**Selective and Divided Listening in a Dual-Language Context**

Emily Allen-Rice1, Sarah Knight2, Angela de Bruin1, and Sven Mattys1

1 Department of Psychology, University of York

2 School of Psychology, Newcastle University

**Author Note**

For the purpose of open access, a Creative Commons Attribution (CC BY) licence is applied to any Author Accepted Manuscript version arising from this submission.

# **Abstract**

Background speech can reduce intelligibility of target speech because of masking effects. However, masking can be mitigated by presenting target and masker dichotically, an effect we refer to as the dichotic advantage. Because masking, and signal degradation in general, is often shown to negatively impact second language (L2) more than first language (L1) processing, we hypothesised that L2 might accrue a greater dichotic advantage than L1 when listener must selectively attend to the target and ignore the masker. Under divided attention, however, dichotic presentation might introduce cognitive costs due to the need for binaural 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 dichotic advantage. Using a dual-language context, Spanish (L1)-English (L2) bilinguals heard one English sentence and one Spanish sentence simultaneously, either diotically or dichotically. They 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. The dichotic advantage (better performance in dichotic than diotic listening) was similar for L1 and L2. It was smaller under divided than selective attention, suggesting that increased the cognitive costs incurred by divided listening reduced the dichotic advantage. The results demonstrate that bilinguals experience a dichotic advantage of a similar size in each language (L1 and L2), even in dual-language contexts and under high cognitive load.

*Keywords: Masking, cognitive load, dual-language context, bilingual language processing, selective attention, divided attention*

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), and in part to informational masking (IM), which includes conditions in which factors beyond EM interfere with target intelligibility (e.g., Pollack, 1975). However, masking is reduced when sound sources are presented dichotically (i.e., one sound in each ear). This is because when either the target or masker is in a given ear, the listener can distinguish the competing streams based on location and interaural cues (e.g., level and timing differences). In this study, we refer to the improved intelligibility resulting from dichotic presentation as the dichotic advantage.

The dichotic advantage can be seen as an extreme example of what is known as spatial release from masking (SRM), which is an improvement in intelligibility when target and masker are separated in space rather than collocated. In 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 benefit from SRM both when they know in advance which speaker to focus on (selective listening, Ihlefeld & Shinn-Cunningham, 2008a) and when they have to 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. For bilingual listeners, masking appears to negatively impact second language (L2) processing more than first language (L1) processing (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 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 word and sentence identification 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, the dichotic advantage (i.e., a reduction in masking) could be expected to benefit L2 more than L1. Only two SRM studies (Ezzatian et al., 2010; Hui et al., 2021) 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) tested L1 and L2 speakers of English. 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. 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 sentences and the noise were collocated (0°; running counter to research demonstrating that masking impacts L2 more than L1), L1 listeners experienced a larger SRM than did L2 listeners. The authors interpreted their results as showing that L1 listeners were better able to make use of binaural cues, such as interaural level and timing differences, to improve their listening accuracy, whereas L2 listeners failed to make use of these cues. These results were supported by a later study (Hui et al., 2022) that used babble masker speech; however, this study involved a comparison of early- and late-immersed English speakers, rather than a comparison between L1 and L2 processing.

However, Ezzatian et al. (2010) did not find an SRM difference between L1 and L2 speakers. In their study, the L2 speakers of English 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 around the listener. Participants were asked to repeat back the target sentences. Both listener groups benefited equally from spatial separation, regardless of masker type.

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) or the task (transcription in Hui et al.; immediate verbal recall in Ezzatian et al.). Additionally, both studies used a between-subjects design, with (likely) monolingual participants in the L1 condition and bilingual participants in the L2 condition. Even when listening to their L1 (English), bilinguals have been shown to use different mechanisms than monolinguals (Blumenfeld & Marian, 2011). Furthermore, between-subjects designs have the disadvantage of allowing variables other than language (L1/L2) to influence the outcome (e.g., immigration status, number of languages known, country of origin, etc.). Therefore, a comparison of language effects within bilinguals is needed.

So far, the focus has been on how release from masking through spatial separation (or dichotic presentation) 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 their 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, L2 processing 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 (or a dichotic advantage) can theoretically be found during both selective and divided listening, cognitive load and the need for attentional control are thought to be higher during divided listening (Treisman, 1971). Indeed, if a listener is attempting to listen to two dichotic talkers, shifting attention between the two ears 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. Listening accuracy increased as the two talkers ranged from diotic to dichotic, reflecting a dichotic advantage. However, when EM was eliminated by playing the two voices in non-overlapping frequency bands, listening accuracy was worse, not better, in the dichotic condition. Thus, when dichotic presentation is not associated with a reduction in masking, processing costs are apparent. This decrease in performance only occurred when the two talkers were presented dichotically. It did not occur at intermediate levels between diotic and dichotic. Therefore, in the current experiment, only diotic and dichotic audio presentations were used. Knight et al. also used the Letter Number Sequencing (LNS) task to measure working memory (WM) capacity and found that LNS performance was positively correlated with listening accuracy in the condition where EM was eliminated. This provides further evidence that attentional control during divided listening is cognitively costly.

Divided listening has been studied primarily in monolingual listeners. However, the potential interplay between the dichotic advantage and attentional costs might differ between L1 and L2 processing. The dichotic advantage 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 dichotic presentation may negatively impact L2 processing more than L1 processing due to the 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 masking alone (and its release through dichotic presentation; Experiment 1) before introducing the cognitive load component via a divided-attention condition (Experiment 2).

# **Experiment 1: Selective Attention**

## **Pre-registration and ethics**

The pre-registrations for Experiment 1 (<https://osf.io/sdj2g>) and Experiment 2 (<https://osf.io/ytbj2>) are available on the Open Science Framework (OSF). Ethical approval for both experiments was granted by the ethics committee at the Psychology Department 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.

## **Hypotheses**

The goal of Experiment 1 was to test whether the magnitude of the dichotic advantage 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 (both sentences played in both ears at the same intensity) or dichotically (one sentence played 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 diotic (i.e., main effect of Audio Presentation, a dichotic advantage). This is based on previous research demonstrating a dichotic advantage in monolingual listeners (e.g., Marrone et al., 2008).

H3: The dichotic advantage will be larger for L2 than L1 listening (i.e., interaction between Language and Audio Presentation). This hypothesis aligns with a substantial body of research showing that masking is more detrimental to L2 than L1 listening when the task involves word or 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 dichotic presentation (and thus the reduction of masking) than L1.

## **Methods**

### ***Participants***

Participants were recruited through Prolific ([www.prolific.co](http://www.prolific.co); Prolific, 2014) and tested online using Gorilla Experiment Builder ([www.gorilla.sc](http://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-Diotic, L1-Dichotic, L2-Diotic, 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 *LexTALE and Lextale-Esp*); reported they did not look up words during the tests. 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)*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 *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 *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.

### ***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: *(((Number of words correct/number of words in total)\*100) + ((number of non-words correct/number of nonwords in total)\*100))/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 either diotically or dichotically. 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 *Analyses*).

The 168 English sentences were taken from the IEEE corpus (Rothauser, 1969). An example IEEE sentence is *The jacket hung on the back of the wide chair* (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. 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 60 dB SPL 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) were generated for each sentence pair to create the diotic 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 audio presentation (diotic or dichotic).

### ***Procedure***

Participants first completed a screening session, which consisted of the Language Background Questionnaire and LexTALEs. Some participants also completed the LNS task in the screening session (N = 19), while others completed it after the selective attention task (N = 79)[[1]](#footnote-1). 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 represented one condition of the selective attention task (L1-Diotic, L1-Dichotic, L2-Diotic, and L2-Dichotic). Participants were told at the beginning of each block whether the sentences would be diotic 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 English written 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%).

## **Analyses**

Participants were excluded from analysis if they: scored fewer than 3 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. 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 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, Audio Presentation, 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), Audio Presentation (diotic and dichotic), and their interaction were included as fixed effects. Both effects were sum coded (0.5 versus -0.5 for dichotic versus diotic and L1-Spanish versus L2-English). The model used the BOBYQA optimiser (Powell, 2009) and a maximum of 109 iterations. The full Hypothesis-Testing Model was as follows:

glmer(score ~ Language\*AudioPresentation + (1 + Language\*AudioPresentation|participant) + (1|item), family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 1e9)))

The data were also analysed using exploratory models to investigate individual differences in listening 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.

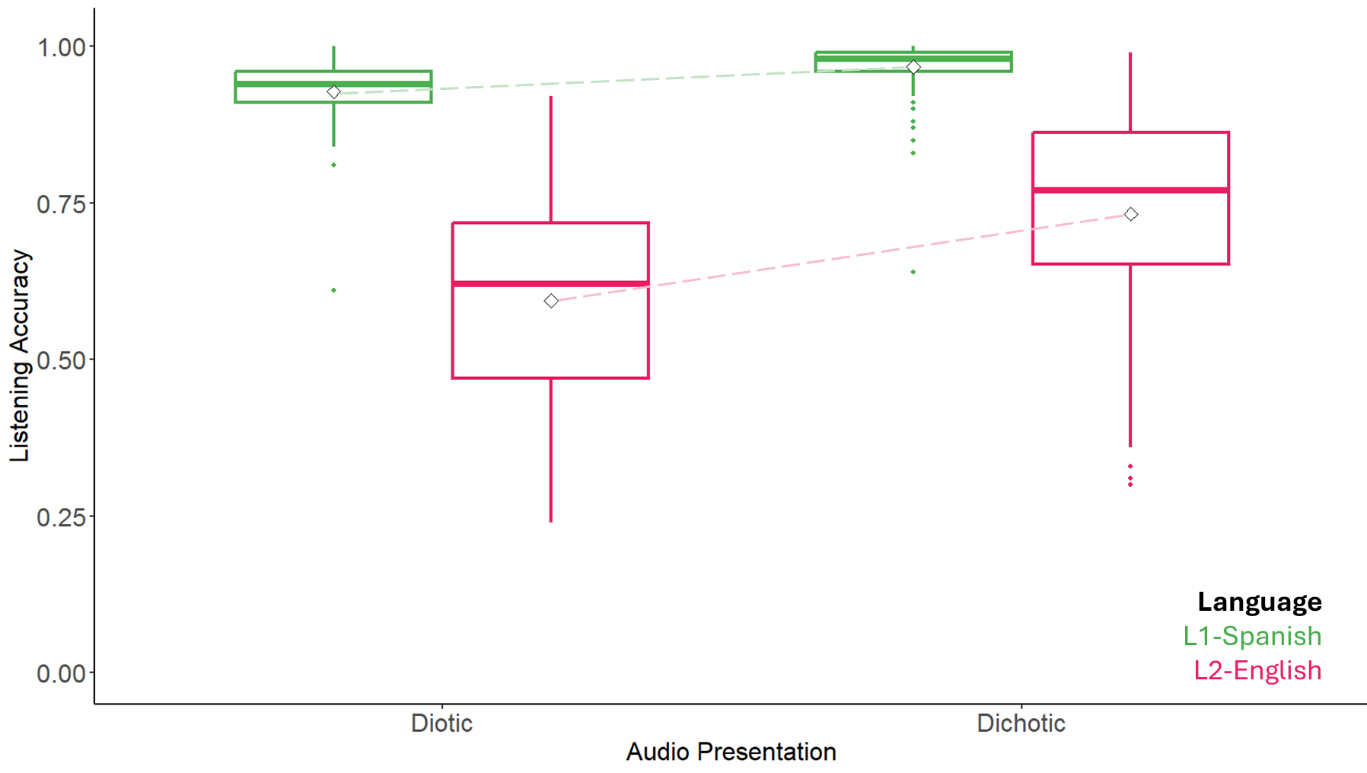
The Language History Model used only L2 performance as the outcome variable, and included L2-English Age of Acquisition, L2-English Proficiency (LexTALE score), and Frequency of Language Use (LSBQ score) as fixed effects to investigate the potential role of individual differences in L2 background. These variables were allowed to interact with Audio Presentation 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.

## **Results**

The results from the selective attention task are displayed in Figure 1. In the Hypotheses-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 Audio Presentation was also found (β = 0.93, SE = 0.15, z = 6.20, *p* < .001), with performance for diotic 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 Audio Presentation (β = 0.28, SE = 0.30, z = 0.92, *p* = .357). This indicates that neither language showed a significantly larger dichotic advantage (average dichotic advantage scores = 0.04 (SD = 0.05) for L1-Spanish; 0.14 (SD = 0.10) for L2-English).

**Figure 1**

*Average listening accuracy for each Language and Audio Presentation level (Experiment 1)*



*Note.* Listening accuracy is the average proportion of keywords correctly transcribed within one condition (L1-Diotic, L1-Dichotic, L2-Diotic, 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 Audio Presentation, English Proficiency, and Frequency of Language Use as fixed effects, and the interaction between Audio Presentation and English Proficiency.

This model showed that, in addition to a main effect of Audio Presentation (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 listening accuracy. A more frequent use of L2-English was associated with higher L2-English listening accuracy. There was also a significant interaction between Audio Presentation 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 Audio Presentation conditions. However, the effect was weaker in the diotic (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 listening accuracy more when the L2-English target and masker were presented dichotically than diotically. Therefore, those with a higher L2-English proficiency demonstrated a larger dichotic advantage 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 Audio Presentation (β = 0.93, SE = 0.15, z = 6.20, *p* < .001) on listening accuracy, but no interaction between Language and Audio Presentation (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 Audio Presentation (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 listening 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 Audio Presentation (β = -0.06, SE = 0.05, z = -1.32, *p* = .188).

## **Discussion**

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

As research has demonstrated that L2 is more negatively impacted by masking than L1 (e.g., Cooke et al., 2008), we expected a larger dichotic advantage for L2 than L1 processing. Contrary to this hypothesis, there was no statistically significant interaction between Language and Audio Presentation. This finding is in line with Ezzatian et al. (2010), who also showed no difference in spatial release from masking for L1 and L2 processing when comparing separate groups of L1 and L2 speakers in single-language two-talker contexts. This suggests that reducing masking through dichotic separation might be qualitatively different from reducing masking through other methods (e.g., reducing the signal-to-noise ratio of a diotic pair of talkers).

As mentioned earlier, research indicates that the L2 disadvantage in noise is not observed when the task consists of identifying phonemes, rather than engaging in word or sentence identification (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 a different impact of dichotic presentation 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 listening accuracy. Similarly, using L2 more frequently was associated with improved overall L2 listening 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 dichotic advantage (with no such interaction observed for L2 use). If a less proficient language is affected more by masking and therefore shows a larger dichotic advantage, 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 listening 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 dichotic presentation (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 dichotic advantage for their L1 than L2 – as was found by Hui et al..

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 the dichotic advantage. 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. 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 demonstrated a similar dichotic advantage. As explained in the Introduction, a selective-attention design was used in Experiment 1 in order to investigate the dichotic advantage in L1 and L2 with a single target voice. However, when listeners must attend to two dichotically presented talkers (during divided rather than selective listening), they must use cognitive control processes to shift their attention between the two ears. Thus, dichotic advantage 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.

# **Experiment 2: Divided Attention**

## **Hypotheses**

The goal of Experiment 2 was to investigate whether the dichotic advantage has a different impact on L1 and L2 processing during divided-attention listening, when dichotic presentation is typically 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 diotic (i.e., a main effect of Audio Presentation).

H3 provides two competing hypotheses regarding the nature of a possible interaction between Language and Audio Presentation:

H3 (i): The dichotic advantage will be larger for L2 than L1 listening, because masking 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, the magnitude of the dichotic advantage 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 ears – a process which requires increased attentional control (Treisman, 1971) and may draw on WM resources (Knight et al., 2023). We therefore hypothesised that higher listening performance in the dichotic condition would be associated with higher LNS scores, more so than in the diotic condition (i.e., an interaction between Audio Presentation 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, Audio Presentation, and LNS Score).

Comparing Experiments 1 and 2, we hypothesised that listening 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, the dichotic advantage was expected to be smaller in Experiment 2 than Experiment 1 because of the larger cognitive cost for dichotic as opposed to diotic listening when tracking two talkers rather than one (i.e., an interaction of Audio Presentation and Task).

## **Methods**

### ***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 the supplementary materials for further participant information.

### ***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 either diotically *or* dichotically across all participants (as in Experiment 1), each sentence pair was presented diotically for half of the participants and dichotically 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 either diotically or dichotically. 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 (diotic 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 *Experiment 1: Selective Attention, 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%).

## **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 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 Hypotheses-Testing Model included Audio Presentation as a by-item random slope (since each item was presented diotically for half of the participants and dichotically 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(score ~ Language\*AudioPresentation\*LNS + (1 + Language\*AudioPresentation|participant) + (1 + AudioPresentation|item), family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”, optCtrl = 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). 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 109 interactions. The Comparison Model was as follows:

glmer(score ~ Language\*AudioPresentation\*Task + (1 + Language\*AudioPresentation\*Task||participant) + (1 + AudioPresentation + Task||item), family = binomial(link = “logit”), glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 1e9)))

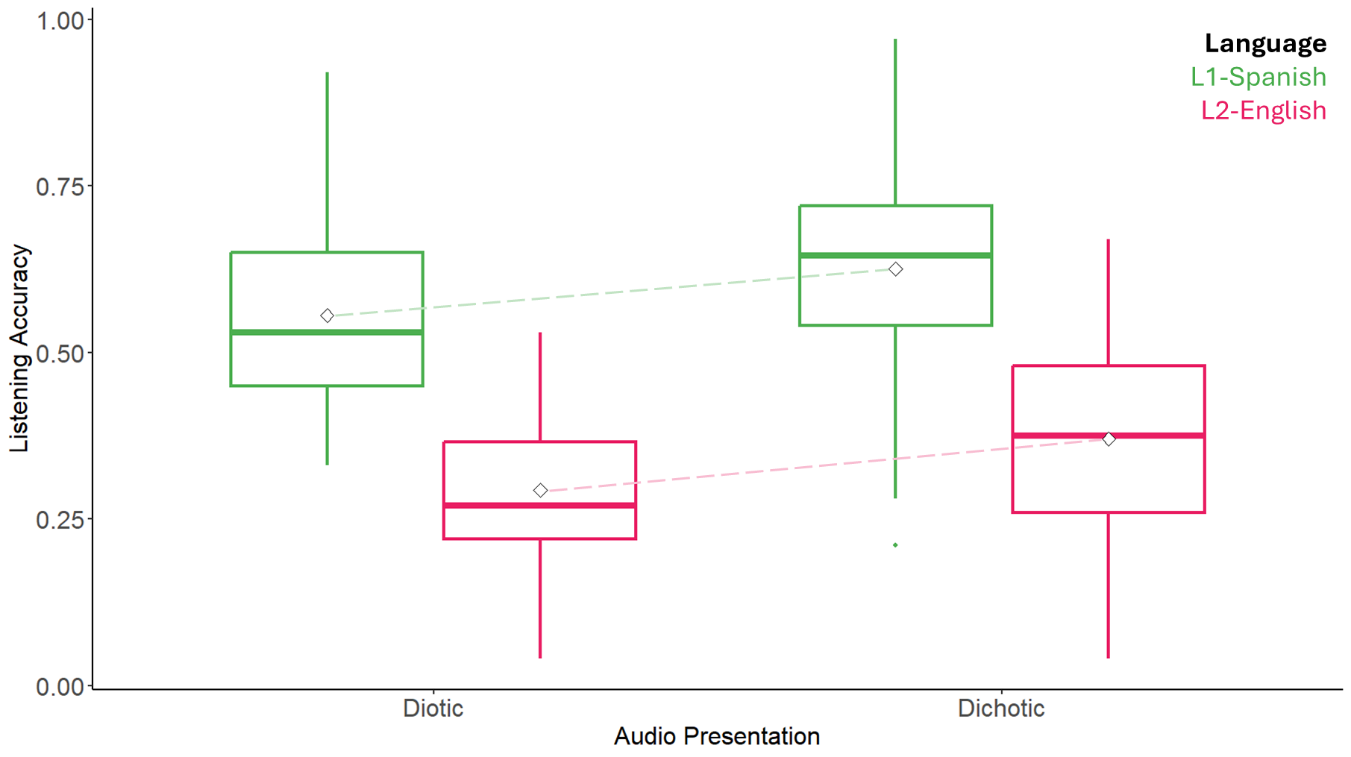
## **Results**

In the Hypotheses-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 Audio Presentation was also found (β = 0.35, SE = 0.06, z = 5.37, *p* < .001), with performance for diotic 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 Audio Presentation (β = -0.05, SE = 0.14, z = -0.32, *p* = .747). This indicates that dichotic presentation was beneficial for listening in both languages, but not one significantly more than the other (average dichotic advantage 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 listening accuracy. However, there was no significant interaction between Language and LNS Score (β = -0.02, SE = 0.10, z = -0.23, *p* = .815), Audio Presentation and LNS Score (β = 0.34, SE = 0.04, z = 0.80, *p* = .424), nor a three-way interaction between LNS Score, Audio Presentation, and Language (β = -0.07, SE = 0.11, z = -0.68, *p* = .495)[[2]](#footnote-2). Figure 3 depicts the relationship between LNS Score and listening accuracy.

**Figure 2**

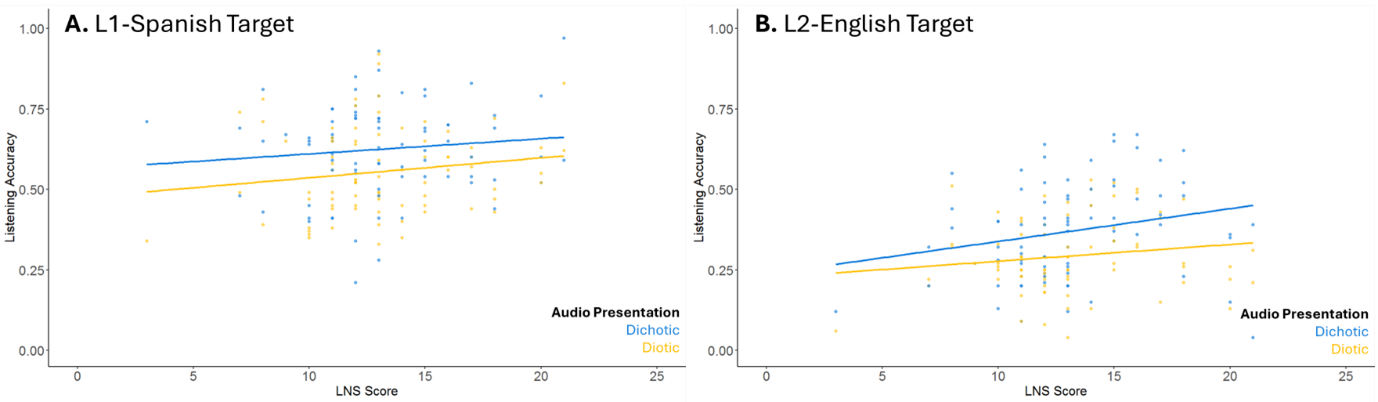
*Average listening accuracy for each Language and Audio Presentation level (Experiment 2)*



*Note.* Listening accuracy is the average proportion of keywords correctly transcribed within one condition (L1-Diotic, L1-Dichotic, L2-Diotic, 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.

**Figure 3**

*Relationship between WM (measured by the LNS) and listening accuracy in Experiment 2*



*Note.* Listening accuracy and LNS Score (using the LNS described in Experiment 2) in the divided attention task. Panel A is listening accuracy for L1-Spanish target, panel B is listening 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 Audio Presentation: β = 0.64, SE = 0.08, z = 8.18, *p* < .001; no interaction between Language and Audio Presentation: 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, Audio Presentation, and Task (β = -0.47, SD = 0.28, z = -1.69, *p* = .091), suggesting that, for both tasks, the dichotic advantage was 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 = 057, 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 Audio presentation 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 lower than in the diotic 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 diotic 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 diotic stimuli (SE = 0.10, z = 4.74, *p* < .001). This demonstrated that the dichotic advantage was larger in the selective attention task than in the divided attention task (average dichotic advantage scores = 0.09 (SD = 0.06) for selective attention; 0.07 (SD = 0.08) for divided attention).

## **Discussion**

Experiment 2 examined whether the magnitude of the dichotic advantage 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. Listening accuracy was higher for L1 than L2, and when the sentences were dichotic compared to diotic. However, contrary to our hypotheses, the dichotic advantage was comparable in L1 and L2.

The presence of a dichotic advantage in divided listening with bilinguals is consistent with research on monolinguals (e.g., Knight et al., 2023) demonstrating that the dichotic advantage outweighs the cognitive costs associated with dividing attention between ears. This outweighing of cognitive costs by dichotic advantage applied equally to L1 and L2. It is possible that both hypothesised interactions were true, but cancelled each other out: that is, that L2 showed a larger dichotic advantage than L1 while also being more negatively impacted by the increased cognitive demands associated with shifting attention between ears. However, in combination with results from Experiment 1, which indicated that L1 and L2 were not differentially affected by the dichotic advantage, it is more parsimonious to conclude that the magnitude of the dichotic advantage did not fundamentally differ for L1 and L2 processing 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 listening 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 listening accuracy, it was not associated with the magnitude of the dichotic advantage. The implications of this result are discussed further in the General Discussion.

# **General Discussion**

Listening in noisy environments can be challenging, partly due to masking between target and masker. Masking can be reduced by presenting the target and masker dichotically (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 experience a dichotic advantage for both L1 and L2, illustrating release from masking; furthermore, this is true whether they must track only one language (selective listening) or both (divided listening). The presence of a dichotic advantage in the divided listening condition suggests that the benefit of masking release through dichotic presentation can outweigh any costs associated with maintaining attention between the two ears (e.g., Knight et al., 2023).

Although both experiments showed a dichotic advantage, the magnitude of this effect was smaller in Experiment 2 (divided attention) than in Experiment 1 (selective attention). This suggests that the benefit of dichotic presentation was attenuated when listeners were asked to track both speakers at the same time. This result is inconsistent with that of Pinto et al. (2020), who found that the benefit of spatially separating talkers was similar in their selective and distributed attention conditions. However, the discrepancy between their results and ours may be due to the different ways in which spatial separation was implemented. Pinto et al.’s stimuli were separated by 160॰ using head-related transfer functions, whereas ours were played dichotically, allowing for a maximal impact of speaker separation. Additionally, Pinto et al.’s task involved target word detection, rather than entire sentence transcription; this may lead to dichotic advantage 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 masking is controlled, dichotic presentation negatively impacts listening accuracy when attending to two stimuli simultaneously. Both Knight et al.’s study and the present one demonstrate that there are costs associated with dichotic presentation during divided attention, even if these are outweighed by the dichotic advantage.

One possible source of costs associated with dichotic presentation during divided-attention listening are the cognitive demands of shifting attention between two ears. 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 auditory stimuli, as illustrated by Best et al. (2006). In Best et al.’s study, stimuli presented close 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 back and forth between two ears was cognitively demanding. Together with our study, these results provide evidence that dichotic presentation is associated with a processing cost due to shifts of 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 listening performance. WM capacity was associated with listening accuracy when attending to both talkers simultaneously (Experiment 2), but not when attending to only one talker (Experiment 1). It is likely that the relationship between LNS score and listening performance 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 the dichotic advantage in bilingual listeners in a dual-language context. We showed that both L1 and L2 demonstrate a dichotic advantage, 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. 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 the lack of interaction between language and WM capacity in our study. However, testing Korean-English bilinguals, Kim et al. (2022) showed that WM is only associated with L2 listening when the stimuli are long passages, rather than short sentences, as they were in this study. Indeed, Fay and Buchweitz’s 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. 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 masking than L1 (e.g., Cooke et al., 2008) and is more cognitively demanding to process (e.g., Spivey & Marian, 1999). We therefore hypothesised that dichotic presentation would impact L1 and L2 differently. We expected that, during divided attention, L2 processing would be more negatively affected by cognitive costs associated with dichotic presentation. However, we found no evidence that L1 and L2 were differently impacted by playing stimuli diotically versus dichotically, indicating that bilinguals use dichotic cues to improve listening to both languages, even in cases where dichotic presentation should be associated with cognitive costs. Although there is a limited amount of literature investigating how stimulus separation 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 show a dichotic advantage, even when cognitive demands are increased during a more demanding divided-listening task.

When considering language differences (regardless of whether the stimuli are played diotically or dichotically), 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 in Experiment 2, 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 focusing more on L2 than L1 (note that this would only apply to the divided listening task because listeners always knew in advance which voice to report during the selective listening task). However, listeners may overcompensate when using this tactic, focusing too much on their L2 and thus unintentionally 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 the role of individual 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 the dichotic advantage: if the Audio Presentation 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 dichotic presentation improves listening performance, even when attention is divided between two talkers. However, the dichotic advantage is smaller during divided attention listening due to the cognitive costs associated with shifting attention between ears. Contrary to expectations, these costs do not differentially affect L1 and L2 performance, suggesting that bilinguals use dichotic cues 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.

# **Statements and Declarations**

## **Ethical Considerations**

Ethical approval for both experiments was granted by the ethics committee at the Psychology Department at 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.

## **Consent to Participate**

All participants were informed of the aims of the study, able to ask questions if they did not understand, and voluntarily provided their consent to participate.

## **Consent for Publication**

Participants were informed that the research that they were participating in, and their anonymous data may be disseminated publicly in publications and on online platforms such as the Open Science Framework, and voluntarily consented to this.

## **Declaration of Conflicting Interest**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding Statement**

The authors received financial support for the research through a studentship from the Department of Psychology, University of York. Participant compensation was supported in part by a grant from the Economic and Social Research Council (ESRC) to S. Mattys (ES/W010488/1).

## **Data Availability**

The data set and R script for the analyses of this study are available on the OSF at <https://osf.io/sdj2g> for Experiment 1 and <https://osf.io/ytbj2> for Experiment 2.

# **References**

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*, 250-263. https://doi.org/10.3758/s13428-017-0867-9

Anywl-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*, 388-407. <https://doi.org/10.3758/s13428-019-01237-x>

Audacity Team (2014). Audacity®: Free Audio Editor and Recorder [Computer program]. Version 3.7.1

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>

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*, 217-218. https://doi.org/10.3758/BF03328599

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package ‘lme4’. *Convergence, 12*(2)

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>

Black, J. W. & Hast, M. H. (1962). Speech reception with altering signal, *Journal of Speech and Hearing Research,* 5(1), 70-75.

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, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.03. <http://www.praat.org/>

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>

Brysbaert, M. & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 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*, 1-19. <https://doi.org/10.1177/23312165231205107>

Cameron, S., Dillon, H., & Newall, P. (2006). Development and evaluation of the listening in spatialized noise test, *Ear and Hearing,* 27(1), 30-42. <https://doi.org/10.1097/01.aud.0000194510.57677.03>

Cooke, M., García 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>

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>

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

Dijkstra, T. & van Heuven, W. J. (2012). Word recognition in the bilingual brain. In: M. Faust (Eds.), *The handbook of the neuropsychology of language.* (pp. 449-471). Blackwell Publishing Ltd.

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.

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>

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

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

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), 244\5-2454. <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>

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>

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of cognitive psychology*, *25*(5), 515-530.

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>

Heine, L., Cornevllie, A., Gorbert, F., Luauté, J., Lavandier, M., Perrin, F. (2021). Virtually spatialized sounds enhance auditory processing in healthy participants and patients with a disorder of consciousness, *Scientific Reports, 11,* 13702. <https://doi.org/10.1038/s41598-021-93151-6>

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*. 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>

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, 35*(1), 49-66.

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.* <https://doi.org/10.1016/j.jml.2023.104427>

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>

Marrone, N., Mason, C. R., & Kidd, G. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America, 124*(2), 1146-1158. <https://doi.org/10.1121/1.2945710>

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.

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>

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.

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, 253*(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.* Retrieved from <https://office.microsoft.com/powerpoint>

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>

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch, *Behavior Research Methods,* 53, 1551-1562. <https://doi.org/103758/s13428-020-01514-0>

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.

Pinto, D., Agmon, G., & Zion Golumbic, E. (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*(S1): S5-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](https://psycnet.apa.org/doi/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.* Retrieved from https://www.prolific.com

Rehebergen, 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

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*(3), 225-246.

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>

The MathWorks Inc., 2022

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

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 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>

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

Wightman, F. L. & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation, *The Journal of the Acoustical Society of America,* 85(2), 868-878. <https://doi.org/10.1121/1.397558>

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

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), EL 129-EL 135. https://doi.org/10.1121/1.5087271

1. 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. [↑](#footnote-ref-1)
2. When the LNS Model for Experiment 1 was rerun using the Experiment 2 LNS scores (N = 66), LNS Score was no longer associated with listening accuracy (See Table S3 for full GLMM output). [↑](#footnote-ref-2)