinguish streams. This idea is supported by research indicating that a more phonologically dissimilar masker facilitates stream segregation more than spatial separation (Zekveld et al., 2014), although in this case the voices differed based on gender rather than language. Thus, the effect of spatial separation in Chapter 2 may have been reduced if the language combination was English-Mandarin compared to the English-Spanish combination, as phonological cues alone could support effective parsing in both spatial conditions.

Additionally, in Chapter 2, any change in target language was accompanied by a change in masker language, making it impossible to determine whether the observed language effect arose because the L1 target was easier to attend to and process than the L2 target, or because the L1 masker produced greater interference than the L2 masker. Both explanations are consistent with the BIA+ (Dijkstra & van Heuven, 2002) and BLINCS (Shook & Marian, 2012) models, which propose that L1 has higher resting activation levels than L2, making L1 both easier to activate as a target and more disruptive as a masker. It is likely that both mechanisms contributed to the observed effects; however, the current design does not allow them to be disentangled.

Future research could address this limitation by incorporating single-language contexts (i.e., L1-L2, L2-L1, L1-L1, and L2-L2 target-masker combinations). These conditions were omitted in Chapter 2 to avoid unnecessary complexity and because the primary focus was on bilingual processing within dual-language contexts. Including single-language context conditions would allow clearer interpretation of the underlying mechanisms. If the language effect is driven primarily by the L1's greater ease of activation, then accuracy should be higher across conditions in which the target is L1 (L1-L2 and L1-L1) than when the target is L2 (L2-L1 and L2-L2). Conversely, if the effect is driven primarily by the L1 being a more disruptive masker, then performance should be higher in conditions where the masker is L2 (L1-L2 and L2-L2) than when the masker is L1 (L1-L1 and L2-L1).

A similar limitation arises in Chapter 3. In Experiment 1, the masker language effect may reflect listeners' use of phonological differences between speech streams (as predicted by the Target-Masker Similarity Hypothesis; Brouwer et al., 2012), or differences in the resting activation levels of the two languages (as predicted by BIA+; Dijkstra & van Heuven, 2002 and BLINCS; Shook & Marian, 2012). Because the target was always L1, which was a deliberate choice to reduce design complexity, these two mechanisms cannot be dissociated.

It is likely that both mechanisms are involved in the better target transcription when the masker is L2 compared to L1. These mechanisms could be distinguished by repeating the paradigm with an L2 target. If the L1 masker continues to be more disruptive than the L2 masker, this would suggest that differing resting activation levels drive the masker language effect. If the L2 masker becomes more disruptive than the L1 masker (when the target is L2), this would indicate that phonological similarity is the predominant driver of the masker language effect. Prior research suggests that this latter pattern is more likely (e.g., Mepham et al., 2022; Van Engen, 2010), implying that while both mechanisms contribute, phonological differences between speech streams are more useful cues for facilitating target perception than differences in resting activation levels.

4.4.3. Talker voice

In Chapter 2, sentence pairs were always one male and one female talker, so that listeners could use gender (voice), rather than language, to identify the target stream, reducing the likelihood of priming a particular language. Using mixed-gender sentence pairs may have reduced SRM in both experiments in Chapter 2, given that SRM is smaller when the target and masker differ in gender (Oh et al., 2021, 2022). Additionally, ceiling effects observed in the attend-L1 conditions in Experiment 1 may have limited the effect of SRM. These ceiling effects may reflect the large phonological differences between the two talkers making them easy to segregate despite being LTAS matched, likely partly due to the sentences being spoken by mis-matched gender voices (e.g., Zekveld et al., 2014). Thus, a larger effect of SRM may have been observed in the Chapter 2 experiments if same-gender sentence pairs were used.

Chapter 3 avoided this issue by using the same talker to record both target and masker sentences. This approach was also necessary because no existing sentence corpus varies semantic relatedness in line with our aims. Instead of relying on gender differences to cue the target, the target and masker were differentiated by F0 manipulation, and participants were familiarised with the target voice at the start of each block. While this approach reduced ceiling effects and minimised the influence of voice characteristics, it introduced a new consideration: the L2-English voice was Spanish accented, an issue not relevant in Chapter 2 because the pre-recorded IEEE sentences (Rothauser, 1969) and Sharvard sentences (Aubanel et al., 2014) were spoken by native speakers. If the accent was too strong, comprehension of the English sentences could have been impaired, potentially limiting variability in IM exerted by the semantic manipulation.

A pilot listening task showed that participants understood the L2-English sentences with 68% accuracy, which was lower than their accuracy for the L1-Spanish sentences (94%). However, reduced L2 accuracy is expected due to proficiency differences and does not necessarily indicate an accent-based disadvantage. In fact, prior research suggests that Spanish listeners often comprehend L2-English speech more effectively when it is spoken with a Spanish accent rather than an English accent (e.g., Major et al., 2002). Given that most participants lived in Spain and grew up there, they were likely familiar with Spanish-accented English, reducing the chance that the accent impaired understanding beyond normal L2 difficulty.

Even if the talker's accent influenced listeners' understanding of the L2-English masker, it is unlikely that this hid a semantic interference effect. Notably, no semantic interference emerged even when the masker was L1 - a condition in which semantic disruption would be expected given the stronger links between words and semantic representations (as suggested by the RHM; Kroll & Stewart, 1994). Thus, the absence of semantic interference seems unlikely to be attributable to accent-related properties of the talker's voice.

4.4.4. Listeners with cochlear implants

Although all participants in this thesis self-reported normal hearing, it is important to consider how the same paradigms might operate in populations with hearing loss. Such considerations provide context for interpreting the generalisability of the findings and for extending the conclusions to applied or clinical settings. Within Chapter 3, it is unlikely that a semantic effect would be found in listeners with hearing loss when no effect was found in listeners with normal hearing. However, Holmes et al. (2018) demonstrate that when listening in noise, target perception is more accurate when it is semantically related to the previous target sentence. This is akin to the idea of listeners using semantic coherence judgements to improve listening when there are limited acoustic cues, although unlike in Chapter 3, these cues proved useful to target identification for Holmes et al.'s participants with hearing impairment.

Cochlear implant (CI) users may experience the tasks in Chapter 2 differently from normal-hearing listeners. While CIs improve SRM and restore localisation ability in patients with unilateral deafness (Grossmann et al., 2016), SRM in CI users is still generally weaker than in normal-hearing populations (e.g., Killan et al., 2015; Misurelli & Litovsky, 2012, 2015), and some studies report no measurable SRM in CI users at all (Goupell et al., 2016). Similarly, bilateral hearing aid users show reduced SRM relative to listeners with normal hearing (Oh et al., 2022), and children using hearing aids often show no SRM (Ching et al., 2011). This could be explained by the fact that hearing-impaired children have a reduced ability to prepare spatial attention for an upcoming talker, even when using acoustic hearing aids (Holmes et al., 2017). Consequently, in Experiment 1 of Chapter 2, CI users might exhibit a substantially smaller SRM effect.

A review by Philips et al. (2023) indicates that CI users generally expend greater listening effort than listeners with normal hearing, suggesting that increasing the cognitive demands of a listening task may disproportionately affect CI users. Children with bilateral hearing aids also show weaker dichotic divided listening than their peers with normal hearing (Ling, 1971), and Shinn-Cunningham and Best (2008) argue that hearing-impaired listeners have a reduced capacity to switch attention rapidly – an ability central to successful divided dichotic listening. If CI users have difficulty shifting attention between ears, they may show a different pattern in Experiment 2 of Chapter 2: performance in the dichotic condition may be poorer than in the collocated condition because the cognitive demands of rapid attention-switching could outweigh the relatively small (or absent) benefits of SRM.

This interpretation remains speculative, but the relationship between SRM and cognitive demand likely differs substantially between CI users and listeners with normal hearing. Future research should therefore examine how these factors trade off in CI populations to determine how they experience spatial listening differently to individuals with normal hearing.

4.5. Relation to real-life comprehension

4.5.1. Memory demands in transcription tasks

Across all listening tasks in this thesis, participants provided their responses only after stimulus offset. As a result, listeners had to retain information in memory for at least a short period, and in Experiment 2 of Chapter 2, they were required to store two sentences simultaneously. This introduces an important consideration: although memory mechanisms do contribute to real-time speech perception, the structure of transcription-based tasks may result in listeners utilising memory-based mechanisms more than they would during real in-the-moment listening.

Phonological short-term memory and speech perception are closely linked (Jacquemot & Scott, 2006); even in everyday communication, phonological short-term memory supports segmentation of continuous speech and the temporary maintenance of phonological representations. Jacquemot et al. (2006) demonstrated that a phonological buffer maintains these representations and re-examines them to resolve phonological ambiguities. However, this is disputed by Lad et al. (2020), who argue that it is not *phonological* working memory that aids speech in noise ability in everyday communication, but rather *auditory* (not specifically speech-based ability) working memory that facilitates tracking speech in noisy environments. Taken together, while there is disagreement about which specific memory systems are involved, some form of memory remains integral to successful speech perception.

Additionally, the Ease of Language Understanding (ELU) model (Rönnberg et al., 2008) posits that in easy listening conditions, speech is matched quickly to stored phonological representations in long-term memory. However, when this process fails, for example due to competing speech masking the target, the listener uses effortful processes, such as working memory, linguistic knowledge, and sentence context to “repair” the target speech. Thus, when there are mismatches between what is perceived and the lexicon (i.e., the listener does not hear an existing word), the listener must retrospectively resolve this. Neurological research supports this claim that when target speech is degraded, effortful, higher-order mechanisms are engaged to modulate auditory processing (e.g., Peelle, 2018; Wild et al., 2012). Additionally, behavioural evidence demonstrates the top-down restoration of degraded speech, and that working memory capacity is a predictor of perceptual restoration ability, as well as lexical knowledge (Burleson & Souza, 2022). Furthermore, cueing the spatial location of a target *after* stimuli offset still benefits understanding of the target (Garnier-Allain et al., 2023), demonstrating the retrospective use of higher-order mechanisms to resolve target uncertainty.

In the experiments within this thesis, the delay between stimulus offset and transcription may allow more time for such re-examination, potentially enabling listeners to recover from the effects of EM. This suggests that in Chapter 2, the collocated condition may have benefited disproportionately from this additional processing time relative to what would occur in real-time listening. As a result, the SRM observed in experimental settings may underestimate the magnitude of SRM in everyday listening. Additionally, according to the ELU model, the delay between stimulus offset and transcription may allow the listener to use high-order mechanisms such as semantic context and lexical information to decide whether or not a masker intrusion is part of the target – supporting the mis-streaming and mis-allocation explanation of Chapter 3's intrusion rate findings.

Treisman's (1964) Attenuation Theory also has implications for transcription-based tasks similar to the ELU model. The theory proposes that both streams receive initial semantic processing, but once the irrelevant stream is identified, its semantic content is subsequently attenuated, meaning that there may only be a short period during which semantic content from *both* streams is stored. Indeed, Hohlfeld and Somer (2005) indicated that the amount of semantic understanding of task-irrelevant speech may decay in short-term memory. If listeners discard the semantics of the irrelevant stream shortly after identification – or if this information decays, then the brief interval between stimulus offset and transcription may reduce the likelihood of semantic interference during the response phase. In natural listening, however, the listener cannot discard irrelevant content before needing to understand and respond to the target talker. Again, in a transcription task, an appropriate response to the target speech is not needed, a substantial departure from everyday listening. This requirement to respond immediately in everyday communication, rather than experiencing an interval before transcription (providing more time for the masker semantic content to be discarded), raises the possibility that semantic interference may be found more often in real-world contexts than in transcription experiments.

A method that provides a more immediate measure of real-life listening is the visual world paradigm, which tracks eye movements while listeners view images corresponding to an auditory target, masker or control item (see Salverda & Tanenhaus, 2017). This results in an “in-the-moment” measure of listening and has been used in speech perception research previously (e.g., Zhou et al., 2023), including during speech-on-speech conditions (Ben-David et al., 2011), although such usage remains uncommon. Abdel-Latif et al. (2025) recently applied the visual world paradigm to a speech-on-speech task and demonstrated its potential as an objective, real-time measure. However, further validation is required, as their study involved only 12 participants. Additionally, the paradigm requires eye tracking, making it impractical for online experiments and therefore of limited applicability to bilingual speech perception studies that rely on remote data collection in a within-subjects design.

4.5.2. Sentence comprehension versus word identification

The requirement to transcribe the target also raises the question of whether transcription accuracy measured in Chapters 2 and 3 reflects comprehension or merely auditory identification – especially since the participants never respond to the targets as they would in everyday conversation. It is possible that listeners perceived the target sentences at the level of individual words – or even phonemes – without fully processing meaning. This issue is especially relevant for Chapter 3: if participants did not comprehend the sentence meaning, then higher-level processes (e.g., semantic processing) may not have occurred, potentially explaining the absence of an effect of semantic relatedness.

It is unlikely that listeners processed the stimuli solely at the phonetic level. Although phonemically correct responses using orthographically incorrect spellings (e.g., *f* instead of *ph* in English; *y* instead of *ll* in Spanish) were counted as correct, such instances were rare. The generally accurate spelling of keywords indicates that listeners processed the stimuli at least at the word level. Evidence from previous research also suggests that listeners typically engage in sentence-level processing in transcription tasks. Pandey and Herrmann (2025) showed that listeners transcribe targets in babble more accurately when provided with semantic context. Similarly, target predictability enhances performance when sentences are degraded (Bhandari et al., 2021). Further evidence using the Speech in Noise corpus (varying final word predictability; Kalikow et al., 1977) and the revised version of this corpus (Bilger et al., 1984) demonstrates that identifying the final word in a sentence is more accurate when it is predictable based on the semantics of the full sentence (e.g., Hunter, 2021; Kalikow et al., 1977; Sheldon et al., 2008; Wilson et al., 2012), including when listening to L2 (Coulter et al., 2020). These findings indicate that comprehension beyond the word level supports successful transcription, implying that listeners are not just recognising isolated words but are integrating them into meaningful sentence representations, demonstrating comprehension.

Furthermore, in Chapter 3, the interpretation of intrusion errors relied on the assumption that listeners judged the semantic coherence of what they heard. It was concluded that listeners produced more semantically competing than neutral intrusions because of mis-streaming rather than semantic co-activation, and that this pattern depended on listeners evaluating the semantic coherence of the sentence they heard. This evaluation of semantic coherence goes beyond phonemic or word-level identification, further supporting the view that participants comprehended the sentences.

Nonetheless, to ensure that target stimuli are processed in a manner closer to real-life conversation, future research could incorporate intermittent comprehension checks throughout the listening task by asking listeners to respond to the target as if in conversation. Accurate responses would provide direct evidence that listeners are not only hearing the target speech but also understanding it. The current transcription-based methodology does not require participants to respond meaningfully to the content of the target, whereas natural conversational listening does, making this a useful addition for future studies.

4.6. Open Questions

Across this thesis, three main questions were addressed. Chapter 2 examined whether bilingual listeners benefit differently from spatial release from masking (SRM) when attending to their L1 or L2, and whether the combined effects of reducing EM and increasing cognitive demand through spatial separation operate differently across L1 and L2. In both selective and divided listening contexts, SRM benefited L1 and L2 to a similar extent, indicating that bilingual listeners use binaural cues similarly across languages and that the cognitive costs of spatial separation reduce the overall magnitude of SRM. Chapter 3 explored whether the semantic content of L1 and L2 maskers affects listening to an L1 target. Although masker semantic content did not influence overall transcription accuracy, they did impact the pattern of errors: in contexts when mis-streaming occurs, listeners use semantic coherence to decide whether a heard word was part of the target or masker, at least in single-language contexts. Together, these findings show that bilingual listeners primarily rely on low-level differences between talkers to parse speech streams, such as spatial and phonological cues. When these are limited, listeners may attempt to use high-level semantic cues for streaming, although the impact of this on target perception is limited.

Although these questions were addressed, several further questions arise from the current findings. Prior research shows that when EM is increased by raising masker intensity, L2 listening is more negatively impacted than L1 listening (e.g., Cooke et al., 2008). This contrasts with the findings from Chapter 2 and those of Ezzatian et al. (2010), that reducing EM via spatial separation does not differentially affect L1 and L2. This discrepancy raises an important question: why does L2 listening suffer more than L1 listening when EM is manipulated through intensity, but not when EM is manipulated through spatial separation? One possibility is that listeners can rely on interaural level differences (ILDs) to segregate streams regardless of language, whereas L2 listeners are less efficient at using non-spatial cues that rely on linguistic proficiency or lexical familiarity. Identifying which mechanisms L2 listeners cannot use as effectively as L1 listeners, and why, remains an important direction for future work.

Regarding Chapter 3, a key open question concerns *when* any semantic interference effect would be expected to arise. Chapter 3 took a holistic approach, assessing whether semantic interference from a masker occurred at all, rather than distinguishing between encoding, rehearsal, or retrieval stages. If semantic interference from the masker occurs primarily during encoding, the experiments within Chapter 3 may not have been optimal for detecting it. In the naturalistic sentences used, corresponding target and masker keywords were not aligned temporally. If the target keyword occurred earlier in time than its semantically related masker counterpart, the target might already have been encoded before semantic interference could occur. Conversely, if target and masker keywords were presented simultaneously, semantic interference at the encoding stage would be more likely to be observed. Relatedly, Hohlfeld and Sommer (2005) conclude that semantic understanding of task-irrelevant speech is modulated by decay in short-term memory, indicating that a lack of alignment between target and masker reduces the likelihood of semantic interference. Determining the stage at which semantic interference occurs – if it occurs at all – would clarify which stages of listening engage higher-level linguistic processes. This could be examined by designing stimuli in which semantically related or unrelated masker keywords are temporally aligned with the corresponding target words.

4.7. Conclusion

This thesis has addressed under-explored questions in bilingual speech processing under adverse listening conditions. It extends previous research by examining bilinguals' processing of their L1 and L2 in a within-subjects design, thereby avoiding the common confounds associated with monolingual and bilingual group comparisons. Chapter 3 also introduced a novel sentence dataset of semantically matched sentence stimuli, validated through latent semantic analysis and subjective participant ratings, providing a resource for future investigations of semantic interference in speech-on-speech tasks. More broadly, the findings contribute to discussions within theories of attention, support aspects of the Target-Masker Similarity Hypothesis, and align with principles of bilingual language processing models.

The key results demonstrate the robust beneficial impact of spatial release from masking (SRM), even when spatial separation introduces cognitive costs, and show that these spatial effects are not modulated by language. Furthermore, the semantic content of background speech does not appear to be co-activated with target speech and does not impair overall listening accuracy. However, in single-language contexts with limited phonological cues for stream segregation, masker semantic content can influence error patterns, with listeners utilising semantic coherence judgements to attempt to aid streaming.

Taken together, this thesis demonstrates that bilinguals rely on multiple cues to facilitate selective listening. Primarily, bilingual listeners depend on low-level phonological differences between languages and spatial cues such as interaural level differences (ILDs). When these cues are limited, listeners may engage higher-level semantic mechanisms (i.e., considering semantic context) to parse speech streams, yet such strategies shape error patterns rather than enhance overall perception. In addition, bilingual listeners are influenced by relative language activation levels between their languages. These findings clarify how bilinguals navigate complex auditory environments and highlight the interplay between acoustic and cognitive factors in adverse listening conditions.

Reference List

Abdel-Latif, K. H. A., Koelewijn, T., Başkent, D., & Meister, H. (2025). Assessment of speech processing and listening effort associated with speech-on-speech masking using the visual world paradigm and pupillometry. *Trends in Hearing*, *29*, 1-13. https://doi.org/10.1177/23312165241306091

Allen-Rice, E., Knight, S., de Bruin, A., & Mattys, S. (in press). Selective and divided listening in a dual-language context. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218251386821

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods, 50*(1), 250-263. https://doi.org/10.3758/s13428-017-0867-9

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Aubanel, V., García Lecumberri, M. L., & Cooke, M. (2014). The Sharvard Corpus: A phonemically-balanced Spanish sentence resource for audiology. *International Journal of Audiology*, *53*(9), 633-638. https://doi.org/10.3109/14992027.2014.907507

Audacity Team (2014). *Audacity®: Free Audio Editor and Recorder* (Version 3.7.1) [Computer Software]. https://www.audacityteam.org/

Axelrod, S., Guzy, L. T., & Diamond, I. T. (1968). Perceived rate of monotic and dichotically alternating clicks. *The Journal of the Acoustical Society of America*, *43*(1), 51-55. https://doi.org/10.1121/1.1910761

Axelrod, S., & Powazek, M. (1972). Dependence of apparent rate of alternating clicks on azimuthal separation between sources. *Psychonomic Science*, *26*(4), 217-218. https://doi.org/10.3758/BF03328599

Bater, L., & Jorda, S. S. (2020). Selective attention. In: V Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual* differences (pp. 4624-4628). Springer. https://doi.org/10.1007/978-3-319-24612-3_1904

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., & Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, *54*(1), 243-262. https://doi.org/10.1044/1092-4388(2010/09-0233)

Bent, T., Buchwald, A., & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *The Journal of the Acoustical Society of America*, *126*(5), 2660-2669. https://doi.org/10.1121/1.3212930

Best, V., Gallun, F. J., Ihlefeld, A., & Shinn-Cunningham, B. G. (2006). The influence of spatial separation on divided listening. *The Journal of the Acoustical Society of America*, *120*(3), 1506-1516. https://doi.org/10.1121/1.2234849

Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*, *12*, Article 714485. https://doi.org/10.3389/fpsyg.2021.714485

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240-250. https://doi.org/10.1016/j.tics.2012.03.001

Bilger, R. C., Nuetzel, J. M., 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. https://doi.org/10.1044/jshr.2701.32

Black, J. W., & Hast, M. H. (1962). Speech reception with altering signal. *Journal of Speech and Hearing Research*, *5*(1), 70-75. https://doi.org/10.1044/jshr.0501.70

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, *118*(2), 245-257. https://doi.org/10.1016/j.cognition.2010.10.012

Boersma, P., & Weenink, D (2023). *Praat: doing phonetics by computer* (Version 6.3.03) [Computer Software]. http://www.praat.org/

Broadbent, D. E. (1958). Immediate memory and the shifting of attention. In D. E. Broadbent (Ed.), *Perception and Communication* (pp. 210-243). Pergamon Press. https://doi.org/10.1037/10037-009

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*(24), 3976-3983. https://doi.org/10.1016/j.cub.2018.10.042

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. https://doi.org/10.1121/1.3675943

Brown, A., Pinto, D., Burgart, K., Zvilichovsky, Y., & Zion-Golumbic, E. (2023). Neurophysiological evidence for semantic processing of irrelevant speech and own-name detection in a virtual café. *Journal of Neuroscience*, *43*(27), 5045-5056. https://doi.org/10.1523/JNEUROSCI.1731-22.2023

Brown, V. A., & Strand, J. F. (2018). Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity. *Language, Cognition and Neuroscience*, *34*(5), 628-640. https://doi.org/10.1080/23273798.2018.1562084

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101-1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2003). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *The Journal of the Acoustical Society of America*, *115*(1), 301-310. https://doi.org/10.1121/1.1628683

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), Article 9. https://doi.org/10.5334/joc.10

Bsharat-Maalouf, D., Degani, T., & Karawani, H. (2023). The involvement of listening effort in explaining bilingual listening under adverse listening conditions. *Trends in Hearing*, *27*, Article 23312165231205107. https://doi.org/10.1177/23312165231205107

Burleson, A. M., & Souza, P. E. (2022). Cognitive and linguistic abilities and perceptual restoration of missing speech: Evidence from online assessment. *Frontiers in Psychology*, *13*, Article 1059192. https://doi.org/10.3389/fpsyg.2022.1059192

Cai, X., Ouyang, M., Yin, Y., & Zhang, Q. (2021). Language proficiency moderates the effect of L2 semantically related distractors in L2 spoken word production. *Brain Research*, *1753*, Article 147235. https://doi.org/10.1016/j.brainres.2020.147231

Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., & Bradlow, A. R. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *American Journal of Audiology*, *22*(1), 157-164. https://doi.org/10.1044/1059-0889(2013/12-0072)

Calandruccio, L., Buss, E., Bencheck, P., & Jett, B. (2018). Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition? *The Journal of the Acoustical Society of America*, *144*(6), 3289-3302. https://doi.org/10.1121/1.5081679

Calandruccio, L., Buss, E., & Bowdrie, K. (2017). Effectiveness of two-talker maskers that differ in talker congruity and perceptual similarity to the target speech. *Trends in Hearing*, *21*, Article 2331216517709385. https://doi.org/10.1177/2331216517709385

Calandruccio, L., & Zhou, H. (2014). Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1087-1097. https://doi.org/10.1044/2013_JSLHR-H-12-0378

Calcus, A., Agus, T., Kolinsky, R., Colin, C., & Deltenre, P. (2015). Isolating informational masking in both pure and complex tone sequences. *Ear and Hearing*, *36*(3), 330-337. https://doi.org/10.1097/AUD.0000000000000116

Chaouch-Orozco, A., Alonso, J. G., & Rothman, J. (2020). Individual differences in bilingual word recognition: The role of experiential factors and word frequency in cross-language lexical priming. *Applied Psycholinguistics*, *42*(2), 447-474. https://doi.org/10.1017/S014271642000082X

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975-979. https://doi.org/10.1121/1.1907229

Ching, T. Y. C., van Wanrooy, E., Dillon, H., & Carter, L. (2011). Spatial release from masking in normal-hearing children and children who use hearing aids. *The Journal of the Acoustical Society of America*, *129*(1), 368-375. https://doi.org/10.1121/1.3523295

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324-345. https://doi.org/10.1016/j.jml.2005.12.004

Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, *14*(4), 699-703. https://doi.org/10.3758/BF03196824

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2)*,* 331-335. https://doi.org/10.3758/BF03196169

Cooke, M. (2003). Glimpsing speech. *Journal of Phonetics*, *31*(3-4), 579-584. https://doi.org/10.1016/S0095-4470(03)00013-5

Cooke, M., Garcia Lecumberri, M. L., & 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. https://doi.org/10.1121/1.2804952

Coulter, K., Gilbert, A. C., Kousaie, S., Baum, S., Gracco, V. L., Klein, D., Titone, D., & Philips, N. A. (2020). Bilinguals benefit from semantic context while perceiving speech in noise in both of their languages: Electrophysiological evidence from the N400 ERP. *Bilingualism: Language and Cognition*, *24*(2), 344-357. https://doi.org/10.1017/S1366728920000516

Culling, J. F., & Stone, M. A. (2017). Energetic masking and masking release. In J. Middlebrooks, J. Simon, A. Popper, & R. Fay. (Eds.), *The auditory system at the cocktail party* (pp. 41-73). Springer Nature. https://doi.org/10.1007/978-3-319-51662-2_3

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264-1268. https://doi.org/10.1121/1.2946707

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. https://doi.org/10.1121/1.1810292

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317-367. https://doi.org/10.1006/cogp.2001.0750

Dalton, P., & Fraenkel, N. (2012). Gorillas we have missed: Sustained inattentional deafness for dynamic events. *Cognition*, *124*(3), 367-372. https://doi.org/10.1016/j.cognition.2012.05.012

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The*

Journal of the Acoustical Society of America, *114*(5), 2913-2922. https://doi.org/10.1121/1.1616924

Declerck, M., & Koch, I. (2023). The concept of inhibition in bilingual control. *Psychological Review*, *130*(4), 953-976. https://doi.org/10.1037/rev0000367

Dekerle, M., Boulenger, V., Hoen, M., & Meunier, F. (2014). Multi-talker background and semantic priming effect. *Frontiers in Human Neuroscience*, *8*, Article 878. https://doi.org/10.3389/fnhum.2014.00878

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881-900. https://doi.org/10.1080/23273798.2019.1591469

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496-518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175-197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the Bilingual Interactive Activation model. *Psychologica Belgica*, *38*(3-4), 177-196. https://doi.org/10.5334/pb.933

Domingo, Y., Holmes, E., Macpherson, E., & Johnsrude, I. S. (2019). Using spatial release from masking to estimate the magnitude of the familiar-voice intelligibility benefit. *The Journal of the Acoustical Society of America*, *146*(5), 3487-3494. https://doi.org/10.1121/1.5133628

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In J. J. O'Haire (Ed), *CHI '88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281-285). Association for Computing Machinery. https://doi.org/10.1145/57167.57214

Edmonds, B. A., & Culling, J. F. (2006). The spatial unmasking of speech: Evidence for better-ear listening. *The Journal of the Acoustical Society of America*, *120*(3), 1539–1545. https://doi.org/10.1121/1.2228573

Ezzatian, P., Avivi, M., & Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? *Speech Communication*, *52*(11-12), 919-929. https://doi.org/10.1016/j.specom.2010.04.001

Ezzatian, P., Li, L., Pichora-Fuller, K., & Schneider, B. A. (2015). Delayed stream segregation in older adults more than just information masking. *Ear and Hearing*, *36*(4), 482-484. https://doi.org/10.1097/AUD.0000000000000139

Fay, A., & Buchweitz, A. (2014). Listening comprehension and individual differences in working memory capacity in beginning L2 learners. *Letrônica*, *7*(1), 113-129. https://doi.org/10.15448/1984-4301.2014.1.16839

Fernandes, M. A., Koji, S., Dixon, M. J., & Aquino, J. M. (2011). Changing the focus of attention: The interacting effect of valence and arousal. *Visual Cognition*, *19*(9), 1191-1211. https://doi.org/10.1080/13506285.2011.618151

Flege, J. E., & Liu, S. (2002). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527-552. https://doi.org/10.1017/S0272263101004041

Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13795-13798. https://doi.org/10.1073/pnas.1401383111

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, & Psychophysics*, *72*(2), 501-516. https://doi.org/10.3758/APP.72.2.501

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences*, *8*(2), 87-93. https://doi.org/10.1016/j.tics.2003.12.011

Gabbay, C., Zivony, A., & Lamy, D. (2019). Splitting the attentional spotlight? Evidence from attentional capture by successive events. *Visual Cognition*, *27*(5-8), 518-536. https://doi.org/10.1080/13506285.2019.1617377

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445-2454. https://doi.org/10.1121/1.2180210

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11-12), 864-886. https://doi.org/10.1016/j.specom.2010.08.014

Garnier-Allain, A., Pressnitzer, D., & Sergent, C. (2023). Retrospective cueing mediates flexible conscious access to past spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *49*(7), 949-967. https://doi.org/10.1037/xhp0001132

Gat, I. B., & Keith, R. W. (1978). An effect of linguistic experience: Auditory word discrimination by native and non-native speakers of English. *Audiology*, *17*(4), 339-345. https://doi.org/10.3109/00206097809101303

Gaudrain, E., Li, S., Ban, V. S., & Patterson, R. D. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. In *Interspeech 2009* (pp. 148-151). Curran Associates. https://hal.science/hal-02144510

Goupell, M. J., Kan, A., & Litovsky, R. Y. (2016). Spatial attention in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, *140*(3), 1652-1662. https://doi.org/10.1121/1.4962378

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515-530. https://doi.org/10.1080/20445911.2013.796377

Grossmann, W., Brill, S., Moeltner, A., Mlynski, R., Hagen, R., & Radeloff, A. (2016). Cochlear implantation improves spatial release from masking and restores localization abilities in single-sided deaf patients. *Otology & Neurotology*, *37*(6), 658-664. https://doi.org/10.1097/MAO.0000000000001043

Har-shai Yahav, P., & Zion Golumbic, E. (2021). Linguistic processing of task-irrelevant speech at a cocktail party. *eLife*, *10*, Article e65096. https://doi.org/10.7554/eLife.65096

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, *29*(1), 87-98. https://doi.org/10.1097/AUD.0b013e31815d638b

Hohlfeld, A., & Sommer, W. (2005). Semantic processing of unattended meaning is modulated by additional task load: Evidence from electrophysiology. *Cognitive Brain Research*, *24*(3), 500-512. https://doi.org/10.1016/j.cogbrainres.2005.03.001

Holmes, E., Folkeard, P., Johnsrude, I. S., & Scollie, S. (2018). Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *International Journal of Audiology*, *57*(7), 483-492. https://doi.org/10.1080/14992027.2018.1432901

Holmes, E., Kitterick, P. T., & Summerfield, A. Q. (2017). Peripheral hearing loss reduces the ability of children to direct selective attention during multi-talker listening. *Hearing Research*, *350*, 160-172. https://doi.org/10.1016/j.heares.2017.05.005

Huggins, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *The Journal of the Acoustical Society of America*, *36*(6), 1055-1064. https://doi.org/10.1121/1.1919151

Hui, C. T. J., Au, E., Xiao, S., Hioka, Y., Masuda, H., & Watson, C. I. (2021). Differences in speech intelligibility in noise between native and non-native listeners under ambisonics-based sound reproduction system. *Applied Acoustics*, *184*, Article 108368. https://doi.org/10.1016/j.apacoust.2021.108368

Hui, C. T. J., Hioka, Y., Masuda, H., & Watson, C. I. (2022). Differences between listeners with early and late immersion age in spatial release from masking in various acoustic environments. *Speech Communication*, *139*, 51-61. https://doi.org/10.1016/j.specom.2022.02.004

Hunter, C. R. (2021). Dual-task accuracy and response time index effects of spoken sentence predictability and cognitive load on listening effort. *Trends in Hearing*, *25*, 1-15. https://doi.org/10.1177/23312165211018092

Ihlefeld, A., & Shinn-Cunningham, B. (2008a). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369-4379. https://doi.org/10.1121/1.2904826

Ihlefeld, A., & Shinn-Cunningham, B. (2008b). Spatial release from energetic and informational masking in a divided speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4380-4392. https://doi.org/10.1121/1.2904825

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica: International Journal of Methodology and Experimental Psychology*, *35*(1), 49-66.

Jacquemot, C., Dupoux, E., Decouche, O., & Bachoud-Lévi, A. C. (2006). Misperception in sentences but not in words: Speech perception and the phonological buffer. *Cognitive Neuropsychology*, *23*(6), 949-971. https://doi.org/10.1080/02643290600625749

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, *10*(11), 480-486. https://doi.org/10.1016/j.tics.2006.09.002

Jin, S. H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *The Journal of the Acoustical Society of America*, *132*(5), 391-397. https://doi.org/10.1121/1.4757730

Johnson, E. K., Bruggeman, L., & Cutler, A. (2017). Abstraction and the (misnamed) Language Familiarity Effect. *Cognitive Science*, *42*(2), 633-645. https://doi.org/10.1111/cogs.12520

Kahneman, D. (1973). *Attention and effort*. Prentice Hall.

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. https://doi.org/10.1121/1.381436

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169-183. https://doi.org/10.1037//0096-3445.130.2.169

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143-189). Springer. https://doi.org/10.1007/978-0-387-71305-2_6

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*(1), 27-39. https://doi.org/10.1016/S0926-6410(01)00085-4

Killan, C. F., Royle, N., Totten, C. L., Raine, C. H., & Lovett, R. E. S. (2015). The effect of early auditory experience on the spatial listening skills of children with bilateral cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *79*(12), 2159-2165. https://doi.org/10.1016/j.ijporl.2015.09.039

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology*, *5*, Article 651. https://doi.org/10.3389/fpsyg.2014.00651

Kim, M., Nam, Y., & Crossley, S. A. (2022). Roles of working memory, syllogistic inferencing ability, and linguistic knowledge on second language listening comprehension for passages of different lengths. *Language Testing*, *39*(4), 593-617. https://doi.org/10.1177/02655322211060076

Knight, S., Rakusen, L., & Mattys, S. (2023). Conceptualising acoustic and cognitive contributions to divided-attention listening within a data-limit versus resource-limit framework. *Journal of Memory and Language*, *131*, Article 104427. https://doi.org/10.1016/j.jml.2023.104427

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174. https://doi.org/10.1006/jmla.1994.1008

Lad, M., Holmes, E., Chu, A., & Griffiths, T. D. (2020). Speech-in-noise detection is related to auditory working memory precision for frequency. *Scientific Reports*, *10*, Article 13997. https://doi.org/10.1038/s41598-020-70952-9

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750-761. https://doi.org/10.1016/j.neuron.2012.11.034

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451-468. https://doi.org/10.1037//0096-1523.21.3.451

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325-343. https://doi.org/10.3758/s13428-011-0146-0

Lew, E., Hallot, S., Byers-Heinlein, K., & Deroche, M. (2024). Navigating the bilingual cocktail party: A critical role for listeners' L1 in the linguistic aspect of informational masking.

Bilingualism: Language and Cognition, *28*(3), 748-756. https://doi.org/10.1017/S1366728924000944

Ling, A. H. (1971). Dichotic listening in hearing-impaired children. *Journal of Speech, Language, and Hearing Research*, *14*(4), 793-803. https://doi.org/10.1044/jshr.1404.793

Litovsky, R. Y. (2012). Spatial release from masking. *Acoustics Today*, *8*(2), 18-25.

Lynch, E. E. (2021). *Effects of perceptual load on dichotic and diotic listening performance* (Publication No. 28830015) [Doctoral dissertation, Ohio University]. ProQuest Dissertations & Theses.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America*, *110*(1), 516-528. https://doi.org/10.1121/1.1377287

Macnamara, B. N., & Conway, A. R. A. (2016). Working memory capacity as a predictor of simultaneous language interpreting performance. *Journal of Applied Research in Memory and Cognition*, *5*(4), 434-444. https://doi.org/10.1016/j.jarmac.2015.12.001

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL quarterly*, *36*(2), 173-190. https://doi.org/10.2307/3588329

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97-115. https://doi.org/10.1017/S1366728903001068

Marrone, N., Mason, C. R., & Kidd, G., Jr. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *The*

Journal of the Acoustical Society of America, *124*(5), 3064-3075. https://doi.org/10.1121/1.2980441

Marrufo-Pérez, M. I., & Lopez-Poveda, E. A. (2025). Speech recognition and noise adaptation in realistic noises. *Trends in Hearing*, *29*, 1-13. https://doi.org/10.1177/23312165251343457

Marsh, J. E., & Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: The new early filter model. *Frontiers in Neuroscience*, *10*, Article 136. https://doi.org/10.3389/fnins.2016.00136

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. https://doi.org/10.1080/01690965.2012.705006

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. https://doi.org/10.1044/jslhr.4003.686

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703-704. https://doi.org/10.1038/nn1264

Mepham, A., Bi, Y., & Mattys, S. L. (2022). The time-course of linguistic interference during native and non-native speech-in-speech listening. *The Journal of the Acoustical Society of America*, *152*(2), 954-969. https://doi.org/10.1121/10.0013417

Mepham, A., Knight, S., McGarrigle, R., Rakusen, L., & Mattys, S. (2025). Pupillometry reveals the role of signal-to-noise in adaption to linguistic interference over time. *Journal of Speech, Language, and Hearing Research*, *68*(5), 2291-2317. https://doi.org/10.1044/2025_JSLHR-24-00658

Microsoft Corporation. (2018). *Microsoft PowerPoint* (Version 2021) [Computer software]. https://office.microsoft.com/powerpoint

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, *42*, 135-159. https://doi.org/10.1146/annurev.ps.42.020191.001031

Middlebrooks, J. C., & Onsan, Z. A. (2012). Stream segregation with high spatial acuity. *The Journal of the Acoustical Society of America*, *132*(6), 3896-3911. https://doi.org/10.1121/1.4764879

Mielicki, M. K., Koppel, R. H., Valencia, G., & Wiley, J. (2018). Measuring working memory capacity with the letter-number sequencing task: Advantages of visual administration. *Applied Cognitive Psychology*, *32*(6), 805-814. https://doi.org/10.1002/acp.3468

Misurelli, S. M., & Litovsky, R. Y. (2015). Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity. *The Journal of the Acoustical Society of America*, *138*(1), 319-331. https://doi.org/10.1121/1.4922777

Misurelli, S. M., & Litovsky, R. Y. (2012). Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *The Journal of the Acoustical Society of America*, *132*(1), 380-391. https://doi.org/10.1121/1.4725760

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56-60. https://doi.org/10.1080/17470215908416289

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*, 309-312. https://doi.org/10.1038/nature01812

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition*, *129*(2), 345-355. https://doi.org/10.1016/j.cognition.2013.07.014

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, *352*, 40-48. https://doi.org/10.1016/j.heares.2017.02.005

Neill, W. T., Valdes, L. A., & Terry, K. M. (1995). Selective attention and the inhibitory control of cognition. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 207-261). Academic Press. https://doi.org/10.1016/B978-012208930-5/50008-8

Oh, Y., Bridges, S. E., Schoenfeld, H., Layne, A. O., & Eddins, D. (2021). Interaction between voice-gender difference and spatial separation in release from masking in multi-talker listening environments. *The Journal of the Acoustical Society of America Express Letters*, 1(8), Article 084404. https://doi.org/10.1121/10.0005831

Oh, Y., Hartling, C. L., Srinivasan, N. K., Diedesch, A. C., Gallun, F. J., & Reiss, L. A. J. (2022). Factors underlying masking release by voice-gender differences and spatial separation cues in multi-talker listening environments in listeners with and without hearing loss. *Frontiers in Neuroscience*, *16*, Article 1059639. https://doi.org/10.3389/fnins.2022.1059639

Pandey, P. R., & Herrmann, B. (2025). The influence of semantic context on the intelligibility benefit from speech glimpses in younger and older adults. *Journal of Speech, Language, and Hearing Research*, *68*(5), 2499-2516. https://doi.org/10.1044/2025_JSLHR-24-00588

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204-214. https://doi.org/10.1097/AUD.0000000000000494

Philips, C., Jacquemin, L., Lammers, M. J. W., Mertens, G., Gilles, A., Vanderveken, O. M., & Van Rompaey, V. (2023). Listening effort and fatigue among cochlear implant users: A scoping review. *Frontiers in Neurology*, *14*, Article 1278508. https://doi.org/10.3389/fneur.2023.1278508

Pigoli, D., Hadjipantelis, P. Z., Coleman, J. S., Aston, J. A. D. (2018). The statistical analysis of acoustic phonetic data: Exploring differences between spoken romance languages. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *67*(5), 1103-1145. https://doi.org/10.1111/rssc.12258

Pinto, D., Agmon, G., & Golumbic, E. Z. (2020). *The role of spatial separation on selective and distributed attention to speech*. bioRxiv. https://doi.org/10.1101/2020.01.27.920785

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S5). https://doi.org/10.1121/1.1995329

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174. https://doi.org/10.1037/0096-3445.109.2.160

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26,* 26-46.

Prolific (2014). *Prolific.* https://www.prolific.com

Regalado, D., Kong, J., Buss, E., & Calandruccio, L. (2019). Effects of language history on sentence recognition in noise or two-talker speech: Monolingual, early bilingual, and late bilingual speakers of English. *American Journal of Audiology*, *28*(4), 935-946. https://doi.org/10.1044/2019_AJA-18-0194

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, 118(3), 1274-1277. https://doi.org/10.1121/1.2000751

Rivenez, M., Darwin, C. J., & Guillaume, A. (2006). Processing unattended speech. *The Journal of the Acoustical Society of America*, *119*(6), 4027-4040. https://doi.org/10.1121/1.2190162

Röer, J. P., Bell, R., Körner, U., & Buchner, A. (2019). A semantic mismatch effect on serial recall: Evidence for interlexical processing of irrelevant speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 515-525. https://doi.org/10.1037/xlm0000596

Röer, J. P., & Cowan, N. (2021). A preregistered replication and extension of the cocktail party phenomenon: One's name captures attention, unexpected words do not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 234-242. https://doi.org/10.1037/xlm0000874

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. https://doi.org/10.1080/14992020802301167

Rossi-Katz, J., & Arehart, K. H. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research*, *52*(2), 435-453. https://doi.org/10.1044/1092-4388(2008/07-0243)

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225-246. https://doi.org/10.1109/TAU.1969.1162058

Salverda, A. P., & Tanenhaus, M. K. (2017). The visual world paradigm. In A. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 89-110). Wiley. https://doi.org/10.1002/9781394259762.ch5

Samuel, A. G. (1991). Perceptual degradation due to signal alternation: Implications for auditory pattern processing. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 392-403. https://doi.org/10.1037/0096-1523.17.2.392

Scharenborg, O., Coumans, J. M. J., & van Hout, R. (2018). The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 233-249. https://doi.org/10.1037/xlm0000441

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489-499. https://doi.org/10.1121/1.2783762

Shinn-Cunningham, B. (2013). Understanding informational masking from a neural perspective. *Proceedings of Meetings on Acoustics*, *19*(1), Article 060143. https://doi.org/10.1121/1.4799846

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Hearing*, *12*(4), 283-299. https://doi.org/10.1177/1084713808325306

Shook, A., & Marian, V. (2012). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism: Language and Cognition*, *16*(2), 304-324. https://doi.org/10.1017/S1366728912000466

Smith, E. D., Holt, L. L., & Dick, F. (2024). A one-man bilingual cocktail party: Linguistic and non-linguistic effects on bilinguals' speech recognition in Mandarin and English. *Cognitive Research: Principles and Implications*, *9*(1), Article 35. https://doi.org/10.1186/s41235-024-00562-w

Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, *122*(6), 3628-3639. https://doi.org/10.1121/1.2799507

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. https://doi.org/10.1111/1467-9280.00151

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., & Moore, B. C. J. (2011). The importance for speech intelligibility of random fluctuations in "steady" background noise. *The Journal of the Acoustical Society of America*, *130*(5), 2874-2881. https://doi.org/10.1121/1.3641371

Szalárdy, O., Tóth, B., Farkas, D., Orosz, G., Honbolygó, F., & Winkler, I. (2020). Linguistic predictability influences auditory stimulus classification within two concurrent speech streams. *Psychophysiology*, *57*(5), Article e13547. https://doi.org/10.1111/psyp.13547

The MathWorks Inc. (2022). *MATLAB* (Version 9.13.0 (R2022b)) [Computer Software]. https://www.mathworks.com

Treisman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, *3*(6), 449-459. https://doi.org/10.1016/S0022-5371(64)80015-3

Treisman, A. M. (1971). Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, *23*(2), 157-167. https://doi.org/10.1080/14640747108400236

Treisman, A., & Geffen, G. (1967). Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, *19*(1), 1-17. https://doi.org/10.1080/14640746708400062

Treisman, A. M., & Riley, J. G. (1969). Is selective attention perception or selective response? A further test. *Journal of Experimental Psychology*, *79*(1), 27-34. https://doi.org/10.1037/h0026890

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453-467. https://doi.org/10.1037/0882-7974.17.3.453

Tzou, Y. Z. (2008). *The roles of working memory, language proficiency, and training in simultaneous interpretation performance: Evidence from Chinese-English bilinguals*. Texas A&M University.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11-12), 943-953. https://doi.org/10.1016/j.specom.2010.05.002

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519-526. https://doi.org/10.1121/1.2400666 van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Van Rullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *The Proceedings of the National Academy of Sciences*, *104*(49), 19204-19209. https://doi.org/10.1073/pnas.0707316104

Villard, S., Yap, I., & Kidd, G, Jr. (2024). Semantic relatedness and the cocktail party problem in aphasia: A hybrid remote/in-lab study. *Aphasiology*, *38*(4), 612-634. https://doi.org/10.1080/02687038.2023.2221997

Wang, Y., Zhang, J., Zou, J., Luo, H., & Ding, N. (2019). Prior knowledge guides speech segregation in human auditory cortex. *Cerebral Cortex*, *29*(4), 1561-1571. https://doi.org/10.1093/cercor/bhy052

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1-25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wechsler, D. (1997). *Wechsler adult intelligence scale – Third Edition*. Frontiers in Psychology.

Wen, Y., & van Heuven, W. J. B. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review*, *24*, 879-886. https://doi.org/10.3758/s13423-016-1151-1

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. *Journal of Neuroscience*, *32*(40), 14010-14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The Revised Speech Perception in Noise Test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *Journal of the American Academy of Audiology*, *23*(8), 590-605. https://doi.org/10.3766/jaaa.23.7.9

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255-260. https://doi.org/10.1037/0278-7393.21.1.255

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*, 2064-2072. https://doi.org/10.3758/s13414-017-1361-2

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*, Article 88. https://doi.org/10.3389/fnins.2014.00088

Zhou, Z., Ding, J., Wang, J., Chen, Y., & Li, X. (2023). The flexibility and representational nature of phonological prediction in listening comprehension: Evidence from the visual world paradigm. *Language and Cognition*, *16*(2), 481-504. https://doi.org/10.1017/langcog.2023.38

Zinszer, B. D., Riggs, M., Reetzke, R., & Chandrasekaran, B. (2019). Error patterns of native and non-native listeners' perception of speech in noise. *The Journal of the Acoustical Society of America*, *145*(2), EL129-EL135. https://doi.org/10.1121/1.5087271

Zobel, B. H., Wagner, A., Sanders, L. D., & Başkent, D. (2019). Spatial release from informational masking declines with age: Evidence from a detection task in a virtual separation paradigm. *The Journal of the Acoustical Society of America*, *146*(1), 548-566. https://doi.org/10.1121/1.5118240

Appendices

Appendix A: Chapter 2 Supplementary Materials

A.1

Table 5 - Language background of the Experiment 2 sample (N = 80)

Language background measure	**L1-Spanish**				**L2-English**			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Age of acquisition (years)[a]	0.00	0.00	0.00	0.00	5.98	2.45	3.00	13.00
LexTale proficiency (%)[b,c]	91.79	6.08	70.83	100.00	79.77	9.01	61.25	96.25
Self-reported understanding (/10)	9.96	0.33	7.00	10.00	8.28	1.01	5.00	10.00
Self-reported speaking (/10)	9.99	0.11	9.00	10.00	7.45	1.18	5.00	10.00
Self-reported reading (/10)	9.96	0.33	7.00	10.00	8.89	0.91	5.00	10.00
Self-reported writing (/10)	9.99	0.25	8.00	10.00	7.80	1.29	6.00	10.00
Language use (/4)[d]	Mean = 2.94, SD = 0.47, Min = 1.21, Max = 3.76							

[a] The L1-Spanish values exclude one participant who stated their age of acquiring Spanish as 34 years but demonstrated high Spanish proficiency throughout the study. Because we assume that they misunderstood the question as how long they had known Spanish, we have included their other scores in analyses.

[b] The LexTALE is a standardised test of language proficiency (see section 2.3.3.2. Materials, LexTALE and Lextale-Esp for details).

[c] The L1-Spanish values excludes one participant who scored 5% in the Lextale-Esp but demonstrated high Spanish proficiency throughout the study. Because we assume that they misunderstood the instructions for the Lextale-Esp, we have included their other scores in the analyses.

[d] This value was calculated from the language and social background questionnaire (see section 2.3.3.2. Materials, Language background questionnaire for details) in which participants were asked to state which language they used in particular situations. 0 = All English, 1 = Mostly English, 2 = Half Each Language, 3 = Mostly Spanish, 4 = All Spanish.

A.2

Table 6 - Language background for the comparison sample (participants who completed Experiment 1 and 2; N = 66)

Language background measure	L1-Spanish				L2-English			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Age of acquisition (years)[a]	0.00	0.00	0.00	0.00	5.85	2.43	3.00	13.00
LexTale proficiency (%)[b,c]	91.62	6.00	70.83	100.00	79.39	8.62	63.75	96.25
Self-reported understanding (/10)	9.95	0.37	7.00	10.00	8.27	1.01	5.00	10.00
Self-reported speaking (/10)	9.98	0.12	9.00	10.00	7.38	1.20	5.00	10.00
Self-reported reading (/10)	9.95	0.37	7.00	10.00	8.85	0.87	6.00	10.00
Self-reported writing (/10)	9.98	0.12	9.00	10.00	7.70	1.29	5.00	10.00
Language use (/4)[d]	Mean = 2.95, SD = 0.47, Min = 1.21, Max = 3.76							

[a] The L1-Spanish values exclude one participant who stated their age of acquiring Spanish as 34 years but demonstrated high Spanish proficiency throughout the study. Because we assume that they misunderstood the question as how long they had *known* Spanish, we have included their other scores in analyses.

[b] The LexTALE is a standardised test of language proficiency (see section 2.3.3.2. Materials, LexTALE and Lextale-Esp for details).

[c] The L1-Spanish values excludes one participant who scored 5% in the Lextale-Esp but demonstrated high Spanish proficiency throughout the study. Because we assume that they misunderstood the instructions for the Lextale-Esp, we have included their other scores in the analyses.

[d] This value was calculated from the Language and Social Background Questionnaire (see section 2.3.3.2. Materials, Language background questionnaire for details) in which participants were asked to state which language they used in different situations. 0 = All English, 1 = Mostly English, 2 = Half Each Language, 3 = Mostly Spanish, 4 = All Spanish.

A.3

Table 7 - GLMM output for Experiment 1: LNS Model (with Experiment 2 LNS)

Fixed Effects	Estimate	Std. Error	Z value	Pr(>|z|)
(Intercept)	2.24	0.11	20.92	< .001***
Language	2.77	0.20	13.96	< .001***
Spatial separation	1.03	0.17	6.11	< .001***
LNS score	0.12	0.08	1.58	.113
Language:Spatial separation	0.58	0.34	1.72	.086
Language:LNS score	0.14	0.13	1.03	.302
Spatial separation:LNS score	0.10	0.08	1.23	.217
Language:Spatial separation:LNS score	0.32	0.17	1.81	.070

Significance codes: *** < .001; ** < .010; * < .050

Appendix B: Chapter 3 Supplementary Materials

B.1

Table 8 - L2-English sentences

	Target (not used, all targets were L1-Spanish)	**Semantically competing masker**	**Neutral masker**
1	The cat casually scratches the chair.	The dog anxiously gnaws the table.	The lily slowly opens its petals.
2	The deer runs quickly across the field.	The horse gallops wildly along the path.	The brother points excitedly at the hamster.
3	The referee definitively resumes the match.	The commentator decisively stops the game.	The knife carefully slices the fruit.
4	The dancer carefully plaits her hair.	The actress abruptly removes her wig.	The tulip slowly dies without water.
5	The dentist carefully straightens the tooth.	The doctor calmly aligns the bone.	The boxer powerfully defeats the opponent.
6	The nurse often jogs from the office.	The surgeon rarely walks from the house.	The mole slowly crawls onto the lawn.
7	The snake glides stealthily along the sand.	The worm wriggles silently through the mud.	The mother works sadly on the computer.
8	The drizzle occasionally lingers in the atmosphere.	The clouds continuously float in the air.	The father attentively monitors the teenagers.
9	The thunder rumbles furiously in the gloom.	The lightning flashes violently in the dark.	The van travels slowly along the street.
10	The hound barks loudly at the walker.	The wolf growls aggressively at the hiker.	The cyclist pedals frantically during the race.
11	The saxophone delicately accompanies the guitarist.	The clarinet melodically harmonises with the pianist.	The nanny cheerfully praises the children.
12	The sister abruptly pushes the boy.	The cousin cruelly slaps the man.	The swimmer gracefully dives into the lake.
13	The bagpipes echo hauntingly across the hills.	The harmonica wails mournfully in the moors.	The moth flaps clumsily in the park.
14	The robin cheerfully chirps in the morning.	The eagle angrily squawks at dawn.	The loser clumsily bestows the prize.
15	The rain drips gently on the roof.	The hail crashes strongly on the shelter.	The coach cycles powerfully to the end.

16	The singer nervously records his song.	The actor fearfully films the movie.	The pigeon carefully traps the beetle.
17	The waves crash loudly against the shore.	The stream laps gently against the shells.	The pupil learns distractedly in the classroom.
18	The mechanic efficiently fixes the bicycle.	The plumber quickly corrects the pipes.	The community abruptly exiles the stranger.
19	The firefighter bravely rescues the kitten.	The paramedic fearfully saves the puppy.	The jury diligently listens in the court.
20	The pilot flies safely in the sky.	The driver speeds recklessly on the road.	The cake smells bitter to the nose.
21	The villain cruelly threatens the mayor.	The superhero kindly protects the councillor.	The friends lazily tidy the bathrooms.
22	The palaeontologist cheerfully excavates the fossil.	The archaeologist sadly uncovers the ruins.	The litigator boldly advises the prosecutor.
23	The lion roars furiously at the zebra.	The tiger purrs contentedly at the giraffe.	The woodworker cuts carefully with the saw.
24	The husband angrily refuses his lunch.	The boyfriend stubbornly avoids the dinner.	The architect easily moves the wood.
25	The aeroplane descends calmly through the fog.	The helicopter ascends clumsily into the haze.	The employee disagrees strongly with their assistant.
26	The boat sails calmly from the harbour.	The yacht embarks gently from the pier.	The follower mumbles timidly during the speech.
27	The grandmother bakes happily in the kitchen.	The grandson fries cheerfully at the stove.	The typhoon rushes frantically through the country.
28	The footballer angrily snatches the trophy.	The basketballer scornfully takes the award.	The pepper definitely improves the soup.
29	The princess unexpectedly betrays the king.	The heir surprisingly deceives the ruler.	The alarm beeps loudly in the room.
30	The wasp painfully stings the aunt.	The ant angrily bites the niece.	The mixologist messily stirs the cocktail.
31	The clown clumsily entertains the crowd.	The juggler skilfully performs before the audience.	The soldier boldly aims his gun.
32	The kettle boils violently next to the fridge.	The toaster heats quickly by the oven.	The toddler whines persistently in the nursery.
33	The analyst angrily stares at the numbers.	The researcher sadly watches the participants.	The tailor attentively alters the suit.
34	The plate loudly splinters in the drawer.	The cup quickly cracks in the cupboard.	The shutters completely muffle the sound.
35	The gladiator fearfully attacks the monster.	The warrior bravely charges towards the beast.	The lady frantically paints the image.

36	The assassin carefully identifies the target.	The murderer cautiously chooses the victim.	The gymnast quickly replaces the trousers.
37	The witch lovingly strokes her raven.	The wizard kindly feeds his owl.	The pig anxiously gobbles the carrot
38	The thief stealthily steals the watch.	The criminal skilfully hides the jewels.	The sheep anxiously drinks the liquid.
39	The queen powerfully commands the army.	The prince boldly fights in the battle.	The author quickly writes the chapter.
40	The florist delicately picks the flowers.	The gardener gently trims the shrubs.	The tourists cheerfully spot the meteor.
41	The plant gradually grows its roots.	The lily slowly opens its petals.	The dog anxiously gnaws the table.
42	The nephew smiles happily at the rabbit.	The brother points excitedly at the hamster.	The horse gallops wildly along the path.
43	The fork gently pierces the vegetable.	The knife carefully slices the fruit.	The commentator decisively stops the game.
44	The daisy gradually wilts in the heat.	The tulip slowly dies without water.	The actress abruptly removes her wig.
45	The wrestler confidently dominates the fight.	The boxer powerfully defeats the opponent.	The doctor calmly aligns the bone.
46	The squirrel quickly climbs up the tree.	The mole slowly crawls onto the lawn.	The surgeon rarely walks from the house.
47	The girl plays happily with the toys.	The mother works sadly on the computer.	The worm wriggles silently through the mud.
48	The uncle carefully supervises the youths.	The father attentively monitors the teenagers.	The clouds continuously float in the air.
49	The car drives timidly along the motorway.	The van travels slowly along the street.	The lightning flashes violently in the dark.
50	The runner sprints quickly along the track.	The cyclist pedals frantically during the race.	The wolf growls aggressively at the hiker.
51	The teacher sternly scolds the students.	The nanny cheerfully praises the children.	The clarinet melodically harmonises with the pianist.
52	The diver elegantly wades into the pool.	The swimmer gracefully dives into the lake.	The cousin cruelly slaps the man.
53	The butterfly flutters elegantly in the garden.	The moth flaps clumsily in the park.	The harmonica wails mournfully in the moors.
54	The winner proudly holds the medal.	The loser clumsily bestows the prize.	The eagle angrily squawks at dawn.
55	The athlete progresses impressively in the competition.	The coach advances powerfully until the end.	The hail crashes strongly on the shelter.

56	The sparrow skilfully ambushes the spider.	The pigeon carefully traps the beetle.	The actor fearfully films a movie.
57	The student studies diligently at the library.	The pupil learns distractedly in the classroom.	The stream laps gently against the shells.
58	The village kindly accepts the visitor.	The community abruptly exiles the stranger.	The plumber quickly corrects the pipes.
59	The lawyer carefully researches the case.	The jury diligently listens in the court.	The paramedic fearfully saves the puppy.
60	The cookie tastes sweet to the mouth.	The cake smells bitter to the nose.	The driver speeds recklessly on the road.
61	The family busily cleans the bedrooms.	The friends lazily tidy the bathrooms.	The superhero kindly protects the councillor.
62	The judge calmly instructs the defendant.	The litigator boldly advises the prosecutor.	The archaeologist sadly uncovers the ruins.
63	The blacksmith strikes firmly with the hammer.	The woodworker cuts carefully with the saw.	The tiger purrs contentedly at the giraffe.
64	The builder effortfully carries the bricks.	The architect easily moves the wood.	The boyfriend stubbornly avoids the dinner.
65	The manager argues angrily with the colleague.	The employee disagrees strongly with their assistant.	The helicopter ascends clumsily into the haze.
66	The leader speaks boldly during his presentation.	The follower mumbles timidly during the speech.	The yacht embarks gently from the pier.
67	The hurricane races chaotically across the earth.	The typhoon rushes frantically through the country.	The grandson fries cheerfully at the stove.
68	The salt almost perfects the salad.	The pepper definitely improves the soup.	The basketballer scornfully takes the award.
69	The bell rings faintly through the building.	The alarm beeps loudly in the room.	The heir surprisingly deceives the ruler.
70	The bartender accidentally spills the wine.	The mixologist messily stirs the cocktail.	The ant angrily bites the niece.
71	The knight bravely lifts her sword.	The soldier boldly aims his gun.	The juggler skilfully performs before the audience.
72	The baby cries desperately from the crib.	The toddler whines persistently in the nursery.	The toaster heats quickly by the oven.
73	The seamstress distractedly mends the dress.	The tailor attentively alters the suit.	The researcher sadly watches the participants.
74	The curtains almost block the view.	The shutters completely muffle the sound.	The cup quickly cracks in the cupboard.
75	The gentleman calmly draws the sketch.	The lady frantically paints the image.	The warrior bravely charges towards the beast.

76	The golfer immediately returns the shoes.	The gymnast quickly replaces the trousers.	The murderer cautiously chooses the victim.
77	The cow quickly devours the apple.	The pig anxiously gobbles the carrot.	The wizard kindly feeds his owl.
78	The donkey hungrily chews the food.	The sheep anxiously drinks the liquid.	The criminal skilfully hides the jewels.
79	The poet slowly reads the words.	The author quickly writes the chapter.	The prince boldly fights in the battle.
80	The adventurers excitedly observe the comet.	The tourists cheerfully spot the meteor.	The gardener gently trims the shrubs.

B.2

Table 9 - L1-Spanish sentences (translations of L2-English sentences)

	Target	**Semantically competing masker**	**Neutral masker**
1	El gato casualmente rasca la silla.	El perro ansiosamente roe la mesa.	El lirio lentamente abre sus pétalos.
2	El ciervo corre rápidamente por el campo.	El caballo galopa salvajemente por el camino.	El hermano señala emocionado al hámster.
3	El árbitro definitivamente reanuda el partido.	El comentarista decisivamente detiene el juego.	El cuchillo cuidadosamente rebana la fruta.
4	La bailarina cuidadosmente se trenza el pelo.	La actriz bruscamente se quita la peluca.	El tulipán lentamente muere sin agua.
5	El dentista cuidadosamente endereza el diente.	El médico tranquilamente alinea el hueso.	El boxeador poderosamente derrota el adversario.
6	La enfermera suele trotar desde la oficina.	El cirujano raramente camina de casa.	El topo lentamente se arrastra hasta el césped.
7	La serpiente se desliza sigilosamente por la arena.	El gusano se retuerce silenciosamente por el barro.	La madre trabaja tristemente en el ordenador.
8	La llovizna ocasionalmente permanece en la atmósfera.	Las nubes continuamente flotan en el aire.	El padre atentamente vigila a los adolescentes.
9	El trueno retumba furiosamente en la penumbra.	Los relámpagos centellean violentamente en la oscuridad.	La furgoneta viaja lentamente por la calle.
10	El sabueso ladra fuertemente al caminante.	El lobo gruñe agresivamente al excursionista.	El ciclista pedalea frenéticamente durante la carrera.
11	El saxofón delicadamente acompaña al guitarrista.	El clarinete melódicamente armoniza con el pianista.	La niñera alegremente alaba a los niños.
12	La hermana bruscamente empuja al niño.	El primo cruelmente abofetea al hombre.	El nadador grácilmente se zambulle del lago.
13	Las gaitas resuenan inquietantes por las colinas.	La armónica gime lúgubremente en los páramos.	La polilla aletea torpemente en el parque.
14	El petirrojo alegremente pía por la mañana.	El águila enfadadamente grazna al amanecer.	El perdedor torpemente entrega el premio.
15	La lluvia gotea suavemente sobre el tejado.	El granizo impacta fuertemente contra el refugio.	El entrenador pedalea poderosamente hasta el final.
16	El cantante nerviosamente graba su canción.	El actor temerosamente filma una película.	La paloma cuidadosamente atrapa al escarabajo.

17	Las olas rompen ruidosamente contra la orilla.	El arroyo chapotea suavemente contra las conchas.	El alumno aprende distraído en el aula.
18	El mecánico eficazmente arregla la bicicleta.	El fontanero rápidamente corrige las tuberías.	La comunidad bruscamente exilia al forastero.
19	El bombero valientemente rescata al gatito.	El paramédico temerosamente salva al cachorro.	El jurado diligentemente escucha en el tribunal.
20	El piloto vuela con seguridad por el cielo.	El conductor acelera temerariamente por la carretera.	El pastel huele amargo al olfato.
21	El villano cruelmente amenaza al alcalde.	El superhéroe amablemente protege el consejero.	Los amigos perezosamente ordenan los baños.
22	El paleontólogo alegremente excava el fósil.	El arqueólogo tristemente descubre las ruinas.	El litigante audazmente aconseja al fiscal.
23	El león ruge furiosamente a la cebra.	El tigre ronronea satisfecho ante la jirafa.	El carpintero corta cuidadosamente con la sierra.
24	El marido enfadadamente rechaza su almuerzo.	El novio obstinadamente evita la cena.	El arquitecto fácilmente mueve la madera.
25	El avión desciende tranquilamente entre la niebla.	El helicóptero asciende torpemente hacia la bruma.	El empleado discrepa fuertemente con su asistente.
26	El barco navega tranquilamente desde el puerto.	El yate embarca suavemente desde el muelle.	El seguidor murmura tímidamente durante el discurso.
27	La abuela hornea feliz en la cocina.	El nieto fríe alegremente en la estufa.	El tifón se precipita frenéticamente el país.
28	El futbolista enfadadamente arrebata el trofeo.	El baloncestista desdeñosamente toma el galardón.	La pimienta definitivamente mejora la sopa.
29	La princesa inesperadamente traiciona al rey.	La heredera sorprendentemente engaña al gobernante.	La alarma pita ruidosamente en la sala.
30	La avispa dolorosamente pica a la tía.	La hormiga enfadadamente muerde a la sobrina.	El mixólogo desordenadamente agita el cóctel.
31	El payaso torpemente entretiene a la multitud.	El malabarista hábilmente actúa ante el público.	El soldado audazmente apunta su arma.
32	La tetera hierve violentamente junto al frigorífico.	La tostadora se calienta rápidamente junto al horno.	La chiquilla se queja persistentemente en la guardería.
33	El analista enfadado mira las	El investigador tristemente	El sastre atentamente altera el

	cifras.	observa a los participantes.	traje.
34	El plato fuertemente se astilla en el cajón.	La taza rápidamente se agrieta en el armario.	Las persianas completamente amortiguan el sonido.
35	El gladiador despavoridamente ataca al monstruo.	El guerrero valientemente carga hacia la bestia.	La dama frenéticamente pinta la imagen.
36	El asesino cuidadosamente identifica al objetivo.	La homicida cautelosamente elige a la víctima.	El gimnasta rápidamente sustituye los pantalones.
37	La bruja cariñosamente acaricia a su cuervo.	El mago amablemente alimenta a su lechuza.	El cerdo ansiosamente engulle la zanahoria.
38	El ladrón sigilosamente roba el reloj.	El delincuente hábilmente esconde las joyas.	La oveja ansiosamente bebe el líquido.
39	La reina poderosamente dirige al ejército.	El príncipe audazmente lucha en la batalla.	El autor rápidamente escribe el capítulo.
40	La florista delicadamente recoge las flores.	El jardinero suavemente poda los arbustos.	Los turistas alegremente divisan el meteoro.
41	La planta gradualmente crece sus raíces.	El lirio lentamente abre sus pétalos.	El perro ansiosamente roe la mesa.
42	El sobrino sonríe feliz al conejo.	El hermano señala emocionado al hámster.	El caballo galopa salvajemente por el camino.
43	El tenedor suavemente perfora la verdura.	El cuchillo cuidadosamente rebana la fruta.	El comentarista decisivamente detiene el juego.
44	La margarita gradualmente se marchita con el calor.	El tulipán lentamente muere sin agua.	La actriz bruscamente se quita la peluca.
45	El luchador confiadamente domina la pelea.	El boxeador poderosamente derrota el adversario.	El médico tranquilamente alinea el hueso.
46	La ardilla rápidamente sube al árbol.	El topo lentamente se arrastra hasta el césped.	El cirujano raramente camina de casa.
47	El niña juega feliz con los juguetes.	La madre trabaja tristemente en el ordenador.	El gusano se retuerce silenciosamente por el barro.
48	El tío cuidadosamente supervisa a los jóvenes.	El padre atentamente vigila a los adolescentes.	Las nubes continuamente flotan en el aire.
49	El coche circula tímidamente por la autopista.	La furgoneta viaja lentamente por la calle.	Los relámpagos centellean violentamente en la oscuridad.
50	El corredor esprinta rápidamente a lo largo de la pista.	El ciclista pedalea frenéticamente durante la carrera.	El lobo gruñe agresivamente al excursionista.
51	El profesor severamente regaña a los alumnos.	La niñera alegremente alaba a los niños.	El clarinete melódicamente armoniza con el pianista.
52	El buceador elegantemente	El nadador grácilmente se	El primo cruelmente abofetea

	se sumerge en la piscina.	zambulle del lago.	al hombre.
53	La mariposa revolotea elegantemente por el jardín.	La polilla aletea torpemente en el parque.	La armónica gime lúgubremente en los páramos.
54	El ganador orgullosamente sostiene la medalla.	El perdedor torpemente entrega el premio.	El águila enfadadamente grazna al amanecer.
55	El atleta progresa impresionantemente en la competencia.	El entrenador avanza poderosamente hasta el final.	El granizo impacta fuertemente contra el refugio.
56	El gorrión hábilmente embosca a la araña.	La paloma cuidadosamente atrapa al escarabajo.	El actor temerosamente filma una película.
57	El estudiante estudia diligentemente en la biblioteca.	El alumno aprende distraído en el aula.	El arroyo chapotea suavemente contra las conchas.
58	El pueblo amablemente acepta al visitante.	La comunidad bruscamente exilia al forastero.	El fontanero rápidamente corrige las tuberías.
59	El abogado cuidadosamente investiga el caso.	El jurado diligentemente escucha en el tribunal.	El paramédico temerosamente salva al cachorro.
60	La galleta sabe dulce en la boca.	El pastel huele amargo al olfato.	El conductor acelera temerariamente por la carretera.
61	La familia afanosamente limpia los dormitorios.	Los amigos perezosamente ordenan los baños.	El superhéroe amablemente protege el consejero.
62	El juez tranquilamente instruye al acusado.	El litigante audazmente aconseja al fiscal.	El arqueólogo tristemente descubre las ruinas.
63	El herrero golpea firmemente con el martillo.	El carpintero corta cuidadosamente con la sierra.	El tigre ronronea satisfecho ante la jirafa.
64	El albañil esforzadamente lleva los ladrillos.	El arquitecto fácilmente mueve la madera.	El novio obstinadamente evita la cena.
65	El director discute enfadado con el colega.	El empleado discrepa fuertemente con su asistente.	El helicóptero asciende torpemente hacia la bruma.
66	El líder habla audazmente durante su presentación.	El seguidor murmura tímidamente durante el discurso.	El yate embarca suavemente desde el muelle.
67	El huracán atraviesa caóticamente la tierra.	El tifón se precipita frenéticamente el país.	El nieto fríe alegremente en la estufa.
68	La sal casi perfecciona la ensalada.	La pimienta definitivamente mejora la sopa.	El baloncestista desdeñosamente toma el galardón.
69	El timbre suena débilmente por toda el edificio.	La alarma pita ruidosamente en la sala.	La heredera sorprendentemente engaña al gobernante.

70	El camarero accidentalmente derrama el vino.	El mixólogo desordenadamente agita el cóctel.	La hormiga enfadadamente muerde a la sobrina.
71	El caballero valientemente levanta su espada.	El soldado audazmente apunta su arma.	El malabarista hábilmente actúa ante el público.
72	El bebé llora desesperadamente desde la cuna.	La chiquilla se queja persistentemente en la guardería.	La tostadora se calienta rápidamente junto al horno.
73	La costurera distraídamente repara el vestido.	El sastre atentamente altera el traje.	El investigador tristemente observa a los participantes.
74	Las cortinas casi bloquean la vista.	Las persianas completamente amortiguan el sonido.	La taza rápidamente se agrieta en el armario.
75	El señor tranquilamente dibuja el boceto.	La dama frenéticamente pinta la imagen.	El guerrero valientemente carga hacia la bestia.
76	El golfista inmediatamente devuelve los zapatos.	El gimnasta rápidamente sustituye los pantalones.	La homicida cautelosamente elige a la víctima.
77	La vaca rápidamente devora la manzana.	El cerdo ansiosamente engulle la zanahoria.	El mago amablemente alimenta a su lechuza.
78	El burro ávidamente mastica la comida.	La oveja ansiosamente bebe el líquido.	El delincuente hábilmente esconde las joyas.
79	El poeta lentamente lee las palabras.	El autor rápidamente escribe el capítulo.	El príncipe audazmente lucha en la batalla.
80	Los aventureros entusiasmados observan el cometa.	Los turistas alegremente divisan el meteoro.	El jardinero suavemente poda los arbustos.

B.3

Description of Intrusion Errors Model

The Intrusion Model included all intrusion data regardless of whether the target keyword was transcribed, whereas the Intrusion Error Model only included data when an error was made (i.e., when the equivalent target word was not correctly transcribed). For example, if the target sentence was *The <u>dancer</u> <u>carefully</u> <u>plaits</u> her <u>hair</u>* and the participant typed *The <u>actress</u> <u>carefully</u> … her <u>hair</u>* (the word *actress* is a masker keyword) then for the Target Accuracy Model, the participant would score 0.50 (for typing *carefully* and *hair*); for the Intrusion Model the participant would score 0.25 (of four possible masker keywords, they only typed one: *actress*). For Intrusion Error Model, the participants were scored by keyword (i.e., each masker keyword is either transcribed (1) or not (0)), rather than by trial, therefore, the participant made two errors (they did not type *dancer* or *plaits*), so would score 1 for typing *actress* (a masker keyword), and 0 for not providing the masker keyword which corresponded with *plaits*. The data from the two correctly transcribed keywords (i.e., the typed target words: *carefully* and *hair*) would not be included in the Intrusion Error Model. The exploratory models used the same fixed and random effects structures as the Target Accuracy Model. Due to a lack of convergence, both exploratory models were run without intercept/slope correlations.

B.4

Table 10 - Experiment 1: GLMM output for Intrusion Errors Model

Fixed Effects	**Estimate**	**Std. Error**	**Z value**	**Pr(>|z|)**
(Intercept)	-2.40	0.12	-19.33	< .001***
Masker semantic content	1.28	0.20	6.33	< .001***
Masker language	4.00	0.21	19.04	< .001***
Trial number	-0.05	0.08	-0.59	.557
Masker semantic content: Masker language	-0.79	0.37	-2.13	.033*
Masker semantic content: Trial number	-0.02	0.16	-0.15	.879
Masker language:Trial number	0.18	0.16	1.12	.263
Masker semantic content:Masker language:Trial number	0.27	0.32	0.84	.403

Significance codes: *** < .001; ** < .010; * < .050

B.5

Table 11 - Experiment 2: L1-Spanish background for participants in the Recognition Task (N = 115)

Language background measure	**SC (N = 53)**				**N (N = 62)**			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Age of acquisition (years)	0.00	0.00	0.00	0.00	0.02	0.13	0.00	1.00
LexTale proficiency (%)[a]	92.50	7.27	63.33	100.00	90.20	9.43	60.00	100.00
Self-reported understanding (/10)	9.92	0.33	8.00	10.00	9.92	0.27	9.00	10.00
Self-reported speaking (/10)	9.94	0.23	9.00	10.00	9.92	0.33	8.00	10.00
Self-reported reading (/10)	9.96	0.19	9.00	10.00	9.98	0.13	9.00	10.00
Self-reported writing (/10)	9.91	0.35	8.00	10.00	9.76	0.59	8.00	10.00
Language frequency of use (/4)[b]	2.75	0.74	1.21	4.00	2.93	0.74	0.94	4.00

[a] The L1-Spanish values exclude the data from one participant who scored 9.17% in the Lextale-Esp (see section 3.3.3.2. Materials, LexTALE and Lextale-Esp). After completing the experiment, they stated that they mixed up the response buttons.

[b] This value is from the language background questionnaire (see section 3.3.3.2. Materials, Language background questionnaire) and determines how frequently participants use each language in their daily lives (0 = All another language, 1 = Mostly another language, 2 = Half each language, 3 = Mostly Spanish, 4 = All Spanish).

B.6

Table 12 - Experiment 2 full GLMM output for Intrusion Errors Model

Fixed Effects	**Estimate**	**Std. Error**	**Z value**	**Pr(>|z|)**
(Intercept)	-0.21	0.09	-2.23	.026*
Masker semantic content	0.88	0.13	6.80	< .001***
Trial number	0.09	0.02	4.69	< .001***
Masker semantic content: Trial number	-0.06	0.04	-1.49	.136

Significance codes: *** < .001; ** < .010; * < .050

B.7

Table 13 - GLMM output for Experiment 2 Recognition Model without participants with d' less than 0.5

Fixed Effects	Estimate	Std. Error	Z Value	Pr(>|z|)
(Intercept)	-0.30	0.11	-2.77	.006**
Masker semantic content	0.76	0.19	3.93	< .001***

Significance codes: *** < .001; ** < .010; * < .050

B.8

Table 14 - GLMM output for Experiment 2 Recognition Transcription Model without participants with d' less than 0.5

Fixed Effects	Estimate	Std. Error	Z Value	Pr(>|z|)
(Intercept)	0.40	0.10	3.82	< .001***
Masker semantic content	0.27	0.19	1.37	.170
Transcription status	2.86	0.15	18.50	< .001***
Masker semantic content:Transcription status	-0.40	0.30	-1.33	.184

Significance codes: *** < .001; ** < .010; * < .050