The Time-Course of Linguistic Interference on Measures of Speech-in-Noise Perception and Listening Effort

Alex Mepham

Doctor of Philosophy

University of York

Department of Psychology

February 2025

ABSTRACT

Speech perception is a dynamic process whereby listeners must attune to a target talker. In adverse listening conditions, listeners must overcome additional challenges associated with energetic and informational masking from competing sound sources. Linguistic interference, a type of informational masking resulting from the linguistic content of competing speech, is the main topic of this thesis. Many studies of linguistic interference assess how different types of competing speech impact speech perception, without considering a listener's ability to adapt to adverse listening conditions over time. This thesis investigates the change in linguistic interference over time in monolingual and multilingual listeners. The first four experiments investigate how knowledge of the linguistic content of competing speech impacts performance and how the patterns of linguistic interference depend on whether listening takes place in native and non-native languages. The last three experiments look specifically at monolingual listeners and investigate not only how speech-in-noise performance changes over time, but also how physiological and subjective measures of listening effort and fatigue change as listeners adapt to adverse listening conditions. Pupillometric measures of listening effort are used to detect changes in physiological state potentially outside of conscious awareness. Effort is measured while listening to a talker in intelligible and unintelligible competing speech and are compared to speech-in-noise performance, and subjective self-reports of listening effort and fatigue. The results from these experiments demonstrate general listener abilities to improve in speech-in-noise perception, and most experiments demonstrate similar rates of improvement over time for both intelligible and unintelligible speech maskers. Task-evoked pupil responses also generally decreased over time. These experiments demonstrate how speech intelligibility, SNR and experimental designs can significantly affect pupillometric measures, and how they can inform empirical and theoretical considerations when measuring effort, fatigue and motivation in speech-in-noise perception.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Sven Mattys for the ongoing support, guidance and mentorship over the course of this PhD programme. Without his continued support, I would not have been able to complete this programme and deliver this thesis. His expertise has been invaluable in the preparation of this thesis, as well as reassurance through practical and moral support even during very difficult periods over my time in the lab. And thanks for hosting the numerous summer and winter parties, reminding me that the process (and my life) is to be enjoyed. Thank you for being an incredible supervisor.

I would also like to thank the current and former members of the Speech Lab during the time I completed my PhD thesis: Yifei Bi, Faith Chiu, Lotte Eijk, Sarah Knight, Georgie Maher, Ronan McGarrigle, Sophie Meekings, Lyndon Rakusen, Emily Rice, and Rachel Yue Zheng. In particular, I would like to thank Sarah Knight for the support, guidance and encouragement over the years I completed the PhD programme. I also would like to thank Angela de Bruin, Gareth Gaskell, and Emma Haiyou-Thomas for input and guidance as part of my thesis advisory panel.

Without the specialist support of Amanda Hall, Jennifer Haviland, and Georgina Yandle, I would also not be in the position I am now, having completed this mammoth task.

I would like to thank the dear friends who over the years of completing my PhD thesis have provided both encouragement and distraction. Thanks to Jonathan and Rob. Thanks to Charlotte and Chloe. Thanks to Liz, my first-reader. Thanks to Alex, Alison, Amy, Anna, Anni, Anthony, Antony, Becca, Ben, Char, Chris, Diane, Dom, Em, Eve, Fran, Hanne, Jasmin, Jennifer, Juliana, Katy, Kit, Lindsay, Lucy, Lydia, Mari, Martin, Matt, Miaomiao, Mirela, Nick, Nicky, Niklas, Okka, Phyllida, Rana, Renée, Robert, Roo, Sam, Sharon, Sofia, Tom, Ursula, Victoria, Ying, and Yujin.

Thank you to Felice, for your generous, insightful and encouraging letters. Thank you for your thoughtful gifts, for the many occasions you have hosted me. Thank you for your support during stressful moments. Thank you for your friendship.

Thank you to my family, who have been of immense support, and who I know are always there for me.

Thank you to Louisa, whose unwavering love has remained with me through every high and low, both during this PhD programme and throughout my life. Thank you, as without you, I would not be here.

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 this University or elsewhere. All experiments were designed by the candidate with assistance from the supervisor Professor Sven L. Mattys, with guidance from Professor Gareth Gaskell and Dr Angela de Bruin on the candidate's thesis advisory panel, and with guidance and assistance from current and former research staff and doctoral students at the University of York, whose guidance and assistance are as follows. Dr Yifei Bi and Dr Lydia Y. Li assisted with the translation of experimental stimuli from English to Mandarin Chinese in Chapter 2. Dr Yifei Bi assisted with the testing of multilingual listeners in Chapter 2. Lyndon Rakusen helped with automating performance scoring in Chapter 2, and the experimental programming in Chapter 3. Dr Sarah Knight and Dr Ronan A. McGarrigle advised on the statistical analyses in Chapter 3. All other testing and final statistical analyses were conducted by the candidate. All sources are acknowledged as references.

Material in Chapter 2 was published in Mepham, A., Bi, Y., & Mattys, S. L. (2022). The time-course of linguistic interference during native and non-native speech-in-speech listening. *J. Acoust. Soc. Am.*, *152*(2), 954-969. https://doi.org/10.1121/10.0013417. Supplementary material for Chapter 2 can be found at https://osf.io/nhcrw/.

At the time of submission, material in Chapter 3 was published in Mepham, A., Knight, S., McGarrigle, R. A., Rakusen, L., & Mattys, L. M. (2025). Pupillometry reveals the role of SNR in adaption to linguistic interference over time. *J. Speech Lang. Hear. Res.*, *68*(5), 2291-2317. https://doi.org/10.1044/2025_JSLHR-24-00658. Supplementary material for Chapter 3 can be found for Experiment 5 at https://osf.io/m4b57/, for Experiment 6 at https://osf.io/pmhsx/, and at Experiment 7 at https://osf.io/6hkp9/.

The research was supported by a research grant from the Leverhulme Trust titled *Cognitive Listening: Investigating speech perception in noise within a cognitive framework*, under grant agreement number RPG-2019-152 awarded to Professor Sven L. Mattys.

1. CHAPTER 1: REVIEW OF THE LITERATURE

1.1 INTRODUCTION

Listening to speech is integral to the spoken transmission of information (for individuals who are not deaf or have severely reduced function of their auditory system). To be able to process information conveyed through speech, one must be able to hear, attend to, and process this speech to act on the information accordingly. However, speech communication often does not happen in optimal listening environments. Listeners might have to exert greater effort to direct their attention to a speaker and overcome reduced intelligibility of the speech signal in adverse listening conditions, which can originate at the speech source (e.g., in non-native speech production, or through reduced motor ability of speech articulators), during transmission (e.g., with competing talkers), within the listener (e.g., with non-native listening or listening with hearing loss), or a combination of adverse listening environments (Mattys et al., 2012).

Both self-reported and physiological measures can be used to assess how much cognitive effort a listener employs to listen to speech in noise. Additionally, these measures of cognitive effort can be used to assess how listening effort in adverse listening conditions changes to the distorted speech over time, and whether changes in listening effort are reflected in behavioural measures of speech perception. The aim of this thesis is to add to the knowledge of perceptual adaptation to speech in adverse conditions, unpacking the underlying mechanisms of the effect of time on informational masking. The empirical chapters of this thesis explore how listeners are able to overcome a specific type of background noise called informational masking by investigating the change over time in both speech perception performance as well as physiological and subjective measures of listening effort.

This literature review critically appraises current research into speech perception in adverse conditions and comprises four main sections. Section 1.2 addresses the relevance of adverse listening conditions for theories of speech perception and describes the difference between *energetic* and *informational masking* with examples of how different types of maskers reduce intelligibility of a speech signal. Section 1.3 discusses the importance of listening effort in understanding speech perception in adverse conditions, as well as detailing physiological measures of listening effort with a particular focus on pupillometric measures. Section 1.4 explores the literature concerning the concept of perceptual adaptation in speech perception, looking at behavioural and pupillometric measures of perceptual and attentional adaptation, what these studies tell us about adaptation to speech perception in adverse conditions, and how the empirical chapters of this thesis build upon this previous research. Finally, this literature review will end by outlining the research questions of this thesis and describing the contents of the following chapters in Section 1.5.

1.2 ADVERSE LISTENING CONDITIONS

When listening to a target talker, it is rare for the speech of the target talker to be spoken in optimal listening environments with an absence of other auditory signals. Speech perception more often occurs in adverse conditions where the speech signal is degraded. This degradation of a target speech signal can in turn impact upon the speech signal's intelligibility, i.e., 'the accuracy with which a message is conveyed by a speaker and recovered by a listener' (Klasner & Yorkston, 2005, p. 127). A speech signal may be degraded at the source, for example when a speaker has a structural or neurogenic speech disorder that causes atypical production of speech, such as dysarthria or apraxia, or if a talker is speaking in a non-native language with a perceived non-native accent (see, e.g., Mattys et al., 2012, for a review). In contrast, a speech

signal may have degraded intelligibility during transmission of the signal; although a speech signal has been produced in a typical fashion by a target talker, the listener is required to direct their attention to the target talker whilst ignoring competing sounds in the environment, be the other sounds speech or non-speech. The challenge of having to attend to a specific talker whilst ignoring competing talkers is commonly referred to as the 'Cocktail Party Problem' (Cherry, 1953). There can also be reduced intelligibility of target speech resulting from factors intrinsic to a listener, rather than from the source of the speech signal or during transmission, for example when a listener is listening to speech in a non-native language or with a reduced hearing sensitivity. These various forms of speech signal degradation at source or during transmission, their effects on speech intelligibility, and how listeners can overcome this degradation, will be explored in this literature review.

Adverse conditions can be broken down further by the type of masking provided by the adverse conditions, namely *energetic* and *informational* masking. Energetic masking results from a target signal degraded due to spectro-temporal overlap with a competing speech stream (Pollack, 1975; Brungart, 2001; see Culling & Stone, 2017 for a review of energetic masking). Informational masking instead refers to how accurate perception and comprehension of the target speech are compromised by non-acoustic features of the masker, such as its semantic overlap with the target or its familiarity to the listener (Cooke et al., 2008; Kidd et al., 2008; Pollack, 1975; see Kidd & Colburn, 2017 for a review of informational masking). When listening to a target talker in the presence of distracting speech, a listener might misallocate components from a masker talker to the target talker, fail to selectively attend to the properties of the target talker, or become distracted by the semantic content in the speech of a competing talker (Cooke et al., 2008; Durlach et al., 2003; Kidd & Colburn, 2017; Summers & Roberts, 2020). Both energetic and informational masking can result in impairments in speech comprehension of a target speech signal albeit through different mechanisms, with the

cognitive resources required to overcome informational masking thought to be greater than for overcoming energetic masking.

1.2.1 ENERGETIC MASKING AND SIGNAL DEGRADATION

Over recent decades, research has looked to explore the effects of energetic masking and signal degradation upon speech perception. The spectro-temporal properties of competing speech provide energetic masking to comprehension of target speech, namely through the overlap of the spectral structures of target and maskers (Brungart, 2001; Brungart et al., 2001; Helfer & Freyman, 2008). However, naturally occurring glimpses (dips in intensity) from a competing speech stream can be employed to improve perception of target speech (Brungart et al., 2006; Cooke, 2006; Festen & Plomp, 1990). Several degradation techniques have been used to investigate the contribution of various acoustic properties on sound perception, and in turn speech intelligibility, including vocoded speech (Bent et al., 2009; Davis et al., 2005), sine-wave speech (Barker & Cooke, 1999; Remez et al., 1981; Roberts et al., 2010), and time-compressed speech (Adank & Janse, 2009; Altmann & Young, 1993; Dupoux & Green, 1997; Golomb et al., 2007; Pelle & Wingfield, 2005).

1.2.1.1 VOCODED SPEECH

Speech that has been vocoded preserves both amplitude and temporal envelope cues of a speech signal but restricts the listener to severely degraded information on the distribution of spectral energy into specific energy bands (Shannon et al., 1995). Vocoded speech aims to simulate speech heard through a cochlear implant, whereby a speech signal is passed through a number of frequency channels to directly stimulate the auditory nerve. Middlebrooks et al.

(2005) details how in a normal ear, the frequencies of an incoming sound are transduced into electrical signals by inner hair cells on the tonotopically organised cochlea, which in turn leads to synaptic activation of the auditory nerve and are processed by the auditory cortex. However, in cochlear implants, the transduction by the cochlea is replaced by a microphone whereby spectral analysis of the incoming signal is processed by a series of band-pass filters, which in turn stimulate the auditory nerve fibres along the cochlea. Cochlear implants have a fixed number of up to 24 frequency channels through which a speech signal is passed. The limited number of frequency channels in a cochlear implant causes the speech signal to degrade, resulting in only the amplitude envelopes of the cochlear frequency channels being extracted as a constant series of electrical pulses, which then stimulates the auditory nerve. (See Middlebrooks et al., 2005 for comprehensive detail on the mechanisms of cochlear implants).

The experience of listening through a cochlear implant can be simulated by dividing a speech signal into a set number of frequency channels and low-pass filtering the signal to obtain the amplitude envelope for each channel, and then applying the amplitude envelope in each frequency range to band-limited noise (Davis et al., 2005). This cochlear implant simulation can allow for comparisons with typical hearing listeners on the effect of vocoding on speech perception.

For example, Bent et al. (2009) had participants listen to 100 sentences band-pass filtered into eight-channel sinewave-vocoded speech with the participants tasked to transcribe as much as they could. Initially, listeners' mean accuracy was around 70% across the first 10 sentences, but listeners were able to improve in their accuracy as they adapted to the vocoded speech, increasing to around 83% after around 60 sentences and maintaining this level of accuracy until the end of the 100 sentences. This study demonstrated that listeners are able to not only perceive vocoded speech, but also improve over time. This improvement over time indicates that listeners are able to adapt to the degraded intelligibility of the target speech signal

through continuous exposure, learning to re-map the new input onto existing phonological representations without explicit training or feedback.

1.2.1.2 SINE-WAVE SPEECH

Unlike vocoded speech, where the amplitude of frequencies within specific frequency bands is modulated, sine-wave speech attempts to mimic the formant structure of speech. Formants are the spectral peaks of a sound spectrum (Fant, 1960) and, in regards to speech, each format corresponds to a resonant overtone in the vocal tract, which gives vowels their distinctive quality and allows for vowels to be differentiated from one another (Ladefoged, 2014). The fundamental frequency (*f0*, or formant zero) is the pitch with which a person speaks, and all higher-order formants provide varying timbre quality to vowels: the higher first formant frequency for 'open' vowels compared to 'closed' vowels, higher second formant frequency for 'front' vowels compared to 'back' vowels, and the third formant adding to more vowel quality distinction (Ladefoged, 2014). Barker and Cooke (1999) suggest that 'natural speech can be reproduced using as few as three time-varying sinusoids' (p.159) emulating the first three formants of natural speech. However, sine-wave speech is not usually recognisable to listeners without training, as all other attributes of speech, such as higher order harmonic structures, including amplitude maxima and minima across the harmonic spectrum, as well as the repeated laryngeal pulse pattern that generates formants in natural speech, are removed from a sine-wave speech signal (Remez et al., 1981). Using sine-wave speech thus provides opportunity to disentangle the contributions of specific speech formants upon perceptual organisation and streaming (Roberts et al., 2010). Moreover, using sine-wave speech allows the contributions of speech intelligibility from top-down processing (i.e., perceptual reorganisation resulting from listener training, using context-specific syntactic or semantic

knowledge) to be dissociated from bottom-up processing of the acoustic signal in real-time (Feng et al., 2012), as well as to compare how the acoustic structure of a speech signal interacts with other language-specific characteristics, including phonological and tonal structure (Rosen & Hui, 2015).

1.2.1.3 TIME-COMPRESSED SPEECH

In addition to the availability of preserved speech formant structure from speech vocoding and modulation, the intelligibility of a talker is also dependent upon speech rate, with naturally slower speech rates associated with more intelligible talkers (Hazan & Markham, 2004). Artificially time-compressing speech allows for the contribution of processing time and working memory to be assessed in addition to the natural variability in talker speech rates. Speech rate manipulation changes the underlying signal, and these changes resulting from speech rate manipulation require *perceptual recalibration* or *normalisation* of a target talker (Miller, 1981, 1987; Miller & Liberman, 1979; Peelle & Wingfield, 2005), with listeners showing the ability to adapt to both naturally fast and time-compressed speech (Dupoux and Green, 1997; Adank and Janse, 2009). Although listeners are able to adapt to time-compressed speech, speech recognition accuracy is much reduced in time-compressed speech compared to clear speech and vocoded speech (O'Leary et al., 2023). However, O'Leary et al. (2023) also identified that the detriment imposed by time-compressed speech can be mitigated by 'time-restored' pauses (i.e., a silence in the remaining time from which a sentence or phrase has been time-compressed), and found improved performance compared to time-compressed speech without time-restored pauses for clear speech, 10-channel and 6-channel vocoded speech. This improved performance for time-compressed speech with time-restored pauses relative to standard time-compressed speech was reduced as the speech became more severely degraded

(i.e., most benefit resulting from time-restored pauses for clear speech, then 10-channel vocoded speech, then 6-channel vocoded speech). These time-restored pauses can be thought of similarly as 'glimpses' in competing speech, which are discussed in Section 1.2.1.5.

1.2.1.4 FUNDAMENTAL FREQUENCY OVERLAP

In addition to a speech signal being degraded at the source, there is a possibility for the speech signal to be degraded during transmission. For example, multiple talkers or distracting sounds can cause spectro-temporal overlap of a target speech signal if these competing speech signals co-occur in the listeners' ambient environment. When listening to a target talker in the presence of competing talkers, one feature of a speaker that can impact intelligibility of the target talker is the gender of the competing talkers. Female talkers tend to have a higher fundamental frequency than male talkers (Bradlow et al., 1996) and also tend to be more intelligible than male talkers (Hazan & Markham, 2004). When comparing the impact of fundamental frequency overlap between competing speech and a target talker, it then becomes possible to evaluate performance differences based on whether the target and masker talkers are of the same or different genders.

Research into the effect of masker gender has shown worse performance in speech-in-noise tasks when the masker talker gender is the same as the target talker gender in the presence of one, two and three competing talkers (Brungart, 2001; Brungart et al., 2001), with results similar for listeners with and without hearing loss (Helfer & Freyman, 2008). This difficulty in listening to target speech in the presence of competing speech from someone with the same gender results from the similar spectral energy present in the target and competing speech. One way to interpret this streaming difficulty is based on the construct of *object-based auditory attention* (Shinn-Cunningham, 2008), whereby target speech perception in the presence of

competing maskers depends upon the listeners' ability to maintain the integrity of the target speech as a different auditory 'object' to the masker. This ability to stream the target and masker talkers into separate 'auditory objects', one to be attended to and the others to be ignored, could result from the increased spectral overlap in talkers of the same gender relative to masker talkers of a different gender to the target, in addition to other components of competing speech that could cause interference, e.g., intelligible linguistic content (see Section 1.2.2 on informational masking).

1.2.1.5 SPEECH OVERLAP AND GLIMPSES

In addition to overlap of spectro-temporal characteristics such as the fundamental frequency and formants, a competing speech signal can contain local amplitude modulations that might result in opportunities to 'glimpse' the target speech (Brungart et al., 2006; Cooke, 2006; Festen & Plomp, 1990). Some examples of these glimpses can be if a competing talker takes a pause in their speech resulting in an extended period where the target speech is not acoustically occluded, or even smaller gaps such as when a competing talker is producing a stop consonant before voice onset. Local glimpsing opportunities may be different depending on the nature of the competing speech, and some competing speech signals provide more opportunities to glimpse the target talker than others (Buss et al., 2020). Howard-Jones and Rosen (1993) found that although listeners were able to exploit glimpses in a masker to hear out target speech, this ability depended on the type of masker, with better performance when glimpses occurred in lower frequency modulation than when they occurred at a higher frequency but shorter duration. Similarly, Rhebergen et al. (2005) noted that when using time-reversed speech as a masker, the unfamiliar spectro-temporal characteristics of reversed phonemes were initially distracting, but the time-reversed signal could be learned over time

and, ultimately, be exploited to hear out target speech in the glimpses of time-reversed phonemes.

Across the experimental chapters in this thesis, competing speech is used as the primary method of target speech signal degradation. However, it is not only the energetic overlap of spectro-temporal components of a competing speech signal that can cause degradation of target speech, but the informational content of the competing speech can also interfere with target speech perception. In the following section, different features of informational masking are presented, divided into sections pertaining to reduced intelligibility at the speech source, during transmission, and at the listener level.

1.2.2 INFORMATIONAL MASKING

In addition to energetic masking, difficulties in target speech perception can also result from non-acoustic features of adverse conditions, termed *informational masking*. Informational masking can arise from various processing failures: misallocation of components from a masker talker to a target talker, failure to selectively attend to the properties of a target talker, heightened cognitive load from selectively tracking a target talker, and semantic interference from known content in the speech of a competing talker (Cooke et al., 2008; Durlach et al., 2003; Kidd & Colburn, 2017; Summers & Roberts, 2020). Informational masking is thus distinct from energetic masking, as informational masking can result from broader issues that are not intrinsic to the speech signal itself. The effect of different informational masking at speech source (compromised speech motor control; non-native accented talkers), informational masking during transmission (presence of competing talkers), and informational masking at listener source (language proficiency; hearing sensitivity) are explored in the following section.

1.2.2.1 REDUCED INTELLIGIBILITY FROM SPEECH SOURCE

1.2.2.1.1 COMPROMISED SPEECH MOTOR CONTROL

To overcome adverse listening conditions, speakers can adapt their speech production to meet communicative and situational demands, such as hyper-articulating and increasing their speech level when maximum acoustic information is required (described by Lindblom's Hypo-Hyper Speech Model, 1990). However, there can be situations where a talker is unable to hyperarticulate their speech to prevent degradation of their speech, for example in talkers with structural or neurogenic speech disorders, like dysarthria, a neurological disorder of the motor speech system that results in compromised integrity of a speech signal and its intelligibility (Borrie et al., 2012a). Processing atypical speech production (e.g., dysarthric speech) can result in heightened cognitive demand for a listener as well as incorrect perception of the target speech due to mismatches between the intended and realised utterances (Liss, 2007; Liss et al., 2000; Samuel & Kraljic, 2009) and may require learning through the reorganisation of perceptual speech categories to accurately perceive the speech of a talker with compromised speech motor control.

Dysarthria frequently co-occurs with other physical, cognitive and memory deficits (Duffy, 2013), and can pose challenges for the target talker in producing their intended utterance accurately. Thus, listener-oriented training with repeated exposure and familiarisation with dysarthric speech has been proposed to help with dysarthric speech intelligibility, without the need for speaker training (Liss, 2007). This type of familiarisation training to dysarthric speech has since been shown to improve speech recognition and intelligibility of dysarthric speech (Borrie et al., 2012b, 2012c). However, even without training, other features of dysarthric speech can be exploited to enhance intelligibility, including a dysarthric talker speaking at a louder level (Fox et al., 1997, 2006; McAuliffe et

al., 2017). The ability for listeners to adapt to atypical speech production provides evidence that, both with and without training, repeated exposure to speech degraded at the source can improve intelligibility of compromised speech production.

1.2.2.1.2 NON-NATIVE SPEECH

Speech intelligibility difficulties can also arise when listeners attend to non-native speech. Challenges in non-native speech perception can occur with both foreign-accented speech (Munro & Derwing, 1995) as well as accents in a native language to which a listener is unfamiliar (Adank et al., 2009; Floccia et al., 2006). Different languages have different phonological inventories. Even if a second-language learner does not have particular phoneme contrasts in their first language, undertaking training of non-native phonetic contrasts can lead to improvement in their identification (Logan et al., 1991; Lively et al., 1993; Lively et al., 1994) and in more accurate phoneme production by these non-native speakers (Bradlow et al., 1997). Even without speaker-oriented training, listeners can passively adapt to non-native speech production by means of perceptual learning (Bradlow & Bent, 2008; Kraljic et al., 2008; Maye et al., 2008; Sidaras et al., 2009). Evidence of perceptual reorganisation by listeners attending to non-native speech indicates that, through sufficient exposure, listeners can adapt to atypical production of speech. Perceptual learning is discussed further in Section 1.3 in the context of speech-in-speech listening.

1.2.2.2 REDUCED INTELLIGIBILITY FROM LINGUISTIC INTERFERENCE

As established above, masking can arise during the transmission of a speech signal, with informational masking resulting from the energetic spectro-temporal characteristics of

competing speech. In this section, and throughout this thesis, the informational masking attributed to the linguistic content of competing speech, such as a listener misallocating components from a masker talker to the target talker (Mepham et al., 2022), is termed *linguistic interference*. Investigations into informational masking resulting from linguistic interference have highlighted several findings about the effect of a known-language masker on the perception of target speech. First, a masker in a language known to the listener is often found to be more disruptive than a masker in an unknown language (Calandruccio et al., 2013; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007; but see Mattys et al., 2010, for an exception). Thus, a known masker language is thought to produce more informational masking compared to an unknown masker language due to linguistic information being available to a listener in a known than unknown masker language. Moreover, and partly overlapping with the previous observation, speech recognition performance is usually worse if the masker language is phonetically similar to the target language, and worse still if the masker and the target are the same language. This finding has led to what is referred to two partly competing underlying mechanisms of target speech perception in the presence of competing talkers: the *known-language interference account* (Brouwer et al., 2012), and the *target-masker linguistic similarity account* (Calandruccio et al., 2013).

Regarding the known-language interference account, Brouwer et al. (2012) found that, when listening to a target in English, both English monolingual speakers and Dutch non-native speakers of English experienced greater disruption when the competing speech was the same language as the target (English) as opposed to another language (Dutch). This is an example of informational masking resulting from the same known language being used as both the target and the masker speech, linguistic interference. However, the magnitude of this linguistic interference was modulated by whether the target speech was native or non-native to the listener. The difference in performance between the Dutch and English masker talkers was less

pronounced for the Dutch than the English listeners, which suggests that linguistic interference is determined by an interaction between bottom-up factors (acoustic/linguistic similarity) and top-down challenges (e.g., automatic activation of the masker language if known, even non-natively, see the BIA+ model of bilingual language activation, Dijsktra & van Heuven, 2002; van Heuven & Dijkstra, 2010).

The target-masker linguistic similarity account differs from the known-language interference account in that the source of interference results from the target and masker being linguistically the same or similar (i.e., competing speech in the same language or phonetically similar to the target language) rather than interference resulting from a competing speech in a language known to the listener. Calandruccio et al. (2013) tested English target speech perception by native monolingual English speakers in the presence of English, Dutch, or Mandarin masker talkers. Similarly to the Brouwer et al. (2012) findings, there was worst performance by the listeners in the English masker condition, suggesting greatest informational interference when masker talkers were in the same language as the target talker. Although there was better performance in the Dutch and Mandarin masker conditions, the listeners performed better in the sentence recognition task in the Mandarin than the Dutch masker conditions. The authors suggested that this benefit resulted from a greater linguistic difference between English and Mandarin than between English and Dutch. They argued that more linguistically different languages have less spectral overlap and thus result in less informational interference compared to linguistically similar languages, even if these masker languages are all unknown to the listeners.

The known language account and the language similarity account are the primary subject of investigation in Chapter 2 of this thesis and explored in depth in the discussion in Chapter 4.

1.2.2.3 REDUCED INTELLIGIBILITY AT LISTENER SOURCE

1.2.2.3.1 LANGUAGE PROFICIENCY AND NON-NATIVE LISTENING

In addition to informational masking at the source of the speech signal and during transmission, there could be specific listener characteristics that result in different speech in noise abilities among listeners. In a review by Scharenborg and van Os (2019), it was established that listening to speech in background noise is more difficult when listening in a non-native language than a native language. This 'non-native disadvantage' has been shown through different approaches, including measuring word recognition accuracy across different signal-to-noise ratios (SNRs; Scharenborg & van Os, 2019) and establishing the SNR needed for native and non-native listeners to achieve the same level of performance, such as by adjusting task SNR (Bradlow & Alexander, 2007; Brouwer et al., 2012; Cooke et al., 2008; Van Engen, 2010). Another approach includes using an adaptive procedure to determine the speech reception threshold where listeners obtain a 50% speech accuracy score (Kaandorp et al., 2016; Kilman et al., 2014; Van Wijngaarden et al., 2002; Warzybok et al., 2015), with both approaches identifying a +4 to +8 dB more favourable SNR for non-native listeners to achieve performance parity with native listeners when listening to speech in noise. This 'non-native disadvantage' effect, where non-native listeners perform worse than native listeners in speech-in-noise tasks, coupled with the informational masking provided by competing speech during transmission, adds to the evidence of increased interference for individuals listening to speech in a non-native language.

.

1.2.2.3.2 REDUCED HEARING SENSITIVITY

Having reduced hearing sensitivity can diminish speech perception ability. Peripheral hearing loss is typically categorised as conductive (caused by impairment of the outer or middle ear), sensorineural (caused by dysfunction in the cochlea or spiral ganglion), or mixed, whereby there is both conductive and sensorineural hearing loss with problems in sound transmission before and after the cochlea (Cunningham & Tucci, 2017). Studies of listeners with hearing loss or a reduced hearing sensitivity show consistently poorer speech perception in noise ability compared to listeners with typical hearing (Ching & Dillon, 2013; Kidd et al., 2002; Koelewijn et al., 2014b). Studies have also shown that although typical hearing listeners experience a benefit in speech in noise perception when target and masker talkers are spatially separated compared to when they are co-located (Litovsky, 2013), listeners with hearing loss experience a much-reduced benefit in release from informational masking, i.e., the amount of improvement when linguistic interference is removed from a competing speech signal, resulting from spatial separation (Arbogast et al., 2005).

1.3 LISTENING EFFORT AND ITS MEASURES

1.3.1 SPEECH IN NOISE PERCEPTION AND LISTENING EFFORT

As discussed in previous sections, when listening to a target talker in the presence of competing talkers, listeners are able to comprehend more of the target talker when the language of the competing talkers is unknown to the listener than when it is known (Brouwer et al., 2012; Calandruccio et al., 2013; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Kilman et al., 2014; Van Engen & Bradlow, 2007). Additionally, listeners are able to adapt over time to speech in the presence of competing auditory stimuli (Bent et al., 2009; Cooke et al., 2022; Erb

et al., 2012, 2013; Mepham et al., 2022; Versfeld et al., 2021). When listening to speech in adverse conditions, there might be situations in which it is necessary for listeners to expend more listening effort, for example, listening to speech in a second language (Borghini & Hazan, 2018), listening to a talker with an unfamiliar accent (McLaughlin & Van Engen, 2020), or listening with a hearing impairment (McGarrigle et al., 2021b; Peelle, 2018). There is increased interest in quantifying the cognitive effort required to accurately perceive speech in noise. This cognitive effort required for speech perception in adverse conditions is defined as *listening effort*; where "deliberate allocation of *resources* to overcome *obstacles* in goal pursuit when carrying out a listening *task*." (Pichora-Fuller et al., 2016, p. 11S).

1.3.2 PHYSIOLOGICAL MEASURES OF LISTENING EFFORT

Recording behavioural performance, such as speech recognition or transcription accuracy only captures the responses of particular tasks, and does not provide insight into other cognitive processes, such as the effort a listener exerts when listening to speech in adverse conditions. Although it is possible to use self-reported measures of listening effort, which can provide important and ecologically valid insights about the subjective experience of effortful listening (McGarrigle et al., 2020), these self-report measures might be liable to bias, might be difficult to interpret or compare, and might not provide sufficient insight into the neurological or cognitive physiological mechanisms involved in listening in adverse conditions (Moore & Picou, 2018).

In addition to subjective and behavioural measures of cognitive effort, changes in physiological state can provide autonomic markers of listening effort (Pichora-Fuller et al., 2016). These physiological measures of listening effort fall primarily into three categories: changes in evoked cortical potential as measured by electroencephalography (EEG; Bernarding

et al., 2014; Haro et al., 2022; Marsella et al., 2017; Wisniewski et al., 2015), changes in blood oxygenation levels as measured by functional magnetic resonance imaging (fMRI; Dimitrijevic et al., 2019), and changes in the autonomic nervous system as measured by skin conductance (Mackersie & Calderon-Moultrie, 2016), heart rate variability (Mackersie & Calderon-Moultrie, 2016; Mazeres et al., 2019; Richter, 2016), and pupillometric measures (Van Engen & McLaughlin, 2018; Zekveld et al., 2018). This thesis will specifically focus on pupillometric measures of listening effort in conjunction with subjective self-report measures of listening effort and behavioural performance as measured by sentence recognition and transcription accuracy.

1.3.3 PUPILLOMETRIC MEASURES OF LISTENING EFFORT

Pupillometric measures can distinguish between tasks requiring more or less effort (Beatty, 1982). This physiological measure of effort is used in the field of auditory cognitive science as a more objective and comparable measure of listening effort as opposed to self-reported ratings of expended effort (for reviews see Van Engen & McLaughlin, 2018; Zekveld et al., 2018). Although pupillometric measures of listening effort are deemed to be more objective than self-reported measures and correlate with task demand (Zekveld et al., 2018), there is inconsistency in whether pupillometric measures of listening effort correlate with subjective ratings of listening effort, with some studies finding correlations (McGarrigle et al., 2020) and others not (Koelewijn et al., 2012; Strand et al., 2018). These conflicting findings are attributed to the method of obtaining the self-reported data, with correlations found when listeners are asked about their subjective ratings of listening effort during an experiment (McGarrigle et al., 2020) and no correlations when listeners are asked at the end of a testing session (Koelewijn et al., 2012; Strand et al., 2018).

1.3.3.1 MECHANISMS OF PUPILLOMETRIC MEASURES

The size of the pupil is controlled by two muscles: sphincter (constrictor) muscles, which reduce the size of the pupil, and dilator muscles, which increases the size of the pupil (Zekveld et al., 2018). Pupil size is modulated by the interplay of the sympathetic and parasympathetic nervous systems (Loewenfeld & Lowenstein, 1999), with pupil size sensitive to a myriad of factors, including illumination level (Wang et al., 2018a), fatigue (Wang et al., 2018b), and cognitive processing (Granholm & Steinhauer, 2004; Steinhauer et al., 2004). It is these differences in pupil size when listeners undertake tasks that can be exploited as a more objective measure of effort, with tasks requiring more cognitive effort resulting in greater pupil dilation (Granholm et al., 1996). However, these differences in pupil size only arise when participants remain engaged and motivated in the task and where processing demands do not exceed available cognitive resources (da Silva Castanheira et al., 2020; Granholm et al., 1996; Wang et al., 2018b; Wendt et al., 2018). The baseline pupil size at a resting state is used as a marker of baseline arousal level (Aston-Jones & Cohen, 2005) or task engagement (Hopstaken, et al., 2015), and the task-evoked pupil response (TEPR) from this baseline arousal level is interpreted as a change in cognitive effort required for resource allocation of a given task. It is important to note that baseline pupil size can differ between individuals based on factors including ambient illumination (Beatty & Lucero-Wagoner, 2000), ethnicity (Quant & Woo, 1992), and age (MacLachlan & Howland, 2002), and thus present challenges when inferring mental effort from TEPR from baseline pupil size. Early studies calculating TEPR from a percentage change from baseline pupil size have been criticised for failing to account for differences in tonic/resting baseline pupil size between participants (e.g., Beatty & Lucero-Wagoner, 2000). Instead TEPR subtraction methods of pupil diameter from baseline pupil size within a participant are advised to account for differences in tonic/resting baseline pupil size resulting from extraneous factors (van der Wel & van Steenbergen, 2018).

1.3.3.2 USE OF PUPILLOMETRIC MEASURES IN AUDITORY COGNITIVE SCIENCE

The use of pupillometric measures in auditory cognitive science has allowed for research into cognitive effort manifested by the sympathetic nervous system whilst undertaking speech-in-noise tasks in different adverse listening conditions, including modulated noise (Koelewijn et al., 2012, 2014a; McLaughlin et al., 2021; Ohlenforst et al., 2018; Paulus et al., 2020; Wendt et al., 2018), time-compressed speech (O'Leary et al., 2023; Paulus et al., 2020), noise-vocoded speech (Paulus et al., 2020), non-native-accented speech (Brown et al., 2020; McLaughlin & Van Engen, 2020), multi-talker babble (Koelewijn et al., 2012, 2014a; Ohlenforst et al., 2018; Wendt et al., 2018), and non-native listening (Borghini & Hazan, 2018). In addition to listening effort, pupillometry techniques have provided implicit measures of tiredness or fatigue from listening (McGarrigle et al., 2020; Wang et al., 2018b) across the lifespan (older adults: McGarrigle et al., 2021a; 2021b*;* children: McGarrigle et al., 2017) and for listeners with hearing impairments (Koelewijn et al., 2014b).

Generally, when a target speech signal is degraded either at source or through transmission, listeners exhibit greater pupil dilation while listening to the target speech, indicating greater effort required while listening in more challenging conditions. An investigation of pupil responses at varying SNRs from easy (+8 dB) to difficult (-20 dB) for fluctuating noise and 1- and 4-talker babble demonstrated that peak pupil dilation was greatest when listeners' speech intelligibility was between 30-70%, between -8 and -4 dB SNR (Wendt et al., 2018). In easier listening conditions (i.e., SNRs greater than -4 dB SNR), the peak pupil dilation was reduced, indicative of less listening effort needed to comprehend and repeat aloud the target sentence. However, there was also reduced peak pupil dilation in the more difficult listening conditions (i.e., SNRs less than -8 dB SNR), with sentence recognition performance at floor. This reduction in peak pupil dilation in the most difficult listening conditions, where

sentence recognition is close to impossible and would, in theory, require the greatest cognitive demands, suggests that the listeners disengage from the task when the effort expended in listening is not compensated by gains in speech recognition (Pichora-Fuller et al., 2016). These results from Wendt et al. (2018) highlight the fine balance needed when using pupillometry as a measure of listening effort for the adverse conditions to be challenging enough to require cognitive effort for speech recognition, but not so difficult that the listeners disengage from the task and expend no effort in trying to comprehend the degraded speech.

When considering signal degradation, the cognitive effort required to recognise speech that has been degraded using noise-vocoding has been assessed using pupillometric measures. Winn et al. (2015) tested listeners' speech recognition accuracy for speech vocoded in 4, 8, 16 and 32 channels, as well as non-vocoded speech, and measured the change in task-evoked pupil response over the time-course of listening to the target sentence. When all trials were aggregated, there was both greater pupil size and higher rate of task-evoked pupil dilation for the lower resolution speech (i.e., the speech signal that had been most degraded by vocoding). This pattern of pupillometric results also persisted when only trials where all keywords were correctly reported were analysed. If listening effort was intrinsically associated with accuracy, one would have expected the rate and magnitude of pupil dilation to be equivalent across all trials where all keywords in a target sentence were reported correctly, regardless of the severity of the signal degradation. However, the fact that the pattern of greater task-evoked pupil dilation was present for the speech signal with most degradation even when participants were correctly reporting the target sentence, indicates that cognitive effort is not intrinsically tied to performance accuracy. Similarly to Wendt et al., (2018), who found an inverted U-shape curved relationship between SNR and peak pupil dilation, the results from Winn et al. (2015) suggest that listening effort is associated more with the severity of signal degradation than with levels of speech recognition performance.

In the context of informational interference, different studies have shown greater peak pupil dilation over time for degraded listening conditions involving an informational component, for example with non-native-accented speech (Brown et al., 2020; McLaughlin & Van Engen, 2020). These studies demonstrate differences between native- and non-native-accented speech conditions, with greater peak pupil dilation for non-native-accented speech over native-accented speech (McLaughlin & Van Engen, 2020). Additionally, these studies also demonstrate a greater reduction in peak pupil dilation over time for non-native-accented speech over native-accented speech (Brown et al., 2020), with differences in peak pupil dilation between conditions indicative of the informational interference caused by non-native phonemes and their mismatch with native phonological representations. Additionally, non-native listeners exhibit greater peak pupil dilation (and hence greater reductions in cognitive effort) while listening to masked speech compared to native listeners (Borghini & Hazan, 2018), with the difference between the listener groups a quantification of the cognitive cost of listening to masked speech in a known but non-native language.

Chapter 3 of this thesis contributes to the field of listening effort caused by informational interference through investigating the differences in mean pupil dilation between multi-talker babble-masked target speech, with intelligible linguistic content present in this type of masker, and time-reversed speech, where there is no available intelligible linguistic content. In the experimental chapters of this thesis, the time-reversed multi-talker babble-masked speech, matched on spectro-temporal frequencies to the intelligible multi-talker babble masker, allows for the quantification of the additional listening effort required to perceptually stream the target speech from the masker, and thus the listening effort cost of informational interference when other interferences from energetic masking have been accounted for.

1.4 PERCEPTUAL ADAPTATION

1.4.1 PERCEPTUAL ADAPTATION TO SIGNAL DEGRADATION

As outlined above, speech signal degradation from energetic and informational masking can result in adverse listening conditions which a listener must overcome. Listening to speech in these adverse conditions can then lead to the listener reorganising their perceptual space to accommodate this new type of speech input (Samuel & Kraljic, 2009). Most studies investigating the known-language interference account (Brouwer et al., 2012) or the target-masker linguistic similarity account (Calandruccio et al., 2013) have based their conclusions on data aggregated over a large number of trials within an experimental session. However, data aggregation ignores the fact that listeners' ability to stream one voice from another can change with practice (Bent et al., 2009; Erb et al., 2012, 2013; Mepham et al., 2022). Using mean sentence recognition over an experimental block as a sole outcome measure might misrepresent the mechanisms underlying informational interference. Assessing the time-course of informational interference is important because it provides insight into the learnability of a masker's characteristics as well as a listener's ability to control the interference of the linguistic content of the masker.

The time-course of speech masking has been explored for word position within sentences (Ezzatian et al., 2012; 2015). There is increasingly more research into how speech recognition performance changes across trials in the course of an experiment. Studies by Bent, et al. (2009), Cooke et al. (2022), Erb et al., (2012, 2013), and Lie et al. (2024) have demonstrated the ability to adapt to masked target sentences. The details of these studies are explored and expanded in the following section.

Bent et al. (2009) demonstrated the ability to perceptually adapt to degraded speech by measuring transcription performance in both six-talker babble at 0 dB SNR and eight-channel

sinewave vocoded speech, simulating a cochlear implant. The authors found an initial improvement in transcription performance for both types of degraded speech. However, performance plateaued after around 40 sentences in the multi-talker babble condition (from 67% to 74%), whereas performance continued to improve up to around 60 sentences for noise-vocoded speech (from 70% to 84%). These results suggest that listeners are not only able to perceptually adapt to degraded speech, but that the ability to learn how to process degraded speech depends on the nature of the degradation (Mattys et al., 2012), with a longer learning window when the degradation consists of systematic and predictable alterations of the signal (noise-vocoded speech) than when the degradation is extrinsic to the signal and is mostly random (multi-talker babble). Similarly, Erb et al. (2012, 2013) found that listeners were able to perceptually adapt to four-band noise-vocoded speech (German low-predictability sentences) over time. Listeners were able to adapt to the noise-vocoded speech over the course of 100 sentences, and this adaptation followed a linear trend.

Perceptual adaptation to distorted speech has also been explored with other masking methods. Lie et al. (2024) investigated how participants improved in speech-reception threshold (SRT) measurements over six lists of 13 sentences in stationary noise, temporally-modulated noise, and spectrally-modulated noise. Results showed that over the course of the six sentence lists, there was lower SRT measurements (i.e., better performance) from the fifth and sixth lists onward compared to the first list for all types of temporally-modulated noise and spectrally-modulated noise, and some of these modulation types having a more rapid improvement with significant improvements at the third or fourth list compared to the first list. There were no differences in the SRT between lists for the stationary noise, indicating that performance did not improve in this masking method.

As well as perceptual adaptation being observed for temporally- and spectrally-modulated maskers, Cooke et al. (2022) explored listeners' abilities to adapt to different types of distorted

speech. In this study, Cooke et al. (2022) assessed speech recognition accuracy in eight types of distorted speech: time-compressed speech (increased by a factor of 2.5, commensurate to a reduction in duration of 40% of the original speech duration), noise-vocoded speech (into six bands filtered through a noise carrier), reversed speech (time-reversing successive nonoverlapping 62 ms of speech), glimpsed speech (resynthesising spectro-temporal regions of a speech-shaped noise-masked signal when mixed at a global SNR of 0 dB, García Lecumberri & Cooke, 2020), sculpted speech (randomly sampled fragments of an operatic work passed through a time-frequency mask, García Lecumberri & Cooke, 2020), narrowband speech (filtered through a third-octave filter centered at 2 kHz), tone-vocoded speech (six bands filtered through a tone carrier), and sine-wave speech (using first and second formant frequencies and their amplitudes).

In each condition of this within-subjects experiment, participants listened to 30 Spanish analogues of the Harvard/IEEE sentences (the 'Sharvard Corpus', Aubanel et al., 2014) in sequence and their accuracy was scored as the percentage of keywords correctly reported. There were differences in intelligibility between the distortion types, with highest intelligibility in the time-compressed speech (82.2%) and lowest intelligibility in the sine-wave speech (40.7%; the list of distortion types in the preceding paragraph depicts the order of intelligibility between the distortion types, from easiest to hardest). In all types of distortion, intelligibility improved across the block. With some distortion types, there was indication that exposure to other distortion conditions influenced intelligibility performance, specifically with the sine-wave speech distortion (where participants scored in the bottom quartile of responses when presented with this condition either first or second) and in the time-compressed speech condition (where over half of the lower quartile scores were present when this condition was the final experimental block), though this facilitation effect was not present in the other distortion conditions. What the authors termed 'rapid adaptation', i.e., fast improvement in

intelligibility over the first few trials of a distortion block, occurred in almost all distortion types except for the glimpsed condition. Taken together, the results from the Cooke at al. (2022) study demonstrate the ability of listeners to adapt to speech impoverished by various distortions, even though the time courses of these adaptation trajectories might differ between distortion types.

Although Cooke et al. (2022) explored various types of distortion on target speech, the authors did not investigate the impact of target speech masked by competing speech. A study by Felty et al. (2009) examined both mean performance of target speech masked by "frozen babble" (i.e., a repeated sample of six-talker babble) and "random babble" (i.e., random sampling of the six-talker babble) and changes in accuracy over the course of a block and at different SNRs [0, 5, 10 dB]. The authors found both better performance overall in the frozen babble compared to the random babble masker, and a steeper learning rate in the frozen babble than the random babble condition, demonstrating that listeners are able to improve faster if the masker is predictable (i.e., repeated) than if it is random. However, the statistical methodology used in this paper compared the difference in rates of increase using Pearson's *r* correlation tests, and thus did not model the improvement trajectories of listening to target speech masked by competing speech.

The experimental chapters of this thesis thus contribute to the research of how speech recognition performance changes over time by investigating the adaptation trajectories in different types of speech maskers. In Chapter 2, the differences in the time-course of perceptual adaptation in known and unknown language maskers are compared between native and non-native speech perception. Similarly, in Chapter 3, the differences in the time-course of perceptual adaptation in known and unknown language maskers are compared at different SNRs and at varying subjective task difficulties derived from listeners' SRTs.

1.4.2 PUPILLOMETRY MEASURES AND CHANGES IN LISTENING EFFORT OVER TIME

Most experiments conducted using pupillometry use a blocked design, with speech recognition performance aggregated over the trials of an entire experimental block (e.g., Borghini & Hazan, 2018; Wendt et al., 2018; Winn et al., 2015). Yet, few experiments have explored how listening effort changes over time during exposure to the adverse listening conditions. A measure of the time-course of effortful listening is important not only because it has been shown that speech perception in adverse conditions is not static over time (Bent et al., 2009), but also because, with pupillometric measures, it is possible to dissociate between listening effort and behavioural performance, allowing for a more accurate description of the involvement of cognition in effortful listening. Some recent studies have explored the change in pupillometric measures when listening to degraded target speech (Brown et al., 2020; Paulus et al., 2020; Versfeld et al., 2021), which will be described in the following section.

Brown et al. (2020) explored the change in peak pupil dilation (PPD) over time when listening to native (American English) and non-native (Mandarin) accented English and found decreases in PPD in both conditions over time (with trials binned into 10-sentence bins), but the decrease was larger in the non-native than native accent condition. Brown et al. (2020) interpreted this greater reduction in the non-native accent condition as a 'levelling out' effect; listeners expending more effort initially in non-native accent listening, but with cognitive effort decreasing faster over the experimental block to similar levels to native-accented speech listening. However, even after extended exposure to the non-native accent, listeners still required more cognitive effort overall for speech comprehension compared to native accent listening. The Brown et al. (2020) study provides insight into how listening to non-native

accented speech requires more cognitive effort than listening to native accented speech. However, this study did not explore how the linguistic content of competing talkers impacts upon the effort required to perceptually stream a target speaker from the competing talkers, nor how these differences in listening effort for intelligible and unintelligible speech change over time. The experiments presented in Chapter 3 explicitly address the question of the additional listening effort cost when listening to intelligible versus unintelligible speech (i.e., the additional contribution of informational masking of speech-in-speech perception), and how this additional listening effort cost changes over time.

In addition to the findings by Brown et al. (2020), Paulus et al. (2020) found reductions over time in both PPD, mean pupil dilation (MPD) and baseline pupil size for noise-vocoded speech, time-compressed speech, speech-shaped noise masking, and in a no-degradation condition. Paulus et al. (2020) interpreted decreases in baseline pupil size as indexing sustained attention, with slower declines relating to adaptation. However, the analysis in change of baseline pupil size as a percentage change from the baseline pupil size of the first trials of each experimental block has been criticised for failing to account for differences in tonic/resting pupil size between participants, and thus the extent to which the pupil can dilate under cognitive demand (van der Wel & van Steenbergen, 2018). The experiments presented in Chapter 3 instead analysed only MPD from the baseline pupil size recorded at the start of each trial (subtracting baseline pupil size from each pupil size sample recorded across the trial, then calculating the MPD across the trial from the onset of the target speech).

Additionally, although in the Paulus et al. (2020) study there are comparisons between degraded speech conditions and a no-degradation condition, these experimental conditions do not allow for comparisons of specific linguistic features of the masked speech pertaining to the issue of listening effort arising from informational masking. The experiments presented in Chapter 3 explore how knowledge of the masker language affects the listening effort required

to successfully stream a target talker from competing talkers, and thus makes direct comparisons between conditions where the speech of competing talkers is intelligible to a listener and condition where the speech is unintelligible but matched on spectro-temporal characteristics. These differences in pupillometric measures thus quantify the contribution of informational masking resulting from the linguistic interference of competing talkers.

Although using a different experimental paradigm, Versfeld et al. (2021) explored the adaptation to informational masking over the course of exposure to various maskers: stationary frozen noise, interrupted frozen noise, randomly sampled competing speech, a repeated (i.e., "frozen") competing speech, randomly sampled time-reversed speech, and frozen time-reversed speech. In this paradigm, participants' SRTs were tested in six lists of 13 sentences, with order of conditions counterbalanced between participants. The 50% SRT for each condition was obtained by scoring as correct a sentence repeated by the participant with no mistakes. These SRTs were analysed as a function of list presentation order within each condition using a repeated-measures ANOVA. Behavioural results showed that SRT decreased (i.e., performance improved) as a function of list presentation across conditions between the first, second, third and fourth lists. There was also an interaction between masker condition and list number, indicating that in the different masker conditions improvement in SRT stopped after the second block (randomly sampled competing speech), the third block (interrupted noise, randomly sampled time-reversed speech, frozen time-reversed speech), the fourth block (frozen competing speech), or in the case of stationary frozen noise had no improvement from the first block. The decrease in 50% SRT (except in the stationary frozen noise condition) demonstrated that listeners are able to improve through exposure to the different types of competing speech, with the SRT required to score 50% correct decreasing from the initial presentation of the condition.

Versfeld et al. (2021) also recorded participants' pupillometric measures during the experiment: peak pupil dilation (PPD), peak pupil latency (PPL), mean pupil dilation (MPD) and baseline pupil size (BPS). Across pupillometric measures, there was no main effect of condition, and with PPD and MPD, there were significant but non-systematic differences between lists (higher PPD in list 1 and 2 compared to list 5; higher MPD in list 2 compared to list 5). BPS differed between lists, with larger BPS in the first list compared to lists 4-6. The authors interpreted a lack of reduction in task-evoked pupil dilation (TEPR) to indicate a true learning effect, whereby performance improved throughout the experiment (reduced SRTs) while maintaining equivalent levels of listening effort. However, the authors did not find any differences in 50% SRT between intelligible (time-forward) and unintelligible (time-reversed) masker conditions, considered to be a typical informational masking effect. Similarly, there was no difference in 50% SRT between the frozen and randomly sampled masker conditions, as found in Felty et al. (2009). This lack of differences in TEPR between informational masking conditions might result from using the 50% SRT experimental paradigm. Using SRT values reduces the number of data points to one value per list (even though lists comprised 13 sentences each). Additionally, the fluctuating nature of an adaptive SRT procedure might have facilitated both the learning of the masked target talker and the dampening of differences in the pupillometric measures to similar extents across experimental conditions that were not granular enough to be identified using repeated-measures ANOVAs, as participants listened to the same target talker across 36 experimental blocks (6 conditions x 6 lists).

Although the studies by Brown et al. (2020), Paulus et al. (2020), and Versfeld et al. (2021) provide some insight into the nature of how pupillometric measures of listening effort change over time, the experimental paradigms and choices of statistical analyses leave gaps in our understanding of how pupil size changes over time when listening to a target talker in competing maskers, and how these pupillometric changes correspond to changes in speech

recognition over time. The experiments presented in Chapter 3 of this thesis thus aim to elucidate the differences between intelligible (time-forward) and unintelligible (time-reversed) competing maskers on a target talker by measuring changes in both speech recognition accuracy and the corresponding changes in pupillometric measures across experimental conditions.

1.5 RESEARCH QUESTIONS AND OUTLINE OF THESIS

The aim of this thesis is to explore how listeners are able to overcome informational masking by investigating the change in speech perception performance over time.

In Chapter 2, a series of four experiments investigates how knowledge of the linguistic content of competing talkers interferes with speech perception as measured by a speech transcription task. The main research questions of this chapter are as follows: (1) Is linguistic interference (i.e., informational masking resulting from the linguistic content of competing speakers) best described by the known-language account or the linguistic similarity account? (See Section 1.2.2 on informational masking) (2) Does linguistic interference change in the course of a test block, an indication of listeners' evolving streaming capacity as familiarity with the input increases? (3) Is linguistic interference affected by whether listeners perform the task in their native language as opposed to a non-native language? Across all experiments, listeners heard either English (Experiments 1, 3 and 4) or Mandarin target speech (Experiment 2) in the presence of English and Mandarin time-forward and time-reversed two-talker babble. Experiments 1 and 2 tested speech-in-noise perception for native listening (English target sentences for native British English speakers, Mandarin target sentences for Mandarin-English bilingual speakers), while Experiments 3 and 4 tested speech-in-noise perception for non-native listening (English target sentences for Mandarin-English bilingual speakers).

In Chapter 3, a series of three experiments investigate how pupillometric measures of listening effort are reflected in behavioural performance changes in speech recognition through continued exposure to adverse listening conditions. These experiments unpack the underlying mechanisms by which listeners improve in speech perception in adverse conditions over time, as well as how both pupillometric and self-reported measures of listening effort required to achieve these levels of performance differ between intelligible and unintelligible competing speech. Across the three experiments, participants listened to English target sentences in the presence of intelligible (time-forward two-talker babble) and unintelligible competing speech (time-reversed two-talker babble). In Experiment 5, an adaptive procedure obtained participants' 50% SRT for each condition to compare improvement trajectories from equivalent starting points. Experiment 6 pinned the SNR to -1.5 dB across all participants to compare the differences in speech recognition performance at different starting performances (and thus different levels of subjective difficulty). Experiment 7 used an adaptive procedure to make the unintelligible time-reversed masker condition (what should theoretically be "easier" to inhibit) the harder condition by lowering the SNR, and make the intelligible time-forward masker (i.e., the "harder" condition) the easier condition by increasing the SNR in order to dissociate whether performance improvement is associated with linguistic interference, or with the starting performance of the adverse listening condition.

Chapter 4 then summarises the results of the seven experiments presented in Chapters 2 and 3, and explores the implications of these results for theories of speech perception in adverse conditions, as well as discussing limitations of the current experiments presented, and potential future directions in exploring perceptual adaptation to speech perception in adverse conditions.

2 CHAPTER 2[1]: THE TIME-COURSE OF LINGUISTIC INTERFERENCE DURING NATIVE AND NON-NATIVE SPEECH-IN-SPEECH LISTENING

2.1 ABSTRACT

Recognising speech in a noisy background is harder when the background is time-forward than time-reversed speech, indicating a *masker direction effect*, and when the masker is in a known than an unknown language, indicating *linguistic interference*. We examined the masker direction effect when the masker was a known versus unknown language, and calculated performance over 50 trials to assess differential masker adaptation. In Experiment 1, native English listeners transcribing English sentences showed a larger masker direction effect with English than Mandarin maskers. In Experiment 2, Mandarin non-native speakers of English transcribing Mandarin sentences showed a mirror pattern. Both experiments thus support the *target-masker linguistic similarity hypothesis*, where interference is maximal when target and masker languages are the same. In Experiments 3 and 4, Mandarin non-native speakers of English transcribing English sentences showed comparable results for English and Mandarin maskers. Non-native listening is therefore consistent with the *known-language interference hypothesis*, where interference is maximal when the masker language is known to

[1] Part of this chapter was published in the following reference: Mepham, A., Bi, Y., & Mattys, S. L. (2022). The time-course of linguistic interference during native and non-native speech-in-speech listening. *J. Acoust. Soc. Am.*, *152*(2), 954-969. https://doi.org/10.1121/10.0013417. Supplementary material for Chapter 2 can be found at https://osf.io/nhcrw/.

the listener, whether or not it matches the target language. A trial-by-trial analysis showed that the masker direction effect increased over time during native listening but not during non-native listening. The results indicate different target-to-masker streaming strategies during native and non-native speech-in-speech listening.

2.2 INTRODUCTION

Experiments investigating the 'Cocktail Party Effect' (Cherry, 1953) have sought to disentangle the effects of different types of maskers on the recognition of target speech. For speech-in-speech listening, it is generally agreed that challenges can arise from energetic masking, whereby a target signal is degraded due to spectro-temporal overlap with competing speech at the cochlear and auditory-nerve levels (e.g., Culling & Stone, 2017) or from informational masking, whereby target recognition is compromised by masking that is non-energetic in nature. Informational masking includes misallocations of acoustic elements from the masker to the target due to perceptual similarity, heightened cognitive load incurred by selective tracking of the target, and interference from the linguistic (e.g., phonetic, semantic) content of the masker (e.g., Cooke et al., 2008; Kidd & Colburn, 2017; Shinn-Cunningham, 2008; Summers & Roberts, 2020). Informational masking resulting from the linguistic content of the masker, which we call "linguistic interference," is the topic of this study.

Investigations into linguistic interference have highlighted two key findings. First, a masker in a language known to the listener is often found to be more disruptive than a masker in an unknown language, which is referred to as the *known-language interference account* (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007; but see Mattys et al., 2010, for an exception). Second, and partly overlapping with the previous account, speech recognition is usually worse if the masker language is phonetically similar to the target

language, whether the masker language is known or unknown to the listener, which is referred to as the *target-masker linguistic similarity account* (Brouwer et al., 2012; Calandruccio et al., 2013; Freyman et al., 1999; Van Engen & Bradlow, 2007). For example, Brouwer et al. (2012) showed that, when listening to English sentences, both English monolingual speakers and Dutch non-native speakers of English experienced greater disruption when the competing speech was the same language as the target (English) than another language (Dutch). However, the effect was smaller for the Dutch than the English listeners, which suggests that linguistic interference is determined by an interaction between challenges with linguistic similarity between target and masker and familiarity with both the target and the masker languages.

Most studies investigating the known-language account and the language-similarity account have based their conclusions on data aggregated over large numbers of trials within an experimental session. However, this overlooks the fact that the ability to stream one voice from another may change with practice and increased familiarity with the masker (Bent et al., 2009; Erb et al., 2012, 2013), and hence, average performance might misrepresent the mechanisms underlying linguistic interference. Although the time-course of speech-in-speech masking has been explored for word position within a sentence (Ezzatian et al., 2012; 2015), less is known about how performance changes across trials in the course of an experiment.

One exception is Bent et al.'s (2009) study, in which the authors compared the recognition of natural speech in multi-talker babble with noise-vocoded speech over 100 sentences. They found an improvement in performance over time for both conditions. However, performance plateaued after around 40 sentences in the babble condition, whereas it continued to improve up to around 60 sentences in the noise-vocoded condition. This result suggests that the ability to learn how to process degraded speech depends on the nature of the degradation (Mattys et al., 2012), with a longer and broader learning window when the degradation consists of systematic and predictable alterations of the signal (noise-vocoded

speech) than when the degradation is extrinsic to the signal and mostly random (multi-talker babble). However, in the Bent et al. (2009) study, the purely linguistic properties of the masker could not be isolated because the energetic component of the masker was not controlled across conditions. Assessing the time-course of linguistic interference is important because it provides an insight into the ease with which a target talker can be streamed from a masker, as well as a listener's ability to overcome the activation of the linguistic content of the masker.

Linguistic interference is usually measured as a decrease in the number of target words correctly transcribed. However, a correlate of linguistic interference (and informational masking in general) is that words from the masker are likely to be erroneously reported as belonging to the target speech through either involuntary incorporation of masker keywords into the target sentence or, in some rare cases, mistaking the masker stream for the target stream. Therefore, the incidence of masker-to-target intrusions ought to be measured alongside correct transcription performance. How these two measures develop over the course of an experiment can refine our understanding of the dynamics of linguistic interference and, specifically, listeners' streaming improvement from trial to trial. Furthermore, comparing the ratio of target word transcription to masker word intrusion in native and non-native listeners can help pinpoint the mechanisms of disruption in these two groups when they engage in speech-in-speech perception.

In sum, the aim of this study was to improve our understanding of how linguistic interference impacts native and non-native listening over time. We ran four speech-in-speech experiments, one with native English speakers (Experiment 1) and three with Mandarin non-native speakers of English (Experiments 2-4). For Experiments 1, 3 and 4, the target speech consisted of English sentences and the competing speech consisted of two-talker babble in English or Mandarin. For Experiment 2, the target speech consisted of Mandarin sentences, which were translations of the English target sentences. To minimise voice differences between

the English and Mandarin babble conditions, a single English-Mandarin bilingual speaker was recorded for both conditions. Furthermore, the two-talker babble maskers in each language were created by digitally altering the fundamental frequency and vocal tract length of that speaker. Thus, all four babble voices (two English and two Mandarin) originated from a single speaker.

Finally, time-reversed versions of the English and Mandarin two-talker babble were used as a way of minimising energetic differences between the two languages. Time-reversed speech preserves the long-term average frequency spectrum of the original signal but removes its semantic content (Licklider & Miller, 1951). Thus, the difference in performance between time-forward and time-reversed maskers provides a measure of the time-forward maskers' ability to interfere with target recognition while controlling as much as possible their long-term average energetic content (but see Rhebergen et al., 2005, for some limitations).

In the following experiments, we aimed to evaluate how known and unknown language maskers interfere with native versus non-native speech recognition and how these effects develop over the course of an experiment. Specifically, we asked: (1) Is linguistic interference best accounted for by a known-language account or a linguistic similarity account? (2) Does linguistic interference change over the course of a test block, an indication of listeners' evolving streaming capacity as familiarity with the input increases? (3) Are the above patterns affected by whether listeners perform the task in their native language as opposed to a non-native language?

2.3 EXPERIMENT 1: Native English listeners – English target sentences

2.3.1 METHODS

2.3.1.1 Participants

Forty native British English speakers (34 female) aged between 18 and 25 years (M = 20.3, SD = 1.9) with no known history of hearing impairments (as determined by self-report) participated in the experiment. Of those, two were excluded due to prior experience with Mandarin. The remaining 38 declared no knowledge of Mandarin or other Sinitic languages. Six of them were excluded due to technical errors during data collection. Thus, 32 participants (27 female) aged between 18 and 25 years (M = 20.1, SD = 1.7) completed the experiment and were included in the analyses. All but three of them declared knowledge of at least one other language (French, n = 16; Spanish, n = 14; German, n = 4; Greek, n = 1; Korean, n = 1; Welsh, n = 1). The University of York Department of Psychology ethics committee approved all experimental procedures for this experiment and Experiments 2-4 (reference number: 747). Listeners either participated for course credit or were compensated at a rate of 6.00 GBP per hour. All participants provided written-informed consent before the start of the study.

2.3.1.2 Materials

2.3.1.2.1 Target stimuli

Two-hundred sentences adapted from the first 20 Harvard/IEEE sentence lists (IEEE, 1969), spoken by a female native British English speaker, were used as target stimuli (see Appendix A; all appendices can be found following the OSF link in the Acknowledgements section). Each target sentence had five keywords (e.g., “The PLAY SEEMS DULL and QUITE

STUPID", keywords capitalised). Sentence duration ranged from 1.59 s to 3.16 s (*M* = 2.20 s, *SD* = .24 s). The fundamental frequency (F0) and associated vocal tract length (VTL) of all target sentences were adjusted to a mean F0 of 210 Hz, which is approximately 15 Hz below and above the F0 of the two maskers. We manipulated the F0 and VTL of the target sentences so that the target sentences could not be distinguishable from the masker sentences solely on the basis of potential sound quality differences associated with the manipulation—by design, the masker sentences necessitated F0 and VTL alteration (see Section 2.3.1.2.2 for details). Manipulating the F0 and VTL of the target sentences also allowed us to use the same F0 and VTL values across all experiments.

2.3.1.2.2 Masker stimuli

A female native Mandarin-English bilingual speaker recorded the English and Mandarin sentences used as maskers. The use of a single speaker for both sets of sentences allowed us to minimise voice variation, and hence, differences in energetic masking across conditions. Although the first language of the bilingual speaker was Mandarin, she grew up in a multilingual environment. At the time of recording, she had lived in the United Kingdom for six years.

A pilot experiment was undertaken to assess the perceived nativeness of the bilingual speaker when speaking English in relation to nine female native monolingual speakers of British English and four female Mandarin-English bilingual speakers. Twenty native English speakers were asked to judge how confident they were that each speaker grew up speaking English in the United Kingdom, using a five-point Likert scale (Not at all confident = 0, Slightly confident = .25, Somewhat confident = .50, Fairly confident = .75, Completely confident = 1.00). All listeners heard the same five sentences spoken by all speakers. Sentences

were presented in a random order in a self-paced online experiment using Gorilla (Anwyl-Irvine et al., 2020). Ratings for the nine monolingual speakers, the four bilingual speakers, and the test speaker were entered in a one-way analysis of variance, which showed a significant effect of the Language Status of the Speaker (monolingual, bilingual, test speaker), $F(2, 57) = 124.3$, $p < .001$. Bonferroni-corrected post-hoc pairwise comparisons indicated that raters were more confident that the test speaker grew up speaking English in the UK ($M = .515$, $SD = .172$) than the bilingual speakers ($M = .133$, $SD = .129$, $p < .001$), a desirable feature for our experiment. However, they judged the test speaker as less likely to have grown up speaking English in the UK than the monolinguals ($M = .842$, $SD = .127$, $p < .001$).

To test the last result further, we asked a new set of twenty native English speakers to rate a single sentence ("Pack the kits and don't forget the salt.") produced by the test speaker and by the nine native English speakers. A single sentence was used to keep the test short. The sentence was chosen randomly. As before, participants were asked to judge how confident they were that each speaker grew up speaking English in the UK. All 10 renditions were presented as clickable icons on a computer screen, next to their corresponding Likert scale. The side-by-side format allowed the listeners to compare the renditions directly, focusing only on accentedness. Their position on the screen was randomised for each rater. Bonferroni-corrected pairwise comparisons did not show significant rating differences (ps $> .05$) between the test speaker and any but one ($p = .009$) of the native British English speakers. Thus, on balance, the results of the two tests suggest that, despite some indication that the test speaker might be detectable by some listeners as not having grown up speaking English in the UK, her speech was perceived as less accented than that of control bilinguals and as native by many listeners in our sample. We therefore judged that the test speaker's voice was suitable to use for the masker stimuli.

The bilingual speaker recorded 64 sentences from Lists 1-4 of the English BKB-R corpus (Bench et al., 1979). BKB-R sentences are simple sentences with three to four keywords (e.g., "The POSTMAN SHUT the GATE", keywords capitalised). A full list of the BKB-R sentences used in this study can be found in Appendix B. The bilingual speaker also recorded Mandarin-translated versions of these sentences. Both sets of sentences were identical to those used in the Calandruccio et al. (2010) study. All sentences were recorded in a single-walled sound-attenuated room at a 44.1 kHz sampling rate with 16-bit resolution using the Audacity© software. Each sentence was recorded a minimum of four times. For each sentence, the two best exemplars were kept. These constituted Set A and Set B. All sentences were manually edited using Praat (Boersma & Weenik, 2019) to remove silences at the beginning and end.

The Set A sentences were concatenated into a continuous stream, henceforth Stream A. The same was done with the Set B sentences, henceforth Stream B. Sentence order within each BKB-R list was the same in both streams, but the order of the lists differed in each stream. Both streams were edited using a version of the manipulation described in Darwin et al. (2003) to adjust the F0 and VTL (Smith et al., 2007; Gaudrain et al., 2009) so that each stream came in a high-F0 version (mean 225 Hz) and a low-F0 version (mean 195 Hz). These values were chosen to be equidistant to those of the target speaker. F0 is known to be a powerful grouping factor for speaker segregation in multi-talker environments (e.g., Bird & Darwin, 1998; Brokx & Nooteboom, 1982; Summers & Leek, 1998). Therefore, this procedure was used to control the long-term energetic overlap between the target and the maskers, i.e., how easy targets and maskers were to stream out from one another based on F0, and between the maskers themselves. As we applied the same manipulations in all four experiments in this study, the results were also more directly comparable across experiments. VTL was manipulated alongside F0 to improve the naturalness of the two streams, as both indices have been shown to contribute to the perception of voice identity (e.g., Skuk & Schweinberger, 2014). For each

language, the high-F0 version of Stream A was combined with the low-F0 version of Stream B to constitute two-talker babble Version 1. Likewise, the low-F0 version of Stream A was combined with the high-F0 version of Stream B to constitute two-talker babble Version 2. These two two-talker babble Versions were counterbalanced between participants in these experiments.

2.3.1.2.3 Long-term average spectra (LTAS)

Figure 1 displays the LTAS for the masker voices relative to the target voice. Note that, although the spectra of the target and masker voices are broadly comparable, this does not preclude the existence of local energetic masking differences between conditions. However, despite these spectral differences, the largely overlapping spectra suggest that our single-speaker procedure was effective in attenuating differences in average energetic profiles across conditions.

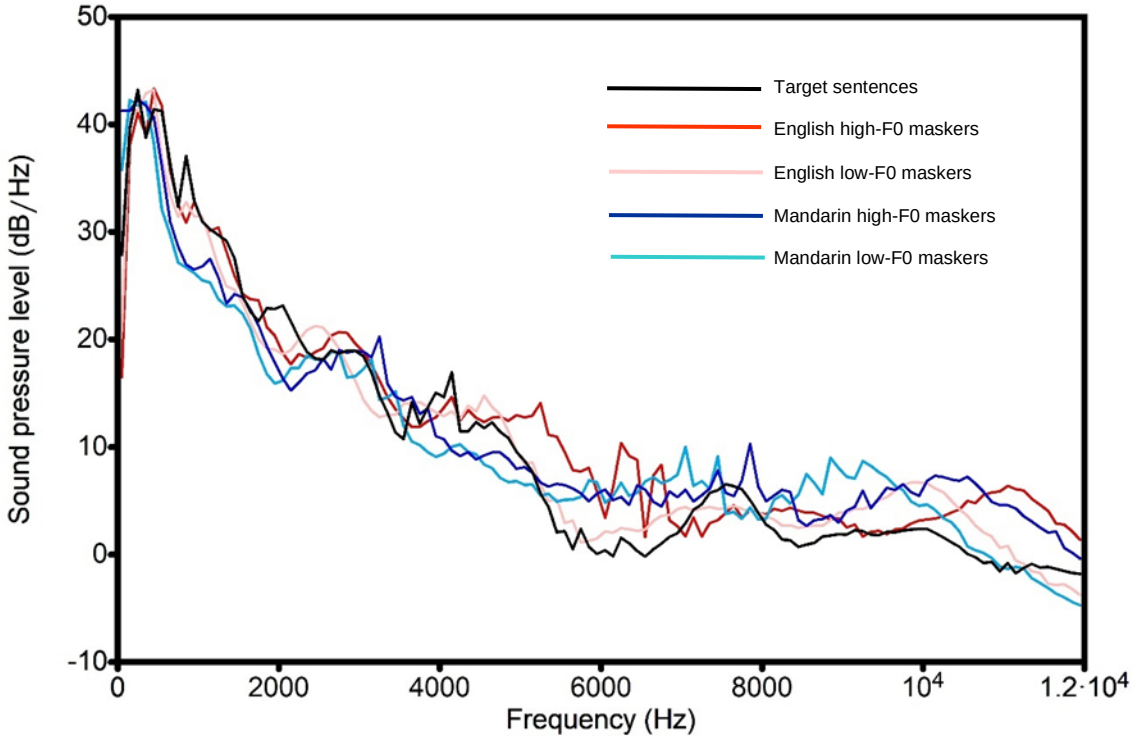


FIG. 1. Long-Term Average Spectra (LTAS) averaged across target sentences, high-F0 and low-F0 English masker sentences, and high-F0 and low-F0 Mandarin masker sentences.

2.3.1.2.4 Target-masker mixtures

A 500 ms silent interval was added to the onset of each target sentence. Each two-talker masker stream was segmented into 4.16 s portions, which covered the 500 ms silent onset interval, the duration of the longest target sentence (3.16 s), and an additional 500 ms silent offset interval. To create time-reversed versions of the maskers, the maskers were time-reversed from offset to onset. The time-reversed maskers allowed us to minimise potential energetic masking differences between the English and Mandarin time-forward masker conditions, and hence, provide a measure of linguistic interference that is not overly

contaminated by energetic masking. Thus, this experiment included four masker conditions: (1) English time-forward, (2) English time-reversed, (3) Mandarin time-forward, and (4) Mandarin time-reversed.

Target and masker sentences were combined using Praat (Boersma & Weenik, 2019). The intensity of the stimuli was normalised to 65 dB SPL for the target sentences and to 68 dB SPL for the masker streams, yielding an SNR of -3 dB, as in one of Calandruccio et al.'s (2010) experiments. The target-masker mixtures were presented diotically.

2.3.1.3 Procedure

Listeners sat in a single-walled sound-attenuated booth. The experiment was conducted using PsychoPy (Pierce et al., 2019). Target-masker mixtures were presented via Denon DJ DN-HP700 headphones. Listeners were instructed to pay attention to the target speaker and to transcribe what she said using a computer keyboard. Before the main experiment started, listeners heard five practice sentences with no masker talkers to familiarise themselves with the target voice.

The main experiment had four blocks, which varied according to the masker language (English vs. Mandarin) and the masker direction (time-forward vs. time-reversed). The order of the four blocks was counterbalanced between the listeners so that, across all listeners, each condition was presented in the same block position an equal number of times, and each condition followed and preceded each other condition an equal number of times. Within each block, 50 target sentences were randomised and the random order was different for each participant. In total, each listener transcribed 200 target sentences.

Listeners transcribed the target sentences using a computer keyboard. Their transcriptions were visible on a computer monitor as they typed and they had the opportunity

to delete and re-enter their responses before proceeding to the next trial. The task was self-paced. The next trial began as soon as a response was submitted.

2.3.2 RESULTS

Listeners' transcriptions were scored by two independent judges (authors A.M. and Y.B., see Author's Declaration) after the experiment. Obvious typographical errors were corrected. Inter-rater discrepancies (< 1% of all trials) were discussed and a score for each discrepancy was agreed upon.

Transcription scoring rules were as follows: (1) Homophones were scored as correct, e.g., "threw" and "through", (2) Verb conjugation changes and noun pluralisation were scored as correct, e.g., "have" and "had", "cat" and "cats", (3) Changes in syntactic category were scored as incorrect, e.g., "the parked lorry" and "the park lorry" (adjective to noun), "apart" and "part" (adverb to noun/verb), (4) Misspelt items that were real words were scored as incorrect, e.g., "rang" and "range", (5) Misspelt items that were homophonous with the target word were scored as correct, e.g., "urge" and "urdge." For each target sentence, listeners' transcriptions were scored as a proportion of keywords correctly transcribed.

Participants were removed from the analysis if their mean score across all four conditions was below 0.2, indicative of generally poor performance (less than one word out of five). In Experiment 1, no participants were removed on this basis.

2.3.2.1 Transcription performance

Mean transcription performance by masker condition is presented in Figure 2. The same results are broken down over time, from trial 1 to trial 50, separately for the English and Mandarin maskers in Figures 3A-B.

The data were analysed with generalised linear mixed-effects regression models (GLMM) and were run in R (version 4.4.1) via RStudio (version 2024.4.2) using the *glmer* function from the *lme4* package (Bates et al., 2015). All models used a binomial distribution and a logit link, with proportion of keywords correct as the dependent variable. The BOBYQA optimiser (Baayen et al., 2008) was used to aid model convergence. The base model included only the random effects, in which both listeners and stimuli were entered as random intercepts. Including slope terms to the random structure was attempted but prevented the model from converging, suggesting that these structures were over-fitted or too complex (Bates et al., 2018). The main effects of Masker Language (coded as English: 0; Mandarin: 1), Masker Direction (coded as time-forward: 0; time-reversed: 1), and Time (trials 1 to 50, rescaled linearly as values between 0 and 1 to assist model convergence, i.e., trial 1 corresponds to 0.02, trial 2 corresponds to 0.04, etc.) were assessed by testing the improvement in model fit when each factor, considered individually, was added to the base model. Each interaction term was assessed by comparing a model with and a model without the interaction term. Improved fit between models was estimated using likelihood ratio tests. The main effect of Masker Language represents the difference between English and Mandarin masker conditions within the time-forward Masker Direction, which serves as the reference level. Similarly, the main effect of Masker Direction reflects the difference between time-forward and time-reversed masker directions within the English masker language condition, the reference for Masker Language. An interaction between Time (as a continuous variable) and the categorical variables Masker Language or Masker Direction indicates how the slope of improvement over time

differs between the levels of that categorical factor relative to the reference group. For example, this interaction reveals whether the improvement slope for Mandarin competing speech differs from that of English competing speech in the time-forward condition. Finally, an interaction between Masker Language and Masker Direction tests whether the effect of Masker Language (English vs. Mandarin) depends on the Masker Direction (time-forward vs. time-reversed), indicating if the Masker Language effect changes when the masker is time-reversed.

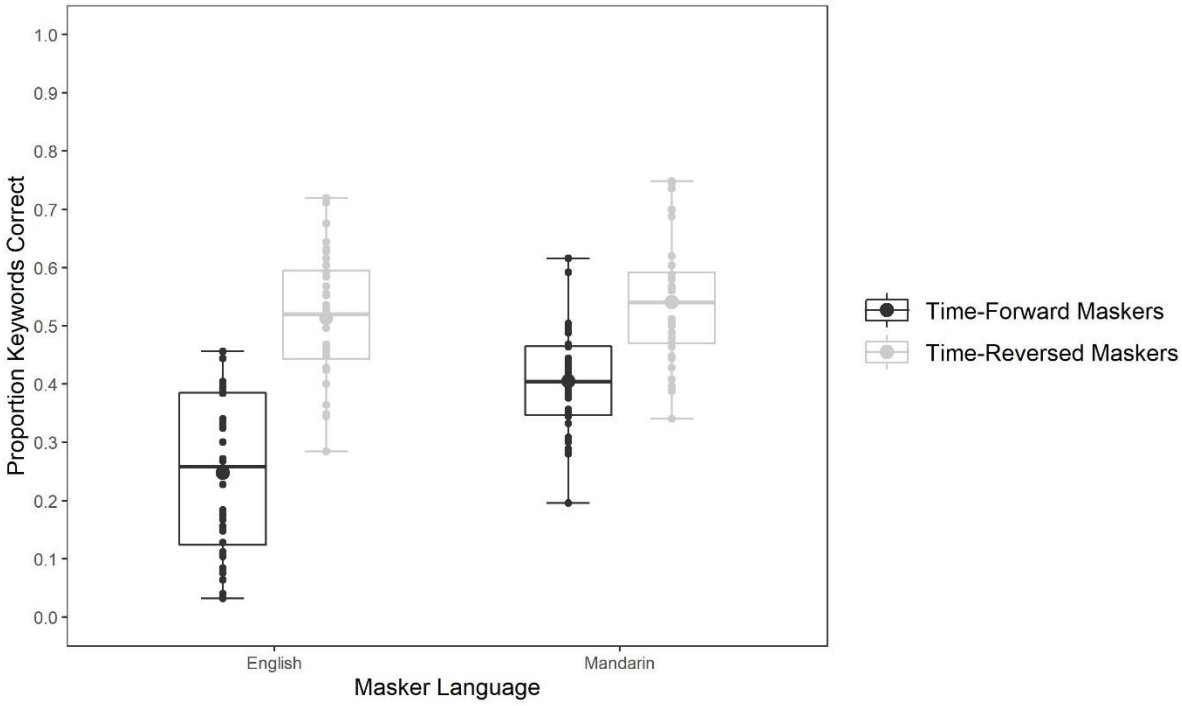


FIG. 2. Boxplot of proportion of keywords transcribed correctly as a function of Masker Language (English, Mandarin) and Masker Direction (time-forward, time-reversed). The boxes represent the interquartile range (IQR), the whiskers show 1.5 IQR over the third quartile (upper) and 1.5 IQR under the first quartile (lower), large dots represent the mean for each condition, thick horizontal bars represent the median values, and smaller dots show individual listeners.

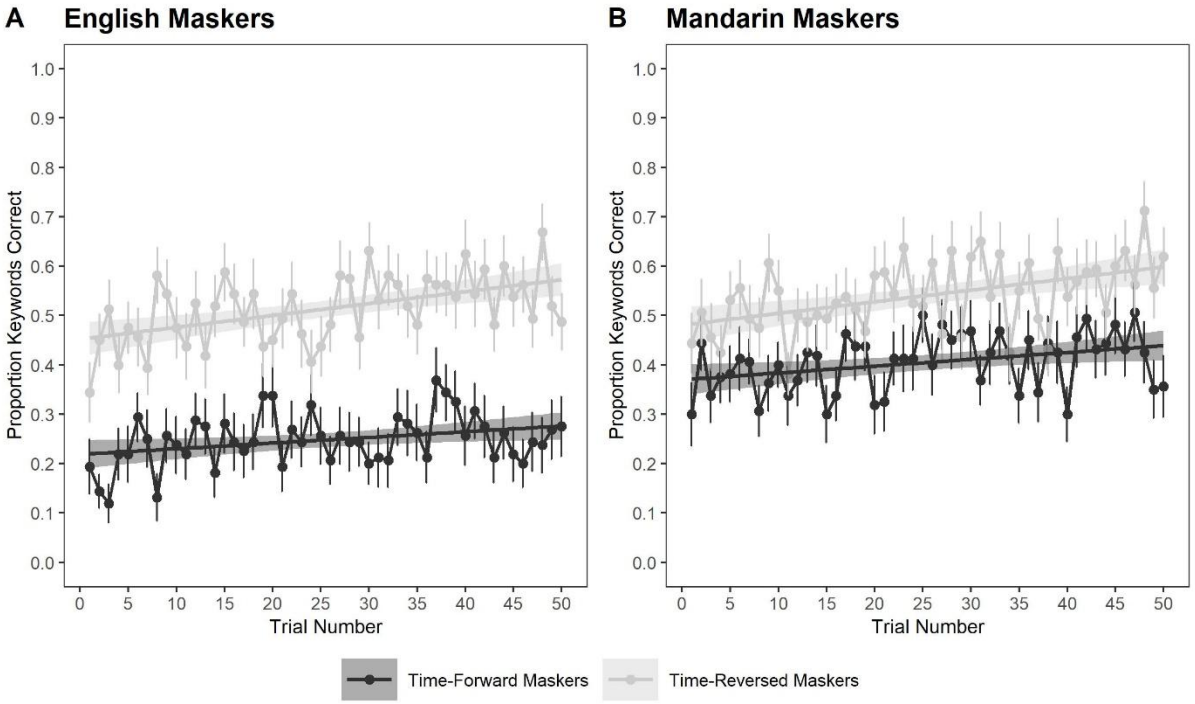


FIG. 3. Mean proportion of correct keywords for the time-forward and time-reversed masker conditions as a function of time (trials 1 to 50) for the English (A) and Mandarin (B) maskers. Error bars represent the standard error from the mean for each trial, and the shaded areas around the linear trend line represent the 95% confidence intervals of the model fit.

Model comparison showed a Masker Language effect, with better transcription performance with Mandarin maskers (M = .473, SD = .342) than English maskers (M = .381, SD = .357), β = .527, SE = .115, $\chi^2(1) = 20.68$, $p < .001$, and a Masker Direction effect, with better performance for time-reversed maskers (M = .527, SD = .343) than time-forward maskers (M = .326, SD = .333), β = -1.06, SE = .104, $\chi^2(1) = 93.24$, $p < .001$. Performance improved over time (Time effect), β = .471, SE = .046, $\chi^2(1) = 105.67$, $p < .001$.

A significant interaction between Masker Language and Masker Direction, β = .727, SE = .199, $\chi^2(1) = 13.17$, $p < .001$, showed that the masker direction effect was larger when the masker was English than Mandarin. This is an indication of linguistic interference: once long-term energetic differences were accounted for (i.e., the masker direction effect), the known

masker (English) was more detrimental to target speech recognition than the unknown masker (Mandarin). Bonferroni-corrected post-hoc pairwise comparisons indicated that all four masker conditions differed from one another (all *p*s < .001), except for time-reversed English versus time-reversed Mandarin ($p > .05$).

A significant interaction between Masker Direction and Time, $\beta = -.214$, $SE = .092$, $\chi^2(1) = 5.41$, $p = .020$, showed that, although both time-forward and time-reversed masker conditions improved over time (time-forward: $\beta = .462$, $SE = .154$, $\chi^2(1) = 8.88$, $p = .003$; time-reversed: $\beta = .734$, $SE = .148$, $\chi^2(1) = 24.85$, $p < .001$), the time-reversed condition did so more steeply. Thus, listeners found it easier to segregate the time-reversed maskers, compared to the time-forward maskers, as they progressed through the block. The three-way interaction between Masker Language, Masker Direction, and Time was not significant, $\chi^2(1) = .05$, $p = .817$, suggesting that this increase was comparable for both masker languages. The Masker Language by Time interaction was also non-significant, $\chi^2(1) = .07$, $p = .800$.

2.3.2.2 Intrusion errors

To analyse intrusions from the masker sentences into the transcription responses, we calculated the proportion of keywords originating from the masker sentences that were incorrectly reported in the listeners' transcriptions. For example, if none of the keywords of the two masker voices were reported in the transcription, the intrusion proportion was zero. If one word out of the eight keywords in the masker voices (four in masker voice 1 + four in masker voice 2) was reported in the transcription, the intrusion proportion was .125, etc. Masker keywords were scored using the BKB-R keyword list (Bench et al., 1979) following the same rules as those used for scoring the target keywords. The intrusion analysis was restricted to the time-forward English masker condition, where the target and masker languages were the same.

Using the model-comparison approach described earlier, we assessed the effect of Time on the proportion of intrusions. The variable coding for intrusions was binary, with 0 representing Correct transcriptions and 1 representing Intrusions. This coding means that the intercept and slope for Time reflect model estimates for Correct transcriptions, and any main effect of Response Type (Correct vs. Intrusions) represents the difference between Intrusions and Correct responses at the start of the time course (Time = 0). The interaction between Time and Response Type captures whether the change in the proportion of responses over time differs between Correct transcriptions and Intrusions. Time (trials 1 to 50) was rescaled linearly to range between 0 and 1 to assist with model convergence, as in the main transcription analysis. Overall, the proportion of intrusions was relatively low in absolute terms (M = .087, SD = .160), but comparatively high when set against the correct transcription scores, which probably reflects the challenge of separating target from masker speech when the masker is in the same language as the target and intelligible to the listener. Intrusions decreased over time, $\chi^2(1) = 126.68$, $p < .001$. This trend is plotted in Figure 4. The correct transcription scores for that condition are plotted as well, but for reference only since the two measures are not independent from each other. The proportion of words correctly transcribed for the time-forward English masker condition did not increase nor decrease significantly over time, β = .266, SE = .241, $\chi^2(1) = 1.12$, p = .290.

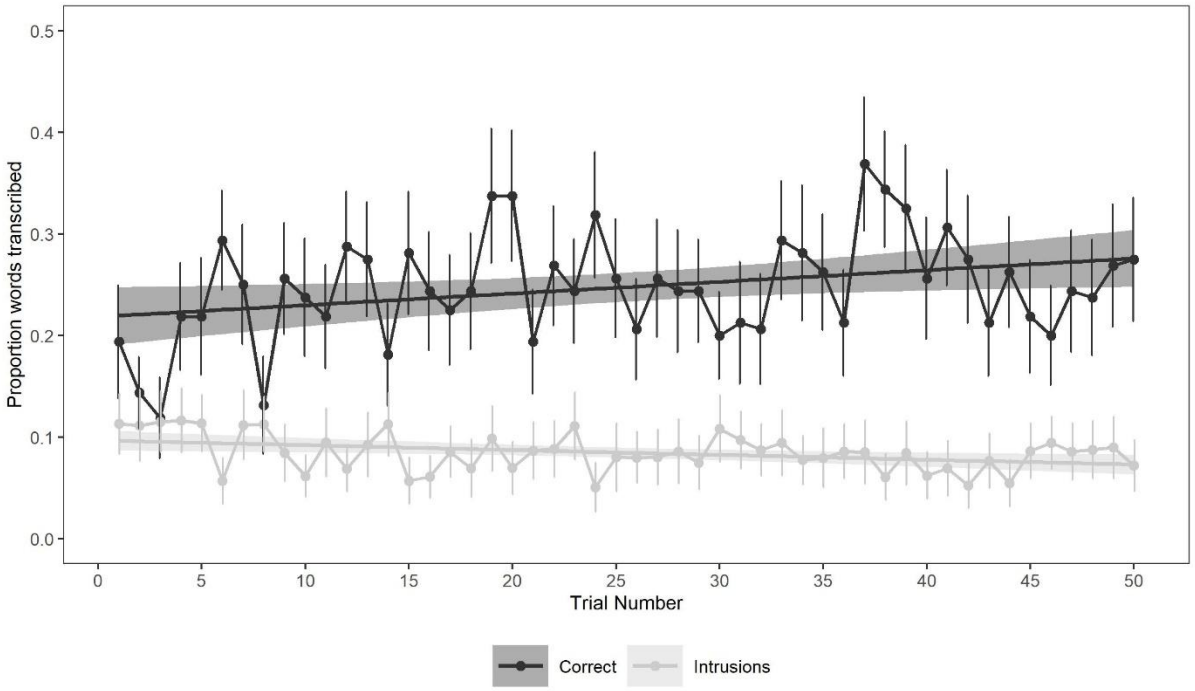


FIG. 4. Proportions of intrusions and correctly transcribed words in the time-forward English masker condition as a function of Time.

2.3.3 DISCUSSION

This experiment shows that time-forward maskers were more disruptive than time-reversed maskers, a clear demonstration that it is harder to inhibit a masker perceived as well-formed speech (whether or not it is a known language) than a speech-like masker with no semantic content that violates some natural speech patterns (e.g., Rhebergen et al., 2005). Critically, this masker direction effect was larger when the masker was in English than in Mandarin, which provides a strong illustration of linguistic interference and is consistent with the finding that a masker in a known language is more detrimental than a masker in an unknown language (e.g., Calandruccio et al., 2013; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007).

Listeners were able to learn how to segregate the target talker from maskers over the course of the experiment, but the learning trajectory varied across conditions. Improvement was faster with time-reversed than time-forward maskers, suggesting that learning to stream and inhibit a masker was easier when the masker did not display natural speech properties (time-reversed speech) than when it did (time-forward speech). The occurrence of intrusion errors in the time-forward English masker condition is further evidence that the informational content of a masker can interfere with perception. However, the lack of change in the rate of intrusions over time while speech transcription accuracy increased indicates that practice can help listeners overcome the distracting nature of the masker speech.

Although Experiment 1 provides a strong demonstration of greater interference from a known than unknown masker language, lending support to the known-language account, it does not necessarily rule out the language-similarity account. Indeed, the masker language that generated most interference (English) was both known to the listeners and matched the target language. Conversely, the Mandarin masker was both unknown to the listeners and different from the target language.

Experiment 2 attempted to tease apart these two accounts. In Experiment 2, the listeners were native Mandarin speakers with non-native knowledge of English. The target sentences were Mandarin translations of the English sentences in Experiment 1. The masker stimuli were the same English and Mandarin sentences (time-forward and time-reversed) as in Experiment 1. According to the known-language account, the masker direction effect should be comparable for English and Mandarin maskers since both languages are known to the listeners—even though one is non-native. In contrast, according to the language similarity account, the masker direction effect should be greater for the Mandarin than the English maskers since the Mandarin maskers are in the same language as the target sentences, whereas the English maskers are not. A strength of this design is that, like in Experiment 1, participants performed the task in their

native language. An additional strength is that the design rests on the expectation that, should the language-similarity account be correct, the results should be a mirror image of those in Experiment 1, thus decoupling the account itself from specific sets of stimuli.

2.4 EXPERIMENT 2: Native Mandarin listeners (with non-native English knowledge) – Mandarin target sentences -3 dB SNR

2.4.1 METHODS

2.4.1.1 Participants

This experiment was conducted online (see Section 2.4.1.3 for details). Thirty-six native Mandarin speakers (29 female) aged between 18 and 35 years (M = 26.9, SD = 4.8, two missing data points), with no known history of hearing impairments, completed the experiment. Listeners' Overall Band Score on the International English Language Testing System (IELTS) or equivalent English proficiency tests (e.g., TOEFL) was collected as a proxy measure of English proficiency. Self-reported IELTS scores ranged from 4.5 (Limited User) to 7.5 (Good User, IELTS, 2020), with a median of 6.5 (Competent User). Thirteen participants also had experience with languages other than English (Japanese, n = 5; French, n = 3; Spanish, n = 3; German, n = 2; Cantonese, n = 1; Catalan, n = 1; Shanghainese, n = 1). At the time of testing, 10 participants were based in the People's Republic of China and the remaining participants were based in other countries (United Kingdom, n = 16; Canada, n = 4; Australia, n = 2; United States, n = 1; Belgium, n = 1; France, n = 1; Spain, n = 1; Sweden, n = 1). Listeners were given the choice to participate in the experiment for either a UK Amazon voucher worth 6.00 GBP or a payment of 6.00 GBP through Prolific. All participants provided written-informed consent to take part in this study.

2.4.1.2 Materials

2.4.1.2.1 Target stimuli

The 200 Harvard/IEEE sentences used in Experiment 1 were translated into Mandarin and spoken by a female native-Mandarin speaker (Y.B., see Author's Declaration). Some keywords were altered to align with everyday Mandarin. A full list of the Mandarin-translated Harvard/IEEE sentences can be found in Appendix C. Each target sentence had five keywords, as in Experiment 1. Sentence duration ranged from 1.43 s to 3.99 s (M = 2.55 s, SD = .48 s). As in Experiment 1, the F0 and VTL of all target sentences were adjusted to a mean F0 of 210 Hz.

2.4.1.2.2 Masker stimuli

These were those of Experiment 1.

2.4.1.2.3 Target-masker mixtures

A 500 ms silent interval was added to the onset of each target sentence. Each two-talker masker stream was segmented into 4.99 s portions, which covered the 500 ms silent onset interval, the duration of the longest target sentence (3.99 s), and an additional 500 ms silent offset interval. The maskers were time-reversed from offset to onset to create the backward masker conditions. The intensity of the target sentences was normalised to 65 dB SPL, and that of the masker streams to 68 dB SPL, yielding an SNR of -3 dB, as in Experiment 1. The target-

masker mixtures were presented diotically. As in Experiment 1, Experiment 2 included four masker conditions: (1) English time-forward, (2) English time-reversed, (3) Mandarin time-forward, and (4) Mandarin time-reversed.

2.4.1.3 Procedure

The experiment was conducted online using Gorilla (Anwyl-Irvine et al., 2020) and all instructions were presented in Mandarin. Listeners were instructed to wear headphones for the duration of the experiment. Two implementations of the experiment were created, one for listeners participating through Prolific, and one for listeners participating for UK Amazon vouchers. Listeners using Prolific were instructed to provide their Prolific ID and the listeners participating for UK Amazon vouchers were instructed to provide their email address. Listeners' email addresses were stored separately from their data and were deleted following receipt of the Amazon voucher. Participants were asked if their first language was Mandarin or Cantonese and the experiment was terminated if they selected Cantonese.

Listeners were then requested to complete a demographic questionnaire and to calibrate their headphones to a comfortable level while listening to the first 20 s of a piece of classical music. To ensure that listeners were wearing headphones, we ran a headphone check following Woods et al.'s (2017) procedure. On each trial, three tones were presented, one of which was quieter than the others. Listeners were asked to select the quietest of the three tones. The use of antiphase audio for some of the tones meant that this task could only be successfully completed with stereo headphones. Six trials were presented, and listeners had to score at least 5/6 to continue with the study and were given two attempts to achieve this score. Following the headphone check, listeners were presented with instructions to allow autoplay of audio files in their web browser before the experiment began.

Listeners were requested to transcribe the target sentences in Chinese characters using their computer keyboard. The rest of the procedure was the same as in Experiment 1.

2.4.2 RESULTS

Listeners' transcriptions were scored using a Python script (available in Appendix D). As in Experiment 1, listeners' transcriptions were scored as a proportion of keywords correctly transcribed and all responses were checked by a native Mandarin speaker (Y.B., see Author's Declaration). No participants were omitted as all of them achieved an average score higher than .2.

2.4.2.1 Transcription performance

Transcription performance by masker condition is displayed in Figure 5 and broken down over time separately for the English and Mandarin maskers in Figures 6A-B.

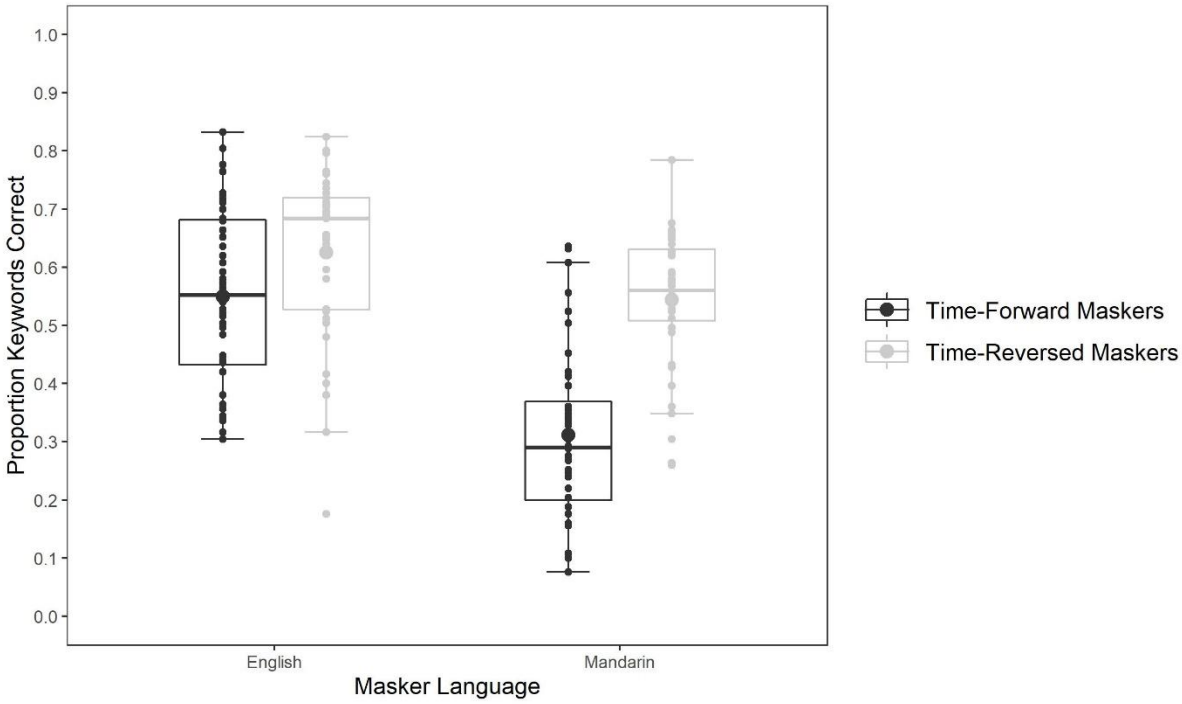


FIG. 5. Boxplot of proportion of keywords transcribed correctly as a function of Masker Language (English, Mandarin) and Masker Direction (time-forward, time-reversed).

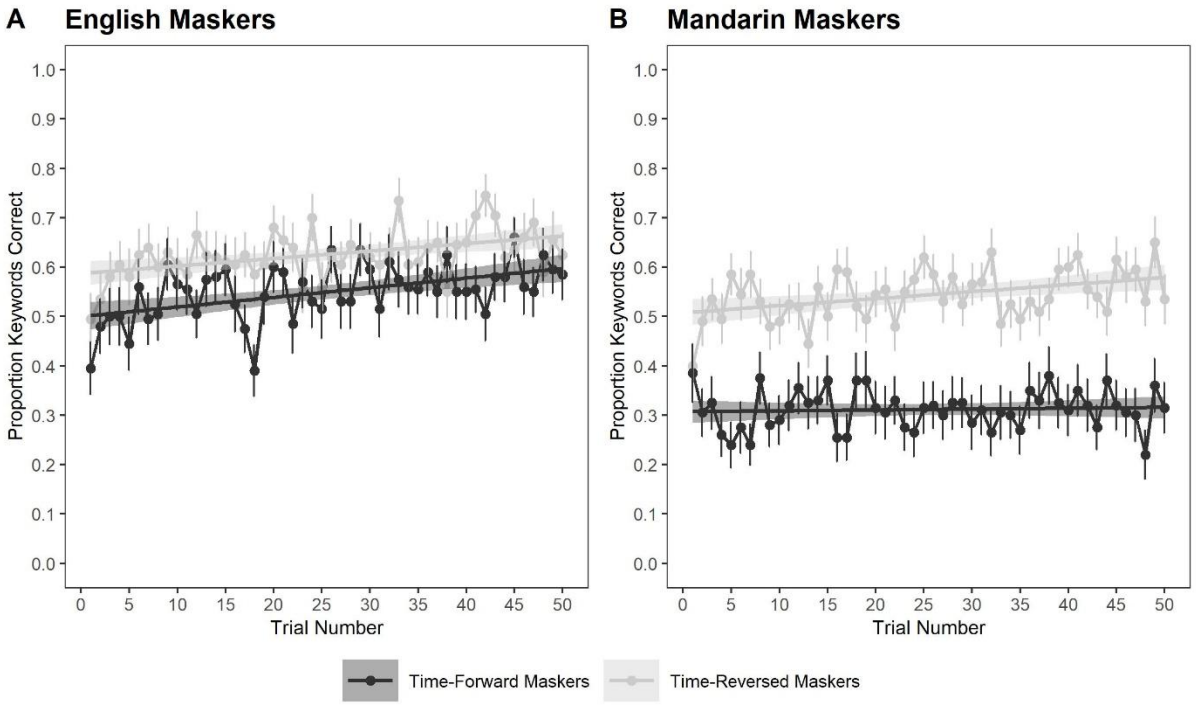


FIG. 6. Mean proportion of correct keywords for the time-forward and time-reversed masker conditions as a function of time (trials 1 to 50) for the English (A) and Mandarin (B) Masker Language conditions.

Transcription performance was analysed following the model-comparison approach used in Experiment 1. We assessed the main effects of Masker Language (coded as English: 0; Mandarin: 1), Masker Direction (coded as time-forward: 0, time-reversed: 1), Time (rescaled linearly on a scale between 0 to 1), and IELTS scores (rescaled on a scale between -0.5 and 0.5 to assist model convergence, i.e., IELTS scores of 5.5, 6.0, 6.5, 7.0, and 7.5 correspond to -0.5, -0.25, 0, 0.25, and 0.5, respectively). The main effect of Masker Language represents the difference between English and Mandarin masker conditions within the time-forward Masker Direction, which serves as the reference level. Similarly, the main effect of Masker Direction reflects the difference between time-forward and time-reversed masker directions within the

English masker language condition, the reference for Masker Language. An interaction between Time (as a continuous variable) and the categorical variables Masker Language or Masker Direction indicates how the slope of improvement over time differs between the levels of that categorical factor relative to the reference group. For example, this interaction reveals whether the improvement slope for Mandarin competing speech differs from that of English competing speech in the time-forward condition. Finally, an interaction between Masker Language and Masker Direction tests whether the effect of Masker Language (English vs. Mandarin) depends on the Masker Direction (time-forward vs. time-reversed), indicating if the Masker Language effect changes when the masker is time-reversed. The centred continuous scale of IELTS Score represents the linear association of IELTS Score on transcription performance. An interaction between IELTS Score and the dummy-coded variables (i.e., Masker Language or Masker Direction) indicates whether the effect (i.e., slope) of IELTS Score differs between the two levels of the dummy-coded variable. Similarly, an interaction between the IELTS Score and Time reflects how the rate of change over time (slope) varies as a function of IELTS Score.

Transcription performance was poorer with Mandarin maskers (M = .424, SD = .329) than English maskers (M = .578, SD = .331), β = -.580, SE = .080, $\chi^2(1) = 49.28$, $p < .001$, and poorer for time-forward maskers (M = .422, SD = .350) than time-reversed maskers (M = .577, SD = .308), β = .584, SE = .080, $\chi^2(1) = 49.88$, $p < .001$. Performance improved over time, β = .418, SE = .043, $\chi^2(1) = 92.83$, $p < .001$. There was no effect of IELTS scores, β = 114, SE = .468, $\chi^2(1) = 0.06$, p = .808.

A significant interaction between Masker Language and Masker Direction, β = .3946, SE = .103, $\chi^2(1) = 14.30$, $p < .001$, showed that, although the masker direction effect was significant in both masker languages (English: β = .479, SE = .130, $\chi^2(1) = 13.25$, $p < .001$; Mandarin: β = 1.130, SE = .133, $\chi^2(1) = 64.05$, $p < .001$), it was larger when the masker was

Mandarin, an indication of linguistic interference, and a mirror image of the results in Experiment 1. Bonferroni-corrected post-hoc pairwise comparisons indicated that all four masker conditions differed from one another (all *p*s < .001), except for time-forward English vs. time-reversed Mandarin ($p > .05$).

A significant interaction between Masker Language and Time, $\beta = -.165$, $SE = .061$, $\chi^2(1) = 7.22$, $p = .007$, showed that performance increased faster with English than Mandarin maskers. There was no interaction between Masker Direction and Time, $\beta = .008$, $SE = .061$, $\chi^2(1) = .02$, $p = .900$. However, a significant three-way interaction between Masker Language, Masker Direction, and Time, $\beta = .286$, $SE = .088$, $\chi^2(1) = 10.78$, $p = .001$, revealed a contrast in how the masker direction effect developed over time in the English versus Mandarin masker conditions. In the Mandarin masker condition, an interaction between Masker Direction and Time showed that the masker direction effect increased over time, $\beta = .220$, $SE = .087$, $\chi^2(1) = 6.35$, $p = .012$. This pattern was driven by an improvement in performance in the time-reversed condition, $\beta = .668$, $SE = .212$, $\chi^2(1) = 10.07$, $p = .002$, but not in the time-forward condition, $\beta = .293$, $SE = .224$, $\chi^2(1) = 1.72$, $p = .190$. In the English masker condition, an interaction between Masker Direction and Time revealed a small decrease in the masker direction effect over time, $\beta = -.183$, $SE = .088$, $\chi^2(1) = 4.35$, $p = .037$. Performance improved in both conditions (time-forward: $\beta = .990$, $SE = .213$, $\chi^2(1) = 22.08$, $p < .001$; time-reversed: $\beta = .728$, $SE = .216$, $\chi^2(1) = 11.50$, $p < .001$), but improvement was slightly faster in the time-forward condition.

2.4.2.2 Intrusion errors

The intrusion analysis was restricted to the time-forward Mandarin masker condition, where the target and masker languages were the same. Using the model-comparison approach described earlier, we assessed the effect of Time on the proportion of intrusions. The variable

coding for intrusions was binary, with 0 representing Correct transcriptions and 1 representing Intrusions. This coding means that the intercept and slope for Time reflect model estimates for Correct transcriptions, and any main effect of Response Type (Correct vs. Intrusions) represents the difference between Intrusions and Correct responses at the start of the time course (Time = 0). The interaction between Time and Response Type captures whether the change in the proportion of responses over time differs between Correct transcriptions and Intrusions. Time (trials 1 to 50) was rescaled linearly to range between 0 and 1 to assist with model convergence, as in the main transcription analysis. The proportions of intrusions and correctly transcribed words are plotted in Figure 7. The proportion of intrusions (M = .108, SD = .114) was similar to that in Experiment 1, but increased over time, β = .155, SE = .070, $\chi^2(1)$ = 4.85, p = .028.

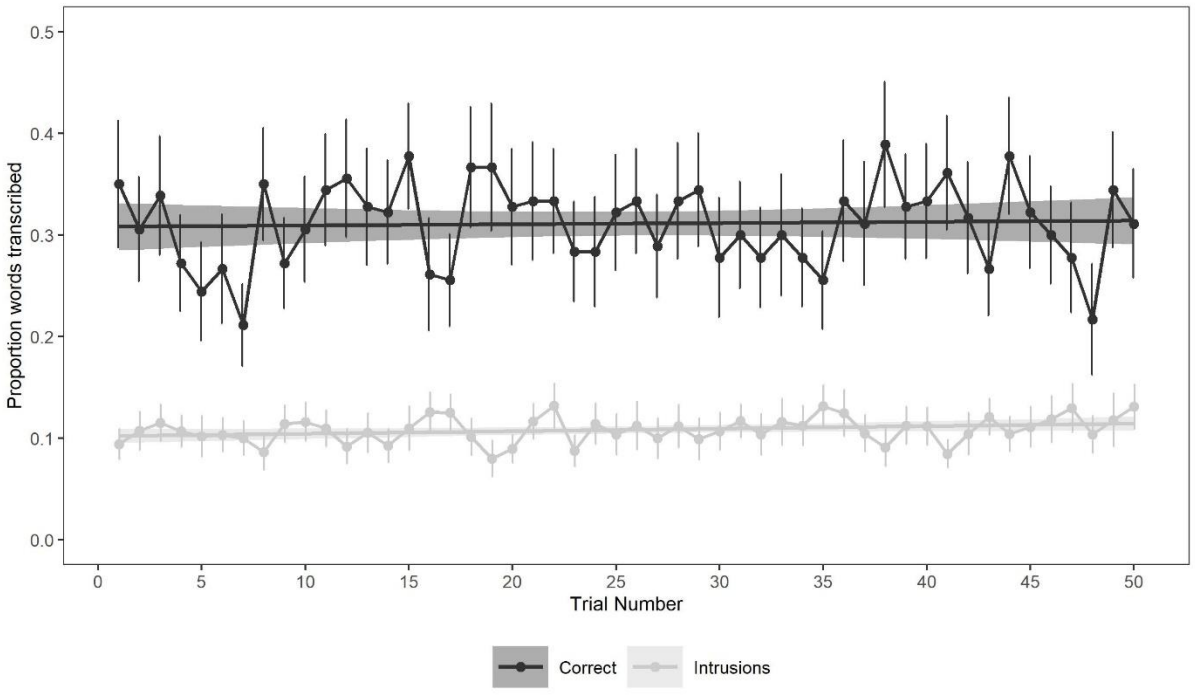


FIG. 7. Proportions of intrusions and correctly transcribed words in the time-forward Mandarin masker condition as a function of Time.

2.4.3 DISCUSSION

The results of Experiment 2 confirm and extend those of Experiment 1. First, the significant interaction between Masker Language and Masker Direction suggests that listeners experienced most linguistic interference when targets and maskers were in the same language (Mandarin). This result, which shows a mirror pattern to that in Experiment 1, provides a strong demonstration of linguistic interference independent of the test languages being used. Second, the results are, by and large, more compatible with the language-similarity account than with the known-language account. Linguistic interference was driven by whether the target and masker languages matched, not by whether the masker language was known to the listeners. Had performance been driven by the listeners' knowledge of the masker language, both maskers should have led to comparable patterns of results. The distinct effect of the two masker languages was also visible in how performance changed over time: the masker direction effect increased over time for the Mandarin masker, whereas it decreased slightly for the English masker. Thus, it was harder to learn how to overcome interference from a target-matched masker (Mandarin) than from a target-mismatched masker (English), lending further support to the language-similarity account.

It could be argued that greater interference from Mandarin than English maskers occurred not because of an overlap between target and masker languages, but because listeners performed the task in their native language, and that their proficiency was better for Mandarin than for English. In Experiment 3, we explored this possibility by testing listeners with non-native knowledge of the target language. Native Mandarin speakers with non-native knowledge of English were tested on the English target sentences of Experiment 1, with the same English and Mandarin maskers as in Experiments 1 and 2. Should Experiment 3 show greater interference from English than Mandarin maskers (as in Experiment 1), this would suggest that the language similarity account holds for any language known to the listener, whether it is

native or non-native. However, should a different pattern emerge, we would have to conclude that the language-similarity account is restricted to native listening and that a more complex model must be considered for non-native listening.

2.5 EXPERIMENT 3: Native Mandarin listeners (with non-native English knowledge) – Mandarin target sentences +1 dB SNR

Experiment 3 was followed the same procedure to Experiment 1, except that the participants were native Mandarin speakers with non-native knowledge of English and that a SNR of +1 dB was used. Following the known-language interference account, both the English and Mandarin time-forward maskers should disrupt transcription performance compared to their respective time-reversed conditions, because Mandarin speakers have access to the linguistic content of both languages. Thus, linguistic interferences from the maskers should be comparable for both masker languages – and possibly somewhat larger for the Mandarin masker since this is the native, and more proficient, language of the listeners. Following the target-masker linguistic similarity account, however, the English time-forward masker should be more disruptive than the Mandarin time-forward masker, relative to their respective time-reversed conditions, because the target and masker languages are the same in the former and different in the latter. Thus, linguistic interference should be greater in the English than Mandarin masker conditions. The effect of Time on the transcription performance and intrusion errors patterns should provide an indication of potential trade-offs between these two mechanisms during learning.

2.5.1 METHODS

2.5.1.1 Participants

Thirty native Mandarin speakers (27 female) with non-native knowledge of English aged between 19 and 33 years (M = 25.12, SD = 3.36) with no known history of hearing impairments participated in the experiment. Fourteen of them declared knowledge of at least one other language in addition to English (Japanese, n = 5; French, n = 3; Spanish, n = 3; German, n = 1; Italian, n = 1; Korean, n = 1). Experience with the additional language ranged from 3 months to 24 years (M = 8.30 years, SD = 9.56 years). The first native language of all but one listener was Mandarin, and all listeners grew up in the People's Republic of China. All 30 participants were included in the analyses. All of them were enrolled at the University of York at the time of testing. Their Overall Band Score on the International English Language Testing System (IELTS, a required test to study at the University of York) was collected as a proxy measure of English proficiency. The median Overall Band Score of the participants was 6.5 (Competent User) and their scores ranged from 5.5 (Modest User) and 7.5 (Good User; IELTS, 2020). Listeners either participated in this experiment for course credit or were compensated for their participation at a rate of 6.00 GBP per hour. All participants provided written-informed consent to take part in this study.

2.5.1.2 Materials

The target and masker stimuli were those of Experiment 1. However, to account for performance differences previously reported between native and non-native listeners (Bradlow & Alexander, 2007; Brouwer et al., 2012), the intensity of the masker sentences was lowered

to 64 dB SPL. The target sentences were played at 65 dB SPL, as in Experiment 1. Thus, in Experiment 3, the SNR was +1 dB SPL, compared to -3 dB SPL in Experiment 1.

2.5.1.3 Procedure

The procedure for Experiment 3 was identical to that of Experiment 1. Inter-rater discrepancies for Experiment 3 were less than 1% of all trials and were resolved in the same way as in Experiment 1.

2.5.2 RESULTS

2.5.2.1 Transcription performance

Transcription performance by masker condition is displayed in Figure 8. The same results are broken down over time (from trial 1 to trial 50) for the English and Mandarin maskers (Figures 9A-B).

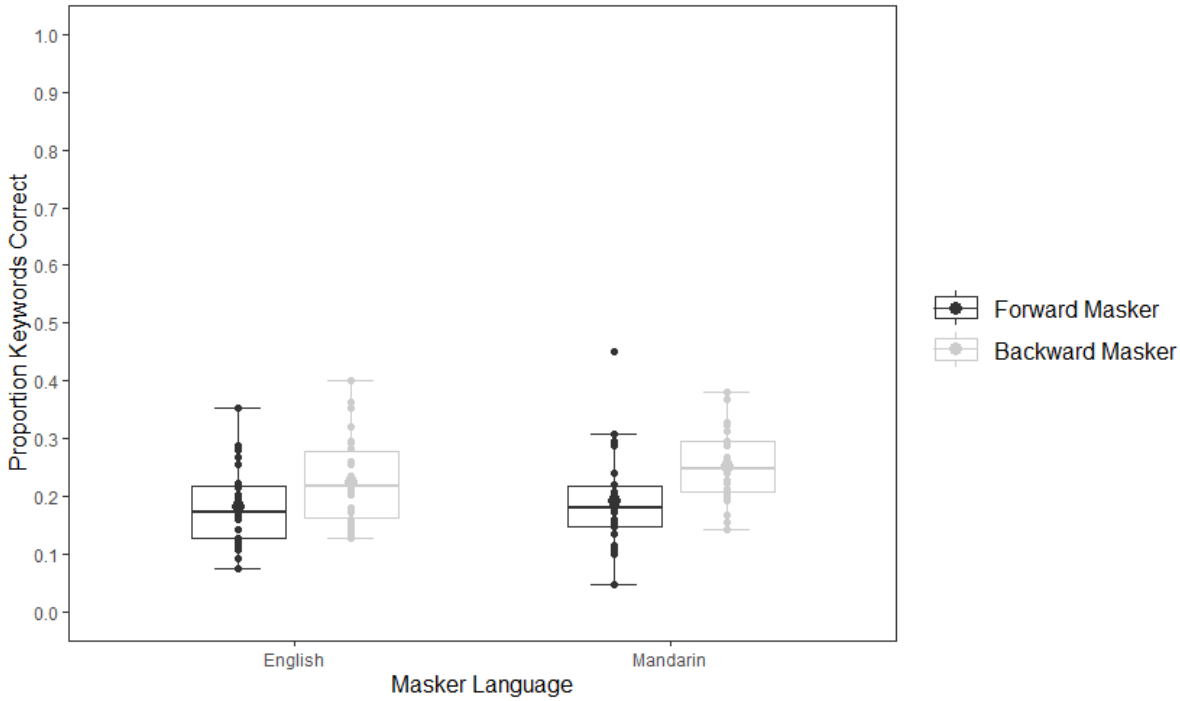


FIG. 8. Boxplot of proportion of keywords transcribed correctly as a function of Masker Language (English, Mandarin) and Masker Direction (time-forward, time-reversed).

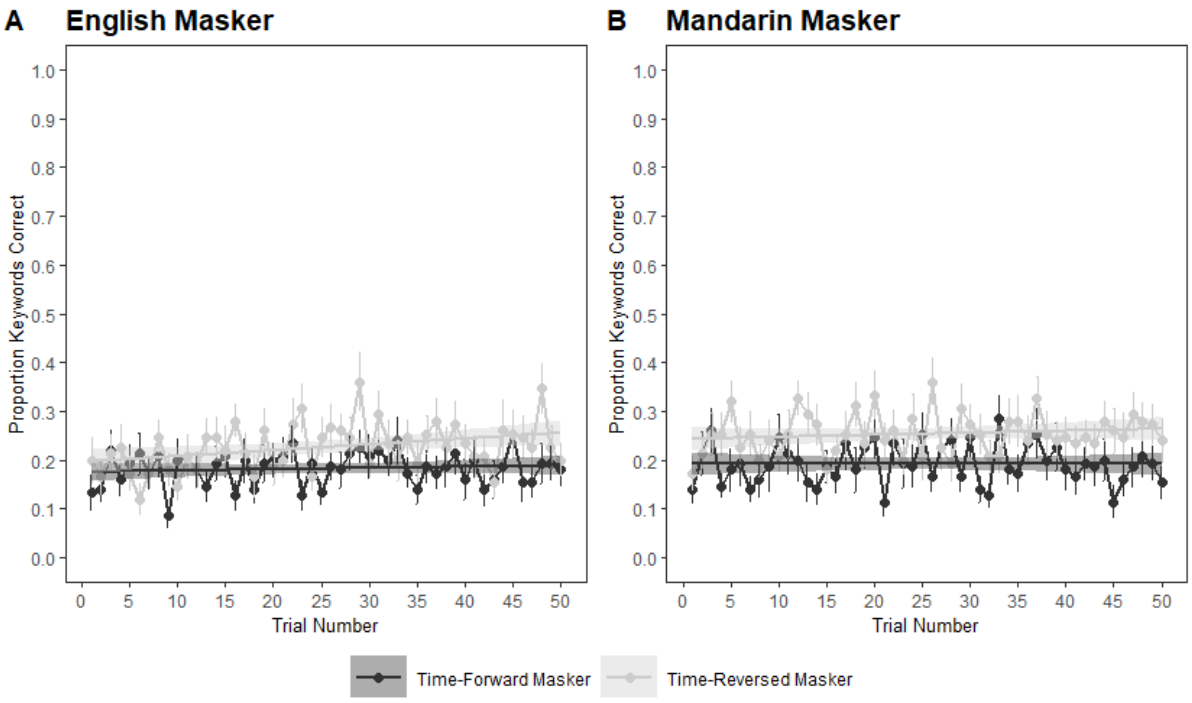


FIG. 9. Mean proportion of correct keywords for the time-forward and time-reversed masker conditions as a function of time (trials 1 to 50) for the English (A) and Mandarin (B) Masker Language conditions.

Transcription performance was analysed following the model-comparison approach used in Experiment 2. We assessed the main effects of Masker Language (coded as English: 0; Mandarin: 1), Masker Direction (coded as time-forward: 0; time-reversed: 1), Time (trials 1 to 50, rescaled as in Experiment 2), and IELTS scores (rescaled on a scale between -0.5 and 0.5 to assist model convergence, i.e., IELTS scores of 5.5, 6.0, 6.5, 7.0, and 7.5 correspond to -0.5, -0.25, 0, 0.25, and 0.5, respectively). Listeners and Stimuli were entered as random intercepts. The main effect of Masker Language represents the difference between English and Mandarin masker conditions within the time-forward Masker Direction, which serves as the reference level. Similarly, the main effect of Masker Direction reflects the difference between time-forward and time-reversed masker directions within the English masker language condition, the reference for Masker Language. An interaction between Time (as a continuous variable)

and the categorical variables Masker Language or Masker Direction indicates how the slope of improvement over time differs between the levels of that categorical factor relative to the reference group. For example, this interaction reveals whether the improvement slope for Mandarin competing speech differs from that of English competing speech in the time-forward condition. Finally, an interaction between Masker Language and Masker Direction tests whether the effect of Masker Language (English vs. Mandarin) depends on the Masker Direction (time-forward vs. time-reversed), indicating if the Masker Language effect changes when the masker is time-reversed. The centered continuous scale of IELTS Score represents the linear association of IELTS Score on transcription performance. An interaction between IELTS Score and the dummy-coded variables (i.e., Masker Language or Masker Direction) indicates whether the effect (i.e., slope) of IELTS Score differs between the two levels of the dummy-coded variable. Similarly, an interaction between the IELTS Score and Time reflects how the rate of change over time (slope) varies as a function of IELTS Score.

Transcription performance was equivalent under the Mandarin (M = .223, SD = .228) and English maskers (M = .205, SD = .218), β = .135, SE = .093, $\chi^2(1)$ = 2.07, p = .150. However, model comparison identified a Masker Direction effect, whereby transcription performance was better with time-reversed (M = .240, SD = .232) than time-forward maskers (M = .187, SD = .211), β = -.358, SE = .092, $\chi^2(1)$ = 14.90, p < .001. Performance also improved over time (Time effect), β = .200, SE = .054, $\chi^2(1)$ = 13.71, p < .001. Additionally, participants with higher IELTS scores showed higher transcription scores, β = 1.28, SE = .237, $\chi^2(1)$ = 20.40, p < .001. None of the interaction terms reached significance (all ps > .05). In particular, release from masking (time-reversed masker minus time-forward masker) was unaffected by either Masker Language or Time (Figure 8).

2.5.2.2 Intrusion errors

It was not possible to model the intrusion data in the English Forward condition due to their extremely low occurrence (less than 1%).

2.5.3 DISCUSSION

The non-native speakers showed a pattern of results broadly consistent with the idea that informational interference is determined by whether the masker language is known or unknown to the listener. Indeed, the non-native participants had comparable linguistic interference in both language masker conditions (i.e., there was no interaction between Masker Language and Masker Direction), suggesting that access to the linguistic content of either masker language impeded transcription performance. In contrast, the results are inconsistent with the idea that informational interference is determined by the similarity (or identity) between the target and masker languages. Had it been so, linguistic interference would have been greater in the English than Mandarin masker condition, as was the case for the native English listeners in Experiment 1.

Another way in which the non-native listeners differed from the native listeners was the relatively minor effect that time had on their performance. For non-native listeners, while transcription accuracy generally improved in the course of the experimental blocks, it did so at the same rate across all conditions. Thus, linguistic interference remained unchanged throughout the experiment, as did the proportion of masker-to-target intrusion errors which were negligible (< 1%). These results suggest that the transcription task might have been sufficiently demanding for non-native listeners to reduce any spare cognitive capacity for controlling their attentional allocation to targets versus maskers, and in turn improve in their ability to segregate the target from masker talkers over the course of the experiment.

However, as the non-native listeners had generally low numbers of intrusion errors, interference from the masker is unlikely to be the primary mechanism by which speech-in-speech masking occurs in non-native listeners. While the proportion of keywords transcribed depended on the participants' IELTS scores, the lack of interaction between IELTS scores and any of the other variables suggests that linguistic interference and the rate of improvement over time were largely independent of the listeners' degree of English proficiency.

In this experiment, the performance of the native Mandarin speakers when listening in a non-native language was much lower across conditions than in Experiments 1 and 2, where participants were listening in their native language. In Experiment 3 we increased the SNR from -3 dB to +1 dB, where an increase of 4 dB SNR has been shown to elicit similar performance between participants listening in their native compared to non-native language (Bradlow & Alexander, 2007; Brouwer et al., 2012). Although the listeners in this experiment had generally high proficiency in their non-native language as measured in their IELTS scores, their proficiency still might have been lower than in the Brouwer et al. (2012) and would require an even more favourable SNR to achieve performance parity to participants listening in their native language. The generally lower scores in Experiment 3 might have prevented any differences to become apparent if performance across all conditions was located at the lower end of the performance spectrum. Other studies have found that non-native listeners need an even more favourable SNR to achieve the same performance as native listeners (+6 dB, Cooke et al., 2008; +8 dB, Van Engen, 2010). Thus, an equivalent experiment was run where the SNR between the target and masker talkers was increased to +6 dB SNR.

2.6 EXPERIMENT 4: Native Mandarin listeners (with non-native English knowledge) – English target sentences +6 dB SNR

2.6.1 METHODS

2.6.1.1 Participants

This experiment was conducted online. Thirty-two native Mandarin speakers (26 female, five male, one did not disclose) aged between 19 and 34 years (M = 26.1, SD = 4.0) with no known history of hearing impairments completed the experiment. All participants had non-native knowledge of English. All but three of them had lived in the UK—from nine months to 10 years 11 months (n = 21, M = 3.7, SD = 3.2). Unlike in Experiments 2 and 3, in which all participants provided an English proficiency score, not all of them did in Experiment 4. For listeners who declared their IELTS score (n = 21), the median was 7.0 (Good User), ranging from 6.0 (Competent User) to 8.5 (Very Good User; IELTS, 2020). IELTS scores were slightly higher in this experiment (M = 6.98, SD = 0.78) than in Experiment 3 (M = 6.33, SD = 0.95), $t(55) = 2.76$, $p = .008$. Thirteen participants declared knowledge of at least one other language in addition to English (Japanese, n = 6; French, n = 3; German, n = 3; Spanish, n = 3; Malay, n = 2; Cantonese, n = 1; Portuguese, n =1, Russian, n = 1). Listeners either participated in this experiment for course credit or were compensated at a rate of 6.00 GBP per hour. All participants provided written-informed consent to take part in the study.

2.6.1.2 Materials

The target and masker stimuli were those of Experiment 1. However, to account for performance differences previously reported between native and non-native listening (Bradlow & Alexander, 2007; Brouwer et al., 2012), the intensity of the masker sentences was lowered

to a more favourable SNR. After piloting various SNRs, the intensity of the masker speech was set to 59 dB SPL. The target sentences were played at 65 dB SPL, as in Experiments 1-3. Thus, in Experiment 4, the SNR was +6 dB, compared to -3 dB in Experiments 1 and 2, and +1 dB in Experiment 3. The target-masker mixtures were presented diotically.

2.6.1.3 Procedure

The procedure was identical to that of Experiment 3.

2.6.2 RESULTS

2.6.2.1 Transcription performance

Transcription performance by masker condition is displayed in Figure 8 and broken down over time separately for the English and Mandarin maskers in Figures 9A-B. Two participants with mean performance lower than 0.2 were removed from subsequent analyses (n = 30).

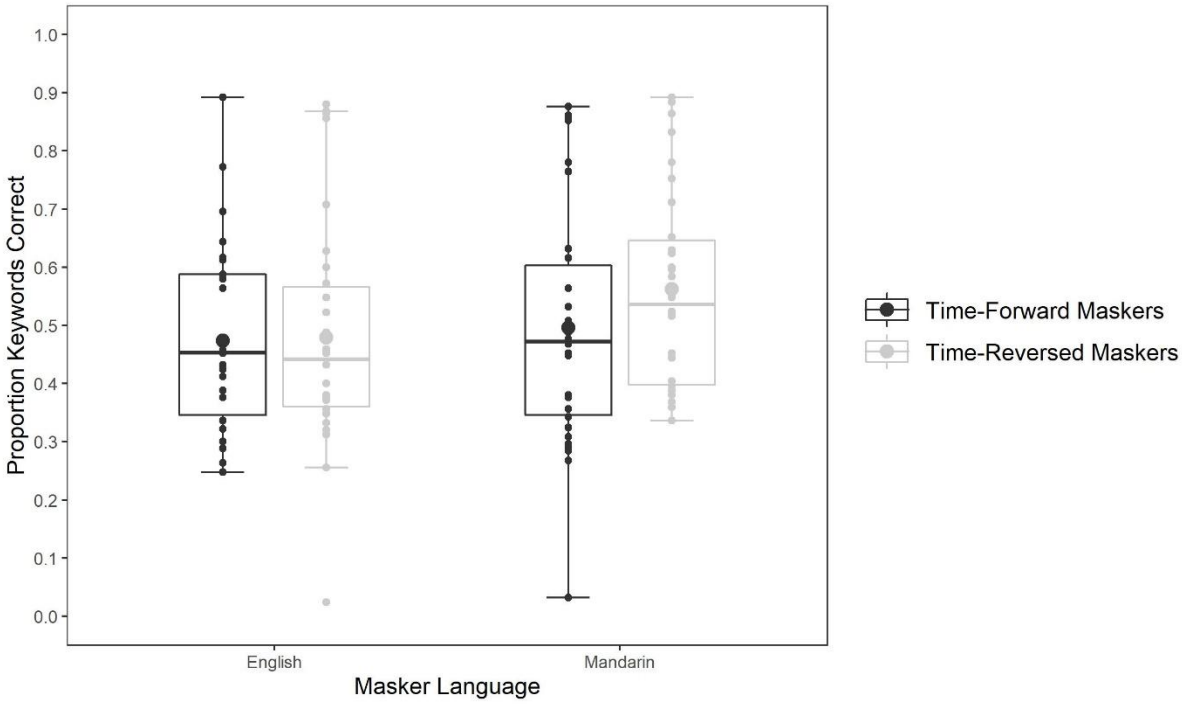


FIG. 10. Boxplot of proportion of keywords transcribed correctly as a function of Masker Language (English, Mandarin) and Masker Direction (time-forward, time-reversed).

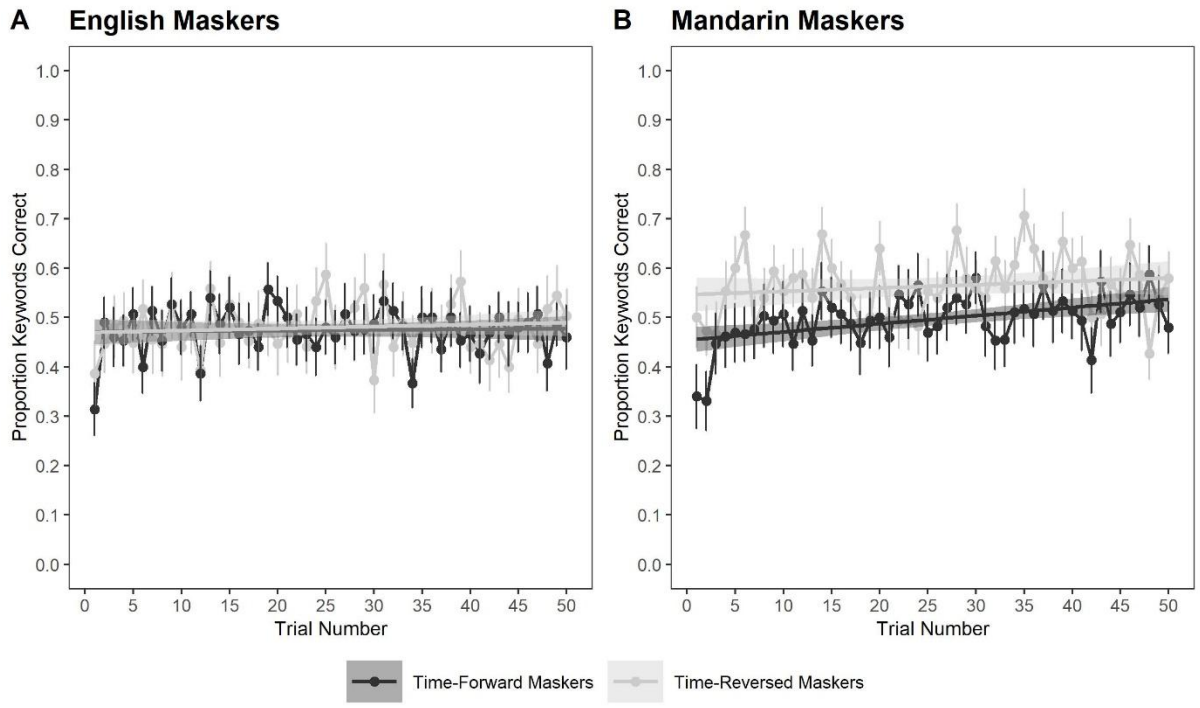


FIG. 11. Mean proportion of correct keywords for the time-forward and time-reversed masker conditions as a function of time (trials 1 to 50) for the English (A) and Mandarin (B) Masker Language conditions.

Transcription performance was analysed following the model-comparison approach used in Experiment 3. We assessed the main effects of Masker Language (coded as English: 0; Mandarin: 1), Masker Direction (coded as time-forward: 0; time-reversed: 1), Time (trials 1 to 50, rescaled as in Experiment 3), and IELTS scores (rescaled on a scale between -0.5 and 0.5 to assist model convergence, i.e., IELTS scores of 5.5, 6.0, 6.5, 7.0, and 7.5 correspond to -0.5, -0.25, 0, 0.25, and 0.5, respectively). Listeners and Stimuli were entered as random intercepts. The main effect of Masker Language represents the difference between English and Mandarin masker conditions within the time-forward Masker Direction, which serves as the reference level. Similarly, the main effect of Masker Direction reflects the difference between time-forward and time-reversed masker directions within the English masker language condition, the reference for Masker Language. An interaction between Time (as a continuous variable)

and the categorical variables Masker Language or Masker Direction indicates how the slope of improvement over time differs between the levels of that categorical factor relative to the reference group. For example, this interaction reveals whether the improvement slope for Mandarin competing speech differs from that of English competing speech in the time-forward condition. Finally, an interaction between Masker Language and Masker Direction tests whether the effect of Masker Language (English vs. Mandarin) depends on the Masker Direction (time-forward vs. time-reversed), indicating if the Masker Language effect changes when the masker is time-reversed. The centered continuous scale of IELTS Score represents the linear association of IELTS Score on transcription performance. An interaction between IELTS Score and the dummy-coded variables (i.e., Masker Language or Masker Direction) indicates whether the effect (i.e., slope) of IELTS Score differs between the two levels of the dummy-coded variable. Similarly, an interaction between the IELTS Score and Time reflects how the rate of change over time (slope) varies as a function of IELTS Score.

Transcription performance was poorer with Mandarin (M = .53, SD = .19) than English maskers (M = .48, SD = .18), β = .300, SE = .091, $\chi^2(1) = 10.81$, p = .001, and poorer with time-forward (M = .48, SD = .19) than time-reversed maskers (M = .52, SD = .19), β = -.221, SE = .091, $\chi^2(1) = 5.80$, p = .016. Performance improved over time, β = .147, SE = .047, $\chi^2(1) = 9.59$, p = .002. None of the two-way interactions reached significance (all ps > .05), but there was a significant three-way interaction, β = .416, SE = .189, $\chi^2(1) = 4.83$, p = .028: The Masker Direction by Time interaction was significant in the Mandarin masker condition, β = .349, SE = .135, $\chi^2(1) = 6.66$, p = .010, but not in the English masker condition, β = -.081, SE = .133, $\chi^2(1) = 0.37$, p = .545. Thus, the masker direction effect decreased over time in the Mandarin masker condition, with performance improving in the time-forward condition, β = .413, SE = .096, $\chi^2(1) = 18.36$, p < .001, but not in the time-reversed condition, β = .048, SE = .096, $\chi^2(1) = 0.25$, p = .617. In the English masker condition, neither the time-forward, β = .036, SE =

.094, $\chi^2(1) = 0.15$, $p = .700$, nor the time-reversed condition, $\beta = .122$, $SE = .097$, $\chi^2(1) = 1.55$, $p = .213$, changed with time.

In an attempt to account for English proficiency, the data were re-analysed for the listeners who provided IELTS scores (21 minus the two participants with performance scores < .2, $n = 19$). Only results involving the IELTS factor are reported. An effect of IELTS scores showed that high IELTS scores were associated with better transcription performance, $\beta = 1.59$, $SE = .401$, $\chi^2(1) = 11.46$, $p < .001$. However, an interaction between IELTS scores and Masker Language, $\beta = .0.312$, $SE = .112$, $\chi^2(1) = 7.73$, $p = .005$, showed that this association was only present in the English masker condition. There was also a significant four-way interaction between Masker Language, Masker Direction, Time, and IELTS scores, $\beta = -1.914$, $SE = .814$, $\chi^2(1) = 5.46$, $p = .019$. However, the patterns arising from this interaction did not lend themselves to a straightforward interpretation, probably due to the small sample size, and are therefore not reported.

2.6.2.2 Intrusion errors

It was not possible to model the intrusion data in the English Forward condition due to their extremely low occurrence (less than 2%).

2.6.3 DISCUSSION

Unlike in Experiments 1 and 2, but similarly to Experiment 3, Experiment 4 did not show a marked contrast between English and Mandarin maskers in terms of the masker direction effect. In other words, for the non-native listeners of Experiment 4, there was no evidence of linguistic interference. Instead, for that group, the results are more consistent with

the idea that the masker direction effect is determined by whether the masker language is known to the listener, as per the known-language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007) rather than by whether the masker language matches the target language, as per the language-similarity account (Brouwer et al., 2012; Calandruccio et al., 2013; Freyman et al., 1999; Van Engen & Bradlow, 2007). Had the latter been true, the masker direction effect would have been greater in the English than Mandarin masker condition, as was the case for the native English listeners in Experiment 1 and (in a mirrored fashion) for the native Mandarin listeners in Experiment 2.

Experiment 4 was run because the results from Experiment 3 might have been confounded due to the generally low performance across conditions. However, across both Experiments 3 and 4 there was a similar pattern of results, whereby the listeners experienced similar levels of linguistic interference in the time-forward conditions irrespective of the language of the masker. This similarity across experiments suggests that the idea still holds that if one can understand the linguistic content of the masker, this will interfere to the same extent when listening in a non-native language, i.e., the known-language account. It could be argued that the wide individual differences in Experiment 4 compared to the previous experiments made the interaction between Masker Language and Masker Direction more difficult to find. However, a visual inspection of the data (and the three-way interaction) in Experiment 4 shows that the language direction effect was, in fact, numerically larger for the Mandarin masker condition, rather than for the English masker condition. Therefore, if anything, there is evidence that the masker more likely to interfere with the task was the masker known natively by the listener rather the masker overlapping with the target language. Moreover, as the results from Experiment 4 mirror the results from Experiment 3 of equivalent interference in known-language maskers, the results from these experiments can be interpreted with confidence as indicating that there is a known-language effect when listening to known

but non-native speech in adverse conditions, whereby any language known to the listener will interfere to the same extent.

Another way in which the non-native listeners in Experiment 4 differed from the native listeners in Experiments 1 and 2 but was similar to those in Experiment 3 was the relatively minor effect that time had on performance. While performance generally improved over the course of Experiment 4, as it did in all previous experiments, this effect was driven mainly by an improvement in the time-forward Mandarin masker condition, that is, the masker direction effect decreased as listeners adapted to the native masker. The lack of masker direction effect in the English masker condition suggests that the non-native listeners managed to inhibit the non-native masker from the very beginning of the experiment to similar extents across, even though the masker was the same language as the target. The very low rate of intrusions from the English masker supports that conclusion and echo the results from Experiment 3.

2.7 GENERAL DISCUSSION

The aim of this study was to investigate how listeners' knowledge of the linguistic content of competing speech impedes target speech perception, and how this effect is amenable to change through exposure over the course of an experiment. We tested two listener groups (native listeners in Experiments 1 and 2 and non-native listeners in Experiments 3 and 4) who differed in their knowledge of the linguistic content of competing speech (English versus Mandarin maskers). English and Mandarin maskers were played in their original format (time-forward) or backward (time-reversed), with the time-reversed condition providing a baseline for any long-term spectral and energetic masking differences between the two masker languages. Linguistic interference was measured as the difference in masker direction effect (time-forward versus time-reversed) between the English and Mandarin masker conditions.

2.7.1 Transcription Performance

For the native listeners, the masker direction effect was largest when the masker was the same language as the target speech and when it was known to the listener (Experiment 1). This result is consistent with both the known-language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007) and the language-similarity account (Van Engen & Bradlow, 2007; Brouwer et al., 2012; Calandruccio et al., 2013, 2017). However, this pattern persisted even when the listeners had knowledge of both masker languages (Experiment 2), which suggests that the results cannot be accounted for entirely by the idea that knowledge of the masker language is the driving force behind linguistic interference, as per the known-language account. Experiment 2 rather fits with the literature showing that speech-in-speech recognition is most impaired when a masker shares speech characteristics (e.g., phonology, prosody) with the target language and, by implication, when the target and masker languages are the same, as per the language similarity account.

Our finding shows both dissimilarities and similarities with Calandruccio et al.'s (2010) study. When Calandruccio et al. (2010) used the same SNR as we did (-3 dB), they did not find greater interference from an English than Mandarin masker during native English listening. However, at a more challenging SNR (-5 dB), their results and ours aligned in showing greater interference from the English masker. Interestingly, our average performance at -3 dB SNR was within the range of their average performance at -5 dB SNR, in fact even lower. Thus, their study and ours converge in showing linguistic interference when the listening conditions are challenging. However, the two studies differ in how energetic differences between the maskers were handled. In our study, we attempted to minimise potential energetic differences by using time-reversed speech as a relative baseline, whereas Calandruccio et al. (2010) compared English and Mandarin maskers with each other directly. However, in subsequent analyses, they considered their data in the context of masker-modulated noise analogues and LTAS profiles.

Those analyses confirmed that, in the easy SNR condition, energetic masking differences could account for linguistic interference, whereas, at the more challenging SNR, they could not.

In contrast with the native listening results of Experiments 1 and 2, the participants who performed the task non-natively (Experiments 3 and 4) did not show greater interference from the target-matched masker (English) than the target-mismatched masker (Mandarin). This suggests that, for non-native listeners, it is the knowledge of the linguistic content of the masker that drives interference, rather than linguistic similarities between target and masker languages. This finding challenges the language similarity account and, instead, supports the known-language account. These data provide an interesting counterpoint to a study by Calandruccio and Zhou (2014), who assessed the recognition of English sentences against English and Greek maskers by monolingual English speakers versus English-Greek bilinguals. They found greater interference from the English masker in both groups, which led them to conclude that it is the similarity between target and masker languages, rather than the listener's knowledge of the masker language, that drives linguistic interference. However, a difference between their experiment and ours was that, while our participants in Experiments 3 and 4 were clearly non-native speakers of English, and could therefore be said to perform the task non-natively, the participants in Calandruccio and Zhou (2014) were simultaneous bilinguals from Greek descent who were born in the USA and started learning both English and Greek from birth or shortly after. Those participants effectively performed the task natively. The finding of a language-similarity effect in that group is therefore consistent with our claim that native listening conforms to the language-similarity account, whereas non-native listening conforms to the known-language account.

Related to this point, it is important to note that, in Experiments 3 and 4, listeners' English proficiency appeared to have influenced transcription accuracy, with high-proficiency listeners performing better than low-proficiency listeners. Although our interpretation of the

contribution of language proficiency to linguistic interference is limited by the relatively small number of IELTS scores available in our experiments, these preliminary patterns confirm the need to consider the language proficiency of listeners performing tasks in their non-native language (see Scharenborg & van Os, 2019, for a review; see von Hapsburg & Peña, 2002, for methodological considerations).

It is worth considering an alternative explanation for the smaller masker direction effect and linguistic interference during non-native than native listening. Recall that, in order to match average performance between the two conditions, we had to set the SNR at -3 dB for native listening in Experiments 1 and 2, and at +1 dB and +6 dB for non-native listening in Experiments 3 and 4 respectively. It has been shown that, for native speakers at least, increased SNR is generally associated with both better performance and lower informational masking, with misallocations of masker content less likely to occur when streaming is made easier by a more advantageous SNR (e.g., Arbogast et al., 2005; Freyman et al., 2008). However, if the higher SNR in the non-native listening condition had made streaming easier, one would also have expected performance to be higher. Still, performance was comparable in both groups. Therefore, although SNR differences should be considered in future research, we do not think that the reduced linguistic interference and lower intrusion rate during non-native listening can be entirely explained by the higher SNR.

2.7.2 Improvement over time

One of the goals of this study was to explore the changes in the masker direction effect and in linguistic interference over the course of the experiment. While transcription performance improved for listeners performing the task either natively (Experiments 1 and 2) or non-natively (Experiments 3 and 4), the change in the masker direction effect differed between the two groups. The native listeners showed an increase in the masker direction effect

over time in most conditions, with a faster rate of improvement for time-reversed than time-forward speech. Thus, they were better at learning to suppress a masker that did not conform to natural speech patterns than compared to one that did. However, for the bilingual listeners in Experiment 2, this pattern was not found when the masker was in English, i.e., the target-mismatched masker. For that condition, improvement was slightly better for time-forward than time-reversed maskers. For listeners performing the task in a native language, experience with bilingualism might therefore confer some ability to inhibit the interference of a known non-native language.

The general increase in masker direction effect for the native listeners is incompatible with the hypothesis that life-long familiarity with time-forward speech (even in an unknown or non-native language) accelerates object formation over the course of an experiment, hence sharpening *object-based auditory attention* (Shinn-Cunningham, 2008) more for time-forward than time-reversed maskers. Instead, the slower improvement with time-forward than time-reversed maskers is consistent with Bent et al.'s (2009) observation that perceptual adaptation is poorer in multi-talker babble than compared to noise-vocoded maskers. Bent et al. (2009) ascribed this difference to the novelty of the noise-vocoded speech, and hence, its learnability over time, compared to the lower potential for learnability of an already familiar signal such as babble noise. Applied to our results, the logic would be that time-reversed speech, which is unfamiliar to most listeners, would have more learnability potential than time-forward speech. For instance, the unfamiliar spectral-temporal characteristics of certain phonemes when played backward (Rhebergen et al., 2005), while distracting at first, could be learned over time and, ultimately, make it easier for the time-reversed speech to be interpreted as an auditory object distinct from the target speech.

Additionally, although both time-forward and time-reversed speech mixtures contained local spectro-temporal modulations that can result in opportunities to 'glimpse' the target

speech (Brungart et al., 2006; Cooke, 2006; Festen & Plomp, 1990), local glimpsing opportunities may be different in time-forward and time-reversed maskers (Buss et al., 2020). Opportunities to glimpse the target speech in the masker stream may be enhanced in the time-reversed masker condition as there is no additional interference from the linguistic content of the masker. Regardless of whether the familiarity or the glimpsing explanation is correct, our results are at odds with the hypothesis that listeners can learn to overcome the masker direction effect with practice. Thus, the distracting nature of a well-formed masker (whether it is a known or unknown language) might be largely automatic, at least during native listening.

In contrast to the listeners performing the task in their native language (Experiments 1 and 2), the non-native listeners in Experiment 3 showed no change in the masker direction effect over time. Similarly, the non-native listeners in Experiment 4 showed no change in the masker direction effect for the English masker and a small decrease for the native but target-mismatched masker (Mandarin). We hypothesise that the better ability to overcome the masker direction effect over time during non-native listening is the unintended consequence of the high level of effort involved in listening to a non-native language. Previous studies have shown that non-native speakers experience greater listening effort than native speakers when listening to speech in both quiet and background noise (Borghini & Hazan, 2018) and when listening to speech spoken by either native or non-native speakers (Song & Iverson, 2018). Increased effort due to non-native listening could have resulted in the listeners undertaking the transcription task at the limit of their cognitive capacity, exhausting the resources that native listeners might otherwise use to involuntarily process distractor information (Lavie et al., 2004, Perceptual Load Theory). In such conditions, whether the masker was time-forward or time-reversed, or whether or not it matched the target language, would have had little impact. Thus, native versus non-native differences in prioritisation of cognitive resources to the main transcription task could explain why there was a large and increasing interference in Experiments 1 and 2, and

minimal linguistic interference in Experiments 3 and 4—but see also considerations about SNR differences in the previous section.

2.7.3 Intrusions

The intrusion of masker words into listeners' responses was analysed in the masker condition where the target and masker languages were the same (time-forward English in Experiments 1, 3 and 4 and time-forward Mandarin in Experiment 2). For the native English listeners (Experiment 1), intrusions neither increased nor decreased throughout the block although accuracy correspondingly increased. For the native Mandarin listeners (Experiment 2), intrusions showed a small increase across trials. The presence of intrusions in these two groups and the fact that they traded off or mirrored the accuracy data suggests that native listeners were unable to fully inhibit the linguistic content of the masker, even though there was some evidence that they learned to partly overcome its effects through practice as target talker transcription increased. In contrast, non-native listening (Experiments 3 and 4) led to almost no intrusion errors at all, consistent with the absence of linguistic interference in the accuracy data.

Lavie et al.'s (2004) perceptual load theory can, again, be drawn upon to explain the difference in intrusions between the two groups. In this framework, high perceptual load reduces distractor interference because it exhausts the resources needed to process the relevant stimuli, leaving little capacity for processing distractor stimuli. This hypothesis has received some support from Francis' (2010) demonstration that increasing perceptual load (from an easy to a hard tone-perception task) reduced the interference of a competing voice on target speech perception. Under the assumption that the phonology and prosody of a non-native target language constitute a situation of high perceptual load, such perceptual load would guard

against interference (i.e., intrusions) from the irrelevant stimulus (maskers). This could explain why the masker direction effect was small and intrusions were almost non-existent for the listeners performing the task in a non-native language. Heightened investment of processing resources towards the transcription task in the non-native listeners (Borghini & Hazan, 2018; Song & Iverson, 2018) and performance being possibly at its peak for that group would also mean that little spare capacity was left for improvement in transcription performance over time.

2.7.4 Limitations

Experiments 1, 2, and 4 were published together as a series of experiments in the *Journals of the Acoustical Society of America* (JASA; Mepham et al., 2022). Experiment 3, although part of the same series of experiments, had performance at too low a level to be interpretable alone and was justification for running Experiment 4 at a higher SNR. Various studies have shown that non-native listeners need a more favourable SNR than native listeners to achieve performance parity (Bradlow & Alexander, 2007; Brouwer et al., 2012; Cook et al., 2008; Van Engen, 2010). Brouwer et al. (2012) found that a higher SNR of +4 dB was needed to achieve performance parity between participants listening to speech-in-noise in their second language compared to their first language. However, this was not observed in Experiment 3 for participants listening in their second language compared to Experiments 1 and 2 where participants listened in their first language. Performance in the time-reversed conditions in Experiment 3, the control conditions in which performance should be easiest as there was no intelligible competing speech, was around 24% compared to Experiments 1 and 2 in which performance in the time-reversed conditions was around 53% and 58% respectively. For Experiment 4, where we increased the SNR, piloting was undertaken to assess at what SNR participants would achieve performance parity in the time-reversed conditions for the non-

native listeners, which was +6 dB SNR in Experiment 4 with performance around 52%. The difference in SNR between Experiment 4, and Experiments 1 and 2 was thus +9 dB SNR, similar to the +8 dB SNR observed in the speech recognition task by Van Engen (2010) and over double the magnitude of the difference found by Brouwer et al. (2012).

One explanation why a much higher SNR was necessary for the non-native listener groups could be that the non-native listener groups were less homogeneous in Experiments 3 and 4 compared to the Brouwer et al. (2012) participants. The Dutch-English bilingual participants in the Brouwer et al. (2012) study were described as having 'on average ten years of English lessons starting at age 11' (p. 1455). Although there was no measure of English proficiency in the Brouwer et al. (2012) study, the proficiency of these participants may have been higher than the English proficiency of the participants in Experiments 3 and 4 presented here. The lower proficiency of some participants may have skewed the participants' performance to be much lower than expected, with no participants scoring above 50% accuracy in any of the conditions in Experiment 3. The University of York, where the participants were recruited in Experiment 3, requires postgraduate students to have an IELTS score of 5.5 or above, corresponding to a 'Modest User' of English (IELTS, 2020), which could be lower than the English proficiency of participants in other experiments. For example, participants in the Bradlow and Alexander (2007) study has a minimum English proficiency score of 560 on the pencil-and-paper TOEFL examination, which corresponds to an IELTS minimum score of 6.5 (ETS, 2010), and thus had a higher proficiency than the participants in Experiments 3 and 4, who had a minimum IELTS score of 5.5 and 6.0 respectively. It is recognised that controlling for language proficiency is necessary when conducting speech perception experiments (Scharenborg & van Os, 2019) and having comparable measures of language proficiency between experiments might allow us to elucidate whether proficiency is an underlying factor in speech perception in noise performance across experiments.

Conversely, it might not be the language proficiency of the listeners that resulted in a more favourable SNR for non-native listeners to achieve performance parity with native listeners, but a facet of the different languages themselves. The participants in the Brouwer et al. (2012) experiments were Dutch-English bilinguals whereas these experiments tested Mandarin-English bilinguals. Dutch and English are more linguistically and phonetically similar than Mandarin and English, and monolingual English listeners experience greater release from masking when listening to English speech in Mandarin than Dutch maskers (Calandruccio et al., 2013). With Mandarin being by nature a more linguistically 'distant' language to English compared to Dutch, this might have reduced the opportunity for the Mandarin-English bilinguals to exploit linguistically and phonetically similar cognates than languages 'closer' to Mandarin, even if the vocabulary of the target sentences had not been encountered by participants before. Additionally, the Mandarin-English bilingual participants in the Van Engen (2010) study had a mean TOEFL English proficiency score of 106.2, equating to IELTS Scores between 7.5 (Good User) and 8.0 (Very Good User; ETS, 2010), but still required a +8 dB more favourable SNR to achieve the same performance as native listeners. Thus, this higher SNR required for non-native listeners to achieve the same performance as native listeners might simply result from the distances between languages known by bilingual listeners rather than primarily the English proficiency of bilingual listeners.

An alternative explanation of the higher SNR needed for non-native listeners compared to native listeners in our experiments than in the Brouwer et al. (2012) study could originate from the test stimuli and procedure. In Brouwer et al. (2012), the target stimuli were BKB-R sentences (Bench et al. 1979), which were compiled for use with partially-hearing children, whereas in our experiments the target stimuli were IEEE sentences (IEEE, 1969) which were initially compiled for use to test the speech quality of audio technology. In the BKB-R sentences, the number of target keywords ranges from 3-4, and were of a vocabulary familiar

to children. In the IEEE sentences used in these experiments, five keywords were needed to be transcribed by participants, some of which might be unfamiliar to listeners where English is not a native language. Additionally, participants in the Brouwer et al. (2012) study were required to repeat aloud the target sentences, whereas in our experiments participants needed to transcribe the sentences that they heard, which may have added additional demands to the task in these experiments compared to repeating the sentences aloud. The additional task demands of reporting the target stimuli as well as the target sentences themselves potentially being unfamiliar might have in turn required the SNR to be higher for the non-native listening group to reach performance parity compared to the native listeners.

One last explanation for why the SNR needed to be higher could have been the heterogeneous listening environments in our experiments. Experiments 1 and 3 were conducted in-person, while Experiments 2 and 4 were conducted online, a result of the SARS-CoV-2 pandemic restrictions. Although we made as much effort as possible to control for listening environment for participants completing the experiment online, such as including headphone checks (Woods et al., 2017) and opportunity for participants to adjust their volume to a comfortable listening level, the listening environments inevitably differed between the listeners. These different listening environments may have resulted in the varying levels of performance observed in Experiment 4. However, if the higher SNR is a result of heterogeneous listening conditions, one would expect to see large variances in performance levels in Experiment 2, which was also conducted online. Instead, we saw similar variances in performance across participants listening in their native language in both Experiment 2 conducted online and Experiment 1 conducted in-person. Additionally, the English proficiency of participants in Experiment 4 was generally higher than the proficiency of participants in Experiment 3, with some listeners having proficiency categorised as ‘Very Good User’ (IELTS, 2020). Taken together, the above explanations suggest that the higher SNR required

for non-native listeners to perform on par with native listeners in speech-in-noise perception is not a result of heterogeneous listening environments, but a result of the nature of the transcription task being potentially more difficult than other speech repetition tasks (Brouwer et al., 2012) and the level of English proficiency of participants in our experiments being not only heterogeneous, but also potentially poorer than in other speech-in-noise experiments comparing native and non-native listening (Bradlow & Alexander, 2007).

2.8 CONCLUSION

This study investigated the effect of masker direction (time-forward vs. time-reversed speech) and masker language over the course of an experiment with native English speakers and native Mandarin speakers with non-native knowledge of English. Better performance with time-reversed maskers than time-forward maskers was found for English and Mandarin listeners performing the task in their native language, and this masker direction effect was particularly pronounced when the masker language was the same as the target language. This result supports the target-masker linguistic similarity hypothesis (Brouwer et al., 2012; Calandruccio et al., 2013; Freyman et al., 1999; Van Engen & Bradlow, 2007), whereby speech-in-speech interference is maximal when the target and masker languages share characteristics (e.g., phonology, prosody) and, by implication, when they are the same language, as was the case here. Furthermore, for listeners performing the task in their native language, the masker direction effect increased over the course of the experiment, which suggests that it is easier to learn to inhibit time-reversed speech than time-forward speech. For listeners performing the task in a non-native language, the masker direction effect was broadly equivalent across the two known masker languages, thus supporting the known-language interference hypothesis (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007).

There were also more intrusion errors during native than non-native listening. We hypothesise that listening in a non-native language might force listeners to engage a large proportion of their cognitive capacity toward the target speech and, as a consequence, reduce opportunities for distraction by the masker.

2.9 ACKNOWLEDGEMENTS

This research was supported by a research grant from the Leverhulme Trust (RPG-2019-152) to S.L.M. We are grateful to Sarah Knight and Lyndon L. Rakusen for their advice concerning the set-up of the experimental paradigm, and to Sarah Knight, Ronan McGarrigle, and Sophie Meekings for feedback on earlier drafts. We also thank Lauren Calandruccio and Ann R. Bradlow for providing the English BKB-R sentences and Mandarin audio files for the BKB-R sentence translations, and Sarah Knight for supplying relevant Praat scripts. We also thank Miaomiao Yu for recording the English and Mandarin BKB-R sentences and for translating the BKB-R sentences alongside Lydia Y. Li. We are also grateful to the speakers who provided speech samples for the initial pilot study. All audio stimuli, analysis scripts, data files, and Appendices can be found at https://osf.io/nhcrw/.

3 CHAPTER 3[2]: THE TIME-COURSE OF PUPILLOMETRIC MEASURES OF LISTENING EFFORT DURING SPEECH-IN-SPEECH LISTENING

3.1 ABSTRACT

Studies of speech-in-speech listening show that intelligible maskers are more detrimental to target speech perception than unintelligible maskers, an effect we refer to as linguistic interference. Research also shows that performance improves over time through adaptation. The extent to which the rate of adaptation differs for intelligible and unintelligible maskers, and whether this pattern is reflected in changes in listening effort, are open questions. In this pre-registered study, native English listeners reported aloud what they could hear of English sentences against an intelligible masker (time-forward English talkers) versus an unintelligible masker (time-reversed English talkers). Over 50 trials, speech recognition accuracy and task-evoked pupil response (TEPR) were recorded, along with self-reported effort and fatigue ratings. In Experiment 5, we used an adaptive procedure to ensure a starting performance of ~50% correct in both conditions. In Experiment 6, we used a fixed signal-to-noise ratio (SNR: -1.5 dB) for both conditions. And in Experiment 7, we used an adaptive procedure to approximate a starting performance at ~66% in the intelligible maskers, and ~41%

[2] At the time of submission, material in Chapter 3 was published in Mepham, A., Knight, S., McGarrigle, R. A., Rakusen, L., & Mattys, L. M. (2025). Pupillometry reveals the role of SNR in adaption to linguistic interference over time. *J. Speech Lang. Hear. Res.*, *68*(5), 2291-2317. https://doi.org/10.1044/2025_JSLHR-24-00658. Supplementary material for Chapter 3 can be found for Experiment 5 at https://osf.io/m4b57/, for Experiment 6 at https://osf.io/pmhsx/, and at Experiment 7 at https://osf.io/6hkp9/.

in the unintelligible maskers. The experiments showed performance patterns consistent with linguistic interference and masker adaptation. However, the rate of adaptation depended on the SNR. When the SNR was higher for the intelligible masker condition, resulting from the adaptive procedure (Experiments 5 and 7), the rate of adaptation was faster for that condition; TEPRs were not affected by trial number or condition in Experiment 5, but were generally higher in the intelligible condition in Experiment 7, even though the SNR was higher in this condition. When the SNR was fixed (Experiment 6), adaptation was similar in both conditions but TEPRs decreased faster in the unintelligible than intelligible masker condition. Self-reported ratings of effort and fatigue were not affected by masker conditions in either experiment. The ease with which listeners learn to segregate target speech from maskers depends on both the intelligibility of the maskers and the SNR. We discuss ways in which these factors interact to promote auditory stream segregation and the extent to which such a process is automatic or requires cognitive resources.

3.1 INTRODUCTION

Listeners face various challenges when listening to speech in a background of competing talkers. The target signal can be degraded due to spectro-temporal overlap from the competing speech, creating interference at the cochlear level (*energetic masking*, e.g., Culling & Stone, 2017). In this case, performance is determined primarily by the extent to which the target signal can be 'glimpsed' through the masker (Barker & Cooke, 2007) in regions of reduced spectro-temporal overlap, which in turn depends in part on the signal-to-noise ratio (SNR) between the target and the masker. Listening can also be compromised by non-energetic properties of the competing signal (*informational masking*, Cooke et al., 2008; Kidd & Colburn, 2017). Informational masking can take various forms, including misallocation of

masker components to the target speaker (phonetic features, segments, words) and attentional capture due to phonological or semantic familiarity with the masker (Cooke et al., 2008; Summers & Roberts, 2020). Attentional capture by the masker is often illustrated by the fact that it is more difficult to understand a target speaker when the language of the competing talkers is known to the listener than when it is unknown (Brouwer et al., 2012; Calandruccio et al., 2013; Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Kilman et al., 2014; Van Engen & Bradlow, 2007). Familiarity with the masker is thought to draw a listener's attention to recognisable features of the masker, and hence, distract them from the to-be-attended signal. We refer to this specific type of informational masking as 'linguistic interference,' which is the topic of the experiments presented in this chapter.

While the above studies have investigated defining characteristics of linguistic interference, less is known about how linguistic interference changes over time. Studies examining listener adaptation to distorted speech indicate that performance can change radically over the course of an experiment as a function of the type of distortion. For instance, Cooke et al. (2022) showed rapid adaptation in the beginning of an experiment, with performance subsequently plateauing for eight different types of masker. Bent et al. (2009) found that the point at which the plateau is reached depends on the type of adverse condition. In their experiment, perception of eight-channel noise-vocoded speech started at 70% and plateaued around ~83% after 60 sentences, whereas speech perception in six-talker babble at 0 dB SNR started at 67% correct and plateaued around 74% after 40 sentences. However not all distortions benefit from repeated exposure. Lie et al. (2024) found learning effects in temporally- and spectrally-modulated noise but less so in stationary noise and for degradations with a low speech reception threshold (SRT; see also Rhebergen et al., 2006; Versfeld et al., 2021).

Critical for the question of linguistic interference, Mepham et al. (2022) showed that improvements in sentence transcription were slower when the competing talkers were intelligible to the listeners (time-forward speech in an unknown language) than when they were unintelligible (time-reversed speech or speech in an unknown language). Thus, learning to ignore an intelligible masker was harder than learning to ignore an unintelligible masker, presumably because of the sustained informational masking caused by familiar intelligible components of the masker speech. Note, however, that Versfeld et al. (2021), who measured changes in SRT over 87 sentences, did not find substantial differences in SRT improvement between time-forward and time-reversed maskers, suggesting that the effect observed by Mepham et al. (2022) might be sensitive to methodological considerations (transcription performance versus SRT) and listener engagement with the task.

It is unclear whether differences in adaptation as a function of masker intelligibility found by Mepham et al. (2022) are reflected in changes in listening effort. Since performance and effort do not necessarily pattern together, measures of effort may provide complementary information about cognitive resource allocation that is not reflected in accuracy scores (Kuchinsky et al., 2013; Winn & Teece, 2021; Winn et al., 2015). Of particular interest is whether Mepham et al.'s (2022) faster improvement for the unintelligible masker condition might come at the cost of increased effort or, alternatively, whether the growing familiarity with the task and stimuli might make the task less rather than more effortful over time. The former would suggest that effort is a compensatory mechanism, with higher performance achieved at the expense of higher effort, while the latter would suggest that effort shows a simple inverse relationship to performance, with higher performance requiring lower effort (e.g., Ohlenforst et al., 2017; Wendt et al., 2018; Winn et al., 2018; Wu et al., 2016; Zekveld et al., 2018).

There are two alternative mechanisms that could be operational when trying to attend to a target talker while ignoring competing talkers. The first is that the segregation of a target talker from distracting maskers occurs automatically and at a low, sensory level, with stream segregation being a 'primitive process', i.e., stream segregation occurring automatically and at the outset of perceiving an auditory scene without requiring attention (Bregman, 1990). Sussman and colleagues (Sussman 2005; Sussman & Winkler, 2001; Sussman et al., 1999, 2002) have used this 'primitive process' hypothesis to describe how changes in tones belonging to distinct streams are processed in auditory scene analysis. However, as speech comprehension is a more high-level process than detecting changes in tone presentation, more cognitive resources might be required for speech-in-noise recognition. One would then expect either a low magnitude of listening effort with very little changes over time, or reductions from an initially high level of listening effort (resulting from the initial adaptation to a new task or environment) to lower levels in measures of listening effort.

An alternative mechanism to the 'primitive process' that accounts for this directed attention is that perceptually segregating a target talker from competing speech requires a listener's conscious effort to attend to the target talker in a top-down manner. Neuroimaging studies have found an interaction between top-down directed attention and bottom-up processing of auditory stimuli, with the prefrontal cortex active during top-down controlled attention shifts (Rule et al., 2002; Salmi et al., 2009) and have provided evidence for the *dynamic filtering theory* (Shimamura, 2000) proposing that the prefrontal cortex is directly involved in the selection, maintenance, updating, and rerouting of information processing. This effect whereby attention is required for rapid perceptual learning has also been shown when listening to speech (Huyck & Johnsrude, 2012). Huyck & Johnsrude (2012) found that listeners benefited from training to vocoded speech when they attended to the training, compared to when listeners experienced no listening training, nor when listeners were distracted by other

auditory or visual stimuli. The authors then suggest that attention to degraded speech is necessary for learning, rather than attending to speech being an automatic or passive process. One would then observe measures of listening effort reflecting this top-down mechanism throughout the time in which a listener is engaged in speech-in-noise perception, either by measures of listening effort remaining stable and at a high magnitude (to cope with the sustained level of conscious effort to stream one talker from other talkers) or by an increase in effort reflecting the cumulative conscious effort to attend to the target talker.

Listening effort can be assessed using pupillometry (for reviews, see Van Engen & McLaughlin, 2018; Zekveld et al., 2018). The extent of pupil dilation is modulated by the interplay of the sympathetic and parasympathetic nervous systems (Loewenfeld & Lowenstein, 1999), with pupil size sensitive to a range of extrinsic factors, including emotional and cognitive processes (Granholm & Steinhauer, 2004; Steinhauer et al., 2004). Importantly for the present purposes, tasks requiring more cognitive effort have been shown to result in greater pupil dilation (Granholm et al., 1996). Furthermore, changes in pupil size appear to correlate with changes in a person's self-reported tiredness from listening in adverse listening conditions (McGarrigle et al., 2021). Pupillometric measures have been used to assess cognitive effort when listening to speech in modulated noise (Koelewijn et al., 2012, 2014a; McLaughlin et al., 2021; Ohlenforst et al., 2018; Paulus et al., 2020; Wendt et al., 2018), time-compressed speech (Paulus et al., 2020), noise-vocoded speech (Paulus et al., 2020; Winn et al., 2015), accented speech (Brown et al., 2020; McLaughlin & Van Engen, 2020), multi-talker babble (Koelewijn et al., 2012, 2014a; Ohlenforst et al., 2018; Wendt et al., 2018), non-native speech (Borghini & Hazan, 2018), and trained versus untrained voices (Biçer et al., 2023). In each case, increased listening demands (caused by, e.g., more adverse SNRs or accented speech) were reflected in greater task-evoked pupil dilation (TEPR), at least when intelligibility was moderate to good.

Most pupillometry experiments aggregate pupil response patterns across many trials in order to capture sensitivity to a particular manipulation such as a challenging SNR or divided attention (e.g., Borghini & Hazan, 2018; Koelewijn et al., 2014a, 2014b; Wendt et al., 2018; Winn et al., 2015). However, alongside the behavioural studies on temporal adaptation to adverse conditions described above, recent pupillometry studies have shown that the effort required to process degraded speech or noise-masked speech decreases over the course of a test block (Brown et al., 2020; Paulus et al., 2020; Versfeld et al., 2021). For instance, Brown et al. (2020) explored changes in peak pupil dilation (PPD) over 50 trials while participants listened to native English or Mandarin-accented English speech. PPD in the early trials was larger in the non-native- than native-accented condition, which suggests that listening to non-native-accented speech was initially more effortful. However, PPD decreased faster in the non-native condition, which indicates that, while listeners initially expended more effort to deal with the difficulty of mapping accented sounds to native phonemic categories, they quickly adapted to the new mapping, hence requiring less effort. This effect can be thought of as a form of 'levelling-out' between the two conditions, with the easy and hard conditions eventually involving comparable levels of effort. Thus, investigating listening effort over time can reveal dynamic processes that aggregated data cannot.

Although the above studies provide clear evidence for a general decrease in listening effort over time, their interpretive value in terms of complementarity or trade-off between performance and effort is limited. Indeed, in the Brown et al. (2020) study, transcription performance was not a factor of interest, with performance well over 90% in both the native- and the non-native-accented conditions. Rather, the study focused on how well participants performed in a concurrent visual task and on their pupil size, two measures of effort. Likewise, while Paulus et al. (2020) compared a range of degraded conditions, none of them allowed an interpretation specific to the informational content of the masker independent of its energetic

content. While Versfeld et al. (2021) found that speech perception in noise performance increased over time (measured as decreases in speech reception threshold, SRT), there was no systematic differences in pupillometric measures of listening effort, potentially resulting from the experimental design of using SRT to measure changes in behavioural performance.

In the present study, following Mepham et al.'s (2022) approach, we investigated how listeners adapt to the informational content of competing speech over time. Across 50 trials, we compared both speech recognition and pupillometric changes in native English speakers listening to target English sentences in English two-talker babble (time-forward intelligible masker) compared to the same English two-talker babble played backwards (time-reversed masker). Specifically, we probed how listeners adapt to 'linguistic interference,' the difference between intelligible (time-forward) and unintelligible (time-reversed) maskers, and how the effort employed by the listeners to manage those two maskers changes over time. Note that the difference between time-forward and time-reversed maskers allows linguistic interference to be assessed while controlling for the long-term average frequency spectra of the two maskers, i.e., their average energetic masking (Mepham et al., 2025). Of interest is whether listeners' effort tracks the ease of speech recognition over time (i.e., high performance associated with low effort) or, rather, reflects compensatory mechanisms (i.e., high performance associated with high effort). We were also interested to see if the levelling-out pattern observed by Brown et al. (2020) between easy and hard conditions generalises to time-reversed and time-forward maskers. Specifically, we asked whether effort would start higher in the time-forward than time-reversed condition, but drop to comparable levels after 50 trials. The number of 50 trials chosen for these experiments were based on similar studies that reported learning effects plateauing between 40 and 50 sentences (Bent et al., 2009), around 40 sentences (Lie et al., 2024), and between 30 and 60 sentences (Versfeld et al., 2021).

Mepham et al. (2022) and Paulus et al. (2020, masking condition) used fixed SNRs across participants and conditions. Although this approach means that long-term energetic masking is controlled across conditions, differences in performance between conditions are likely to be present at the start of each block, which makes the true effect of time difficult to compare between conditions. To address this limitation, our study included an initial adaptive procedure which established participants' individual 50% SRT in the time-forward and time-reversed conditions. This meant that participants started the two conditions at equivalent performance levels. We chose 50% because it is the performance level at which effort has been shown to peak (Wendt et al., 2018), in addition to helping mitigate the risk of floor or ceiling effects. Unlike the Brown et al. (2020) and Paulus et al. (2020) studies, we also included subjective measures of effort and fatigue. McGarrigle et al. (2020) found that, although subjective ratings of effort did not change over time, subjective ratings of fatigue did, with reported fatigue increasing over the course of an experiment. Self-report measures of effort and fatigue can reveal complementary information to physiological measures (Alhanbali et al., 2019; Strand et al., 2018) by shedding light on the perceived costs of adaptation to linguistic interference.

Our hypotheses are as follows: First, we predict that sentence transcription will improve over time in both conditions, reflecting listeners' ability to better stream targets from maskers as familiarity with the task and stimuli increases (Bent et al., 2009; Cooke et al., 2022; Erb et al., 2012, 2013; Mepham et al., 2022). In particular, following Mepham et al. (2022), we expect that the improvement will be more pronounced in the time-reversed condition because this condition does not elicit linguistic interference. Second, we predict that TEPR will decrease linearly in both conditions reflecting decreasing effort as participants become familiar with the task and stimuli, which would be in line with previous research (Brown et al., 2020; Paulus et al., 2020) and would align with the 'primitive process' account posited by Bregman (1990). A

key question, however, is whether the decrease in TEPR will be less or more pronounced for the time-forward than the time-reversed condition. A less pronounced decrease for the time-forward condition would reflect the sustained cognitive demands imposed by linguistic interference and would be in line with the expected performance pattern. Alternatively, a more-pronounced decrease could occur because the familiarity of the time-forward masker, while initially a disadvantage, could make it easier over time to identify the masker as a competing object, hence facilitating listeners' ability to ignore it as the block progresses. The latter pattern would be consistent with the levelling-out results reported by Brown et al. (2020). Third, we predict that subjective effort ratings will be higher in the time-forward condition, due to saliency of the intelligible masker's linguistic content, but will not show significant changes over time in either condition (McGarrigle et al., 2020), whereas fatigue ratings will increase linearly in both conditions, as per McGarrigle et al. (2021b). However, the increased cognitive demands imposed by sustained linguistic interference may result in a steeper increase in subjective fatigue over time in the time-forward than the time-reversed condition.

3.2 EXPERIMENT 5

3.2.1 METHODS

3.2.1.1 Participants

Forty native British English listeners (10 male, 29 female, one non-binary) aged between 18 and 30 years (M = 21.10, SD = 3.48) with no known history of hearing impairments participated in the experiment. Four listeners described their language status as bilingual from birth, speaking British English and an additional language. One of the 40 participants was excluded from the pupil analyses due to a high proportion of missing pupil data. Using the

Westfall et al. (2014) power analysis approach, it was determined that 39 participants were required to achieve statistical power of 0.9 to reach an effect size d = .4, with 100 stimuli in a counterbalanced design (n = 50 stimuli in each condition). Details can be found in the preregistration documents referenced in the Acknowledgements section. All participants had pure-tone audiometry (PTA) measures ≤ 20 dB HL at 0.5, 1, 2, and 4 kHz (M = 4.63, SD = 3.33). The University of York Department of Psychology ethics committee approved all experimental procedures for this and the following experiments (ethics reference number: 747). Listeners either participated in this experiment for course credit or were compensated for their participation at a rate of 6.00 GBP per hour. All participants provided written informed consent to take part in this study.

3.2.1.2 Equipment

Listeners completed the experiment in a single-walled sound-attenuated room. PTA testing was conducted using a Kamplex Diagnostic Audiometer AD 25. During the main listening task, listeners were positioned 65 cm away from a 24-inch LCD flat monitor, which displayed a fixation cross. The listener's head was stabilised on a head-and-chin-rest which was secured to the edge of a table. Stimulus presentation was programmed using a bespoke Python script in PsychoPy (Pierce et al., 2019). Auditory stimuli were presented via Denon DJ DN-HP700 headphones. A microphone was positioned inside the test booth so that verbal responses could be heard and scored online by the experimenter who listened via headphones. Pupil size was recorded using the EyeLink 1000 Plus at a sampling rate of 500 Hz.

Pupil size was recorded for the right eye only. It was captured as a continuous recording for each trial. Pupil size was recorded as an integer number corresponding to the number of thresholded pixels in the camera's pupil image. Typical pupil area can range between 100 and

10,000 units, with a precision of 1 unit, corresponding to a resolution of 0.01 mm for a 5 mm pupil diameter. The desktop-mounted eye-tracker camera was positioned between the listener and the computer monitor at a distance of 55 cm from the listener (at 0° azimuth angle). The eye-tracker camera was aligned to the centre of the computer monitor and was positioned just below the bottom of the monitor to maximise the trackable range without obscuring the listener's view of the monitor.

3.2.1.3 Materials

3.2.1.3.1 Target stimuli

The target stimuli were taken from Mepham et al. (2022). These were two-hundred sentences from the first 20 Anglicised-Modernised Harvard/IEEE sentence lists (IEEE, 1969), spoken by a female native British English speaker (a full list of the Harvard/IEEE sentences used in this study is available in Appendix A, and Appendices can be found following the OSF link provided in the Acknowledgements section.). Each target sentence had five keywords (e.g., "The PLAY SEEMS DULL and QUITE STUPID", keywords capitalised). All sentences were recorded in a single-walled sound-attenuated room at a 44.1 kHz sampling rate with 16-bit resolution using Audacity© using a Shure SM58 microphone and a RME Fireface UFX II built-in soundcard. Sentence duration ranged from 1.59 s to 3.16 s (M = 2.20 s, SD = 0.24 s). The mean fundamental frequency (F0) of the target sentences was 203 Hz (see Section 3.2.1.3.2 for further details). The F0 and VTL of all sentences were adjusted using the same method as that used for the masker stimuli to a mean F0 of 210 Hz. This value was 15 Hz below and above the high-F0 and low-F0 maskers, respectively (see Section 3.2.1.3.2). The F0 and VTL were edited following the procedure described in Darwin et al. (2003; see also Smith et al., 2007; Gaudrain et al., 2009). VTL was manipulated alongside F0 to improve the naturalness of the

speaker, as both indices have been shown to contribute to the perception of voice identity (e.g., Skuk & Schweinberger, 2014).

3.2.1.3.2 Masker stimuli

The masker stimuli were also taken from Mepham et al. (2022). They consisted of 64 sentences from Lists 1-4 of the English BKB-R corpus (Bench et al., 1979) spoken by a female native British English speaker. These BKB-R sentences are simple sentences with three to four keywords (e.g., "The POSTMAN SHUT the GATE", keywords capitalised). A full list of the BKB-R sentences used in this study is available in Appendix B. These sentences were identical to those used in the Calandruccio et al. (2010) study. All sentences were recorded in a single-walled sound-attenuated room at a 44.1 kHz sampling rate with 16-bit resolution using Audacity©. Each sentence was recorded a minimum of four times. For each sentence, the two best exemplars were kept. All sentences were manually edited using Praat (Boersma & Weenik, 2019) to remove silences at the beginning and end of the sentences. This was done through visual inspection of the spectrogram. One of the exemplars of each sentence was used to create Set A, and the other exemplar was used to create Set B. The Set A sentences were concatenated into a continuous stream, henceforth Stream A. The same was done with the Set B sentences, henceforth Stream B. The mean F0 of the Stream A sentences was 208 Hz and the mean F0 of the Stream B sentences was 205 Hz. Sentence order within each BKB-R list was the same in both streams, but the order of the lists differed in each stream.

Following the same procedure as the one used for the target sentences, the F0 and VTL of each sentence within the streams were edited to create a high-F0 (mean of 225 Hz) and a low-F0 version (mean of 195 Hz) of each stream. These two values are approximately 15 Hz above and below the average F0 of the target sentences (210 Hz), respectively. The high-F0

version of Stream A was combined with the low-F0 version of Stream B to constitute the two-talker masker. The use of a single voice to create the two masker speakers was designed to avoid idiosyncratic dominance of one masker voice over the other, as was done in Mepham et al. (2022; see also Smith et al., 2024, for a similar procedure).

3.2.1.3.3 Experimental mixtures

The target and masker stimuli were mixed online during the experiment. For each trial, the two-talker masker speech stream was randomly sampled for the duration of the target sentence, plus two seconds preceding it and two seconds following it. The masker level was fixed at 65 dB SPL and the level of the target sentence was determined by the SRT adaptive procedure (see Section 3.2.1.4). The masker speech stream was sampled randomly without replacement, resulting in a different masker speech stream for each trial and for each participant.

3.2.1.3.4 Subjective assessment of effort and fatigue

Two questions were used to explore subjective effort and fatigue. For effort, we adapted a question from the NASA task load index assessing mental demands (Hart & Staveland, 1988), a commonly used subjective measure of effort (e.g., Dimitrijevic et al., 2019; McGarrigle et al., 2017, 2020; Pals et al., 2019; Peng & Wang, 2019; Picou et al., 2017; Strand et al., 2018): "How hard did you have to work to understand what was said for the previous ten sentences? (0: Very low; 20: Very high.)". For fatigue, we used the Oncology Nursing Society (ONS) Brief Fatigue Inventory: English (Burke & Naylor, 2020; Picou & Ricketts, 2014): "Please rate

your fatigue (weariness, tiredness) by choosing the one number that best describes your fatigue right NOW. (0: No fatigue; 10: As bad as you can imagine.)".

3.2.1.4 Procedure

At the beginning of the experiment, listeners underwent audiometric threshold testing. The main experiment then comprised two blocks, one for each listening condition (time-forward masker and time-reversed masker). Each block comprised two parts. The first part was an adaptive procedure to obtain the listener's 50% SRT for the target-masker mixtures in that condition. The second part used the listener's 50% SRT for the 50-trial speech recognition task. The order of the two blocks was counterbalanced across participants. In both the adaptive procedure and the speech recognition task, listeners were asked to repeat aloud as much of the target talker as they could. To familiarise the listeners with the target voice and minimise the chances that they would accidentally track one of the masker voices instead, three practice trials were played before the adaptive procedure began. These were three target sentences from the Harvard/IEEE (IEEE, 1969) corpus, unused in the adaptive or recognition tasks, spoken by the target speaker. The practice sentences were played at an intensity level of 65 dB SPL.

For the adaptive procedure, listeners heard the target-masker mixtures at varying SNRs. The intensity of the masker was fixed at 65 dB SPL, and the intensity of the target talker changed according to the adaptive procedure. All listeners started with an SNR of +10 dB, with step sizes of 6 dB at the start, 4 dB after the first reversal, then 2 dB after the second reversal and for the remaining reversals. The adaptive procedure followed a one-up one-down staircase for eight reversals. The 50% SRT was calculated by fitting a logistic function to the performance and corresponding SNRs for each reversal during the adaptive procedure. If the logistic function failed to fit, or returned an infinite value, an approximate 50% accuracy SNR

was used instead, calculated by taking the mean SNR value over all eight reversals (n = 15 occasions used the back-up procedure, i.e., 18.75%). The 50% SRT value was then used for the main speech recognition task.

At the beginning of each block of the main speech recognition task and after every 10 trials thereafter, the listeners were presented with two questions, one asking about their subjective rating of effort and the other asking about their subjective rating of fatigue (see above). The questions were presented on a monitor and participants scored their responses on a sliding scale using a computer mouse.

For both the adaptive procedure and the speech recognition task, listeners were instructed to focus on a fixation cross that appeared on the monitor for the duration of the trial. They were cued to respond at the end of the masker, which itself finished 2.0 s after the end of the target sentence (see Section 3.2.1.3.3). There was a 4.0 s gap between the end of their answer and the fixation cross for the next trial. The listeners' responses were scored online by the experimenter against the five keywords for each sentence. Listeners were offered a break between the first and the second blocks. The eye-tracking equipment was recalibrated at the beginning of the second block for all participants. The entire experiment lasted under an hour.

3.2.1.5 Analysis

3.2.1.5.1 Pupillometry

Following recommendations from Winn et al. (2018), pupil data were pre-processed to remove noise. Any missing values in pupil size (e.g., caused by blinks or pupil non-detection) were coded as *NA* and linearly interpolated using the *gazeR* package (Version 0.1, Geller et al., 2020). Trials containing > 20% missing data were removed from the analysis. This resulted in

the removal of 11 trials across all participants (0.25% of all trials in the data set). One participant had 21 trials with > 20% missing data and was removed from the pupillometric analysis following procedures outlined in the pre-registration.

Baseline-correction was performed on a trial-by-trial basis. Of the 2.0 s of masker speech preceding the onset of the target sentence, we only used the 1.0 s immediately preceding the onset of the target sentence to avoid pupil responses that might reflect sensory onset adaptation rather than a genuine dilation baseline. The mean pupil size value recorded during this 1.0 s window was then subtracted from every sample recorded after target speech onset. We chose to use the mean pupil dilation (MPD) as the TEPR measure instead of PPD or latency to mean size (e.g., Zekveld et al., 2010) because MPD is sensitive to masker manipulations and time (McGarrigle et al., 2021a, 2021b) and because MPD and PPD indices have been shown to converge on similar patterns (Neagu et al., 2023). MPD was calculated as the relative change in baseline-corrected pupil size from the 1.0 s baseline throughout the target sentence and subsequent 2.0 s speech offset, i.e., [mean pupil size throughout the target sentence and 2.0 s speech offset] minus [mean pupil size throughout the 1.0 s preceding baseline].

Linear mixed-effect modelling (LMEM) using the *lmer* function in the *lme4* package (Bates et al., 2015) was conducted to examine TEPR as a function of Masker (coded as time-reversed speech: 0; time-forward speech: 1), and Time (trials 1-50) as fixed effects. The main effect of Masker reflects the difference in the effect of the time-forward masker from the time-reversed masker, the reference level for Masker.

Listener and Sentence were used as random intercepts. Masker|Listener and Masker|Sentence were used as random slopes, following Barr et al.'s (2013) recommendation to use the most complex random-effects structure supported by the data. Initially, a full model of all main effect and interaction terms was used, and the contribution of each term was tested using likelihood ratio testing (i.e., comparing the full model to a reduced version with the term

of interest removed). Where models failed to converge, random slopes and intercepts that were highly correlated or where the variance could not be estimated were removed from the model. This resulted in the Masker|Sentence slope to be removed from the analyses. Additionally, the BOBYQA optimiser was used to aid model convergence (Powell, 2009).

3.2.1.5.2 Speech recognition performance

Speech recognition performance was calculated as the proportion of keywords (out of 5) correctly reported for each target sentence (as in Mepham et al., 2022). Generalised linear mixed-effect models (GLMEM) with a logit link and binomial distribution were run using the *glmer* function from the *lme4* package (Bates et al. 2015). The models assessed mean differences in proportion of keywords correctly reported as a function of Masker (coded as time-reversed: 0; time-forward: 1) and Time (1 to 50). The main effect of Masker reflects the difference in the effect of the time-forward masker from the time-reversed masker, the reference level for Masker. Listener and Sentence were used as random intercepts, Masker|Listener and Masker |Sentence as random slopes, and the BOBYQA optimiser was used to aid model convergence. As for the TEPR data, a full model of all main effect and interaction terms was used and the contribution of terms assessed using likelihood ratio testing.

3.2.1.5.3 Subjective measures

Due to the small number of data points for the subjective ratings of effort and fatigue, a repeated-measures analysis of variance (ANOVA, *aov* function from the *stats* package) was run instead of linear mixed-effects models. The dependent variables were the effort and fatigue ratings, which were rescaled as a subtraction from the baseline effort/fatigue rating at the start

of the condition before participants undertook any trials. The analyses were calculated using the rescaled values. The independent variables were Masker (time-forward vs. time-reversed) and Time (ordinal variable with 5 levels corresponding to trials 10, 20, 30, 40, and 50). A by-participant error term [Error(participant/(Masker*Time))] was included to analyse the data as repeated-measures.

3.2.2 RESULTS

3.2.2.1 50% Speech Reception Threshold (SRT)

The average SNR required to achieve 50% correct transcription was higher in the time-forward condition (M = -0.38 dB, SD = 2.15 dB) than in the time-reversed condition (M = -2.65 dB, SD = 1.82 dB), $t(39) = -7.88$. $p < .001$. This SNR difference (2.27 dB) illustrates the cost of ignoring an intelligible (time-forward) masker compared to an unintelligible (time-reversed) masker, a hallmark of linguistic interference.

3.2.2.2 Speech recognition performance

Figure 12 displays speech recognition performance as a function of Masker and Time. As expected from the adaptive procedure, performance started around 50% correct in both masker conditions. There was no significant effect of Masker, $B = 0.154$, $SE = 0.200$, $X^2(1) = 0.59$, $p = .443$. A significant effect of Time, $B = 0.790$, $SE = 0.085$, $X^2(1) = 87.58$, $p < .001$, indicated that performance improved over the course of the blocks. However, a significant interaction between Masker and Time, $B = -0.394$, $SE = 0.118$, $X^2(1) = 11.17$, $p < .001$, showed that the improvement was faster in the time-forward than time-reversed condition; the effect of

Time was nevertheless significant in both conditions, B = 0.804, SE = 0.085, $X^2(1)$ = 89.65, p < .001, and B = 0.405, SE = 0.082, $X^2(1)$ = 24.25, p < .001, respectively.

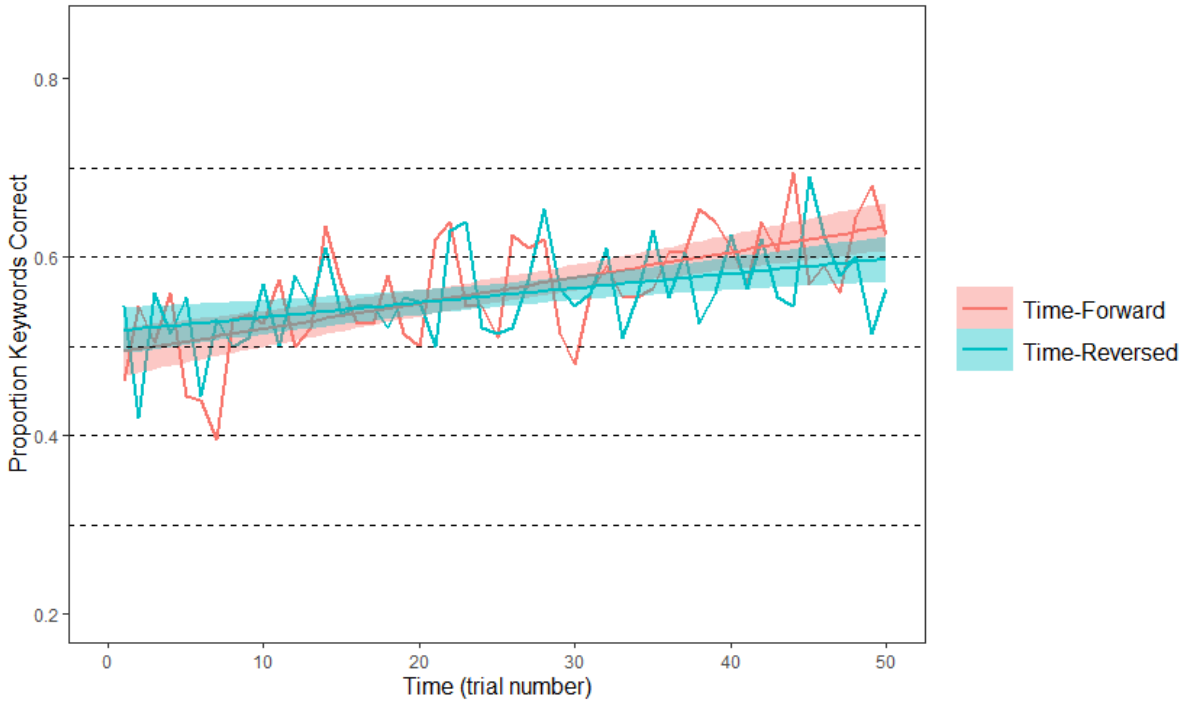


FIG. 12. Mean proportion of keywords correctly reported for each trial in each Masker condition as a function of Time. Each dot is the mean proportion of keywords correctly reported for each trial. The shaded area represents 95% confidence intervals.

3.2.2.3 Pupillometry measures

Figure 13 shows TEPR as a function of Masker and Time. There was no significant effect of Masker, B = -0.478, SE = 14.38, $X^2(1)$ < 0.01, p = .974, nor a significant effect of Time, B = -0.388, SE = 0.265, $X^2(1)$ = 2.14, p = .144. There was also no significant interaction between Masker and Time, B = -0.012, SE = 0.376, $X^2(1)$ < 0.01, p = .975. Figure 14 shows the average TEPR over the course of a trial for each masker condition broken down by bins of ten

trials within a listening block (e.g., trials 1-10, 11-20, etc.) and displays the TEPR growth curves over the first 4.0 s averaged for each bin of ten sentences.

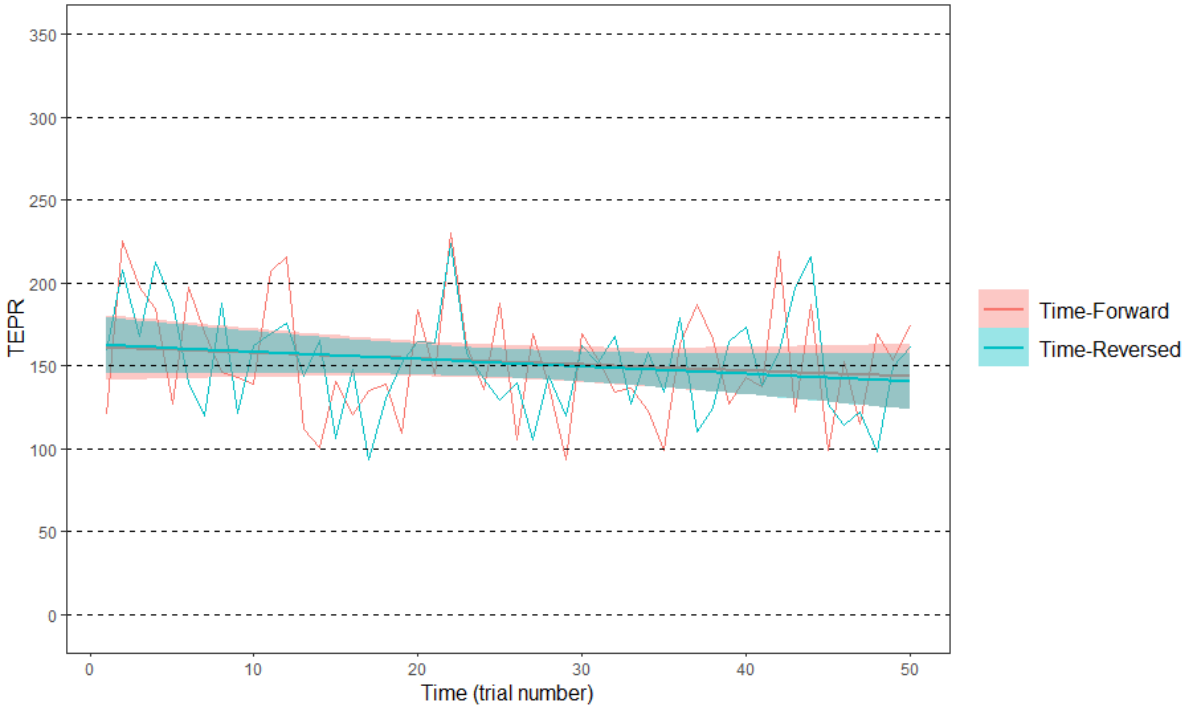


FIG. 13. TEPR for each trial in each Masker condition as a function of Time. For each trial, the TEPR value is the mean TEPR across participants. The shaded area represents 95% confidence intervals.

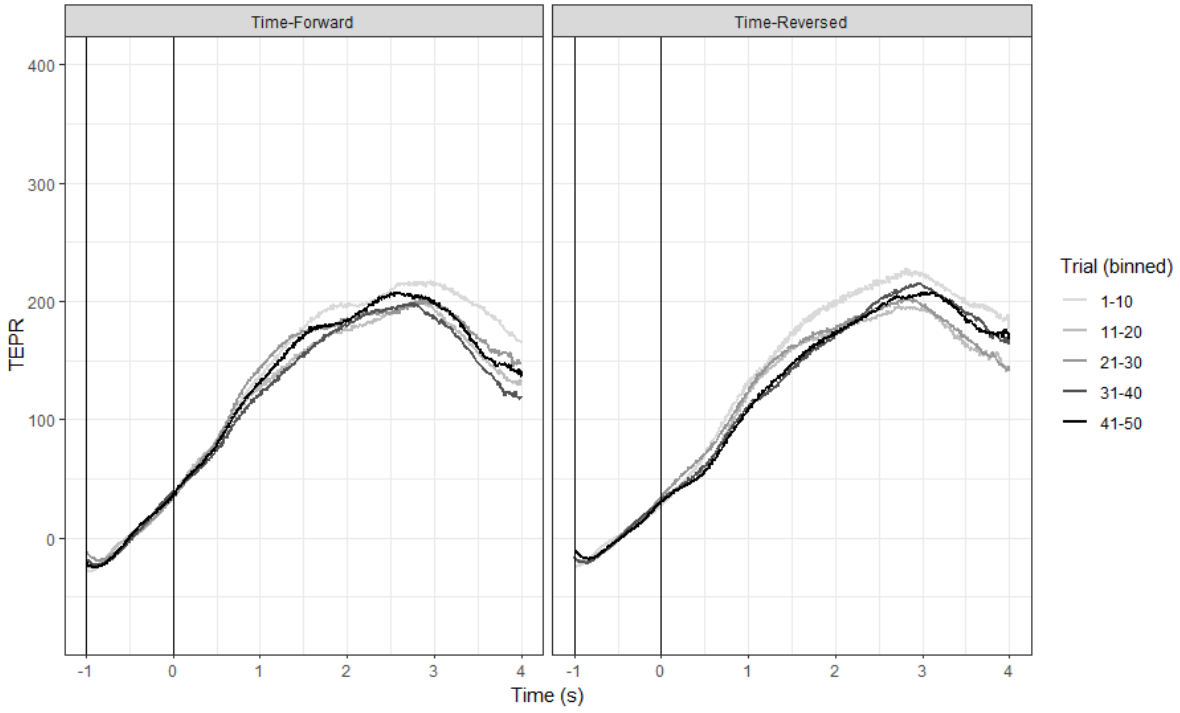


FIG. 14. TEPR over the first four seconds of target sentence onset, binned by groups of ten sentences in order of presentation, as per Brown et al. (2020). The vertical lines indicate the start of the baseline pupil size (t = -1 s before the start of the target sentence) and of the target sentence (at t = 0 s).

3.2.2.4 Subjective Measures

Figure 15 presents the subjective ratings of effort and fatigue, calculated as change from the baseline ratings collected prior to the first trial of the main task. For effort, there were no significant main effects or interactions involving Masker or Time (all $F < 0.93$, all $p > .34$, all $\eta_p^2 < .02$). For fatigue, there was a significant effect of Time, $F(2.35, 89.40) = 14.42$, $p < .001$, $\eta_p^2 = .27$, indicating that fatigue ratings increased as the blocks progressed. There was no effect of Masker, $F(1, 38) = 0.38$, $p = .541$, $\eta_p^2 = .01$ nor a significant Masker by Time interaction. $F(3.17, 120.32) = 0.26$, $p = .866$, $\eta_p^2 < .01$.

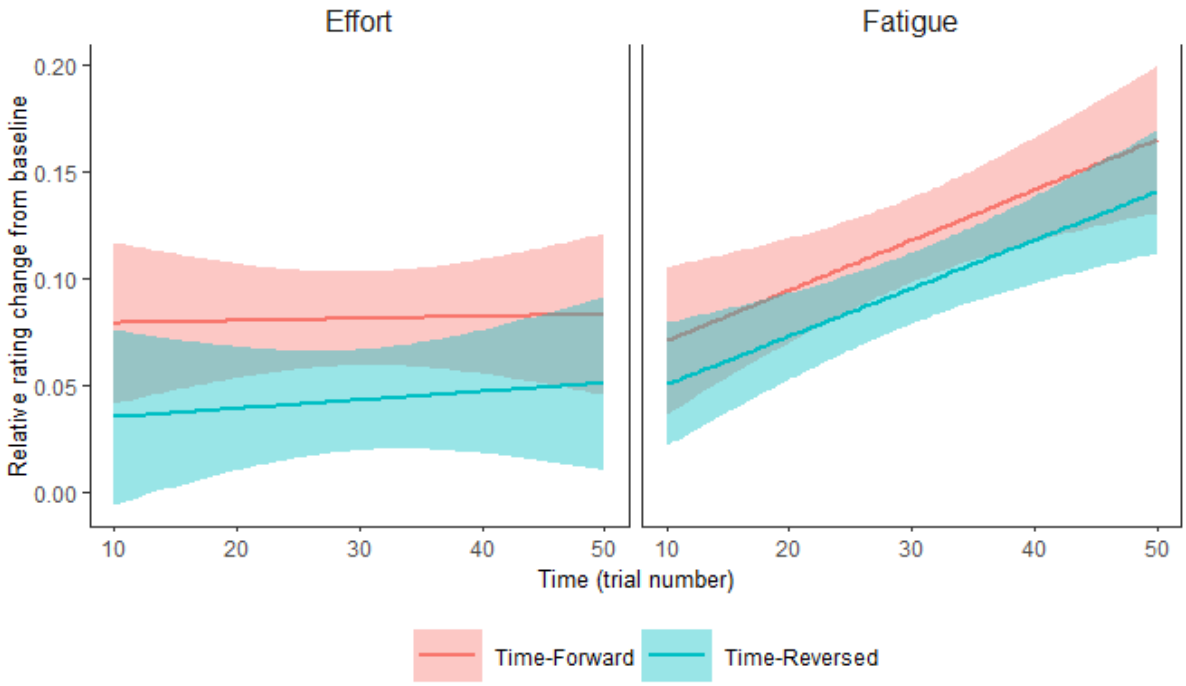


FIG. 15. Mean ratings of effort and fatigue after trials 10, 20, 30, 40, and 50 relative to baseline ratings. The scale on the y-axis corresponds to the re-scaling of the original effort and fatigue questions (effort/20, fatigue/10).

3.2.3 DISCUSSION

Our results replicate previous findings that speech recognition in adverse conditions improves over the course of an experiment (Bent et al., 2009; Cooke et al., 2022; Erb et al., 2012, 2013; Mepham et al., 2022). However, contrary to our expectation and Mepham et al.'s (2022) results, performance improved more, rather than less, in the time-forward than time-reversed masker condition. Mepham et al. (2022) claimed that the slower improvement for the time-forward condition in their study reflected the sustained linguistic interference in that condition relative to the easier segregation of a 'neutral' masker, i.e., a masker with no

linguistic content. Our present results suggest that, on the contrary, the linguistic, and hence familiar nature of the time-forward masker made it easier for listeners to identify it as a separate auditory object (Shinn-Cunningham, 2008) and therefore learn to ignore it.

An important methodological difference between the present study and Mepham et al.'s (2022) is that, in the Mepham et al. (2022) study, the SNR was fixed at -3 dB for both the time-forward and time-reversed conditions. In the present study, the SNR was adjusted so that performance in both masker conditions was around 50% at the start of the blocks. As a consequence, the listeners in the present study required a higher (i.e., more favourable) SNR in the time-forward than the time-reversed condition. It is therefore possible that the more favourable SNR in the time-forward condition made it easier to glimpse the target sentences through the masker and learn to ignore the masker over time. In contrast, learning to ignore the masker in the time-reversed condition might have been harder because the masker was louder than the target, hence acting as a more effective distractor.

Although the pupillometric data showed a small numerical TEPR decrease over time (cf. Brown et al., 2020; Paulus et al., 2020; but see Versfeld et al., 2021), this trend was not significant. Since the adaptive procedure could have provided opportunity for familiarisation with the task and stimuli, this might have made the decrease in effort over the course of the block less pronounced than expected.

Taken together, these results depart from expectations in two ways. First, we did not find an effect of masker type on pupil size, which is in contrast with Koelewijn et al.'s (2012) finding of larger pupil dilation in a single-talker masker condition (intelligible) than in control noise conditions (unintelligible). Second, we did not find a difference in the rate of pupil size decrease between the time-forward and time-reversed conditions. This is inconsistent with the levelling-out pattern reported by Brown et al. (2020), where the pupil response decrease over time was steeper for hard than easy listening conditions. However, our study and the studies

by Koelewijn et al. (2012) and Brown et al. (2020) differed methodologically in several important ways. In the Koelewijn et al. (2012) study, pupillometric measures were taken while listeners underwent an SRT adaptive procedure, whereas, in this experiment, pupillometric measures were taken with a fixed SNR, after completion of a SRT procedure. In the Brown et al. (2020) study, a dual-task paradigm was used, whereas we used a single task in our study. Third, performance in the Brown et al. (2020) study was near ceiling by design, whereas performance was in the 50-60% range in this experiment. Finally, we did not find a main effect of time in the pupillometric data, which was observed in both the Brown et al. (2020) and Paulus et al. (2020) studies. One can interpret the pupillometric and subjective measures of listening effort patterning together in their lack of a significant change over time. However, there might be other factors pertaining to the experiment that could have resulted in no significant decrease in TEPR, which would have been consistent with studies demonstrating reduced TEPR over time. One alternative explanation for the lack of main effect of time could arise from the adaptive procedure prior to each experimental condition. Although the SNR between target and masker sentences fluctuated throughout the adaptive procedure to obtain each participant's 50% SRT, this could have resulted in the participants becoming familiar with the task procedure, and therefore not eliciting higher TEPRs while acclimating to the experimental paradigm. This would then result in reduced rapid adaptation at the onset of the experiment and more consistent levels of TEPRs throughout the experiment. Although the lack of a main effect of time suggests that speech perception in noise is not automatic and requires sustained effort (contrasting Bregman's, 1990, 'primitive process'), the pupil dilation levels in this experiment are similar to those in the Brown et al. (2020) native accented speech condition. Taken together, this suggests that the pupil dilation here is more related to lower effort for speech-in-noise perception, compared to the higher effort expected in the Brown et al. (2020)

non-native accented speech condition, and higher levels of TEPR one would expect with the speech stream segregation requiring directed attention (Huyck & Johnsrude, 2012).

As mentioned earlier, the faster improvement in transcription performance in the time-forward than the time-reversed condition could be due to the higher SNR, and thus higher audibility of the target, for the intelligible than unintelligible maskers. Therefore, in Experiment 6, we used a single SNR for both conditions, which is similar to the procedure in the Mepham et al. (2022) study. If the higher SNR for the time-forward condition in Experiment 5 was responsible for its faster improvement over time, this advantage should disappear when the SNR is the same for both maskers and would echo the results in Mepham et al. (2022), as audibility and opportunities for glimpses would be identical in both conditions. However, if the faster improvement for the time-forward condition truly demonstrates listeners' ability to better learn to stream and suppress an intelligible masker than an unintelligible masker, the pattern in Experiment 6 should replicate that in Experiment 5.

3.3 EXPERIMENT 6

3.3.1 METHODS

3.3.1.1 Participants

Forty native British English listeners (11 male, 26 female, three non-binary) aged between 18 and 31 years (M = 20.52, SD = 2.86) with no known history of hearing impairments participated in the experiment. All listeners described their language status as monolingual, speaking British English from birth, and had pure-tone audiometry (PTA) measures ≤ 20 dB HL at 0.5, 1, 2, and 4 kHz (M = 5.59, SD = 2.94). Listeners either participated in this experiment

for course credit or were compensated for their participation at a rate of 6.00 GBP per hour. All participants provided written informed consent to take part in this study.

3.3.1.2 Equipment

The equipment and set-up were as in Experiment 5.

3.3.1.3 Materials

The target and masker stimuli were the same as in Experiment 5, except that, for both the time-forward and time-reversed conditions, the target sentences were played at 63.5 dB SPL (the average of the target levels in the time-forward and time-reversed conditions of Experiment 5) and the maskers were played at 65.0 dB SPL, as in Experiment 5, resulting in a -1.5 dB SNR throughout the experiment. The materials used to explore the subjective ratings of effort and fatigue were the same as in Experiment 5.

3.3.1.4 Procedure

The procedure was the same as in Experiment 5 except that the adaptive procedure at the start of each block in Experiment 5 was replaced with a practice session. In each practice session, participants listened without responding to five target sentences played without a masker and the same five sentences in the presence of a masker at -1.5 dB SNR. Participants were then presented with five new target and masker mixtures at that SNR, and were asked to repeat as much of the targets as they could. Feedback was provided on how many keywords they reported correctly. Initially, as per our preregistration, we had planned to use a pseudo-

adaptive procedure using randomly sampled SNRs to emulate the adaptive procedure of Experiment 5. However, we found that, if listeners were presented with the time-forward condition first, they often erroneously reported the content of the masker rather than the target. For this reason, we used the procedure described above. Additionally, if the listener consistently reported the masker rather than the target in the five practice trials, the experimenter entered the testing room and encouraged them to pay attention to the quieter talker until they could reliably distinguish targets from maskers. The procedure for the condition then restarted. The rest of the procedure was the same as in Experiment 5.

3.3.1.5 Analysis

Analysis procedures of pupillometric and behavioural data were identical to those in Experiment 5.

3.3.2 RESULTS

3.3.2.1 Speech recognition performance

Figure 16 shows speech recognition performance as a function of Masker and Time. There was a significant effect of Masker, $B = 1.024$, $SE = 0.096$, $X^2(1) = 66.78$, $p < .001$, with better performance in the time-reversed ($M = 0.685$, $SD = 0.307$) than time-forward condition ($M = 0.466$, $SD = 0.364$). This effect is a direct consequence of using the same SNR in the two conditions and it provides evidence for linguistic interference. A significant effect of Time, $B = 0.407$, $SE = 0.080$, $X^2(1) = 25.99$, $p < .001$, indicated an improvement in correctly reporting keywords over the course of a block. The interaction between Masker and Time was not

significant, B = 0.109, SE = 0.116, $X^2(1)$ = 0.89, p = .346, suggesting that the rate at which listeners' performance improved did not differ between the time-reversed and time-forward conditions.

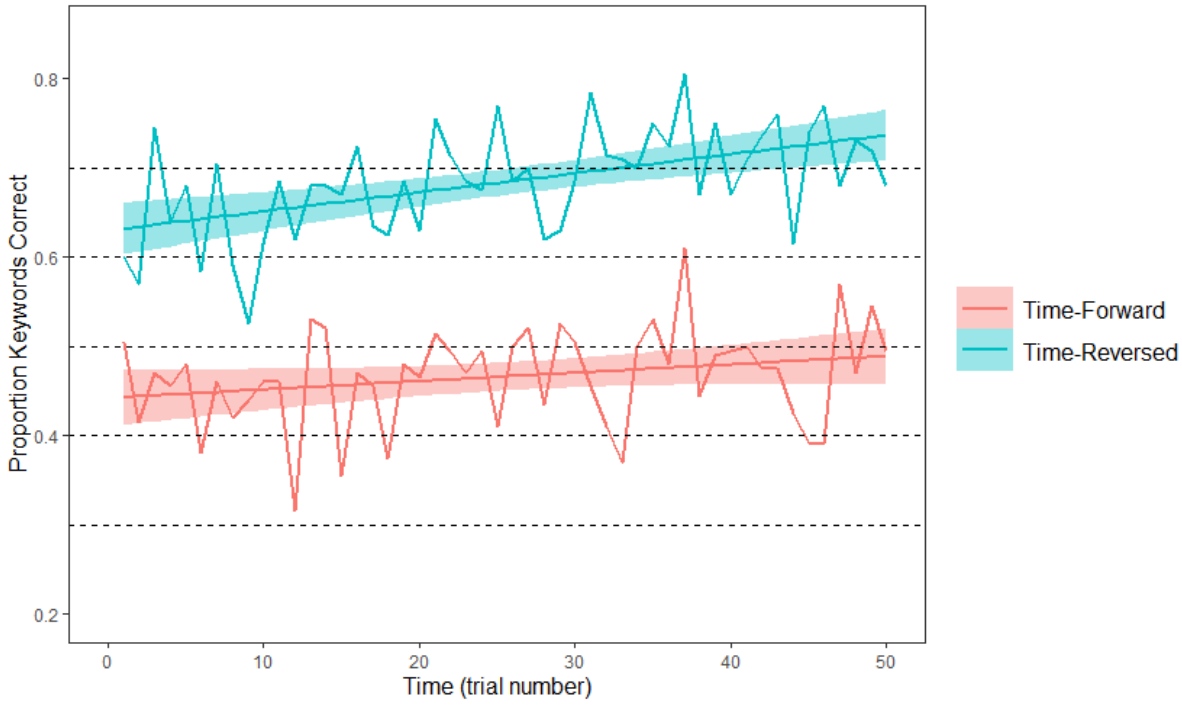


FIG. 16. Mean proportion of keywords correctly reported for each trial in each Masker condition as a function of Time. Each dot is the mean proportion of keywords correctly reported for each trial. The shaded area represents 95% confidence intervals.

3.3.2.2 Pupillometry measures

Figure 17 shows TEPR as a function of Masker and Time. There was no significant effect of Masker, B = -15.79, SE = 15.60, $X^2(1)$ = 1.02, p = .312, but an effect of Time showed that TEPR significantly decreased over time, B = -1.966, SE = 0.272, $X^2(1)$ = 51.84, p < .001. An interaction between Masker and Time, B = -0.100, SE = 0.385, $X^2(1)$ = 6.744, p = .009,

indicated that this decrease was steeper for the time-reversed than time-forward condition. This interaction can also be seen in Figure 18, which plots trial-level TEPR broken down by bins of 10 trials. Using separate GLMMs for the first half of the block (trials 1 to 25) and for the second half (trials 26 to 50), the effect of Masker showed significance for the first half, B = -28.12, SE = 13.14, $X^2(1) = 4.34$, $p = .037$, and higher significance for the second half, B = -54.33, SE = 13.06, $X^2(1) = 14.40$, $p < .001$ (both GLMMs had the Sound intercept removed as the variance associated with this item trended towards zero). Figure 18 displays the TEPR growth curves over the first 4.0 s averaged for each bin of ten sentences.

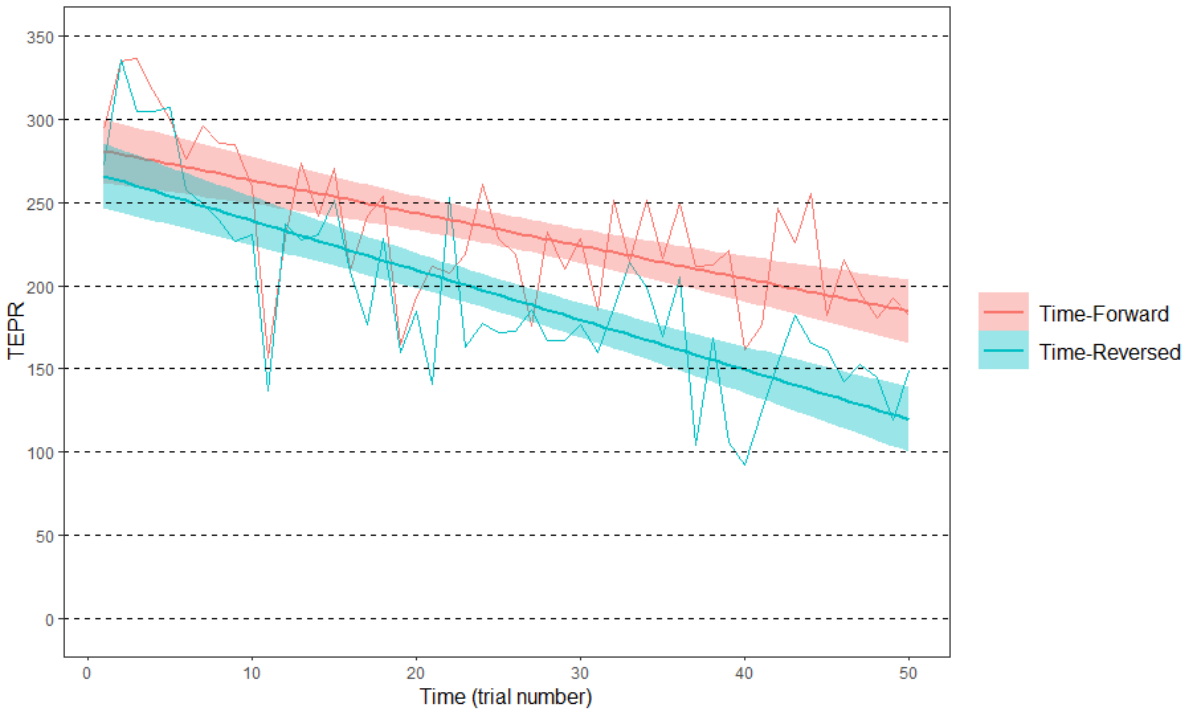


FIG. 17. TEPR for each trial in each masker Condition as a function of Time. For each trial, the TEPR value is the mean TEPR across participants. The shaded area represents 95% confidence intervals.

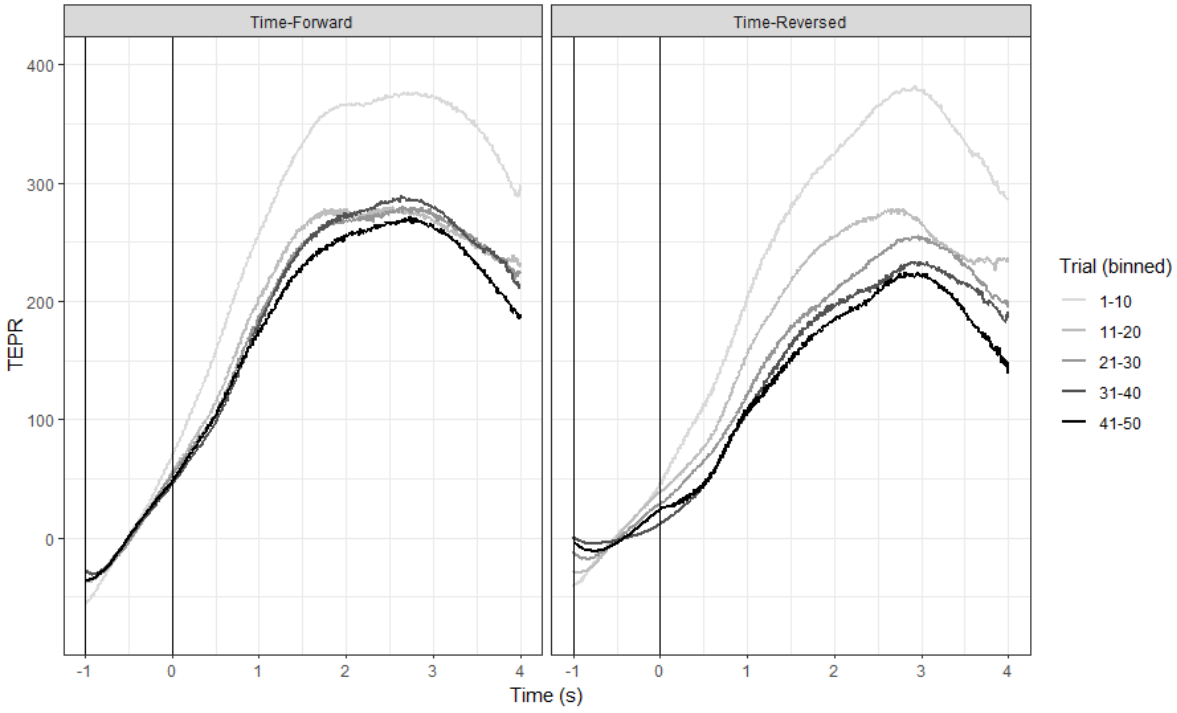


FIG. 18. TEPR over the first four seconds of stimulus onset, binned by groups of ten sentences in order of presentation, as per Brown et al. (2020). The vertical lines indicate the start of the baseline pupil size (t = -1 s before the start of the target sentence) and of the target sentence (at t = 0 s).

3.3.2.3 Subjective Measures

Figure 19 presents the subjective ratings of effort and fatigue as a change from baseline ratings. For effort, there was no significant effect of Masker, $F(1, 39) = 0.12$, $p = .728$, $\eta_p^2 = .003$, Time, $F(3.02, 117.68) = 0.26$, $p = .859$, $\eta_p^2 = .007$, and no significant interaction between Masker and Time, $F(3.24, 126.31) = 0.61$, $p = .620$, $\eta_p^2 = .02$.

For fatigue, there was a significant effect of Time, $F(2.21, 86.25) = 25.27$, $p < .001$, $\eta_p^2 = .39$. with greater reported fatigue as the block progressed. There was no effect of Masker,

$F(1, 39) = 0.289$, $p = .594$, $\eta_p^2 = .007$, and no significant interaction between Masker and Time, $F(2.72, 106.15) = 0.06$, $p = .973$, $\eta_p^2 = .002$.

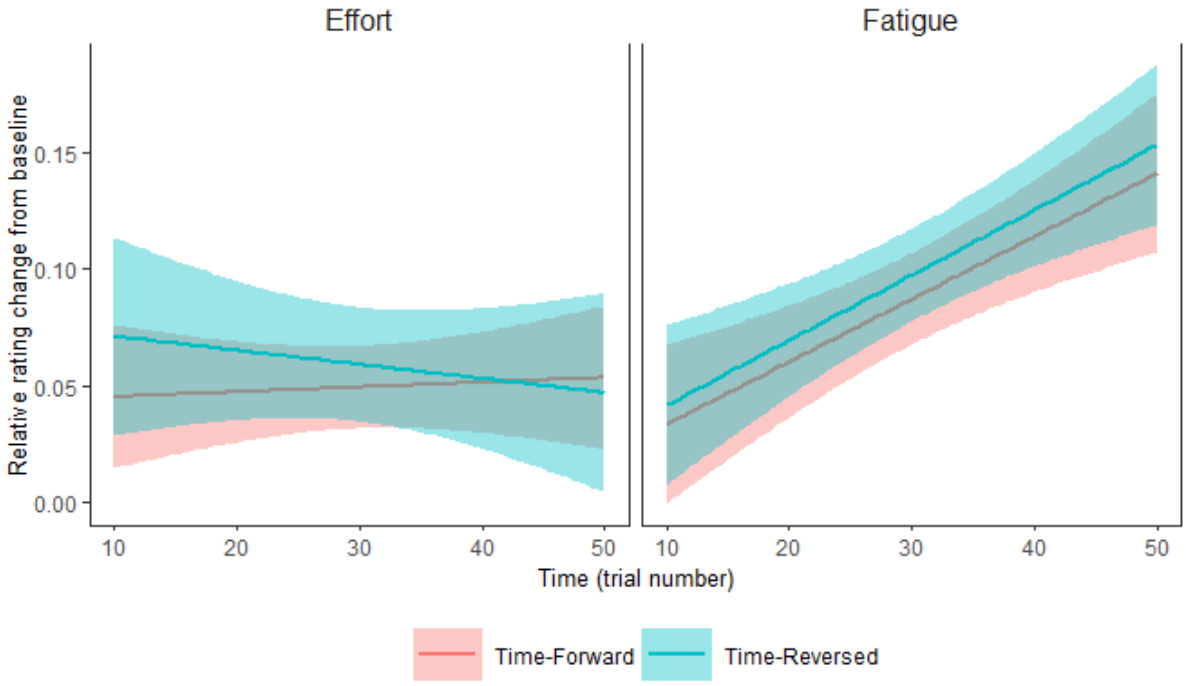


FIG. 19. Mean ratings of effort and fatigue after trials 10, 20, 30, 40, and 50. The scale on the y-axis corresponds to the re-scaling of the original effort and fatigue questions (effort/20, fatigue/10).

3.3.3 DISCUSSION

Speech recognition performance in Experiment 6 did not replicate the interactive pattern found in Experiment 5. In Experiment 5, performance improved more steeply for the intelligible (time-forward) than unintelligible (time-reversed) masker condition. In Experiment 6, this interaction was non-significant and, numerically, it resembled the performance pattern in the Mepham et al. (2022) study, with faster improvement in the unintelligible than

intelligible masker condition. Thus, when glimpses are controlled by using a fixed SNR (as in Experiment 6 and by Mepham et al., 2022), the intelligibility of a masker might actually hinder learning rather than facilitate it.

This interpretation is supported by the pupillometry data, where TEPRs decreased more slowly in the intelligible masker condition (the 'hard' condition) than in the unintelligible masker condition (the 'easy' condition). Therefore, learning to ignore a meaningful masker might be more effortful than learning to ignore a meaningless one. This finding broadly supports the finding of Koelewijn et al. (2012) that intelligible maskers require more cognitive effort to ignore than unintelligible ones. However, this result is in contrast with Brown et al.'s (2020) observation that the effort associated with adaptation to a native accent (the 'easy' condition) decreased more slowly than to a non-native accent (the 'hard' condition), which suggests that radically different adaptation mechanisms might be at play for speech degraded by an external masker, as in our study, and speech degraded at the source, as in the Brown et al. (2020) study (cf. Mattys et al., 2012). However, this pattern of decreases in mean pupil dilation between intelligible and unintelligible masker conditions is similar to the pattern of results from Winn et al. (2015) who found that participants experienced higher magnitudes and rates of pupil dilation when listening to noise-vocoded speech with fewer channels compared to vocoded speech with more channels, even when participants were accurately reporting entire target sentences correctly.

As in Experiment 5, the subjective ratings of effort showed no significant difference between the two masker conditions, with subjective effort remaining constant over time, again suggesting a lack of association between subjective and physiological measures of listening effort (Koelewijn et al., 2012; McGarrigle et al., 2014, 2021a; Pichora-Fuller et al., 2016; Strand et al., 2018). These results hint to listeners' potential lack of awareness of the changes in cognitive effort associated with the learning process. With respect to the fatigue ratings,

these, too, were similar to those in Experiment 5: subjective fatigue increased over time and it did so similarly for the intelligible and unintelligible maskers. Increased fatigue ratings over the course of an experimental task suggest that listeners are more attuned to tiredness associated with sustained cognition rather than to the effort required to complete a challenging task.

The pupillometric results differed between Experiments 5 and 6. In Experiment 5, TEPR did not significantly decrease over the course of the masker conditions, while in Experiment 6 there was a main effect of Time and a significant interaction between Masker and Time, demonstrating not only that TEPR decreased across the experimental conditions, but that there were faster decreases in TEPR in the unintelligible maskers compared to the intelligible competing speech. The pupillometric results of Experiment 6 are more aligned to previous experiments demonstrating decreases in TEPR over time (Brown et al., 2020; Paulus et al., 2020), and are likely driven by the main effect of Masker: less cognitive effort required for listening to speech among unintelligible maskers as opposed to competing talkers with intelligible linguistic content. However, as in Experiment 5, there was no change over time in subjective measures of listening effort, suggesting that the relationship between pupillometric and self-reported measures of listening effort might not necessarily align, consistent with the majority of studies comparing correlations between pupil dilation, performance, and subjective effort (Koelewijn et al., 2012; McGarrigle et al., 2014, 2021b; Pichora-Fuller et al., 2016; Strand et al., 2018). Instead, there were increases in subjective measures of fatigue while TEPR decreased (McGarrigle et al., 2021b), although the rate of increase in fatigue was the same while listening to speech in both intelligible and unintelligible maskers. As in Experiment 5, this would suggest both that listeners are more attuned to changes in their physiological state associated with fatigue rather than with cognitive effort, and that the results of Experiment 5 and 6 add to the inconsistent relationship between subjective measures of effort and fatigue and changes in pupil size when listening to speech in noise.

Taking together the results of Experiments 5 and 6 indicate that masker intelligibility, which we manipulated to assess linguistic interference, might not be the only factor determining the speed at which listeners learn to stream a target from a masker. Target audibility (opportunities for glimpses) through the masker might also play a role. In Experiment 5, equating the initial performance level across conditions by manipulating the SNRs led to different improvement rates for intelligible and unintelligible maskers to those observed in Experiment 6, where the SNR was kept constant. It may therefore be the case that audibility determines the rate of improvement to a greater extent than the intelligibility of the maskers.

Experiment 7 aimed to adjudicate between an audibility explanation and a linguistic interference explanation. In Experiment 7, the SNRs were manipulated such that the initial performance levels of the time-forward and time-reversed conditions were the mirror image of those in Experiment 6. Using an adaptive procedure, the starting point of the time-forward condition was set to 66% (the approximate intercept value of the time-reversed condition in Experiment 6) and the starting point of the time-reversed condition was set to 41% (the approximate intercept value of the time-forward condition in Experiment 6).

A replication of Experiment 6, (i.e., similar rates of improvement in performance in the time-forward and time-reversed maskers despite the higher initial performance level) would suggest a dominance of linguistic interference over and above SNR levels. However, a replication of Experiment 5 (i.e., faster improvement in the time-forward than time-reversed condition) would suggest that linguistic interference is significantly modulated by or even reducible to SNR and its associated target audibility.

3.4 EXPERIMENT 7

3.4.1 METHODS

3.4.1.1 Participants

Thirty-nine native British English listeners (six male, 30 female, four non-binary) aged between 18 and 21 years (M = 19.52, SD = 0.82) with no known history of hearing impairments participated in the experiment and were included in the analyses. All listeners described their language status as monolingual, speaking British English from birth, and had pure-tone audiometry (PTA) measures ≤ 20 dB HL at 0.5, 1, 2, and 4 kHz (M = 5.35, SD = 3.35). Listeners either participated in this experiment for course credit or were compensated for their participation at a rate of 6.00 GBP per hour. All participants provided written informed consent to take part in this study.

3.4.1.2 Equipment and materials

The equipment and materials were as in Experiment 5.

3.4.1.3 Procedure

The procedure was the same as in Experiment 5, except that after the adaptive procedure which targeted the 50% SRT, the 41% SRT and the 66% SRT for the time-reversed and time-forward conditions respectively were calculated by fitting a logistic function to the accuracy performance and corresponding SNRs. If the logistic function failed to fit, or returned an infinite value, the mean difference between the 50% and 66% SRT in the time-forward

condition in Experiment 5 (0.85 dB) was added to the mean SNR value over all eight reversals, and the mean difference between the 50% and 41% SRT in the time-reversed condition in Experiment 5 (0.64 dB) was subtracted from the mean SNR over all eight reversals. There were no instances in this experiment where the logistic function failed to fit. These SRT values were then used for the main speech recognition task.

3.4.1.4 Analysis

Analysis procedures of pupillometric and behavioural data were almost identical to those in Experiment 5, except in the pupillometry analyses, the random intercept of Sentence was removed to avoid singular model fit.

3.4.2 RESULTS

3.4.2.1 Speech Reception Thresholds (SRT)

The average SNR required to achieve 50% correct transcription was higher in the time-forward condition (M = 0.70 dB, SD = 2.69 dB) than in the time-reversed condition (M = -0.30 dB, SD = 2.86 dB), $t(38) = 3.25$. p = .002. This SNR difference (1.00 dB) again illustrates linguistic interference, i.e., greater masking from an intelligible (time-forward) than unintelligible (time-reversed) masker. In this experiment the 66% SRT obtained from the logistic regression for the time-forward condition was M = 1.33 dB, SD = 2.74, and the 41% SRT obtained from the logistic regression for the time-reversed condition was M = -0.77 dB, SD = 2.87 dB.

3.4.2.2 Speech recognition performance

Figure 20 shows speech recognition performance as a function of Masker and Time. As expected by the adaptive manipulation, there was a significant effect of Masker, B = -0.435, SE = 0.189, $X^2(1) = 5.04$, p = .025, with better performance in the time-forward (M = 0.613, SD = 0.363) than time-reversed condition (M = 0.548, SD = 0.355). There was also a significant main effect of Time, B = 0.313, SE = 0.086, $X^2(1) = 13.07$, $p < .001$, with performance increasing over the course of the experimental conditions. However, there was no significant interaction between Masker and Time, $B = 0.124$, $SE = 0.118$, $X^2(1) = 1.09$, $p = .296$, suggesting that increases in speech recognition performance were similar across intelligible and unintelligible maskers.

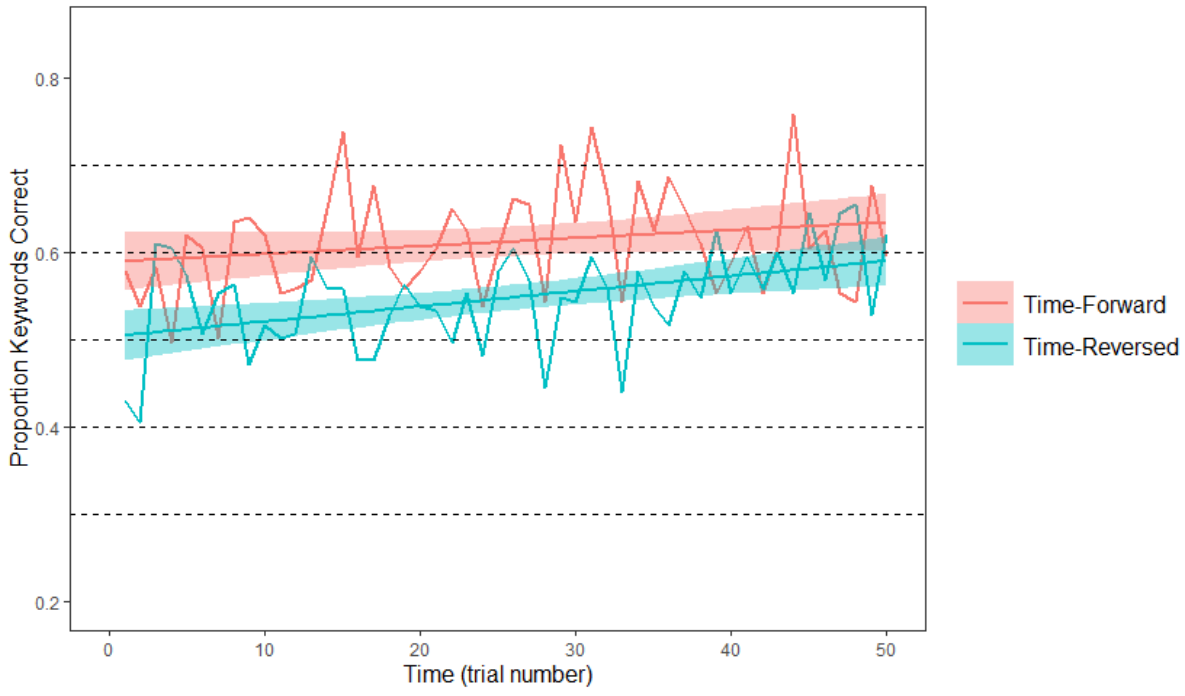


FIG. 20. Mean proportion of keywords correctly reported for each trial in each Masker condition as a function of Time. Each dot is the mean proportion of keywords correctly reported for each trial. The shaded area represents 95% confidence intervals.

3.4.2.3 Pupillometry measures

Figure 21 shows TEPR as a function of Masker and Time. TEPR significantly decreased over time, B = -1.763, SE = 0.284, $X^2(1)$ = 38.39, p < .001. There was also a significant effect of Masker, B = -32.37, SE = 15.03, $X^2(1)$ = 4.56, p = .033, with higher average TEPR in the time-forward masker (M = 159.85, SD = 212.48) than the time-reversed masker (M = 136.37, SD = 210.97). However, there was no significant interaction between Masker and Time, B = 0.331, SE = 0.402, $X^2(1)$ = 0.68, p = .410. Figure 22 displays the TEPR growth curves over the first 4.0 s averaged for each bin of ten sentences.

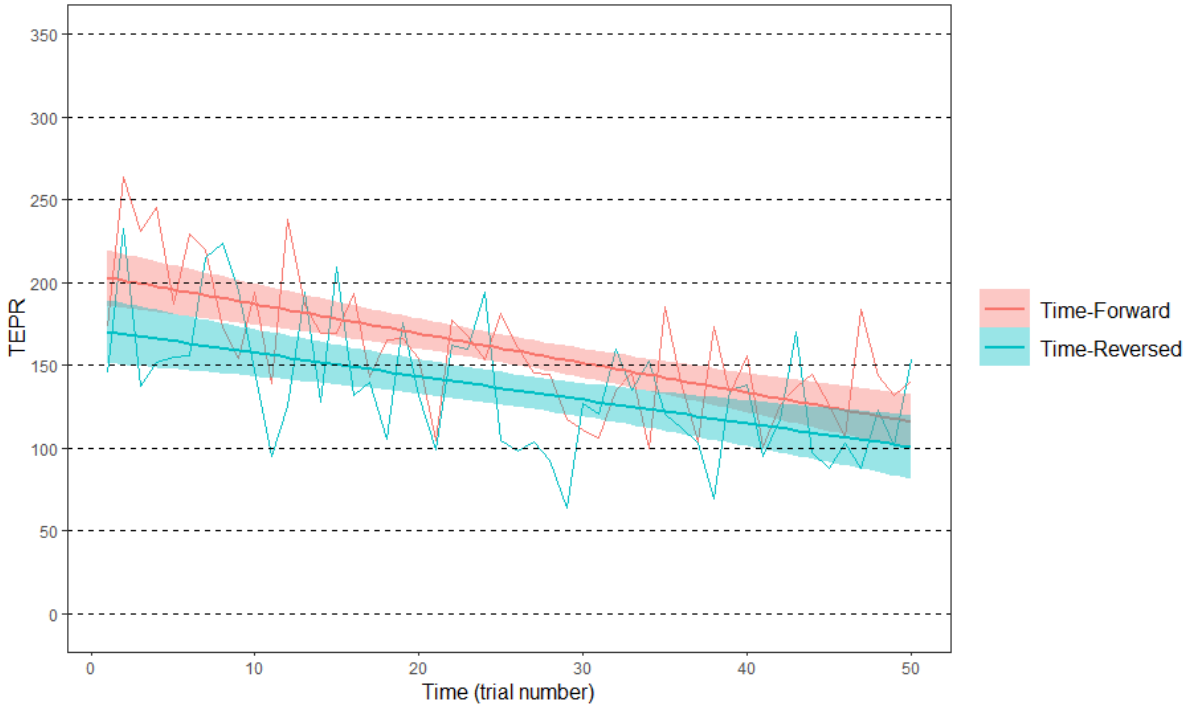


FIG. 21. TEPR for each trial in each masker Condition as a function of Time. For each trial, the TEPR value is the mean TEPR across participants. The shaded are represents 95% confidence intervals.

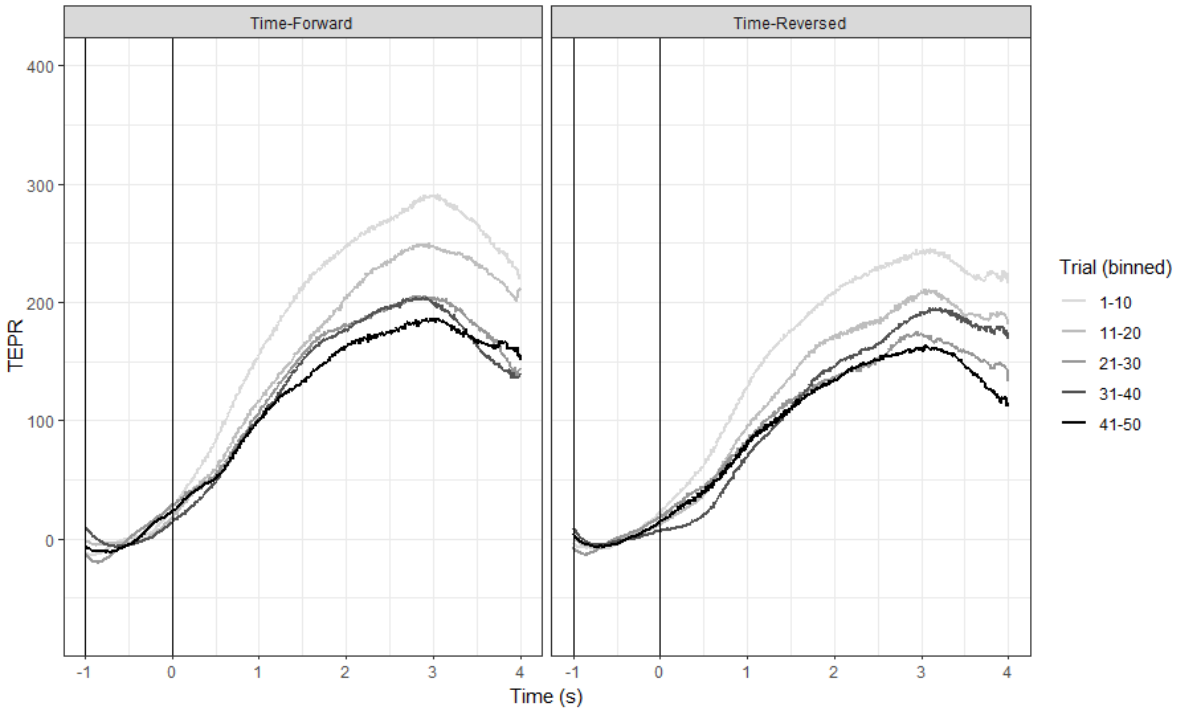


FIG. 22. TEPR over the first four seconds of stimulus onset, binned by groups of ten sentences in order of presentation, as per Brown et al. (2020). The vertical lines indicate the start of the baseline pupil size (t = -1 s before the start of the target sentence) and of the target sentence (at t = 0 s).

3.4.2.4 Subjective Measures

Figure 23 presents the subjective ratings of effort and fatigue as a change from baseline ratings. For effort, there were no significant effects of Masker, $F(1, 37) = 2.44$, $p = .126$, Time, $F(3.17, 117.28) = 0.59$, $p = .631$, and no significant interaction between Masker and Time, $F(4, 148) = 2.23$, $p = .069$.

For fatigue, there was a significant effect of Time, $F(2.08, 77.05) = 15.59$, $p < .001$, with greater reported fatigue as the block progressed. There was no significant main effect of

Masker, $F(1, 37) = 2.80$, $p = .103$, and no significant interaction between Masker and Time, $F(2.53, 93.56) = 0.47$, $p = .670$.

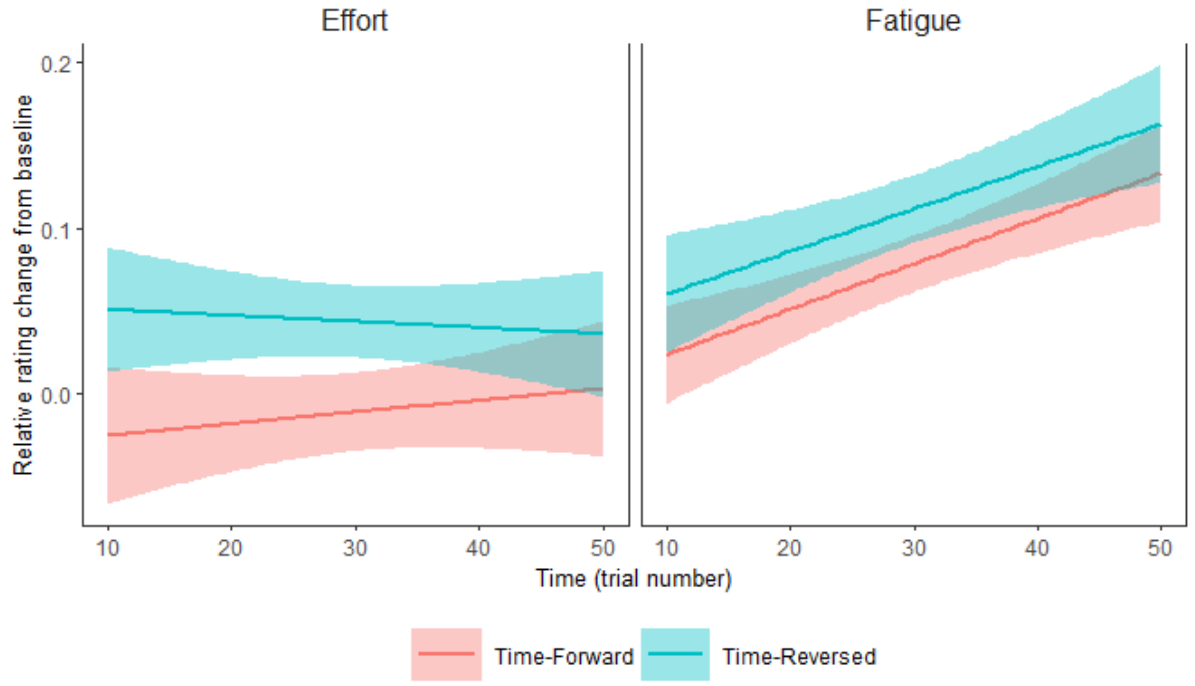


FIG. 23. Mean ratings of effort and fatigue after trials 10, 20, 30, 40, and 50. The scale on the y-axis corresponds to the re-scaling of the original effort and fatigue questions (effort/20, fatigue/10).

3.4.3 DISCUSSION

In Experiment 7, we used an adaptive procedure to obtain the SRTs at approximately 66% among intelligible maskers, and SRTs at approximately 41% among unintelligible maskers. The SNRs required for these SRTs were approximately the mean starting performance levels of the unintelligible and intelligible masker conditions respectively in Experiment 6.

The behavioural results of Experiment 7 showed a difference in mean performance across experimental conditions; mean performance was higher in the time-forward compared to the time-reversed condition, which is not surprising considering we ensured that starting performance was higher for the time-forward condition than for the time-reversed condition. Similar to Experiments 5 and 6, we found a significant effect of Time, suggesting that performance increased across the experimental conditions. However, we found no Masker by Time interaction, suggesting that these performance trajectories were similar across experimental conditions. This lack of significant difference in improvement over time suggests that speech recognition performance might be a function of the SNR, and in turn the initial starting performance, in conjunction with the linguistic interference typically observed in the literature (Calandruccio et al., 2013; Brouwer et al., 2012; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007).

If speech recognition performance was dependent upon linguistic interference alone, we would have observed faster improvement in the unintelligible than intelligible masker conditions, as in Experiment 6. If speech recognition performance was dependent upon SNR and starting performance alone, we would have observed faster improvement in the intelligible masker condition where we artificially created a more favourable listening condition compared to the unintelligible maskers. Note that in Experiment 5 we found an interaction between Masker and Time when starting performance was set around 50%, but that there was no equivalent significant interaction in Experiment 6 where starting performance between the masker conditions naturally differed at a fixed SNR. It could be the case that differences in performance trajectory only emerge when listeners start at equivalent performance levels as opposed to when starting performance differs between conditions, either naturally due to differing linguistic interference as in Experiment 6, or when difficulty is artificially manipulated as in Experiment 7. These findings suggest that performance trajectory might also

be bound to SNR and starting performance in conjunction with linguistic interference, which could result in a ceiling as found in Bent et al. (2009). (Note that in Bent et al., 2009, each condition comprised 100 sentences and improvement plateaued after 40 and 60 trials for two-talker babble and noise-vocoded speech respectively.)

The pupillometric results of Experiment 7, like in Experiment 6, showed that TEPR decreased over time, here suggesting that speech perception in noise is an automatic 'primitive process' not requiring sustained levels of listening effort (Bregman, 1990). There was also a significant effect of Masker, with generally higher TEPR in the intelligible (time-forward) masker than the unintelligible (time-reversed) masker, broadly supporting the finding of Koelewijn et al. (2012) that intelligible maskers require more cognitive effort to ignore than unintelligible maskers. However, there was no difference in the decrease in TEPR between masker conditions. Even though performance was higher in the intelligible than unintelligible maskers, the physiological measures of listening effort was also higher in the intelligible than unintelligible masker. These results suggest that even though participants were performing better in the condition at a higher SNR, the amount of effort needed to overcome the linguistic interference present in this condition was higher than the effort needed to listen to speech in unintelligible maskers. Even when the SNR between target and maskers was made more favourable, higher levels of cognitive effort, as measured by pupil dilation, were required to achieve the higher performance. These results indicate that linguistic interference, even at more favourable SNRs, requires more listening effort than maskers with no linguistic interference, even if the unintelligible maskers are at a relatively lower SNR than intelligible maskers.

Similar to Experiments 5 and 6, subjective ratings of fatigue increased over time across conditions whereas self-reported effort did not increase over time, suggesting that listeners are attuned to fatigue from sustained listening to speech in noise. However, like in Experiments 5 and 6, there was no differences over time nor between maskers in ratings of effort, suggesting

that listeners are not attuned to changes in cognitive effort required to complete a task, even though physiological pupillometric measures in listening effort identify changes over time and between different masker conditions in Experiments 6 and 7.

The results of Experiment 7 taken together with the results of Experiments 5 and 6 thus suggest that speech recognition performance is dependent upon the starting performance and SNR between target and masker talkers in conjunction with any linguistic interference effect resulting from the distracting intelligible content from competing talkers.

3.5 GENERAL DISCUSSION

The aim of this study was to investigate how listeners learn to ignore competing talkers over time. We asked whether adaptation to a masker is easier or harder if the masker is intelligible compared to being unintelligible. We also asked whether changes in performance are reflected in physiological (pupillometric) and self-reported measures of listening effort. We measured native English speakers' recognition accuracy of target English sentences in the presence of intelligible (time-forward English two-talker babble) versus unintelligible (time-reversed English two-talker babble) maskers over 50 trials. Trial-by-trial changes in task-evoked pupil responses were calculated and subjective ratings of effort and fatigue were collected every ten trials. Experiment 5 used an adaptive procedure to set the starting performance at approximately 50% across masker conditions. Experiment 6 used a fixed SNR across masker conditions. Experiment 7 used an adaptive procedure to make the SNR for the intelligible masker condition easier (targeting 66% SRT), and the SNR for the unintelligible masker condition more difficult (targeting 41% SRT). We also measured mean pupil dilation during each trial, and self-reported measures of effort and fatigue after every ten trials.

3.5.1 Discussion of Behavioural and Pupillometric Results

In Experiment 5, there was no difference in overall performance between the intelligible and unintelligible masker conditions, resulting from the 50% SRT procedure applied for each participant in each condition before the main experimental block occurred. Across maskers, participants improved over the course of the block, but there was faster improvement in the intelligible than unintelligible maskers. Pupil dilation did not decrease over time, and were at similar levels for both the intelligible and unintelligible maskers. The behavioural results were surprising, as one would expect faster improvement in the unintelligible maskers due to the presence of linguistic content, and thus informational interference in the intelligible maskers, as reported by Mepham et al. (2022). However, these results might have arisen from a higher SNR needed to achieve the 50% starting performance in the intelligible than unintelligible maskers, and could thus have been easier to perceptually segregate the target from masker talkers in the intelligible condition.

Experiment 6 was run to dissociate whether linguistic interference or SNR results in faster improvements in performance. Using a fixed SNR of -1.5 dB across conditions, we found a typical informational masking effect, with higher performance in the unintelligible than intelligible maskers, which we attribute to linguistic interference present in the intelligible maskers (Mepham et al., 2022). Participants' performance improved across both masker conditions, though there was no difference in the rate of improvement between masker conditions. This lack of interaction between Masker and Time differs from the results of the Mepham et al. (2022) results, that found when native English listeners were listening to time-forward and time-reversed English and Mandarin maskers, there was faster improvement in the time-reversed than time-forward maskers. Even though the pattern was not significant, the tendency of the Masker by Time interaction mimicked the interaction found in Mepham et al. (2022).

The difference in the sentence recognition accuracy results in Experiment 6 compared to the Mepham et al. (2022) results may be due to multiple factors. First, Mepham et al. (2022) had four experimental masker conditions in a 2 x 2 design (Masker Language: English, Mandarin; Masker Direction: time-forward, time-reversed), whereas in Experiment 6 there were only two masker conditions (time-forward and time-reversed English two-talker babble). The Mepham et al. (2022) significant interaction thus comprises both the Mandarin and English maskers into each Masker Direction category, and demonstrates that listeners improve faster in time-reversed speech than in natural language time-forward speech. There was no three-way interaction between Masker Direction, Masker Language and Time in the Mepham et al. (2022) results, indicating that when the effect of Masker Language is teased apart from Masker Direction, the rate of improvement is similar across all four masker conditions, echoing the lack of significant interaction found in Experiment 6. (Although the lack of significant three-way interaction might be a result of lack of statistical power, the very low X^2 and high p values indicate that even with a larger participant sample there would still not be an interaction between Masker Language, Masker Direction and Time).

Alternative explanations why a significant interaction between Masker and Time was not found in Experiment 6 could result from differences in the experiment design. The first of these methodological differences in the SNR used in Experiment 6, -1.5 dB, and in the Mepham et al. (2022) study, -3 dB SNR. The more favourable SNR used in Experiment 6 could have meant that there was greater potential for improvement in the 'harder' intelligible maskers than in the equivalent conditions in the Mepham et al. (2022) study, and thus result in similar improvement trajectories for the intelligible and unintelligible maskers in Experiment 6. Undertaking the sentence recognition task at multiple SNRs, similar to the approach taken by Wendt et al. (2018) appraising pupillometric responses at varying SNRs, could elucidate not only the SNR where there is maximal difference in performance improvement between speech

recognition in intelligible and unintelligible maskers, but also the SNR range where differences in improvement trajectories are observed in speech recognition between maskers. The second methodological difference is the speech-in-noise task used in the two experiments, a sentence recognition task with verbal responses in Experiment 6 compared to a sentence transcription task with written responses in Mepham et al. (2022). These different findings between Experiment 6 and the Mepham et al. (2022) study might have arisen due to the nature of the tasks, potentially requiring different cognitive processes to respond in a verbal and written manner.

The pupillometric data differed between Experiments 5 and 6, with pupil dilation in Experiment 6 decreasing across masker conditions. The decreases in pupil dilation in this experiment align with the suggestion that speech stream segregation is an automatic 'primitive process' that does not require sustained levels of listening effort (Bregman, 1990). Additionally, there was a faster decrease in the unintelligible maskers compared to the intelligible masker condition. The pupillometric results align with the informational interference effect found in the behavioural data, with faster decreases in effort in the 'easier' unintelligible maskers (i.e., higher performance) than in the more challenging intelligible maskers (i.e., lower performance). These results also suggest that decreases in pupillometric measures of listening effort are not necessarily a mirror of increases in behavioural performance. Instead, they might reflect the behavioural Masker effect: higher levels of listening effort maintained in the intelligible masker condition with linguistic interference than the unintelligible maskers with no available linguistic content (Koelewijn et al., 2012). This pattern of decreases in mean pupil dilation between intelligible and unintelligible masker conditions can be compared to the pattern of results from Winn et al. (2015) who found that participants experienced higher magnitudes and rates of pupil dilation when listening to noise-

vocoded speech with fewer channels compared to vocoded speech with more channels, even when participants were accurately reporting entire target sentences correctly.

Experiment 7 was run where the starting performance for the two masker conditions in Experiment 6 was used for the opposite masker condition in Experiment 7, achieved using a 66% adaptive procedure for the intelligible masker condition, and a 41% SRT for the unintelligible masker condition. The results of Experiment 7 showed both a Masker effect (from the manipulated SNR between the conditions) and improvement over time across conditions. However, in this experiment, there was no difference in the rate of improvement over time between the masker conditions, even when the intelligible masker condition had a more favourable SNR than the unintelligible masker condition. This might mean that when starting performance is very different between two similar masker conditions that differ only in the linguistic content available to a listener, the different pattern in performance trajectories might be too fine to tease apart. Alternatively, the reasons why listeners' improvement trajectories differ across the masker conditions could have resulted from different mechanisms inherent to overcome the adverse conditions, e.g., linguistic interference, low SNR, diminished dynamic range.

Regarding the pupillometric measures in Experiment 7, there was a decrease across masker conditions, as seen in Experiment 6, demonstrating less cognitive effort employed over the course of the listening conditions. There was also an effect of Masker, with higher TEPRs in the intelligible than unintelligible masker conditions. Even with a more favourable SNR and a higher performance in the intelligible masker condition, participants still had higher TEPRs elicited in this condition with linguistic interference. These results suggest that linguistic interference still requires more cognitive effort than listening to speech in unintelligible maskers (Koelewijn et al., 2012), even when the SNR for the former is more favourable and results in higher performance than the latter (and can be interpreted in light of the results of

Winn et al., 2015, whereby even at matched performance accuracy, greater speech signal degradation elicited higher magnitudes and rates of pupil dilation). In Experiment 7 there was no difference in the rate of TEPR decrease over time, suggesting that although the levels of TEPRs, and thus cognitive effort, were different between the masker conditions, rates of TEPR decrease might pattern similarly even when the adverse conditions themselves are different, i.e., linguistic interference for the intelligible maskers, low SNR for the unintelligible maskers.

Taken together, the results from Experiments 5-7 demonstrate a complex picture of adaptation to speech in noise over time. The findings suggest that it is only when the initial performance level is equated by manipulating extrinsic conditions of the listening environment (e.g., SNR) with sufficient dynamic range for improvement in performance that differences in the rate of adaptation to competing speech emerge. These differences in improvement in auditory object formation and suppression of competing talkers only arise when the SNR is more favourable, which then leads to the competing maskers becoming 'easier' to group and inhibit.

Synthesising the results of the pupillometric data is equally challenging. Experiment 7 suggests that linguistic interference from intelligible maskers results in higher levels of cognitive effort compared to unintelligible maskers even when performance is higher and SNR more favourable with intelligible maskers. This pattern of results is not reflected in Experiment 5, where pupil responses were similar across conditions and did not decrease over time. If linguistic interference from informational masking results in higher levels of pupil dilation, this pattern should have been observed in Experiment 5, instead of the equivalent levels of pupil dilation observed across conditions. Alternatively, if pupil dilation is inversely proportional to speech recognition performance, we should have observed higher levels of pupil dilation in the unintelligible masker condition compared to the intelligible masker condition. Explanations for these discrepancies in the pupillometric findings, including the role of motivation in listening

to speech in noise and the operationalisation of pupillometric measures to assess changes in physiological state, are explored in more detail in Sections 4.2.4 and 4.2.5 respectively.

3.5.2 Subjective and Pupillometric Measures of Effort

In Experiments 6 and 7, there was a decrease in TEPR over the course of the experimental block. These results are in line with the existing literature demonstrating a decrease in pupillometric measures of effort over the course of an experimental task, and aligns with the body of evidence that participants employ less effort, as measured by pupil dilation, as they become familiar with an experimental task (Bregman, 1990; Brown et al., 2020; Paulus et al., 2020). In contrast, there were no decreases in TEPR over time in Experiment 5. Additionally, the self-reported measures of effort and fatigue do not necessarily align with the pupillometric data. There was no change in participants' self-report of effort for any of the masker conditions across the experiments. But there were consistent increases in self-reported fatigue as the experimental conditions progressed across the experiments.

Considering the patterns of TEPR decreases over time across conditions, these decreases could be interpreted in various ways. The first is that decreases in listening effort as measured by pupil dilation in Experiments 6 and 7 are the result of an automatic learning to separate the target and masker talkers as distinct auditory objects, and demonstrate the 'primitive process' of auditory scene segregation (Bregman, 1990; Sussman, 2017). The lack of an effect of time in Experiment 5 can also be interpreted using this paradigm when considering the TEPRs in Experiment 5 were at similar levels to TEPRs at the end of Experiments 6 and 7, as well as in the native accented speech condition in the Brown et al. (2020) study. These lower TEPRs in Experiment 5 could then be interpreted as the task not requiring as high levels of listening effort compared to Experiments 6 and 7, rather than high

levels of sustained listening effort required for adaptation to the maskers (Huyck & Johnsrude, 2012). In Experiment 5, we used an adaptive procedure to find the 50% SRT for each participant in each condition, both to compare the change in performance over time and also because starting an experiment where performance is approximately 50% has been shown to elicit the most pupil dilation (Wendt et al., 2018). But these high levels of TEPR were not observed in Experiment 5. An alternative explanation for not observing these high levels of TEPR could be due to the nature of the adaptive procedure prior to each experimental condition. Although the SNR between target and masker sentences fluctuated throughout the adaptive procedure to obtain each participant's 50% SRT, this could have resulted in the participants becoming familiar with the task procedure, and therefore not eliciting higher TEPRs while acclimating to the experimental paradigm. This would then result in reduced rapid adaptation at the onset of the experiment and more consistent levels of TEPR throughout the experiment. However, Experiment 7 also used an adaptive procedure to obtain participants' 41% and 66% SRT in the unintelligible and intelligible maskers respectively, and higher TEPRs were observed both in the intelligible maskers compared to the unintelligible competing speech, as well as in the start of the experimental condition compared to the end. Although the 41% and 66% SRTs were extracted using a logistic function on the adaptive procedure data, this potential 'practice effect' from the adaptive procedure did not spill over into the TEPRs of the experimental conditions, with higher TEPR levels at the start of the experimental conditions. The absence of a decrease in TEPR over time in Experiment 5 is then unlikely to result from exposure to the target and masker talkers during the adaptive procedure.

Across the experiments, the different pattern of results in subjective self-reported measures compared to the pupillometric data suggest that participants might be more attuned to increases in task-related fatigue than actual cognitive effort. The design of our experiments included subjective self-report questions about both effort and fatigue, which allows us to

assess how a listener's subjective perception of changes in their physiological state compare to the physiological changes measured by pupillometry. However, if questions about effort were asked in isolation, without also asking about perceived changes in fatigue, this could conflate the distinct pattern of results between effort and fatigue observed in this study (McGarrigle et al., 2014), and we might have observed changes in self-reported effort that were not present in Experiments 5-7. But the stable self-reported ratings of effort and the increased ratings in self-reported fatigue across experiments still differ from physiological measures of listening effort as interpreted from the pupillometric data: the pupillometric measures in Experiment 5 pattern with the self-reported effort ratings in a lack of change over time, whereas TEPRs in Experiments 6 and 7 show decreases in TEPR across masker conditions. What can be deduced from these inconsistent patterns between self-reported effort, fatigue, as well as the pupillometric data is that the measures of pupil dilation are potentially identifying decreases in cognitive effort that are below the level of consciousness in the participants (evidence of lack of correlations between subjective ratings and pupillometric measures of effort, e.g., Strand et al., 2018; Moore & Picou, 2018) or that these decreases in pupil dilation are identifying changes in participants' physiological states that correspond neither to effort nor fatigue. However, as the lack of correlation between self-report and pupillometric measures of effort and fatigue has been documented (McGarrigle et al. 2017), this argument of an inconsistent relationship between subjective and physiological measures, potentially resulting from methodological differences, is more likely than an alternative explanation for this distinction between self-report and pupil dilation measures of effort and fatigue. Alternative explanations why no differences were observed in subjective measures of listening effort in Experiments 5-7 are discussed in more detail in Section 4.2.5 in the context of the operationalisation of pupillometric measures and subjective self-reports to assess changes in a listener's physiological state.

3.5.3 Limitations and Remaining Questions

We used an adaptive procedure to identify the 50% SRT for both conditions in Experiment 5, and the 66% SRT and 41% SRT for the intelligible and unintelligible masker conditions respectively in Experiment 7. However, in Experiment 7, the 50% SRTs for the experimental conditions were considerably higher than the 50% SRTs used in Experiment 5 for both the intelligible and unintelligible masker conditions. Although this should not have any functional consequences (an adaptive procedure is used to tailor the SNR to the right level for each participant in each condition), the question remains whether the higher SRTs used in Experiment 7 would have any repercussions on the results for this experiment. In Experiment 7, the mean performance in the unintelligible condition overall was 54.8%, which is much higher than the intended starting point of 41% taken from Experiment 6. One could argue that the more favourable SNR in Experiment 7 resulted in higher performance than anticipated, even with the SNR manipulation. However, the mean performance in the intelligible masker condition was lower than the intended starting point of 66%. Generally higher performance across both masker conditions could indicate that the higher SNRs in Experiment 7 compared to Experiment 5 for the 50% SRT could call into question the validity of these results. However, because the higher SNR did not result in generally higher performance than anticipated from the adaptive SRT procedures, this difference in SNRs between experiments is likely to result from differences in the participant samples, with participants in Experiment 7 needing generally higher SNRs to achieve the same 50% SRT across masker conditions than the participants in Experiment 5.

3.6 CONCLUSIONS

In conclusion, the results of the three experiments in this study taken together demonstrate clear effects of linguistic interference, both on sentence recognition performance (Experiment 5: 50% SRT, Experiment 6: sentence recognition accuracy) and in pupillometric measures of listening effort (Experiment 6 and 7). Across all experiments, sentence recognition accuracy increased during sustained exposure to competing talkers, and pupillometric measures of listening effort decreased over time in Experiments 6 and 7 but not in Experiment 5. Taken together, these results demonstrate the ability of listeners to both perceptually segregate competing speech streams, and improve in speech recognition in noise. Different rates of performance improvement between conditions were found when initial performance was matched between conditions, with faster improvement in intelligible maskers at a more favourable SNR than unintelligible maskers (Experiment 5), but not when starting performance differed between masker conditions (Experiments 6 and 7). Decreases in pupillometric measures of listening effort differed between maskers at a fixed SNR (faster decreases in unintelligible maskers, Experiment 6), than when SNRs were obtained using adaptive SRT procedures. Across experiments, participants' subjective ratings of fatigue increased across conditions, but there was no difference in subjective ratings of effort between conditions in any of the experiments.

3.7 ACKNOWLEDGEMENTS

This research was supported by a research grant from the Leverhulme Trust (RPG-2019-152) to S.L.M. We are grateful to Lauren Calandruccio and Ann R. Bradlow for providing the English BKB-R sentences and Mandarin audio files for the BKB-R sentence translations. We also thank Miaomiao Yu for recording the English sentences. We are also grateful to the speakers who provided speech samples for the initial pilot study. All Appendices can be found at https://osf.io/nhcrw/. All audio stimuli, analysis scripts and data files can be found at the following links for Experiment 5 (https://osf.io/m4b57/), Experiment 6 (https://osf.io/pmhsx/), and Experiment 7 (https://osf.io/6hkp9/).

4. GENERAL DISCUSSION

The preceding experimental chapters have sought to investigate the effect of informational masking on speech recognition in noise, exploring how the language status of a listener (i.e., listening to target speech in a native versus non-native language) impacts on transcription performance of masked speech and how listeners improve over time (Chapter 2), as well as how speech recognition performance is reflected in pupillometric measures of listening effort and the time-course of adaptation to masked speech (Chapter 3). The final chapter of this thesis will discuss the major findings from the experimental research presented in the previous chapters, the theoretical implications of these findings, as well as their limitations and directions for further research.

Within this chapter, Section 4.1 presents the summary of findings across the empirical experiments presented in Chapters 2 and 3. Section 4.2 then reflects on how the findings of the empirical chapters of this thesis relate to theories of informational masking, multilingual listening, and pupillometric and subjective measures of listening effort. Section 4.3 concludes the discussion of the entire thesis.

4.1 SUMMARY OF FINDINGS

The experiments in Chapter 2 sought to assess a number of questions. Firstly, there were two competing accounts (albeit somewhat overlapping) of how a listener overcomes linguistic interference when attending to a target talker and ignoring competing speakers. One account, termed the *known-language account*, hypothesises that a competing speech masker in a language known to the listener will be more disruptive to target speech perception than a

masker in an unknown language, i.e., any intelligible linguistic content will disrupt speech recognition to the same extent (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). An alternative account, termed the *linguistic similarity account*, hypothesises that speech recognition should be worse if the masker language is phonetically similar to the target language, irrespective of whether the masker language is known or unknown to the listener (Brouwer et al., 2012; Calandruccio et al., 2013, 2017; Van Engen & Bradlow, 2007). The four experiments in Chapter 2 sought to disentangle these two mechanisms of overcoming linguistic interference by assessing speech recognition amidst intelligible and unintelligible competing speech.

A second question that the experiments in Chapter 2 sought to address was the extent to which linguistic interference changes through adaptation over the course of a test block. There is evidence that speech recognition performance often increases over time as participants adapt to the distorted target signal, the competing speech to be ignored, as well as the task procedure (Bent et al., 2009; Brown et al., 2020; Cooke et al., 2022; Erb et al., 2012, 2013; Lie et al., 2024; Paulus et al., 2020; Versfeld et al., 2021). The experiments in this thesis sought to identify whether these improvements in performance differed between intelligible and unintelligible competing speech; a distinction between rates of improvement would point to inherent differences in how intelligible competing speech is segregated from a target talker and ignored compared to unintelligible maskers.

The third question that the experiments in Chapter 2 sought to explore is whether linguistic interference and improvement over time are affected by whether listeners perform the task in their native language as opposed to a non-native language. Previous studies have demonstrated that non-native listeners experience greater informational masking than native listeners (Cooke et al., 2008), but that both native and non-native listeners experience most informational masking when the target and masker languages are matched (Van Engen, 2010).

The experiments in Chapter 2 assessed whether the mechanisms of overcoming linguistic interference were similar between listeners attending to a target talker in native and non-native languages, and whether any changes in speech recognition accuracy over time was similar between native and non-native listening.

The experiments in Chapter 3 sought to build upon the results of Chapter 2 by assessing how listening effort changes over the course of an experiment when listening to a target talker amidst intelligible and unintelligible competing speech. When considering speech-in-noise perception, there are two alternative mechanisms that could be operational. The first is that the segregation of a target talker from distracting maskers occurs automatically and at a low, sensory level, not requiring a listener's conscious effort to attend to one talker and ignore the rest. This account aligns with the hypothesis posited by Bregman (1990) that stream segregation is a 'primitive process', i.e., stream segregation occurring automatically and at the outset of perceiving an auditory scene without requiring attention. This 'primitive process' account has been used by Sussman and colleagues (Sussman 2005; Sussman & Winkler, 2001; Sussman et al., 1999, 2002) to describe the process of auditory scene analysis. These studies primarily used event-related brain potentials to assess how changes in tone presentation from distinct sound streams are processed when tone presentation changes in to-be-ignored streams, with the authors concluding that initial stream segregation is automatic and does not require directed attention (Sussman, 2005; Sussman & Winkler, 2001; Sussman et al., 1999, 2002).

However, speech comprehension is a more high-level process than detecting changes in tone presentation, and it might require more cognitive resources for speech-in-noise recognition. In the context of the experiments in this thesis, while there may be changes in speech recognition accuracy as listeners adapt to an adverse multi-talker babble listening environment, one would observe either a low magnitude of listening effort with little change

over time, or reductions from an initially high level of listening effort (resulting from the initial adaptation to a new task or environment) to lower levels in measures of listening effort.

In a review appraising research for both passive and directed attention in auditory scene analysis, Sussman (2017) described how after automatic stream segregation, attention can be used to process an attended-to speech stream. If the task at hand is sufficiently demanding to require constant attention, this automatic 'primitive process' might be eclipsed by the need for sustained attention, and could result in high levels of listening effort. An alternative mechanism to the 'primitive process' account is that segregating a target talker from competing speech requires a listener's conscious effort to attend to the target talker in a top-down manner. Neuroimaging studies have found an interaction between top-down directed attention and bottom-up processing of auditory stimuli, with the prefrontal cortex active during top-down controlled attentional shifts (Rule et al., 2002; Salmi et al., 2009). These results provide evidence for the *dynamic filtering theory* proposed by Shimamura (2000) that the prefrontal cortex is directly involved in the selection, maintenance, updating, and rerouting of information processing. This effect whereby attention is required for rapid perceptual learning has also been shown when listening to speech (Huyck & Johnsrude, 2012) whereby listeners benefit from training to vocoded speech when attending to perceptual training, compared to when listeners experienced no listening training or were distracted by other auditory or visual stimuli. Huyck and Johnsrude (2012) interpret this finding as attending to degraded speech being necessary for learning, rather than an automatic or passive process. If speech recognition in noise is primarily undertaken by top-down processes of directed attention, one would then observe measures of listening effort reflecting this top-down mechanism throughout the time in which a listener is engaged in speech-in-noise perception, either by measures of listening effort remaining stable and at a high magnitude (to cope with the sustained level of conscious effort

to stream one talker from other talkers) or by an increase in effort reflecting the cumulative conscious effort to attend to the target talker.

To compare these two accounts of how listening effort changes over time, the experiments in Chapter 3 had similar speech recognition task procedures to those in Chapter 2. In the experiments in Chapter 3, in addition to behavioural performance, listening effort was measured using pupillometry, as well as self-reported measures of effort and fatigue.

4.1.1 INFORMATIONAL MASKING AND SPEECH PERCEPTION IN NOISE

The results from the experiments in Chapter 2 showed that participants experienced greatest informational masking when listening to target speech in the presence of intelligible competing speech that matched the language of the target talker. There was worse performance in English competing speech for native English listeners in Experiment 1, and worse performance in Mandarin competing speech for Mandarin-English bilingual listeners in Experiment 2. Across all experiments in Chapter 2, there was greatest performance in the time-reversed masker conditions, where no intelligible linguistic content could be gleaned from the masker speech. These results are consistent with the literature on informational masking, with greatest interference on speech perception in noise in the presence of intelligible compared to unintelligible maskers (Brouwer et al., 2012; Calandruccio et al., 2013, 2017; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007).

Taken together, these results support the linguistic similarity account: both listener groups listened to the target speech in their native language, and while the Mandarin-English bilingual listeners had access to the linguistic content of both the Mandarin and English competing speech, both groups experienced the most linguistic interference in the experimental

condition where the target and masker languages were matched. This finding is inconsistent with the known-language account, as the Mandarin-English bilingual speakers would have had similar performance in the time-forward English and time-forward Mandarin maskers under the known-language mechanistic account. Thus, when listening to target speech in a native language, greater linguistic similarity of the masker speech relative to the target speech causes the most linguistic interference, irrespective of whether the listener knows the linguistic content of the competing speech.

In contrast, in Experiments 3 and 4, participants who attended to target speech in a non-native but known language experienced a different pattern of results: listeners experienced interference to similar extents in any known language masker, regardless of whether the masker language matched the language of the target talker. There was still better performance in the unintelligible time-reversed conditions. In contrast to the pattern in Experiments 1 and 2, the pattern for Experiments 3 and 4 is consistent with the known-language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007), whereby any language known to a listener will cause interference regardless of whether the language is native or non-native to the listener. These differing patterns of results between the participants listening in their native language (following the linguistic similarity account) and those listening in a non-native language (following the known-language account) suggest that the underlying mechanisms while attending to a target talker in the presence of competing talkers differ between native and non-native listening.

The difference in how listeners dealt with intelligible and unintelligible maskers was also evident in the experiments in Chapter 3. The effect of informational interference manifested differently in these experiments depending upon the experimental paradigm. In Experiment 5, listeners underwent an adaptive procedure to identify their individual 50% SRT for each masker condition. Listeners required a more favourable SNR to obtain 50% correct in

the intelligible maskers compared to the unintelligible maskers, evidence of the impact of masker intelligibility on speech recognition in noise. In Experiment 6, the SNR was fixed, as in the experiments in Chapter 2, with a similar pattern of results emerging in the Chapter 2 experiments: higher speech recognition performance in the unintelligible maskers than in intelligible competing speech.

The results of both of these experiments align with the literature on informational masking, whereby a more favourable SNR is required to overcome the informational content of competing speech (Rhebergen et al., 2005), and that there is greater interference on speech perception in noise with intelligible maskers compared to unintelligible competing speech (Brouwer et al., 2012; Calandruccio et al., 2013, 2017; Van Engen & Bradlow, 2007). These results illustrate the greater disruption caused by intelligible linguistic content of speech by competing talkers compared to unintelligible competing speech, and that the effect of linguistic interference is robust enough to appear with different behavioural measures of speech recognition depending on the experimental paradigm. In Experiment 7, the impact of informational masking was most evident in the pupillometric measures of listening effort, which is discussed in Section 4.1.3.

4.1.2 ADAPTATION TO ADVERSE LISTENING CONDITIONS

In addition to investigating the main effect of linguistic interference on speech recognition, the series of experiments in Chapters 2 and 3 explored how listeners improved over the course of an experimental block. There is evidence that both speech recognition performance (Bent et al., 2009; Cooke et al., 2022; Erb et al., 2012, 2013; Felty et al., 2009; Lie et al., 2024; Versfeld et al., 2021) and pupillometric measures of listening effort (Brown et al., 2020; Paulus et al., 2020) change over the course of an experiment, with behavioural

performance increasing over time, suggesting adaptation to the distorted signal or improvements in selective attention, and pupil dilation decreasing as participants learn to ignore the competing speech as well as adapting to the task procedure. The experiments in this thesis sought to identify whether these improvements in performance differed between intelligible and unintelligible competing speech.

Overall, the results from Chapters 2 and 3 showed that performance improved regardless of whether or not the masker language was intelligible to the listener. Again, these findings are consistent with the literature exploring perceptual learning and adaptation over time: better speech recognition performance at the end of an experimental block compared to the start of the block through continuous exposure to adverse listening conditions or degraded speech (Bent et al., 2009; Cooke et al., 2022; Erb et al., 2012, 2013; Felty et al., 2009; Lie et al., 2024).

Although there was a general improvement over time in all experiments, there were differences in the rates of improvements depending upon the masker speech. In Experiment 1 with the monolingual English listeners, there was faster improvement in the time-reversed maskers than the time-forward maskers, demonstrating that intelligible linguistic content impedes the rate with which listeners can adapt to and ignore competing speech. Note, however, that this pattern was the same regardless of the masker language, suggesting that any language of time-reversed speech impedes speech recognition of a target talker to similar extents. However, the results of this experiment alone were not able to distinguish between the known language account and the linguistic similarity account, as the results of this experiment are consistent with both mechanistic accounts: greater linguistic interference when the masker is an intelligible language compared to an unintelligible language (linguistic similarity account), but also greater linguistic interference when the target and masker languages are matched (known language account).

In Experiment 2, with the Mandarin-English bilingual speakers listening to Mandarin target sentences, there were differences in improvement between the English and Mandarin maskers. With the Mandarin maskers, there was improvement in the time-reversed but not in the time-forward speech, which demonstrates the impact of informational interference on the rate of improvement. However, with the English maskers, there was improvement in both the time-forward and time-reversed conditions, with faster improvement in the time-forward condition.

Taken together, the results of Experiments 1 and 2 demonstrate that the improvement in speech transcription accuracy over time is generally present in all speech-on-speech conditions, except when the native language of the listener is also the language of the masker talkers, where less or no improvement over time is observed. With the monolingual English speakers in Experiment 1, there was generally lower improvement in the time-forward maskers, indicating that any speech-like maskers can interfere to a greater extent than time-reversed non-speech maskers. This pattern could have resulted from overlapping or mismatched phonemes from the masker talkers to the target (Kidd & Colburn, 2017), or from the bilingual Mandarin-English listeners in Experiment 2 having greater experience with suppressing known but irrelevant speech as a separate auditory object when listening in their native language (Shinn-Cunningham, 2008).

Regarding non-native listening, in Experiment 3, there were no significant interactions, showing that, although speech transcription accuracy improved across maskers, there were no differences in the rate of improvement between maskers. However, because the level of overall performance was generally low, Experiment 4 aimed to replicate Experiment 3 with a higher SNR to assess whether this pattern of results remained at a higher overall level of performance. Experiment 4 showed a significant interaction: in the Mandarin maskers (the native known but to-be-ignored language), there was faster improvement in the time-forward maskers than the

time-reversed maskers, whereas in the English maskers (the non-native language matched to the target talker), there was no difference in the rate of improvement between the time-reversed and time-forward maskers.

The greater improvement in the time-forward than time-reversed maskers could suggest that experience with bilingualism might confer some ability to inhibit the interference of a known non-native language, and could result from high levels of effort required for non-native listening compared to listening in a native language (Borghini & Hazan, 2018; Song & Iverson, 2018). However, without measures of listening effort present to assess why this difference in adaptation occurred between native and non-native listeners, it is not possible to determine whether differences in effort needed to undertake the speech recognition task were the cause of these different adaptation trajectories. The results of Experiment 4 are difficult to reconcile with the current literature about informational masking, and are discussed in Section 4.2.

The experiments in Chapter 3 demonstrated a more complicated picture of adaptation to speech in noise over time. These experiments sought to investigate both how speech recognition accuracy changes over the course of exposure to competing talkers and how listening effort required to undertake the task changes over time, as measured by pupil dilation and self-reported ratings of effort and fatigue. In Experiment 5, when speech recognition accuracy was initially equated between masker types at the beginning of each condition for each listener, there was a faster improvement in speech recognition accuracy in the intelligible maskers compared to the unintelligible maskers. This was counter to our expectations, as we had predicted, based on Mepham et al. (2022), that if there was informational interference from an intelligible linguistic content, this interference would inhibit the rate at which listeners would adapt to the adverse listening condition. Instead, we found the opposite pattern.

Experiment 6 investigated whether this faster improvement in the intelligible maskers would be present when the SNR was fixed across conditions, or whether we would find faster improvements in the unintelligible than intelligible competing speech, as observed in the experiments in Chapter 2, but we did not find any differences in the rate of improvement across masker conditions. We considered whether this could be because of the higher initial starting performance in the unintelligible masker condition compared to the intelligible competing speech, reducing the dynamic space within which listeners could improve.

Therefore, Experiment 7 investigated whether this difference in improvement would manifest when the intelligible masker condition was made 'easier' by a more favourable SNR, compared to maskers with no informational masking made 'harder' with a lower SNR. However, we found no difference in the rate of improvement between the intelligible and unintelligible masker conditions. Taken together, the results from the three experiments in Chapter 3 suggest that it is only when the initial performance level is equated by manipulating SNR such that a sufficient dynamic range for improvement is available that differences in the rate of adaptation to competing speech emerge. This difference in improvement in auditory object formation and suppression of competing talkers only then arises when the SNR is more favourable, which then leads to the competing maskers becoming 'easier' to group and inhibit.

4.1.3 PUPILLOMETRIC AND SUBJECTIVE MEASURES OF LISTENING EFFORT

In Chapter 3, pupillometric measures of listening effort were used to assess the impact of masker intelligibility on both speech-in-noise recognition and the listening effort associated with overcoming informational masking. Measuring listening effort with pupil dilation and self-reported ratings of effort and fatigue allows for investigation into whether adaptation to

competing speakers is automatic and effortless, or requires conscious effort on the part of the listener (see Section 4.1.1).

In Experiment 5, when starting performance between the intelligible and unintelligible masker conditions was equated using an adaptive procedure to obtain listeners' 50% SRTs for each condition, there was no difference in the magnitude of pupil dilation between intelligible and unintelligible maskers. These similar levels of TEPR likely arose from matching starting performance across masker conditions: if the SNR is manipulated to achieve equivalent performance across experimental conditions, one would expect listeners to put in the same level of effort to achieve this equated starting performance. There was also no decrease in TEPR over time, unlike what is commonly observed as a habituation response to the familiarity with the task and stimuli (Brown et al., 2020; Paulus et al., 2020). The lack of a main effect of time could have arisen from the adaptive procedure prior to each experimental condition. Although the SNR between target and masker sentences fluctuated throughout the adaptive procedure to obtain each participant's 50% SRT, this exposure to the target and masker talkers could have resulted in the participants becoming familiar with the task procedure, and therefore not eliciting higher TEPRs while acclimating to the experimental paradigm. This would then have resulted in a reduced need for rapid adaptation at the onset of the experiment and more consistent levels of TEPRs throughout the experiment. Although the lack of a main effect of time initially suggests that speech perception in noise is not automatic and requires sustained effort (Huyck & Johnsrude, 2012), the pupil dilation levels in this experiment were similar to those seen in the 'easier' native accented speech condition compared to the 'harder' non-native accented speech condition in the Brown et al. (2020) study which did elicit higher levels of TEPR. Taken together, this suggests that the pupil dilation here is more related to lower effort for speech-in-noise perception, compared to the higher effort expected in the Brown et al.

(2020) non-native accented speech condition, and higher levels of TEPR one would expect if sustained attention was required for perceptual learning (Huyck & Johnsrude, 2012).

In Experiment 6, when the SNR was fixed across masker conditions, there was a faster decrease in pupil dilation in the unintelligible competing speech compared to the intelligible maskers. This faster decrease likely reflects the lower level of effort required to achieve the higher recognition performance, resulting from the fixed SNR across experimental conditions: the 'easier' condition, where higher behavioural performance was observed, in turn resulted in less effort required to achieve this level of performance. There was no adaptive procedure in Experiment 6, unlike in Experiment 5, and though there were practice trials at the beginning of each condition at the experiment's fixed SNR for participants to familiarise themselves with the target talker speaking among the maskers, this different familiarisation approach might have resulted in the higher TEPR at the beginning of the block compared to the end of the block, with participants potentially being less familiar with the target talker at the beginning of experimental conditions in Experiment 6 than Experiment 5.

In Experiment 7, the SNR was manipulated such that the condition with informational masking was made 'easier' through a more favourable SNR, whereas the condition with no informational masking was made 'harder' through a lower SNR. This manipulation was conducted using an adaptive procedure to find a participant's 50% SRT for each masker condition, then the 41% SRT for the unintelligible maskers and the 66% SRT for the intelligible maskers were extracted from the logistic function plotted from participants' responses in the adaptive procedure. This SRT manipulation aimed to assess whether it was the release from linguistic interference that resulted in better performance in the unintelligible maskers in Experiment 6, or whether it was a direct result of the initial performance in the unintelligible masker condition starting at a higher level than in the intelligible maskers. In Experiment 7, even though performance was higher in the intelligible than unintelligible maskers, there were

still higher levels of TEPR in the intelligible maskers compared to the unintelligible maskers. This dissociation suggests that an intelligible masker still requires more effort to segregate and ignore than an unintelligible masker, even when the target talker is 'easier' to hear out (i.e., at a higher SNR). There were also decreases in TEPR over time across experimental conditions, as observed in Experiment 6, though there were no differences in the rate of TEPR decrease over time. The decreases in TEPR over time in Experiments 6 and 7 then suggest that the lack of main effect of time in Experiment 5 resulted in lower levels of effort in the experiment, and not necessarily an artefact of the adaptive procedure, as no main effect of time would then have been observed in Experiment 7.

Across Experiments 5-7, there were no differences in the levels of subjective listening effort across listening conditions, and self-reported ratings of effort neither increased nor decreased over time. In contrast, listeners' self-reported fatigue increased consistently across all conditions in Experiments 5-7. These results suggest that not only are listeners more attuned to changes in fatigue during speech perception in noise, but that the pattern of these subjective measures not mirroring the physiological TEPR results means that TEPR is not solely measuring neither cognitive effort nor fatigue in isolation. Taken together, the TEPR results across Experiments 5-7 paint an inconsistent picture of the nature of TEPR as a measure of listening effort or fatigue.

4.2 THEORETICAL IMPLICATIONS

4.2.1 INFORMATIONAL MASKING AND NATIVE LISTENING

In Chapter 2, there was a marked difference between patterns of results between native and non-native listeners. Overall, listeners attending target speech in their native language

experienced greater levels of interference when the competing speech was in the same language to the target speech, compared to unknown or non-language-matched maskers. Even when listeners had knowledge of the linguistic content of competing speech, performance was still better in these masker conditions than when the languages of the target and masker were matched. This pattern of results aligns with the target-masker linguistic similarity account, whereby listeners experience greatest interference when the target and masker languages are matched (Brouwer et al., 2012; Calandruccio et al., 2013, 2017). However, the results were different for listeners attending to target speech in a non-native language in that they aligned with the known-language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007), whereby listeners experience similar levels of linguistic interference from any known-language intelligible maskers compared to unintelligible time-reversed maskers.

These differences in the pattern of results between native and non-native listening could reflect different underlying mechanisms. Across experiments in Chapter 2, time-reversed speech was used as a control condition to English and Mandarin time-forward speech. The intention of this time-reversed manipulation was to have a 'baseline' masker that had similar features to natural speech (i.e., preserved spectral elements with an audible speech-like prosody, albeit with reversed temporal elements affecting speech sounds like plosives, Rhebergen et al., 2005; Rosen, 1992) but had no intelligible content. It could be argued that time-reversed speech does not completely remove informational masking since some phonemic features are preserved in the time-reversed signal (e.g., frication). However, for the purpose of discussing the linguistic components of informational masking, I refer to the time-reversed maskers as 'energetic maskers'. These maskers can then be compared to the informational masking that results from the time-forward maskers.

In Experiment 1, the difference in performance between the conditions, i.e., lowest performance in the language-matched time-forward maskers and best performance in

unintelligible time-reversed speech, could be explained simply by informational masking: no linguistic content of the Mandarin speech was intelligible to native English listeners, and thus could not interfere with the English target speech for the monolingual English listeners. Still, the lower performance in this condition compared to the time-reversed conditions could result from various factors, including overlapping or misallocated phonemes between the target and masker (Cooke et al., 2008), differences in glimpsing opportunities (Buss et al., 2020), natural fluctuations in the intensity level of the masker speech stream (Festen & Plomp, 1990), or greater familiarity with a natural but unknown language than with an artificial time-reversed speech-like masker. However, the arguments supporting glimpses and intensity level fluctuations do not hold, as one would then expect similar performances between the time-forward and time-reversed Mandarin masker conditions, as the average temporal aspects of the glimpses and the intensity level fluctuations were the same overall across these conditions (though note that 'ramped envelopes', similar to the structure of plosive phonemes in time-reversed speech, are perceived to be of longer duration than 'damped envelopes', similar to plosives in normal speech, Carlyon, 1996; Irino & Patterson, 1996; Rhebergen et al., 2005; Schlauch et al., 2001; Stecker & Hafter 2000). The phoneme misallocation hypothesis is the most likely explanation for the difference between time-forward matched maskers, time-forward unmatched maskers, and time-reversed maskers, in addition to a potential greater familiarity with natural but unknown Mandarin speech compared to an artificial and unintelligible speech-like masker.

Although the Mandarin-English bilinguals in Experiment 2 had access to the linguistic content of the competing English speech, they still showed a similar pattern to Experiment 1: best performance in unintelligible time-reversed maskers, then in the non-language-matched English time-forward masker, and worse performance in the language-matched time-forward Mandarin masker. Even with access to the linguistic content of the unmatched language

maskers, these listeners had better performance than the matched language maskers. Still, performance in the time-forward English condition was lower than in the unintelligible time-reversed conditions. These performance differences result from similar mechanisms to those hypothesised in Experiment 1, that it is easier to supress unintelligible speech-like maskers than natural speech (Rhebergen et al., 2005). However, this argument does not take into consideration the access to intelligible linguistic content in the time-forward English masker which needs to be supressed by listeners in Experiment 2, which is not present for the time-forward Mandarin masker for listeners in Experiment 1. This means that there is most interference when listening in one's native language *and* when the languages of the target and masker talkers, rather than simply having access to any linguistic content of competing speech.

Taken together, these results suggest that participants listening in their native language are able to suppress competing speech even if the masker is known to them. Performance is still not as high as when listening to an artificial speech-like masker, potentially resulting from misallocation of target and masker speech or from other factors such as familiarity or novelty of unknown but natural competing speech. Further research would need to be conducted to dissociate these different explanations of monolingual and bilingual speech perception patterning in similar ways, and to uncover the underlying mechanisms active in monolingual and bilingual native speech perception in known and unknown language maskers.

4.2.2 INFORMATIONAL MASKING AND NON-NATIVE LISTENING

What remains unclear from Experiments 1 and 2 is whether native listening, both monolingually and bilingually, is consistent with the known-language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007), whereby listeners experience similar levels of linguistic interference for any known-language intelligible maskers compared

to unintelligible maskers, or with the target-masker language similarity account, whereby listeners experience greatest interference when the languages of the target and the masker are matched (Brouwer et al., 2012; Calandruccio et al., 2013, 2017).

The results from Experiments 3 and 4 provide evidence that non-native listening is different from native listening, with non-native listeners in both experiments experiencing similar levels of interference in both known language maskers, compared to the unintelligible time-reversed maskers. Taken together, Experiments 1 to 4 suggest that the underlying mechanisms during native listening follow the target-masker language similarity account, whereas during non-native listening, the underlying speech perception mechanisms follow the known-language account. One explanation for the difference between native and non-native listening might be linked to the level of language activation when undertaking the task, as proposed by the Bilingual Interaction Activation Plus (BIA+) model (Dijsktra & van Heuven, 2002; van Heuven & Dijkstra, 2010). The BIA+ model describes the mechanisms by which bilingual listeners process language, and assumes that a bilingual listener has an integrated (i.e., combined) and non-selective access to lexicons across all their known languages, and that word identification and task or decision execution are distinct subsystems for goal accomplishment (i.e., using the right language to complete a task). For the native listeners in Experiment 2, task instructions and procedure were in Mandarin, as was the target language, with their non-native language, English, not being referred to in the experiment. It is therefore possible that their non-native language was not 'active' enough to cause linguistic interference on target speech perception compared to the native language competing speech, and thus have a pattern of results similar to the linguistic similarity account (Brouwer et al., 2012).

In contrast, the participants listening to target speech in their non-native language might have both the non-native language and the native language activated, causing interference to similar extents in the Mandarin and English masker conditions. One issue with this explanation

is that it assumes that listeners always have their native language ‘active’, and any non-native languages ‘inactive’, while the BIA+ model assumes that the bilingual lexicon is integrated and that access to it is language non-specific (van Heuven & Dijkstra, 2010), which would lead to the expectation of interference from all known languages, as described by the known language account (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007). There might still be linguistic interference from a non-native masker when listening in a native language, but this interference could be lower compared to interference from a native language. The fact that the monolingual listeners exhibited a similar pattern of interference to the bilinguals listening to their native language suggests that, although there might be some linguistic interference from the known non-native competing speech, this interference is more likely to be from energetic or phonological overlap between Mandarin maskers and English targets for the native listeners, and would explain why there was still lower speech recognition performance in competing speech when the masker language was not matched to the target talker, compared to when the maskers were time-reversed competing speech.

An issue with testing cognitive processes in multilingual listeners that might impact speech-on-speech performance is that of language proficiency: although a group of listeners might have the same language background, the proficiency of each individual’s languages and the context in which they are used can pose a challenge to synthesising a general theory of multilingual speech perception. The review of non-native listening from Scharenborg and van Os (2019) highlights how studies using participant groups with a range of non-native proficiency make comparisons between studies difficult. In addition, they highlight a tension to be balanced: whether to use standardised homogenous groups with similar non-native language proficiency or more heterogenous groups with varying non-native proficiency. The former allows for more controlled experimental comparisons to a control monolingual listener group, but might reduce the variability in performance that is not characteristic of a wider non-

native listening population or of non-native listeners with other language use statuses. The latter instead allows for a more general comparison of listeners from various language backgrounds which might be more reflective of the non-native listening population, but could result in wider variability compared to a control monolingual group that dampens any small effects between native and non-native listening. Although the synthesis of a general model of speech perception across native and non-native listening is outside the scope of this thesis, the results from the experiments in this thesis illustrate the need for different theories of the underlying mechanisms of speech-in-speech perception for native and non-native listening.

4.2.3 LANGUAGE STATUS AND ADAPTATION

Although the findings from Chapter 2 give evidence to linguistic interference resulting from informational masking, listeners appear able to not only supress the interference caused by informational masking, but are also able to adapt to adverse listening conditions and improve in speech-in-noise performance. As detailed in Section 4.1.2, improvement in speech recognition depended on whether a listener was listening in a native or non-native language: native listeners improved in speech-in-noise performance faster in unintelligible (Experiment 1) and unmatched maskers (Experiment 2), whereas for the non-native listeners, there were no improvements in speech-in-noise perception over time (Experiment 3) and faster improvement in the known unmatched competing speech (Experiment 4). One might expect that knowing the linguistic content of competing speech would slow down the rate of improvement in speech recognition compared to unintelligible speech maskers, due to cumulative high cognitive load resulting from linguistic interference (Lavie et al., 2004). However, in these experiments, we found either faster improvement in the intelligible time-forward Mandarin maskers compared to the time-reversed Mandarin maskers, or no significant improvement in the time-reversed

and time-forward English maskers. This faster increase in the intelligible masker conditions could arise from a very low starting speech recognition accuracy as listeners initially attuned to the adverse listening conditions, then rising to higher levels of performance, compared to generally higher levels of performance in the unintelligible competing speech with lower increases over time from the initially higher starting performance.

However, there are also other explanations that could account for these differences in improvement between the native and non-native listening experiments. It has been demonstrated that listening to non-native speech requires more cognitive effort than listening to native speech (Borghini & Hazan, 2018; Song & Iverson, 2018) and this increased cognitive effort might need to be sustained over the course of an experiment to maintain a high level of performance, with a reduced capacity for improvement in speech-in-noise perception over time if there is a persistent high cognitive load (Lavie et al., 2004). It has also been demonstrated that if a task is too difficult, the amount of cognitive effort invested in that task is reduced as listeners disengage from the task (Wendt et al., 2018). These cumulative high levels of cognitive load might in turn result in increased fatigue (Alhanbali et al., 2019; McGarrigle et al., 2021b; Strand et al., 2018) and limit any capacity for improvements to adapt to the competing speech.

An additional explanation of these results might lie in the broad non-native language proficiency range between the participants. It is critical to assess the language abilities of participants when undertaking experiments in a non-native language (Scharenborg & van Os, 2019) and although we attempted to account for language proficiency using IELTS scores, the heterogeneous language abilities of the non-native listeners might have increased statistical variance, making it difficult to observe any significant improvements over time with the number of participants in these experiments. Again, experiments specifically assessing the impact of non-native language proficiency on adaptation to speech maskers in non-native

listening will need to be run to elucidate whether these performance patterns are intrinsic to the nature of non-native listening, and are indeed distinct from informational interference patterns in native speech perception.

4.2.4 PUPILLOMETRIC MEASURES OF LISTENING EFFORT: TOP-DOWN VERSUS BOTTOM-UP PROCESSING

The results from the experiments in Chapter 3 assessing changes in listening effort using pupillometric measures provided an inconsistent pattern of results. Across Experiments 6 and 7 in Chapter 3, pupil dilation decreased over the course of the experiment as listeners adapted to the task procedure and stimuli, which aligns with the literature on how pupil dilation changes over the course of an experiment (Brown et al., 2020; Paulus et al., 2020). However, there was no main effect of time in Experiment 5, suggesting that the physiological state of the listeners did not change across the experiment, even when behavioural performance improved. As mentioned in Section 4.1.3, the TEPR in Experiment 5 was consistently lower, suggesting that the participants in this experiment did not employ as much effort in the experimental task compared to the other experiments, rather than these results in Experiment 5 arising from elements of the task design or procedure. The differences in the rate of decrease between maskers across Experiments 6 and 7, and the low levels for TEPR in Experiment 5 suggest that pupil responses do not necessarily covary with the difficulty of the task as measured by behavioural performance, and, taken together, these results can be interpreted that segregating target from masker talkers might be an automatic 'primitive process' (Bregman, 1990), with pupil dilation decreasing as listeners adapt to adverse listening conditions.

In Experiment 5, there were no differences between the rate of decrease in pupil dilation between the intelligible and unintelligible maskers, even though there were differences in the

rate of improvement in speech recognition performance. This pattern suggests that behavioural performance is not simply in inverse proportion with pupil dilation, otherwise these increases in behavioural performance would be reflected in different rates of decreases in pupillometric response. In Experiment 6, there was faster pupil size decrease in the unintelligible than intelligible masker condition, which was an expected difference resulting from the fixed SNR across participants and conditions. This pattern demonstrates that the impact of informational masking on pupillometric measures of listening effort differs when listening to speech in intelligible versus unintelligible maskers: i.e., listeners expend less effort for speech perception in unintelligible competing speech compared to speech perception in intelligible competing speech. In Experiment 7, there was no difference in the rate of decrease in pupil dilation, but there was greater average pupil dilation in the intelligible than unintelligible maskers, even though there was higher speech recognition performance in the intelligible than unintelligible maskers (resulting from the experimental manipulation of the SNRs to make the intelligible masker condition 'easier' and the unintelligible masker condition 'harder').

There are two aspects of the results of the pupillometric measures that are difficult to reconcile across these experiments. The first is the lack of a direct relationship between listening performance and listening effort. One might expect the level of listening effort to be similar when performance is similar, as observed in Experiment 5, but there was faster improvement in performance in the intelligible maskers than the unintelligible maskers, even though there was no decrease in listening effort as measured by pupil dilation. The second concerns the differences in the pattern of results between Experiment 5 and Experiment 7 that both used an adaptive procedure to obtain SRT values for listeners in each condition of the experiment. In both these experiments, an adaptive procedure was used to set the SRT for each participant at 50% in Experiment 5 and, in Experiment 7, at 66% in the intelligible masker condition and 41% in the unintelligible masker condition. However, while there was no

difference in magnitude in pupil dilation in Experiment 5 when the SRT was set to 50%, there was higher magnitudes of pupil dilation in Experiment 7 when there was a higher (i.e., 'easier') SRT for the intelligible masker condition compared to the unintelligible masker condition with a lower SRT.

Synthesising the results of the pupillometric results is challenging. Experiment 7 suggests that linguistic interference from intelligible maskers results in higher levels of cognitive effort compared to unintelligible maskers even when performance is higher and SNR more favourable with intelligible maskers, and supports previous findings that intelligible maskers require more cognitive effort to ignore than unintelligible maskers (Koelewijn et al., 2012). This pattern of results is not reflected in Experiment 5, where pupil responses were similar across conditions. If linguistic interference results in higher levels of pupil dilation, this pattern should have been observed in Experiment 5, instead of the equivalent levels of pupil dilation observed across conditions. Alternatively, if pupil dilation is inversely proportional to speech recognition performance, we should have observed higher levels of pupil dilation in the unintelligible masker condition compared to the intelligible competing speech in Experiment 7.

This discrepancy between the patterns of pupillometric results in Experiments 5 and 7 could be interpreted by understanding listener motivation to engage with the task. The Framework for Understanding Effortful Listening (FUEL, Pichora-Fuller et al., 2016) incorporates not only a listener's hearing difficulties and task demands, but also a listener's motivation to expend cognitive effort. In Experiment 5, a 50% SRT was used across masker conditions, resulting in similar speech recognition accuracy in the competing speech conditions. As this subjective difficulty was equated, one could then expect the level of motivation for the listeners to engage in speech recognition to be similar between conditions and thus result in similar levels of listening effort. In Experiment 7, the difference in SRT used

between the conditions might have resulted in different levels of motivation: although listeners had better speech recognition accuracy in the intelligible than unintelligible maskers because of the higher SRT, this better performance might have in turn conferred greater motivation to engage in speech recognition, and thus resulted in higher levels of pupil dilation. Considering motivation to engage in effortful listening then poses a problem when using pupil dilation as a measure of listening effort; if pupil dilation can reflect both listening effort and listener motivation, pupil dilation becomes a confounded measure, making it impossible to assess the independent contribution of each factor. In our experiments, there were no subjective measures of a listener's motivation to engage in the speech recognition task, which could have helped in interpreting whether the differences in pupil dilation between Experiments 5 and 7 resulted from differences in listener motivation for speech perception in intelligible and unintelligible competing speech.

4.2.5 PUPILLOMETRIC AND SUBJECTIVE MEASURES OF LISTENING EFFORT

Although there was no main effect of time in the pupillometric data in Experiment 5, there were consistent decreases in mean pupil dilation over the course of the experimental blocks in Experiments 6 and 7. However, the self-reported measures of effort and fatigue did not align consistently with the pupillometric data. There was no change in participants' self-report of effort for any of the masker conditions across the experiments, but there were consistent increases in self-reported fatigue as the experimental conditions progressed.

The subjective self-reported measures suggest that participants might be more attuned to increases in task-related fatigue than actual cognitive effort. The design of our experiments included subjective self-report questions about both effort and fatigue, which allows us to assess how a listener's subjective perception of changes in their physiological state compare to

the physiological changes measured by pupillometry. However, if questions about effort were asked in isolation, without also asking about perceived changes in fatigue, this could conflate the distinction between effort and fatigue observed in this study (McGarrigle et al., 2014), and we might have observed changes in self-reported effort that were not present in Experiments 5-7. But the stable self-reported ratings of effort and the increased ratings in self-reported fatigue across experiments still differ from physiological measures of listening effort as interpreted from the pupillometric data, which broadly show decreases across masker conditions. What can be deduced from these differences is that the measures of pupil dilation are potentially identifying decreases in cognitive effort that are below the level of consciousness in the participants. This explanation also aligns with Bregman's (1990) hypothesis of speech segregation as an automatic 'primitive process', not requiring top-down cognitive effort on the part of the listener.

Although subjective ratings of effort remained stable across experiments, listeners reported fatigue increasing as they progressed through the listening conditions. Returning to the FUEL model (Pichora-Fuller et al., 2016) discussed in Section 4.2.4, this model proposes that listener motivation is a key component in understanding why listeners decide to engage in effortful listening. The model also proposes that listener fatigue might expound the level of motivation to engage in a task, with subjective fatigue potentially manifesting as decreased motivation to continue doing a task. If listeners are becoming less motivated to engage with the task, this could be reflected in increased subjective listener fatigue, even if there are no detriments to speech recognition accuracy. However, to interpret these consistent increases in fatigue as resulting in decreases in listener motivation, self-report ratings of listener motivation or task engagement would need to be obtained. Future research exploring changes in physiological state over time would need to include self-report measures of motivation, not only to assess whether changes in pupil dilation are associated with decreased listening effort

or decreased listener motivation, but also to assess whether increases in fatigue are associated with decreases in listener motivation, or are independent from a listener's motivation to engage with a task.

Another explanation why these differences between pupillometric and subjective measures of listening effort arose could relate to the relative difficulty in answering questions about one's perceived internal state, and the point during the experiment at which these subjective measures were obtained. Although pupillometric measures of listening effort are deemed to be more objective than self-reported measures and correlate with task demand (Zekveld et al., 2018), there has been inconsistency whether pupillometric measures of listening effort correlate with subjective ratings of listening effort. This inconsistency is usually attributed to the method of obtaining the self-reported data: correlations are found when listeners are asked about their subjective ratings of listening effort during a testing session (McGarrigle et al., 2020), whereas no correlations are found when listeners are asked at the end of a testing session (Koelewijn et al., 2012; Strand et al., 2018).

Testing how subjective ratings of mental effort can be influenced by the number and type of subjective questions asked, Moore and Picou (2018) found that participants substituted their ratings on cognitive effort (a more 'computationally expensive' question) with ratings on their perceived performance, with lower ratings of performance corresponding to high ratings of effort. Although we did not include explicit questions about perceived performance in our experiments, listeners might still have substituted their ratings of effort with perceived accuracy: performance did not decrease over the course of experimental blocks, and this might have been perceived by the listeners as suggesting that a constant level of listening effort was needed to maintain performance, while pupillometric measures showed either stable or decreasing levels of physiological measures of listening effort across all conditions in all experiments.

Likewise, increases in ratings of fatigue might have been due to listeners finding it easier to respond to questions about fatigue than to tease apart changes in physiological state relating to effort, changes relating to fatigue, and changes relating to other aspects of one's physiological state. The number and type of questions investigating subjective ratings of effort, and how these questions on internal states interact with one another during an experiment, might be the reason why there is an inconsistent relationship between self-report and pupillometric measures of effort (Koelewijn et al., 2012; McGarrigle et al., 2020; Strand et al., 2018) and fatigue (McGarrigle et al. 2017; Wang et al., 2018b), and why in the experiments in this thesis there were no similarities in the pattern of subjective and pupillometric measures of listening effort.

To determine if the difference between pupillometric measures of listening effort and subjective ratings of effort and fatigue is due to pupillometric measures capturing a more complex physiological state, or the types and number of questions asked, a more comprehensive study would be needed to explore the relationship between pupil dilation and the types of self-reported questions about changes in physiological state. A systematic review by Shields et al. (2023) assessed the correlations between measures of listening effort from 48 papers and found that measures of listening effort are poorly correlated, potentially arising from the range of physiological, behavioural and subjective measures of effort and fatigue used across studies. As a result, Shields et al. (2023) proposed using composite measures of previously validated physiological, behavioural and subjective measures that incorporate the fatigue and stress of effortful listening. In the context of pupillometric measures of listening effort, this could be assessed using techniques such as exploratory and confirmatory factor analysis: different types of questions addressing the change of a listener's physiological state can be appraised by assessing how each question targeting subjective assessment of a listener's change of physiological state loads onto different factors, e.g., task-related cognitive effort,

task-induced fatigue, etc., in conjunction with behavioural measures (e.g., speech recognition accuracy, reaction time), physiological measures (e.g., pupillometric, electroencephalogram, functional near-infrared spectroscopic, Richter et al., 2023), or other cognitive tasks.

Using these composite measures and multivariate analysis techniques will not only allow for discrepancies between pupillometric and subjective measures of listening effort to be appraised, but also potentially resolve the current conflicting evidence of a relationship between physiological and self-reported measures of listening effort to be consolidated (Koelewijn et al., 2012; Moore & Picou, 2018; Shields et al., 2023; Strand et al., 2018). However, assessing effortful listening as prescribed by Shields et al. (2023) is likely to be incompatible with designs of experiments like those presented in this thesis, which investigated more granular changes in behavioural performance and listening effort. The proposition by Shields et al. (2023) for this type of comprehensive assessment of listening effort could be made in conjunction with multiple physiological measures of listening effort, as discussed by Richter et al. (2023), to provide a more holistic picture of how speech recognition accuracy and the corresponding magnitude of listening effort changes when listeners adapt to adverse listening conditions over time. Using composite subjective and physiological measures of listening effort in experiments with procedures similar to those presented in this thesis would allow an assessment of the relative contributions to changes in listening effort, fatigue, motivation, and other physiological states.

4.3 CONCLUSIONS

The empirical findings of this thesis have contributed to the study of speech perception in adverse listening conditions. One of the primary contributions relates to how speech in noise perception changes over time. The results from this thesis provide evidence that listeners

generally improve in speech in noise accuracy over time through adaptation to adverse listening conditions. Critically, the results also highlight intrinsic differences in rates of adaptation based on the language status of the listener, and whether a listener is processing target speech in a native or non-native language. The findings from this thesis have also demonstrated that there are different mechanisms by which listeners adapt to competing speech: linguistic interference in native listening follows the linguistic similarity account, whereas linguistic interference in non-native listening follows the known language account.

This thesis has also demonstrated how both physiological pupillometric measures of listening effort and subjective self-reported ratings of listening effort and fatigue change over time. As listeners adapted to speech in adverse conditions, listening effort, as measured by pupil dilation, largely decreased over time (except in Experiment 5 where no changes were observed), whereas subjective ratings of listening effort remained unchanged across Experiments 5-7. These dissimilar patterns between physiological and subjective measures of listening effort add to evidence that the two measures might index different constructs, and that pupil dilation might not be as robust a measure of listening effort as others claim.

The evidence in these experiments also confirms that, across behavioural and pupillometric measures, informational masking resulting from linguistic interference has a detrimental impact not only on average speech recognition performance, but also on how listeners adapt to adverse listening conditions. Where linguistic interference was present from the linguistic content of competing speech, and where there were no SNR manipulations to mitigate the impact of informational interference, there was worse overall performance, generally slower improvements over time, and slower decreases in pupil dilation. Where there were SNR manipulations to compensate for the effect of informational interference, listeners achieved equivalent or higher speech recognition performance, but still equivalent or even higher levels of listening effort among intelligible competing speech to achieve these higher

performance levels compared to unintelligible competing speech. Taken together, the results in this thesis evidence the impact of informational masking resulting from linguistic interference on overall behavioural performance and how listeners improve in speech perception in adverse conditions over time, as well as how physiological pupillometric measures of listening effort can elucidate the differences in physiological states when listening to speech against an intelligible or unintelligible masker.

There are differences in results presented across the chapters in this thesis, such as the differences between native and non-native listening in Chapter 2, with linguistic interference in native listening following the linguistic similarity account and non-native listening following the known language account, as well as differences between pupillometric and self-reported ratings of listening effort in Chapter 3, suggesting that our current understanding of multilingual speech perception and the interplay between physiological and subjective measures of listening effort requires further investigation. However, the conflicting evidence presented in this thesis which at times is difficult to synthesise provides an opportunity for further targeted investigation into the differences in adaptation to adverse conditions and resulting physiological measures of listening effort for native and non-native listening, and the intricacies of the relationship between physiological and subjective measures of cognitive effort.

5. APPENDICES

5.1 APPENDIX A: Anglicised-Modernised Harvard/IEEE Sentences (IEEE, 1969)

List 1

1. A large size in shoes is hard to sell.
2. The boy's canoe slid on the smooth planks.
3. Glue the paper to the dark blue background.
4. It's quite easy to tell the depth of a well.
5. These days a chicken leg is not a rare dish.
6. Rice is often served in round bowls.
7. The juice of lemons makes fine punch.
8. The box was thrown beside the parked lorry.
9. The pigs were fed chopped corn and cabbage.
10. We faced four hours of steady work.

List 2

1. The girl at the booth sold fifty tickets.
2. The boy was there when the sun rose.
3. A rod is used to catch pink salmon.
4. The source of the huge river is a clear spring.
5. Kick the ball straight and follow through.
6. Help the woman get back on her feet.
7. A pot of tea helps to pass the evening.
8. Smoky fires lack flame and heat.
9. A soft cushion broke the man's fall.
10. The salt breeze came across the sea.

List 3

1. It is a pleasure to read verse out loud.
2. The small pup chewed a hole in the sock.
3. The fish twisted and turned on the bent hook.
4. Press the trousers and sew on the shirt button.
5. The swan dive was far short from perfect.
6. The beauty of the view stunned the young boy.
7. Two blue fish swam in a tank.
8. Her purse was full of useless rubbish.
9. The horse reared and threw the tall rider.
10. It snowed, rained and hailed all in the same morning.

List 4

1. What joy there is in living.
2. Hoist the load onto your left shoulder.
3. Take the winding path to reach the lack.
4. Note closely the size of the fuel tank.
5. Wipe the grease of his dirty face.

6. Mend the coat before you go out.
7. The wrist was badly sprained and hung limp.
8. The stray cat gave birth to kittens.
9. The young girl gave no clear response.
10. The meal was cooked before the bell rang.

List 5

1. Mesh wire keeps the chicks inside.
2. In the early days a king ruled the state.
3. The ship was torn apart by the sharp reef.
4. Sickness kept him home in the third week.
5. The wide road shimmered in the hot sun.
6. The lazy cow lay in the cool grass.
7. Lift the square stone over the fence.
8. A rope will bind seven books together.
9. Hop over the fence and plunge in.
10. The friendly group left the local store.

List 6

1. Both lost their lives in the raging storm.
2. The frosty air passed through her coat.
3. The crooked maze failed to fool the mouse.
4. Counting quickly leads to wrong answers.
5. The show was a flop from the very beginning.
6. A saw is a tool used for cutting wood.
7. The wagon moved on well-oiled wheels.
8. March the soldiers past the next hill.
9. A lot of sugar makes tea sweet.
10. Place the rosebush near the porch steps

List 7

1. Those last words were a strong statement.
2. We talked about the side show at the circus.
3. Use a pencil to write the first draft.
4. He ran halfway to the hardware store.
5. The bell rang to end third period.
6. A small stream cut across the field.
7. Cars and busses were stuck in the snow drifts.
8. A set of china hit the floor with a crash.
9. This is a great time for walks on the moors.
10. The dune rose up from the water’s edge.

List 8

1. The fruit peel was cut into thick slices.
2. The yacht slid around the point into the bay.
3. The two met while playing in the sand.

4. The ink stain dried on the finished page.
5. The walled town was seized without a fight.
6. The lease runs out in sixteen weeks.
7. A tame squirrel makes a nice pet.
8. The horn of the car woke the sleeping man.
9. The heat strongly with firm strokes.
10. The pearl was worn in a thin silver ring.

List 9

1. He lay still and hardly moved a muscle.
2. The navy attacked the big task force
3. See the cat glaring at the scared mouse.
4. There are more than two factors here.
5. The hat brim was wide and too floppy.
6. The lawyer tried to lose his case.
7. The grass curled around the fence post.
8. Cut the pie into large pieces.
9. Many try but seldom get rich.
10. Always close the barn door tightly.

List 10

1. The bill was paid every third week.
2. The slush lay deep along the street.
3. A wisp of cloud hung in the blue air.
4. A pound of sugar costs more than eggs.
5. The fin was sharp and cut the clear water.
6. The play seems dull and quite stupid.
7. Bail the boat to stop it from sinking.
8. The term ended in late June that year.
9. A tusk is used to make costly gifts.
10. Ten cards were laid in order.

List 11

1. Move the pot over the hot fire.
2. An oak tree is strong and gives shade.
3. Cats and dogs hate each other
4. The pipe began to rust from new.
5. Open the crate but don’t break the glass.
6. Add the sum to the product of these three.
7. Thieves who rob friends deserve jail.
8. A ripe taste in cheese improves with age.
9. Act on these orders with great speed.
10. The dog crawled under the high fence.

List 12

1. We find joy in the simplest things.

2. The bark of a pine tree is shiny and dark.
3. Leaves turn brown and yellow in the autumn.
4. The flag waves when the wind blew.
5. Split the log with a quick sharp blow.
6. Burn the peat after the logs are finished
7. He ordered apple with ice cream.
8. Clean the carpet on the right hand side only.
9. Hemp is a weed found in parts of the tropics.
10. A bad back kept his score low.

List 13

1. The tiny girl took off her hat.
2. Type out three lists of orders.
3. The harder he tried the less he got done.
4. The boss ran the firm with a watchful eye.
5. The cup cracked and spilled its contents.
6. Effort can cleanse the most dirty dishes.
7. The slang word for all alcohol is booze.
8. It caught its hind paw in a rusty trap.
9. The wharf could be seen from the opposite shore.
10. Feel the heat from the weak dying flame.

List 14

1. Port is a strong with a smoky taste.
2. A cramp is no small danger on a swim.
3. He said the same phrase thirty times.
4. Pick the bright rose without leaves.
5. Two plus seven is less than ten.
6. The glow deepened in the eyes of the sweet girl.
7. Bring your problems to the wise woman.
8. Write a nice note to a friend you cherish.
9. Clothes and lodgings are nothing to a free man.
10. We frown when events take a bad turn.

List 15

1. The cigar burned a hole in the desk top.
2. The young child jumped the rusty gate.
3. Guess the results from the first scores.
4. A sweet pickle tastes good with ham.
5. The just case got the right verdict.
6. These thistles bend in high wind.
7. Pure bred poodles have curls.
8. The tree top waved in a graceful way.
9. The blot on the book was made with green ink.
10. Mud splattered on his white shirt.

List 16

1. The sofa cushion was red and light weight.
2. The empty flask stood on the tin tray.
3. The speedy man beat his track record.
4. He broke a new racket that day.
5. The coffee table was too high for the sofa.
6. The urge to write short stories is not rare.
7. The pencils have all been used up.
8. The pirates seized the crew of the lost ship.
9. He tried to replace the coin but failed.
10. She sewed the torn coat quite neatly.

List 17

1. A shower of dirt fell from the hot pipes.
2. The jacket hung on the back of the wide chair.
3. At that high level the air is pure.
4. Drop the two when you add the figures.
5. It is hard to buy a filing case now.
6. An abrupt start will not with the prize.
7. Wood is best for making toys and blocks.
8. The office pain was a dull sad tan.
9. He knew the skill of the great young actress.
10. A rag will soak up spilled water.

List 18

1. Add the store’s account to the last penny.
2. The steam hissed from the broken valve.
3. The dog almost hurt the small child.
4. There is the sound of dry leaves outside.
5. The sky that morning was clear and bright blue.
6. Scraps of torn paper littered the floor.
7. Sunday is the best part of the week.
8. The doctor cured him with these pills.
9. The new girl was fired today at noon.
10. She felt happy when the ship arrived in the port.

List 19

1. She has a way of wearing smart clothes.
2. Acids burns holes in wool cloth.
3. Fairy tales should be fun to write.
4. Eight miles of woodland burned to waste.
5. The third act was dull and tired the players.
6. A young child should not suffer fright.
7. Add the column and put the sum here.
8. We admire and love a good cook.
9. Over there the flood mark is ten inches.
10. He carved a head from the round block of marble.

List 20

1. The farmers came in to thresh the oat crop.
2. The fruit of the fig tree is apple shaped.
3. Corn cobs can be used to kindle fire.
4. Where were they when the noise started.
5. The paper box was full of thumb tacks.
6. Sell your gift to a buyer for good gain.
7. The tongs lay beside the ice pail.
8. The petals will fall with the next puff of wind.
9. Bring your best compass to the third class.
10. They could laugh although they were sad.

List 21 (practice trials)

1. Even the worst will beat his low score.
2. The attic of the brown house was on fire.
3. A rod is used to catch trout and eels.
4. Float the soap on top of the bath water.
5. A blue crane is a tall wading bird.

5.2 APPENDIX B: BKB-R Masker sentences (Bench et al., 1979)

List 1
The **CLOWN** had a **FUNNY FACE.**
The **CAR ENGINE'S RUNNING.**
SHE CUT with her **KNIFE.**
CHILDREN LIKE STRAWBERRIES.
The HOUSE had NINE ROOMS.
THEY'RE BUYING some **BREAD.**
The **GREEN TOMATOES** are **SMALL.**
HE PLAYED with his **TRAIN.**
The **POSTMAN SHUT** the **GATE.**
THEY'RE LOOKING AT the **CLOCK.**
The **BAG BUMPS** on the **GROUND.**
The **BOY DID** a **HANDSTAND.**
A **CAT SITS ON** the **BED.**
The **TRUCK CARRIED FRUIT.**
The **RAIN CAME DOWN.**
The **ICE CREAM** WAS **PINK.**

List 2
The **LADDER'S NEAR** the **DOOR.**
THEY had a **LOVELY DAY.**
The **BALL WENT** into the **NET.**
The **OLD GLOVES** are **DIRTY.**
HE CUT his **FINGER.**
The **THIN DOG** was **HUNGRY.**
The **BOY KNEW** the **GAME.**
The **GRASS GROWS** in **SUMMER.**
SHE'S TAKING her **COAT.**
The **POLICE CHASED** the **CAR.**
A **MOUSE RAN DOWN** the **HOLE.**
The **LADY'S MAKING** a **TOY.**
Some **STICKS WERE UNDER** the **TREE.**
The **LITTLE BABY SLEEPS.**
THEY'RE WATCHING the **TRAIN.**
SCHOOL FINISHED EARLY.

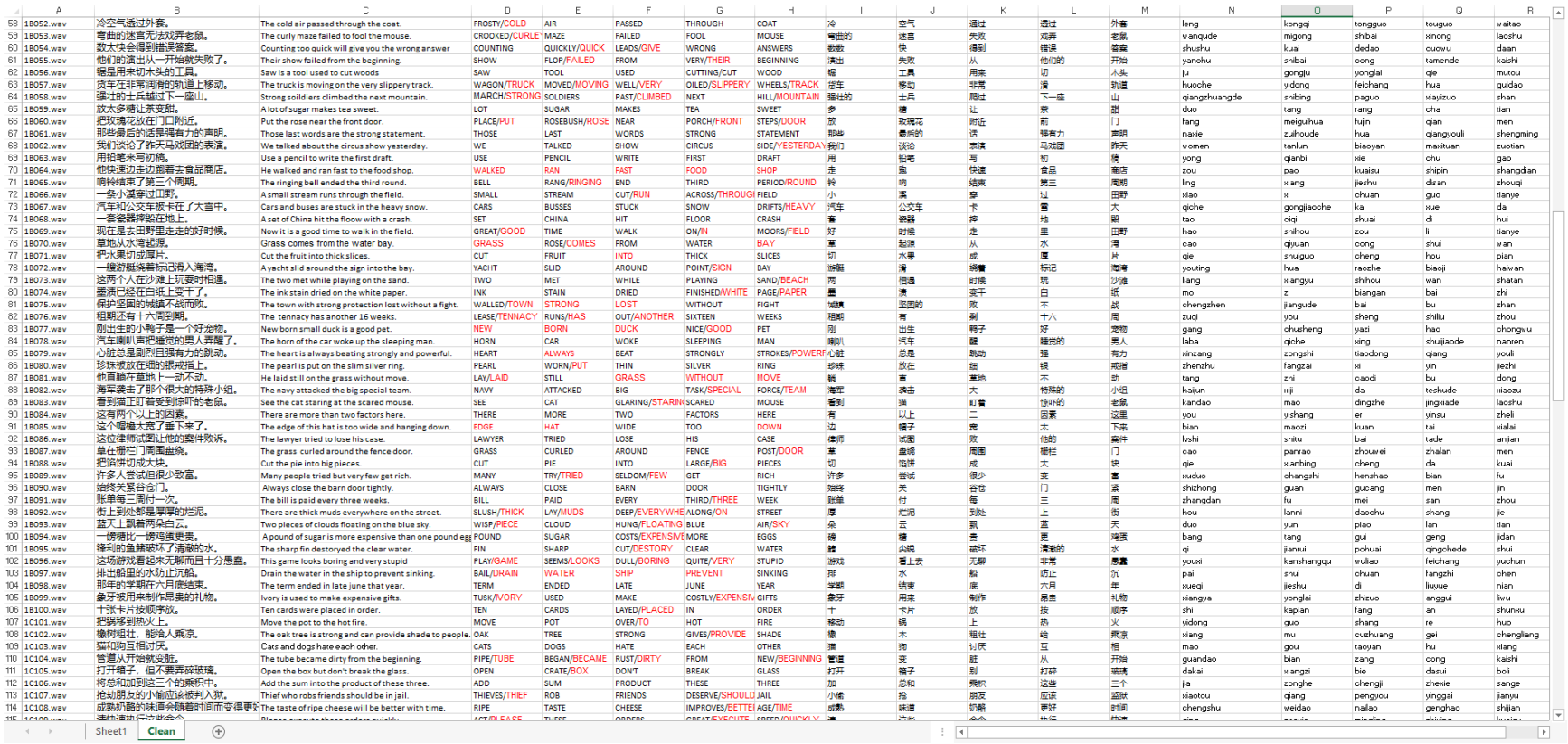
List 3
The **GLASS BOWL BROKE.**
The **DOG PLAYED** with a **STICK.**
The **KETTLE'S QUITE HOT.**
The **FARMER FEEDS** a **LAMB.**
THEY SAY some **SILLY THINGS.**
The **LADY WORE** a **COAT.**
The **CHILDREN** are **WALKING HOME.**
HE NEEDED his **HOLIDAY.**

The **MILK CAME** in a **BOTTLE.**
The **MAN CLEANED** his **SHOES.**
THEY ATE the **LEMON JELLY.**
The **BOY'S RUNNING AWAY.**
FATHER LOOKED at the **BOOK.**
SHE DRINKS from her **CUP.**
The **ROOM'S GETTING COLD.**
A **GIRL KICKED** the **TABLE.**

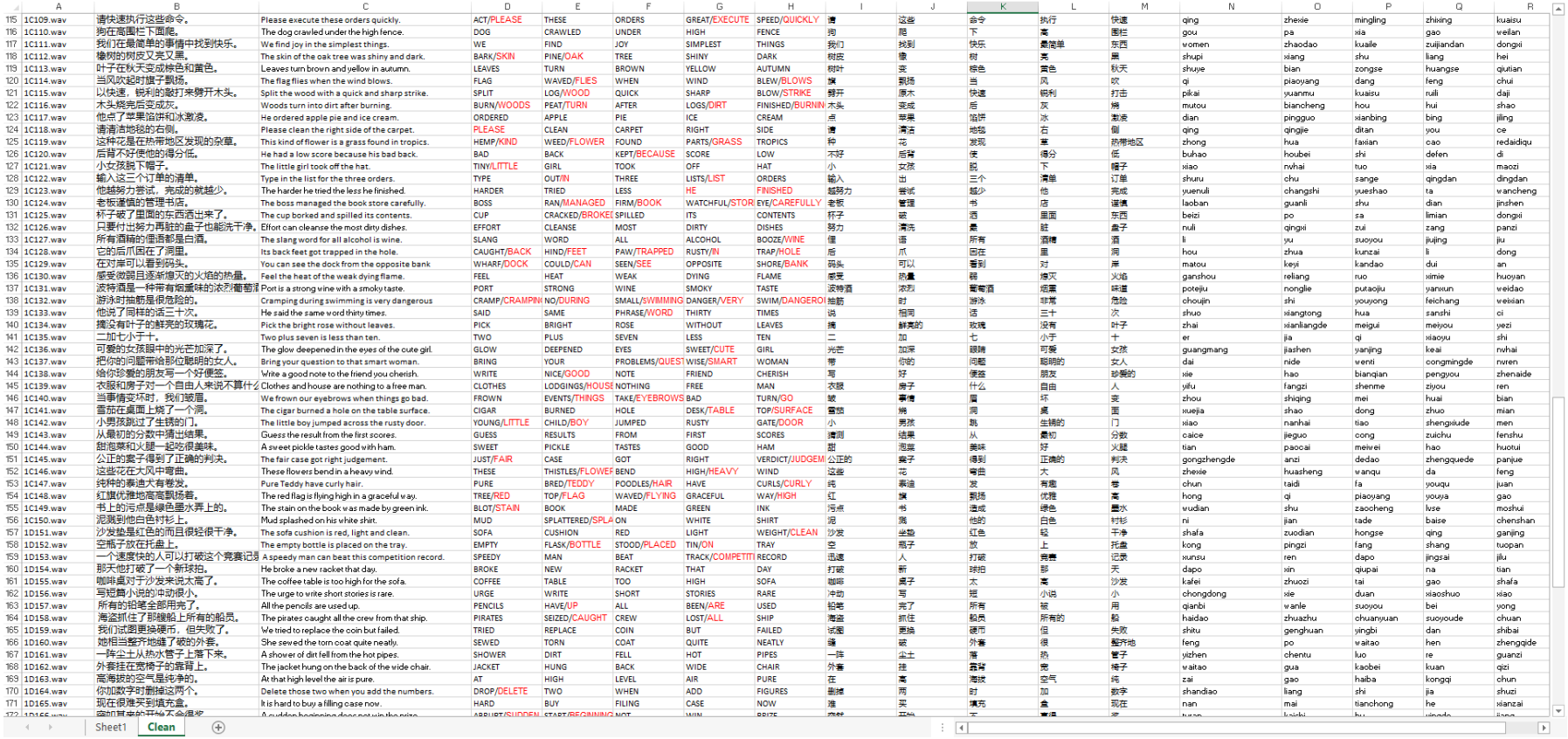
List 4

The **WIFE HELPED** her **HUSBAND.**
The **MUSIC** was **VERY LOUD.**
The **OLD MAN WORRIES.**
A **BOY RAN DOWN** the **PATH.**
The **HOUSE** had a **NICE GARDEN.**
SHE SPOKE TO her **SON.**
THEY'RE CROSSING the **STREET.**
LEMONS GROW on **TREES.**
HE FOUND his **BROTHER.**
Some **ANIMALS SLEEP** on **STRAW.**
The **JAM JAR** was **FULL.**
THEY'RE KNEELING DOWN.
The **GIRL LOST** her **DOLL.**
The **COOK'S MAKING** a **CAKE.**
The **CHILD DROPS** the **TOY.**
The **MUD STUCK** on his **SHOE.**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	FILENAME	Mandarin Sentence	English Sentence	KEYWORD1	KEYWORD2	KEYWORD3	KEYWORD4	KEYWORD5	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEYWORD1_Pin	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEYWO	Mandarin_KEY
2	Practice1.wav	最差的人也能超过他的低分。	Even the worst will beat his low score.															
3	Practice2.wav	棕色房子的阁楼着火了。	The attic of the brown house was on fire.															
4	Practice3.wav	用鱼竿钓鳟鱼和鳗鱼。	A rod is used to catch trouts and eels.															
5	Practice4.wav	让肥皂漂在洗澡水上。	Float the soap on top of the bath water.															
6	Practice5.wav	蓝鹤是一只高涉水鸟。	A blue crane is a tall wading bird.															
7	1A001.wav	大尺寸的鞋子很难卖。	A large size in shoes is difficult to sell.	LARGE/BIG	SIZE	SHOES	HARD/DIFFICULT	SELL	大	尺寸	鞋	难	卖	da	chicun	xie	nan	mai
8	1A002.wav	男孩的独木舟在光滑的木板上滑动。	The boy's canoe slid on the smooth wood.	BOY'S	CANOE	SLID	SMOOTH	PLANKS /WOOD	男孩的	独木舟	滑	光滑的	木板	nanhaide	dumuzhou	hua	guanghuade	muban
9	1A003.wav	把纸用胶水贴在深蓝色背景上。	Glue the paper to the dark blue background.	GLUE	PAPER	DARK	BLUE	BACKGROUND	胶水	纸	深	蓝色	背景	jiaoshui	zhi	shen	lanse	beijing
10	1A004.wav	井的深度容易被测量出来。	It's easy to measure the depth of a well.	QUITE	EASY	TELL/MEASURE	DEPTH	WELL	相当	容易	测量	深度	井	xiangdang	rongyi	celiang	shendu	jing
11	1A005.wav	现在鸡腿不是一盘罕见的菜。	Now chicken legs is not a rare dish.	DAYS/NOW	CHICKEN	LEG	RARE	DISH	现在	鸡	腿	罕见	菜	xianzai	ji	tui	hanjian	cai
12	1A006.wav	米饭通常装在圆碗里。	Rice is often put in round bowls.	RICE	OFTEN	SERVED/PUT	ROUND	BOWLS	米饭	通常	装	圆	碗	mifan	tongchang	zhuang	yuan	wan
13	1A007.wav	新鲜的柠檬汁打得很好。	The fresh lemon juice is made well.	FRESH	JUICE	LEMONS/LEMON	MAKES/MADE	PUNCH/WELL	新鲜的	果汁	柠檬	打	好	xinxiande	guozhi	ningmeng	da	hao
14	1A008.wav	箱子被扔在停着的卡车旁边。	The box was thrown beside the parked truck.	BOX	THROWN	BESIDE	PARKED	LORRY/TRUCK	箱子	扔	旁边	停着的	卡车	xiangzi	reng	pangbian	tingzhede	kache
15	1A009.wav	给猪喂切碎的玉米和卷心菜。	The pigs were fed chopped corn and cabbage.	PIGS	FED	CHOPPED	CORN	CABBAGE	猪	喂	切碎的	玉米	卷心菜	zhu	wei	qiesuide	yumi	juanxincai
16	1A010.wav	我们仍然剩下四个小时的工作要做。	We still have four hours' work left.	STILL	FOUR	HOURS	WORK	LEFT	仍然	四	小时	工作	剩下	yiran	si	xiaoshi	gongzuo	shengxia
17	1A011.wav	商店里的女孩卖了五十张门票。	The girl at the shop sold fifty tickets.	GIRL	BOOTH/SHOP	SOLD	FIFTY	TICKETS	女孩	商店	卖了	五十	门票	nvhai	shangdian	maile	wushi	menpiao
18	1A012.wav	当太阳升起时男孩在那。	The boy was there when the sun rose.	BOY	THERE	WHEN	SUN	ROSE	男孩	那里	当	太阳	升起	nanhai	nali	dang	taiyang	shengqi
19	1A013.wav	用一根棍子来钓粉色的三文鱼。	A stick is used to catch pink salmon.	ROD/STICK	USED	CATCH	PINK	SALMON	棍子	用来	钓	粉色	三文鱼	gunzi	yonglai	diao	fense	sanwenyu
20	1A014.wav	大河的水源是清澈的泉水。	The source of the huge river is the clear spring.	SOURCE	HUGE	RIVER	CLEAR	SPRING	水源	大	河	清澈的	泉水	shuiyuan	da	he	qingchede	quanshui
21	1A015.wav	踢直线球然后拿踢着球。	Kick the ball with straight line and stick with it.	KICK	BALL	STRAIGHT	FOLLOW/LINE	THROUGH/STICK	踢	球	直	线	踢着	ti	qiu	zhi	xian	genzhe
22	1A016.wav	帮助那个女人重新站起来。	Help that woman to stand up again.	HELP	WOMAN	GET/STAND	BACK/UP	FEET/AGAIN	帮助	女人	站	起来	重新	bangzhu	nvren	zhan	qilai	chongxin
23	1A017.wav	一壶茶有助于度过这个夜晚。	A pot of tea helps to spend the night.	POT	TEA	HELPS	PASS/SPEND	EVENING/NIGHT	壶	茶	帮助	度过	夜晚	hu	cha	bangzhu	duguo	yewan
24	1A018.wav	冒烟的火缺少火焰和高温。	Smoky fires lack flame and heat.	SMOKY	FIRES	LACK	FLAME	HEAT	烟	火	缺少	火焰	高温	yan	huo	queshao	huoyan	gaowen
25	1A019.wav	软垫把男人拌倒了。	The soft cushion made the man fall down.	SOFT	CUSHION	BROKE/MADE	MANS/MAN	FALL	柔软的	垫子	让	男人	摔倒	rouruande	dianzi	rang	nanren	shuaidao
26	1A020.wav	咸湿的海风从海上而来。	The salty and wet wind is blowing from the sea.	SALT/SALTY	BREEZE/WIND	CAME/BLOW	ACROSS/FROM	SEA	咸	微风	吹	从	大海	xian	weifeng	chui	cong	dahai
27	1A021.wav	很高兴大声朗读诗歌。	Read the poetry out loudly and happily.	PLEASURE/HAPPIL	READ	VERSE/POETRY	OUT	LOUD/LOUDLY	高兴	读	诗	出来	大声	gaoxing	du	shi	chulai	dasheng
28	1A022.wav	小狗在袜子上咬了个洞。	The little dog bited a hole on the sock.	SMALL/LITTLE	PUP/DOG	CHEWED/BITE	HOLE	SOCK	小	狗	咬	洞	袜子	xiao	gou	yao	dong	wazi
29	1A023.wav	鱼在弯曲的钩子上扭曲转动。	The fish is twisting and turning on the curled hook.	FISH	TWISTED/TWISTIN	TURNED/TURNING	BENT/CURLED	HOOK	鱼	扭曲	转动	弯曲	钩	yu	niuqu	zhuandong	wanqu	diaogou
30	1A024.wav	把裤子熨平并缝上一个衬衫的纽扣。	Iron the trousers and sew a button for the shirt.	PRESS/IRON	TROUSERS	SEW	SHIRT	BUTTON	熨平	裤子	缝	衬衫	纽扣	yunping	kuzi	feng	chenshan	niukou
31	1A025.wav	天鹅潜水远远不够完美。	Swan diving is far from perfect.	SWAN	DIVE/DIVING	FAR	SHORT/FROM	PERFECT	天鹅	潜水	远	不够	完美	tiane	qianshui	yuan	bugou	wanmei
32	1A026.wav	美景让小男孩感到震撼。	The little boy is shocked by the beautiful view.	BEAUTY/BEAUTIFU	VIEW	STUNNED/SHOCK	YOUNG/LITTLE	BOY	美丽	景色	震惊	小	男孩	meili	jingse	zhenjing	xiao	nanhai
33	1A027.wav	两条蓝色的鱼在水箱里游泳。	Two blue fishes are swimming in the tank.	TWO	BLUE	FISH	SWAM/SWIMMING	TANK	两	蓝色	鱼	游泳	水箱	liang	lanse	yu	youyong	shuixiang
34	1A028.wav	她的钱包里塞满了没用的垃圾。	Her purse is full of useless rubbish.	HER	PURSE	FULL	USELESS	RUBBISH	她的	钱包	满	没用的	垃圾	tade	qianbao	man	meiyongde	laji
35	1A029.wav	马咆哮着把高个子骑手甩了出去。	The horse roared and threw the tall rider.	HORSE	ROARED	THREW	TALL	RIDER	马	咆哮	扔	高	骑手	ma	paoxiao	reng	gao	qishou
36	1A030.wav	同一天早上又下雪，又下雨，又下冰雹	It snowed, rained and hailed in the same morning.	SNOWED	RAINED	HAILED	SAME	MORNING	雪	雨	冰雹	同一天	早上	xue	yu	bingbao	tongyitian	zaoshang
37	1A031.wav	生活中有什么快乐的事情。	What is happy things about life.	WHAT	JOY/HAPPY	THERE/THINGS	IS/ABOUT	LIVING/LIFE	什么	快乐	事情	关于	生活	shenme	kuaile	shiqing	guanyu	shenghuo
38	1A032.wav	把重量放到你左肩上。	Put the weight on your left shoulder.	HOIST/WEIGHT	LOAD/PUT	YOUR	LEFT	SHOULDER	重量	放到	你的	左	肩膀	zhongliang	fangdao	nide	zuo	jianbang
39	1A033.wav	沿着小路到达湖泊。	Arrive at the lake along the winding path.	TAKE/ALONG	WINDING	PATH	REACH/ARRIVE	LAKE	沿着	蜿蜒的	路	到达	湖	yanzhe	wanyande	lu	daoda	hu
40	1A034.wav	仔细注意汽油罐的尺寸。	Pay close attention to the size of the gas tank.	NOTE/ATTENTION	CLOSELY/CLOSE	SIZE	FUEL/GAS	TANK	注意	仔细	尺寸	汽油	罐	zhuyi	zixi	chicun	qiyou	guan
41	1A035.wav	擦去他脏脸上的油。	Wipe off the oil from his dirty face.	WIPE	GREASE/OIL	OFF	DIRTY	FACE	擦	油	掉	脏	脸	ca	you	diao	zang	lian
42	1A036.wav	外出之前穿上外套。	Put on the jacket before going out.	MEND/PUT	COAT/JACKET	BEFORE	GO	OUT	穿	外套	之前	外	出	chuan	waitao	zhiqian	wai	chu
43	1A037.wav	他的手腕严重扭伤，虚弱地垂下来。	His wrist was badly strained and falling down weakly.	WRIST	BADLY	SPRAINED	HUNG	limp/WEAKLY	手腕	严重	扭伤	垂着	虚弱地	shouwan	yanzhong	niushang	chuizhe	xuruode
44	1A038.wav	流浪猫今天生了小猫。	The street cat gave birth to a small cat today.	STRAY/STREET	CAT	GAVE/TODAY	BIRTH	KITTENS/SMALL	流浪	猫	今天	生	小	liulang	mao	jintian	sheng	xiao
45	1A039.wav	小女孩没有给出明确答复。	The little girl did not give a clear answer.	YOUNG/LITTLE	GIRL	GAVE	CLEAR	RESPONSE/ANSWE	小	女孩	给出	明确	答复	xiao	nvhai	geichu	mingque	dafu
46	1A040.wav	铃响之前饭就做好了。	The meal was cooked before the bell rang.	MEAL	COOKED	BEFORE	BELL	RANG	饭	做	之前	铃	响	fan	zuo	zhiqian	ling	xiang
47	1A041.wav	金属线把小鸡困在了盒子里面。	The metal wire keeps the chicken inside the box.	MESH/METAL	WIRE	KEEPS	CHICKS/CHICKEN	INSIDE	金属	电线	困在了	鸡	里面	jinshu	dianxian	kunzaile	ji	limian
48	1A042.wav	早期国王统治着国家。	A king ruled the country in the early days.	EARLY	DAYS	KING	RULED	STATE/COUNTRY	早	时期	国王	统治	国家	zao	shiqi	guowang	tongzhi	guojia
49	1A043.wav	船在锋利的礁石上撞裂了。	The ship was torn apart on the sharp rock.	SHIP	TORN	APART	SHARP	REEF/ROCK	船	撞	裂	锋利	礁石	chuan	zhuang	lie	fengli	jiaoshi
50	1A044.wav	这是他因病在家待的第三周。	It's the third week that he stayed at home due to illness	SICKNESS/ILLNESS	KEPT/STAY	HOME	THIRD	WEEK	病	待	家	第三	周	bing	dai	jia	disan	zhou
51	1A045.wav	宽阔的道路在烈日下闪闪发光。	The wide road shimmered in the hot sun.	WIDE	ROAD	SHIMMERED	HOT	SUN	宽	路	闪光	烈	日	kuan	lu	shanguang	lie	ri
52	1A046.wav	懒牛躺在凉快的草地上。	The lazy cow lay in the cool grass.	LAZY	COW	LAY	COOL	GRASS	懒	牛	躺	凉快的	草	lan	niu	tang	liangkuaide	cao
53	1A047.wav	把正方形石头搬到栅栏右边。	Move the square stone to the right side of the fence.	LIFT/MOVE	SQUARE	STONE	OVER/RIGHT	FENCE	搬	正方形	石头	右边	栅栏	ban	zhengfangxing	shitou	youbian	zhalan
54	1A048.wav	绳子可以把七本书一起捆起来。	The rope can bind seven books together.	ROPE	BIND	SEVEN	BOOKS	TOGETHER	绳	捆	七	书	一起	sheng	kun	qi	shu	yiqi
55	1A049.wav	跳过栅栏走进去。	Jump over the fence and walk inside.	HOP/JUMP	OVER	FENCE	PLUNGE/WALK	IN/INSIDE	跳	过去	栅栏	走	进去	tiao	guoqu	zhalan	zou	jinqu
56	1A050.wav	一群友好的人离开了小商店。	A group of friendly people left the small shop.	FRIENDLY	GROUP	LEFT	LOCAL/SMALL	STORE/SHOP	友好的	一群	离开	小	商店	youhaode	yiqun	likai	xiao	shangdian
57	1B051.wav	双方在凶猛的风暴中丧生。	Both lost their lives in the terrible storm.	BOTH	LOST	LIVES	RAGING/TERRIBLE	STORM	双方	丧	生	凶猛的	风暴	shuangfang	sang	sheng	xiongmengde	fengbao
58	1B052.wav	冷空气通过外套	The cold air passed through the coat	FROSTY/COLD	AIR	PASSED	THROUGH	COAT	冷	空气	通过	外套		leng	kongqi	tongguo	waitao	



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
58	1B052.wav	冷空气透过外套。	The cold air passed through the coat.	FROSTY/COLD	AIR	PASSED	THROUGH	COAT	冷	空气	通过	通过	外套	leng	kongqi	tongguo	touguo	waitao
59	1B053.wav	弯曲的迷宫无法戏弄老鼠。	The curly maze failed to fool the mouse.	CROOKED/CURLE	MAZE	FAILED	FOOL	MOUSE	弯曲的	迷宫	失败	戏弄	老鼠	wanqude	migong	shibai	xinong	laoshu
60	1B054.wav	数太快会得到错误答案。	Counting too quick will give you the wrong answer	COUNTING	QUICKLY/QUICK	LEADS/GIVE	WRONG	ANSWERS	数数	快	得到	错误	答案	shushu	kuai	dedao	cuowu	daan
61	1B055.wav	他们的演出从一开始就失败了。	Their show failed from the beginning.	SHOW	FLOP/FAILED	FROM	VERY/THEIR	BEGINNING	演出	失败	从	他们的	开始	yanchu	shibai	cong	tamende	kaishi
62	1B056.wav	锯是用来切木头的工具。	Saw is a tool used to cut woods	SAW	TOOL	USED	CUTTING/CUT	WOOD	锯	工具	用来	切	木头	ju	gongju	yonglai	qie	mutou
63	1B057.wav	货车在非常润滑的轨道上移动。	The truck is moving on the very slippery track.	WAGON/TRUCK	MOVED/MOVING	WELL/VERY	OILED/SLIPPERY	WHEELS/TRACK	货车	移动	非常	滑	轨道	huoche	yidong	feichang	hua	guidao
64	1B058.wav	强壮的士兵越过下一座山。	Strong soldiers climbed the next mountain.	MARCH/STRONG	SOLDIERS	PAST/CLIMBED	NEXT	HILL/MOUNTAIN	强壮的	士兵	爬过	下一座	山	qiangzhuangde	shibing	paguo	xiayizuo	shan
65	1B059.wav	放太多糖让茶变甜。	A lot of sugar makes tea sweet.	LOT	SUGAR	MAKES	TEA	SWEET	多	糖	让	茶	甜	duo	tang	rang	cha	tian
66	1B060.wav	把玫瑰花放在门口附近。	Put the rose near the front door.	PLACE/PUT	ROSEBUSH/ROSE	NEAR	PORCH/FRONT	STEPS/DOOR	放	玫瑰花	附近	前	门	fang	meiguihua	fujin	qian	men
67	1B061.wav	那些最后的话是强有力的声明。	Those last words are the strong statement.	THOSE	LAST	WORDS	STRONG	STATEMENT	那些	最后的	话	强有力	声明	naxie	zuihoude	hua	qiangyouli	shengming
68	1B062.wav	我们谈论了昨天马戏团的表演。	We talked about the circus show yesterday.	WE	TALKED	SHOW	CIRCUS	SIDE/YESTERDAY	我们	谈论	表演	马戏团	昨天	women	tanlun	biaoyan	maxituan	zuotian
69	1B063.wav	用铅笔来写初稿。	Use a pencil to write the first draft.	USE	PENCIL	WRITE	FIRST	DRAFT	用	铅笔	写	初	稿	yong	qianbi	xie	chu	gao
70	1B064.wav	他快速边走边跑着去食品商店。	He walked and ran fast to the food shop.	WALKED	RAN	FAST	FOOD	SHOP	走	跑	快速	食品	商店	zou	pao	kuaisu	shipin	shangdian
71	1B065.wav	响铃结束了第三个周期。	The ringing bell ended the third round.	BELL	RANG/RINGING	END	THIRD	PERIOD/ROUND	铃	响	结束	第三	周期	ling	xiang	jieshu	disan	zhouqi
72	1B066.wav	一条小溪穿过田野。	A small stream runs through the field.	SMALL	STREAM	CUT/RUN	ACROSS/THROUGH	FIELD	小	溪	穿	过	田野	xiao	xi	chuan	guo	tianye
73	1B067.wav	汽车和公交车被卡在了大雪中。	Cars and buses are stuck in the heavy snow.	CARS	BUSSES	STUCK	SNOW	DRIFTS/HEAVY	汽车	公交车	卡	雪	大	qiche	gongjiaoche	ka	xue	da
74	1B068.wav	一套瓷器摔裂在地上。	A set of China hit the floow with a crash.	SET	CHINA	HIT	FLOOR	CRASH	套	瓷器	摔	地	裂	tao	ciqi	shuai	di	hui
75	1B069.wav	现在是去田野里走走的好时候。	Now it is a good time to walk in the field.	GREAT/GOOD	TIME	WALK	ON/IN	MOORS/FIELD	好	时候	走	里	田野	hao	shihou	zou	li	tianye
76	1B070.wav	草地从水湾起源。	Grass comes from the water bay.	GRASS	ROSE/COMES	FROM	WATER	BAY	草	起源	从	水	湾	cao	qiyuan	cong	shui	wan
77	1B071.wav	把水果切成厚片。	Cut the fruit into thick slices.	CUT	FRUIT	INTO	THICK	SLICES	切	水果	成	厚	片	qie	shuiguo	cheng	hou	pian
78	1B072.wav	一艘游艇绕着标记滑入海湾。	A yacht slid around the sign into the bay.	YACHT	SLID	AROUND	POINT/SIGN	BAY	游艇	滑	绕着	标记	海湾	youting	hua	raozhe	biaoji	haiwan
79	1B073.wav	这两个人在沙滩上玩耍时相遇。	The two met while playing on the sand.	TWO	MET	WHILE	PLAYING	SAND/BEACH	两	相遇	时候	玩	沙滩	liang	xiangyu	shihou	wan	shatan
80	1B074.wav	墨渍已经在白纸上变干了。	The ink stain dried on the white paper.	INK	STAIN	DRIED	FINISHED/WHITE	PAGE/PAPER	墨	渍	变干	白	纸	mo	zi	biangan	bai	zhi
81	1B075.wav	保护坚固的城镇不战而败。	The town with strong protection lost without a fight.	WALLED/TOWN	STRONG	LOST	WITHOUT	FIGHT	城镇	坚固的	败	不	战	chengzhen	jiangude	bai	bu	zhan
82	1B076.wav	租期还有十六周到期。	The tennacy has another 16 weeks.	LEASE/TENNACY	RUNS/HAS	OUT/ANOTHER	SIXTEEN	WEEKS	租期	有	剩	十六	周	zuqi	you	sheng	shiliu	zhou
83	1B077.wav	刚出生的小鸭子是一个好宠物。	New born small duck is a good pet.	NEW	BORN	DUCK	NICE/GOOD	PET	刚	出生	鸭子	好	宠物	gang	chusheng	yazi	hao	chongwu
84	1B078.wav	汽车喇叭声把睡觉的男人弄醒了。	The horn of the car woke up the sleeping man.	HORN	CAR	WOKE	SLEEPING	MAN	喇叭	汽车	醒	睡觉的	男人	laba	qiche	xing	shuijiaode	nanren
85	1B079.wav	心脏总是剧烈且强有力的跳动。	The heart is always beating strongly and powerful.	HEART	ALWAYS	BEAT	STRONGLY	STROKES/POWERF	心脏	总是	跳动	强	有力	xinzang	zongshi	tiaodong	qiang	youli
86	1B080.wav	珍珠被放在细的银戒指上。	The pearl is put on the slim silver ring.	PEARL	WORN/PUT	THIN	SILVER	RING	珍珠	放在	细	银	戒指	zhenzhu	fangzai	xi	yin	jiezhi
87	1B081.wav	他直躺在草地上一动不动。	He laid still on the grass without move.	LAY/LAID	STILL	GRASS	WITHOUT	MOVE	躺	直	草地	不	动	tang	zhi	caodi	bu	dong
88	1B082.wav	海军袭击了那个很大的特殊小组。	The navy attacked the big special team.	NAVY	ATTACKED	BIG	TASK/SPECIAL	FORCE/TEAM	海军	袭击	大	特殊的	小组	haijun	xiji	da	teshude	xiaozu
89	1B083.wav	看到猫正盯着受到惊吓的老鼠。	See the cat staring at the scared mouse.	SEE	CAT	GLARING/STARIN	SCARED	MOUSE	看到	猫	盯着	惊吓的	老鼠	kandao	mao	dingzhe	jingxiade	laoshu
90	1B084.wav	这有两个以上的因素。	There are more than two factors here.	THERE	MORE	TWO	FACTORS	HERE	有	以上	二	因素	这里	you	yishang	er	yinsu	zheli
91	1B085.wav	这个帽檐太宽了垂下来了。	The edge of this hat is too wide and hanging down.	EDGE	HAT	WIDE	TOO	DOWN	边	帽子	宽	太	下来	bian	maozi	kuan	tai	xialai
92	1B086.wav	这位律师试图让他的案件败诉。	The lawyer tried to lose his case.	LAWYER	TRIED	LOSE	HIS	CASE	律师	试图	败	他的	案件	lvshi	shitu	bai	tade	anjian
93	1B087.wav	草在栅栏门周围蜷缩。	The grass curled around the fence door.	GRASS	CURLED	AROUND	FENCE	POST/DOOR	草	蜷缩	周围	栅栏	门	cao	panrao	zhouwei	zhalan	men
94	1B088.wav	把馅饼切成大块。	Cut the pie into big pieces.	CUT	PIE	INTO	LARGE/BIG	PIECES	切	馅饼	成	大	块	qie	xianbing	cheng	da	kuai
95	1B089.wav	许多人尝试但很少致富。	Many people tried but very few get rich.	MANY	TRY/TRIED	SELDOM/FEW	GET	RICH	许多	尝试	很少	变	富	xuduo	changshi	henshao	bian	fu
96	1B090.wav	始终关紧谷仓门。	Always close the barn door tightly.	ALWAYS	CLOSE	BARN	DOOR	TIGHTLY	始终	关	谷仓	门	紧	shizhong	guan	gucang	men	jin
97	1B091.wav	账单每三周付一次。	The bill is paid every three weeks.	BILL	PAID	EVERY	THIRD/THREE	WEEK	账单	付	每	三	周	zhangdan	fu	mei	san	zhou
98	1B092.wav	街上到处都是厚厚的烂泥。	There are thick muds everywhere on the street.	SLUSH/THICK	LAY/MUDS	DEEP/EVERYWHE	ALONG/ON	STREET	厚	烂泥	到处	上	街	hou	lanni	daochu	shang	jie
99	1B093.wav	蓝天上飘着两朵白云。	Two pieces of clouds floating on the blue sky.	WISP/PIECE	CLOUD	HUNG/FLOATING	BLUE	AIR/SKY	朵	云	飘	蓝	天	duo	yun	piao	lan	tian
100	1B094.wav	一磅糖比一磅鸡蛋更贵。	A pound of sugar is more expensive than one pound eg	POUND	SUGAR	COSTS/EXPENSIVE	MORE	EGGS	磅	糖	贵	更	鸡蛋	bang	tang	gui	geng	jidan
101	1B095.wav	锋利的鱼鳍破坏了清澈的水。	The sharp fin destoryed the clear water.	FIN	SHARP	CUT/DESTORY	CLEAR	WATER	鳍	尖锐	破坏	清澈的	水	qi	jianrui	pohuai	qingchede	shui
102	1B096.wav	这场游戏看起来无聊而且十分愚蠢。	This game looks boring and very stupid	PLAY/GAME	SEEMS/LOOKS	DULL/BORING	QUITE/VERY	STUPID	游戏	看上去	无聊	非常	愚蠢	youxi	kanshangqu	wuliao	feichang	yuchun
103	1B097.wav	排出船里的水防止沉船。	Drain the water in the ship to prevent sinking.	BAIL/DRAIN	WATER	SHIP	PREVENT	SINKING	排	水	船	防止	沉	pai	shui	chuan	fangzhi	chen
104	1B098.wav	那年的学期在六月底结束。	The term ended in late june that year.	TERM	ENDED	LATE	JUNE	YEAR	学期	结束	底	六月	年	xueqi	jieshu	di	liuyue	nian
105	1B099.wav	象牙被用来制作昂贵的礼物。	Ivory is used to make expensive gifts.	TUSK/IVORY	USED	MAKE	COSTLY/EXPENSIV	GIFTS	象牙	用来	制作	昂贵	礼物	xiangya	yonglai	zhizuo	anggui	liwu
106	1B100.wav	十张卡片按顺序放。	Ten cards were placed in order.	TEN	CARDS	LAYED/PLACED	IN	ORDER	十	卡片	放	按	顺序	shi	kapian	fang	an	shunxu
107	1C101.wav	把锅移到热火上。	Move the pot to the hot fire.	MOVE	POT	OVER/TO	HOT	FIRE	移动	锅	上	热	火	yidong	guo	shang	re	huo
108	1C102.wav	橡树粗壮，能给人乘凉。	The oak tree is strong and can provide shade to people.	OAK	TREE	STRONG	GIVES/PROVIDE	SHADE	橡	木	粗壮	给	乘凉	xiang	mu	cuzhuang	gei	chengliang
109	1C103.wav	猫和狗互相讨厌。	Cats and dogs hate each other.	CATS	DOGS	HATE	EACH	OTHER	猫	狗	讨厌	互	相	mao	gou	taoyan	hu	xiang
110	1C104.wav	管道从开始就变脏。	The tube became dirty from the beginning.	PIPE/TUBE	BEGAN/BECAME	RUST/DIRTY	FROM	NEW/BEGINNING	管道	变	脏	从	开始	guandao	bian	zang	cong	kaishi
111	1C105.wav	打开箱子，但不要弄碎玻璃。	Open the box but don't break the glass.	OPEN	CRATE/BOX	DON'T	BREAK	GLASS	打开	箱子	别	打碎	玻璃	dakai	xiangzi	bie	dasui	boli
112	1C106.wav	将总和加到这三个的乘积中。	Add the sum into the product of these three.	ADD	SUM	PRODUCT	THESE	THREE	加	总和	这些	三个	jia	zonghe	chengji	zhexie	sange	
113	1C107.wav	抢劫朋友的小偷应该被判入狱。	Thief who robs friends should be in jail.	THIEVES/THIEF	ROB	FRIENDS	DESERVE/SHOULD	JAIL	小偷	抢	朋友	应该	监狱	xiaotou	qiang	pengyou	yinggai	jianyu
114	1C108.wav	成熟奶酪的味道会随着时间而变得更好	The taste of ripe cheese will be better with time.	RIPE	TASTE	CHEESE	IMPROVES/BETTEI	AGE/TIME	成熟	味道	奶酪	更好	时间	chengshu	weidao	nailao	gengbao	shijian
115	1C109.wav	请快速执行这些命令。	Please execute these orders quickly.	ACT/PLEASE	THESE	ORDERS	GREAT/EXECUTE	SPEED/QUICKLY	请	这些	命令	执行	快速	qing	zhexie	mingling	zhixing	kuaisu



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
115	1C109.wav	请快速执行这些命令。	Please execute these orders quickly.	ACT/PLEASE	THESE	ORDERS	GREAT/EXECUTE	SPEED/QUICKLY	请	这些	命令	执行	快速	qing	zhexie	mingling	zhixing	kuaisu
116	1C110.wav	狗在高围栏下面爬。	The dog crawled under the high fence.	DOG	CRAWLED	UNDER	HIGH	FENCE	狗	爬	下	高	围栏	gou	pa	xia	gao	weilan
117	1C111.wav	我们在最简单的事情中找到快乐。	We find joy in the simplest things.	WE	FIND	JOY	SIMPLEST	THINGS	我们	找到	快乐	最简单	东西	women	zhaodao	kuaile	zuijiandan	dongxi
118	1C112.wav	橡树的树皮又亮又黑。	The skin of the oak tree was shiny and dark.	BARK/SKIN	PINE/OAK	TREE	SHINY	DARK	树皮	橡	树	亮	黑	shupi	xiang	shu	liang	hei
119	1C113.wav	叶子在秋天变成棕色和黄色。	Leaves turn brown and yellow in autumn.	LEAVES	TURN	BROWN	YELLOW	AUTUMN	树叶	变	棕色	黄色	秋天	shuye	bian	zongse	huangse	qiutian
120	1C114.wav	当风吹起时旗子飘扬。	The flag flies when the wind blows.	FLAG	WAVED/FLIES	WHEN	WIND	BLEW/BLOWS	旗	飘扬	当	风	吹	qi	piaoyang	dang	feng	chui
121	1C115.wav	以快速，锐利的敲打来劈开木头。	Split the wood with a quick and sharp strike.	SPLIT	LOG/WOOD	QUICK	SHARP	BLOW/STRIKE	劈开	原木	快速	锐利	打击	pikai	yuanmu	kuaisu	ruili	daji
122	1C116.wav	木头烧完后变成灰。	Woods turn into dirt after burning.	BURN/WOODS	PEAT/TURN	AFTER	LOGS/DIRT	FINISHED/BURNIN	木头	变成	后	灰	烧	mutou	biancheng	hou	hui	shao
123	1C117.wav	他点了苹果馅饼和冰激凌。	He ordered apple pie and ice cream.	ORDERED	APPLE	PIE	ICE	CREAM	点	苹果	馅饼	冰	激凌	dian	pingguo	xianbing	bing	jiling
124	1C118.wav	请清洁地毯的右侧。	Please clean the right side of the carpet.	PLEASE	CLEAN	CARPET	RIGHT	SIDE	请	清洁	地毯	右	侧	qing	qingjie	ditan	you	ce
125	1C119.wav	这种花是在热带地区发现的杂草。	This kind of flower is a grass found in tropics.	HEMP/KIND	WEED/FLOWER	FOUND	PARTS/GRASS	TROPICS	种	花	发现	草	热带地区	zhong	hua	faxian	cao	redaidiqu
126	1C120.wav	后背不好使他的得分低。	He had a low score because his bad back.	BAD	BACK	KEPT/BECAUSE	SCORE	LOW	不好	后背	使	得分	低	buhao	houbei	shi	defen	di
127	1C121.wav	小女孩脱下帽子。	The little girl took off the hat.	TINY/LITTLE	GIRL	TOOK	OFF	HAT	小	女孩	脱	下	帽子	xiao	nvhai	tuo	xia	maozi
128	1C122.wav	输入这三个订单的清单。	Type in the list for the three orders.	TYPE	OUT/IN	THREE	LISTS/LIST	ORDERS	输入	出	三个	清单	订单	shuru	chu	sange	qingdan	dingdan
129	1C123.wav	他越努力尝试，完成的就越少。	The harder he tried the less he finished.	HARDER	TRIED	LESS	HE	FINISHED	越努力	尝试	越少	他	完成	yuenuli	changshi	yueshao	ta	wancheng
130	1C124.wav	老板谨慎的管理书店。	The boss managed the book store carefully.	BOSS	RAN/MANAGED	FIRM/BOOK	WATCHFUL/STOR	EYE/CAREFULLY	老板	管理	书	店	谨慎	laoban	guanli	shu	dian	jinshen
131	1C125.wav	杯子破了里面的东西洒出来了。	The cup borked and spilled its contents.	CUP	CRACKED/BROKE	SPILLED	ITS	CONTENTS	杯子	破	洒	里面	东西	beizi	po	sa	limian	dongxi
132	1C126.wav	只要付出努力再脏的盘子也能洗干净。	Effort can cleanse the most dirty dishes.	EFFORT	CLEANSE	MOST	DIRTY	DISHES	努力	清洗	最	脏	盘子	nuli	qingxi	zui	zang	panzi
133	1C127.wav	所有酒精的俚语都是白酒。	The slang word for all alcohol is wine.	SLANG	WORD	ALL	ALCOHOL	BOOZE/WINE	俚	语	所有	酒精	酒	li	yu	suoyou	jiujing	jiu
134	1C128.wav	它的后爪困在了洞里。	Its back feet got trapped in the hole.	CAUGHT/BACK	HIND/FEET	PAW/TRAPPED	RUSTY/IN	TRAP/HOLE	后	爪	困在	里	洞	hou	zhua	kunzai	li	dong
135	1C129.wav	在对岸可以看到码头。	You can see the dock from the opposite bank	WHARF/DOCK	COULD/CAN	SEEN/SEE	OPPOSITE	SHORE/BANK	码头	可以	看到	对	岸	matou	keyi	kandao	dui	an
136	1C130.wav	感受微弱且逐渐熄灭的火焰的热量。	Feel the heat of the weak dying flame.	FEEL	HEAT	WEAK	DYING	FLAME	感受	热量	弱	熄灭	火焰	ganshou	reliang	ruo	ximie	huoyan
137	1C131.wav	波特酒是一种带有烟熏味的浓烈葡萄酒	Port is a strong wine with a smoky taste.	PORT	STRONG	WINE	SMOKY	TASTE	波特酒	浓烈	葡萄酒	烟熏	味道	puteiju	nonglie	putaojiu	yanxun	weidao
138	1C132.wav	游泳时抽筋是很危险的。	Cramping during swimming is very dangerous	CRAMP/CRAMPIN	NO/DURING	SMALL/SWIMMING	DANGER/VERY	SWIM/DANGERO	抽筋	时	游泳	非常	危险	choujin	shi	youyong	feichang	weixian
139	1C133.wav	他说了同样的话三十次。	He said the same word thirty times.	SAID	SAME	PHRASE/WORD	THIRTY	TIMES	说	相同	话	三十	次	shuo	xiangtong	hua	sanshi	ci
140	1C134.wav	摘没有叶子的鲜亮的玫瑰花。	Pick the bright rose without leaves.	PICK	BRIGHT	ROSE	WITHOUT	LEAVES	摘	鲜亮的	玫瑰	没有	叶子	zhai	xianliangde	meigui	meiyou	yezi
141	1C135.wav	二加七小于十。	Two plus seven is less than ten.	TWO	PLUS	SEVEN	LESS	TEN	二	加	七	小于	十	er	jia	qi	xiaoyu	shi
142	1C136.wav	可爱的女孩眼中的光芒加深了。	The glow deepened in the eyes of the cute girl.	GLOW	DEEPENED	EYES	SWEET/CUTE	GIRL	光芒	加深	眼睛	可爱	女孩	guangmang	jiashen	yanjing	keai	nvhai
143	1C137.wav	把你的问题带给那位聪明的女人。	Bring your question to that smart woman.	BRING	YOUR	PROBLEMS/QUEST	WISE/SMART	WOMAN	带	你的	问题	聪明的	女人	dai	nide	wenti	congmingde	nvren
144	1C138.wav	给你珍爱的朋友写一个好便签。	Write a good note to the friend you cherish.	WRITE	NICE/GOOD	NOTE	FRIEND	CHERISH	写	好	便签	朋友	珍爱的	xie	hao	bianqian	pengyou	zhenaide
145	1C139.wav	衣服和房子对一个自由人来说不算什么	Clothes and house are nothing to a free man.	CLOTHES	LODGINGS/HOUSE	NOTHING	FREE	MAN	衣服	房子	什么	自由	人	yifu	fangzi	shenme	ziyou	ren
146	1C140.wav	当事情变坏时，我们皱眉。	We frown our eyebrows when things go bad.	FROWN	EVENTS/THINGS	TAKE/EYEBROWS	BAD	TURN/GO	皱	事情	眉	坏	变	zhou	shiqing	mei	huai	bian
147	1C141.wav	雪茄在桌面上烧了一个洞。	The cigar burned a hole on the table surface.	CIGAR	BURNED	HOLE	DESK/TABLE	TOP/SURFACE	雪茄	烧	洞	桌	面	xueqia	shao	dong	zhuo	mian
148	1C142.wav	小男孩跳过了生锈的门。	The little boy jumped across the rusty door.	YOUNG/LITTLE	CHILD/BOY	JUMPED	RUSTY	GATE/DOOR	小	男孩	跳	生锈的	门	xiao	nanhai	tiao	shengxiude	men
149	1C143.wav	从最初的分数中猜出结果。	Guess the result from the first scores.	GUESS	RESULTS	FROM	FIRST	SCORES	猜测	结果	从	最初	分数	caice	jieguo	cong	zuichu	fenshu
150	1C144.wav	甜泡菜和火腿一起吃很美味。	A sweet pickle tastes good with ham.	SWEET	PICKLE	TASTES	GOOD	HAM	甜	泡菜	美味	好	火腿	tian	paocai	meiwei	hao	huotui
151	1C145.wav	公正的案子得到了正确的判决。	The fair case got right judgement.	JUST/FAIR	CASE	GOT	RIGHT	VERDICT/JUDGEM	公正的	案子	得到	正确的	判决	gongzhengde	anzi	dedao	zhengquede	panjue
152	1C146.wav	这些花在大风中弯曲。	These flowers bend in a heavy wind.	THESE	THISTLES/FLOWE	BEND	HIGH/HEAVY	WIND	这些	花	弯曲	大	风	zhexie	huasheng	wanqu	da	feng
153	1C147.wav	纯种的泰迪犬有卷发。	Pure Teddy have curly hair.	PURE	BRED/TEDDY	POODLES/HAIR	HAVE	CURLS/CURLY	纯	泰迪	发	有趣	卷	chun	taidi	fa	youqu	juan
154	1C148.wav	红旗优雅地高高飘扬着。	The red flag is flying high in a graceful way.	TREE/RED	TOP/FLAG	WAVED/FLYING	GRACEFUL	WAY/HIGH	红	旗	飘扬	优雅	高	hong	qi	piaoyang	youya	gao
155	1C149.wav	书上的污点是绿色墨水弄上的。	The stain on the book was made by green ink.	BLOT/STAIN	BOOK	MADE	GREEN	INK	污点	书	造成	绿色	墨水	wudian	shu	zaocheng	lvse	moshui
156	1C150.wav	泥溅到他白色衬衫上。	Mud splashed on his white shirt.	MUD	SPLATTERED/SPLA	ON	WHITE	SHIRT	泥	溅	他的	白色	衬衫	ni	jian	tade	baise	chenshan
157	1C151.wav	沙发垫是红色的而且很轻很干净。	The sofa cushion is red, light and clean.	SOFA	CUSHION	RED	LIGHT	WEIGHT/CLEAN	沙发	坐垫	红色	轻	干净	shafa	zuodian	hongse	qing	ganjing
158	1C152.wav	空瓶子放在托盘上。	The empty bottle is placed on the tray.	EMPTY	FLASK/BOTTLE	STOOD/PLACED	TIN/ON	TRAY	空	瓶子	放	上	托盘	kong	pingzi	fang	shang	tuopan
159	1C153.wav	一个速度快的人可以打破这个竞赛记录	A speedy man can beat this competition record.	SPEEDY	MAN	BEAT	TRACK/COMPETIT	RECORD	迅速	人	打破	竞赛	记录	xunsu	ren	dapo	jingsai	jilu
160	1C154.wav	那天他打破了一个新球拍。	He broke a new racket that day.	BROKE	NEW	RACKET	THAT	DAY	打破	新	球拍	那	天	dapo	xin	qiupai	na	tian
161	1C155.wav	咖啡桌对于沙发来说太高了。	The coffee table is too high for the sofa.	COFFEE	TABLE	TOO	HIGH	SOFA	咖啡	桌子	太	高	沙发	kafei	zhuozi	tai	gao	shafa
162	1C156.wav	写短篇小说的冲动很小。	The urge to write short stories is rare.	URGE	WRITE	SHORT	STORIES	RARE	冲动	写	短	小说	小	chongdong	xie	duan	xiaoshuo	xiao
163	1C157.wav	所有的铅笔全部用完了。	All the pencils are used up.	PENCILS	HAVE/UP	ALL	BEEN/ARE	USED	铅笔	完了	所有	被	用	qianbi	wanle	suoyou	bei	yong
164	1C158.wav	海盗抓住了那艘船上所有的船员。	The pirates caught all the crew from that ship.	PIRATES	SEIZED/CAUGHT	CREW	LOST/ALL	SHIP	海盗	抓住	船员	所有的	船	haidao	zhuazhu	chuanyuan	suoyoude	chuan
165	1C159.wav	我们试图更换硬币，但失败了。	We tried to replace the coin but failed.	TRIED	REPLACE	COIN	BUT	FAILED	试图	更换	硬币	但	失败	shitu	genghuan	yingbi	dan	shibai
166	1C160.wav	她相当整齐地缝了破的外套。	She sewed the torn coat quite neatly.	SEWED	TORN	COAT	QUITE	NEATLY	缝	破	外套	很	整齐地	feng	po	waitao	hen	zhengqide
167	1C161.wav	一阵尘土从热水管子上落下来。	A shower of dirt fell from the hot pipes.	SHOWER	DIRT	FELL	HOT	PIPES	一阵	尘土	落	热	管子	yizhen	chentu	luo	re	guanzi
168	1C162.wav	外套挂在宽椅子的靠背上。	The jacket hung on the back of the wide chair.	JACKET	HUNG	BACK	WIDE	CHAIR	外套	挂	靠背	宽	椅子	waitao	gua	kaobei	kuan	qizi
169	1C163.wav	高海拔的空气是纯净的。	At that high level the air is pure.	AT	HIGH	LEVEL	AIR	PURE	在	高	海拔	空气	纯	zai	gao	haiba	kongqi	chun
170	1C164.wav	你加数字时删掉这两个。	Delete those two when you add the numbers.	DROP/DELETE	TWO	WHEN	ADD	FIGURES	删掉	两	时	加	数字	shandiao	liang	shi	jia	shuzi
171	1C165.wav	现在很难买到填充盒。	It is hard to buy a filling case now.	HARD	BUY	FILING	CASE	NOW	难	买	填充	盒	现在	nan	mai	tianchong	he	xianzai
172	1C166.wav	突如其来的开始不会很多	A sudden beginning does not win the prize.	ABRUPT/SUDDEN	START/BEGINNING	NOT	WIN	PRIZE	突然	开始	不	赢得	奖	turan	kaishi	bu	yingde	jiang

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
172	1D166.wav	突如其来的开始不会得奖。	A sudden beginning does not win the prize.	ABRUPT/SUDDEN	START/BEGINNING	NOT	WIN	PRIZE	突然	开始	不	赢得	奖	turan	kaishi	bu	yingde	jiang
173	1D167.wav	木头最适合制作玩具和积木。	Wood is best for making toys and blocks.	WOOD	BEST	MAKING	TOYS	BLOCKS	木头	最好	制造	玩具	积木	mutou	zuihao	zhizao	wanju	jimu
174	1D168.wav	办公室的油漆是暗淡、让人伤感的棕色。	The office paint was a dull, sad brown.	OFFICE	PAINT	DULL	SAD	TAN/BROWN	办公室	油漆	暗淡	伤感	棕色	bangongshi	youqi	andan	shanggan	zongse
175	1D169.wav	他知道这位伟大的年轻女演员的演技。	He knew the skill of the great young actress.	KNEW	SKILL	GREAT	YOUNG	ACTRESS	知道	演技	伟大的	年轻	女演员	zhidao	yanji	weidade	nianqing	nvyanyuan
176	1D170.wav	布会把溅出的水吸干。	Cloth will soak up spilled water.	RAG/CLOTH	SOAK	UP	SPILLED	WATER	布	吸收	干	溅出	水	bu	xishou	gan	jianchu	shui
177	1D171.wav	将商店的账户添加到最后一分钱。	Add the store's account to the last penny.	ADD	STORE'S	ACCOUNT	LAST	PENNY	加	商店的	帐户	最后	一分钱	jia	shangdiande	zhanghu	zuihou	yifenqian
178	1D172.wav	蒸汽从破裂的门发出噪音。	The steam is making noise from the broken door.	STEAM	HISSED/NOISE	FROM	BROKEN	VALVE/DOOR	蒸汽	噪音	从	破裂的	门	zhengqi	zaoyin	cong	poliede	men
179	1D173.wav	那条狗差点把小男孩和小女孩弄伤了。	The dog almost hurt the boy and girl.	DOG	ALMOST	HURT	SMALL/BOY	CHILD/GIRL	狗	差点	伤	男孩	女孩	gou	chadian	shang	nanhai	nvhai
180	1D174.wav	外面有干树叶的声音。	There was a sound of dry leaves outside.	THERE	SOUND	DRY	LEAVES	OUTSIDE	那里	声音	干	树叶	外	nali	shengyin	gan	shuye	wai
181	1D175.wav	那天早晨的天空清澈而且蔚蓝。	The sky was clear and bright blue that morning.	SKY	MORNING	CLEAR	BRIGHT	BLUE	天空	早上	晴朗	亮	蓝	tiankong	zaoshang	qinglang	liang	lan
182	1D176.wav	撕破的纸散落在地板上。	The torn paper was scattered on the floor.	SCRAPS/SCATTER	TORN	PAPER	LITTERED/ON	FLOOR	散落	撕破	纸	上	地板	sanluo	sipo	zhi	shang	diban
183	1D177.wav	周日是一周中最好的一天。	Sunday is the best day of a week.	SUNDAY	BEST	PART/DAY	OF	WEEK	星期日	最好	天	的	周	xingqiri	zuihao	tian	de	zhou
184	1D178.wav	医生用这些药治好了他。	The doctor cured him with these pills.	DOCTOR	CURED	HIM	THESE	PILLS	医生	治疗	他	这些	药	yisheng	zhiliao	ta	zhexie	yao
185	1D179.wav	新来的女孩今天中午被解雇。	The new girl was fired today at noon.	NEW	GIRL	FIRED	TODAY	NOON	新	女孩	解雇	今天	中午	xin	nvhai	jiegu	jintian	zhongwu
186	1D180.wav	当船到达港口时，他们感到很高兴。	They felt happy when the ship arrived in port.	FELT	HAPPY	SHIP	ARRIVED	PORT	感到	高兴	船	到达	港口	gandao	gaoxing	chuan	daoda	gangkou
187	1D181.wav	她很会穿衣服。	She is very good at wearing clothes.	HAS/SHE	WAY/VERY	WEARING	SMART/GOOD	CLOTHES	她	非常	穿	会	衣服	ta	feichang	chuan	hui	yifu
188	1D182.wav	酸会在羊毛布上烧洞。	Acid burns holes in wool cloth.	ACID	BURNS	HOLES	WOOL	CLOTH	酸	烧	洞	羊毛	布	suan	shao	dong	yangmao	bu
189	1D183.wav	写童话应该很有趣。	Writing fairy tales should be fun.	FAIRY	TALES	SHOULD	FUN	WRITE/WRITING	虚构的	故事	应该	有趣	写	xugoude	gushi	yinggai	youqu	xie
190	1D184.wav	八英里的林地被烧毁。	Eight miles of woodland burned to waste.	EIGHT	MILES	WOODLAND	BURNED	WASTE	八	英里	林地	烧	毁	ba	yingli	lindi	shao	hui
191	1D185.wav	第三轮很无聊让球员很累。	The third round was dull and tired the players.	THIRD	ACT/ROUND	DULL	TIRED	PLAYERS	第三	轮	无聊	累	球员	disan	lun	wuliao	lei	qiuyuan
192	1D186.wav	婴儿和幼儿不应受到惊吓。	Babies and children should not be scared.	YOUNG/BABIES	CHILD/CHILDREN	NOT	SUFFER/SHOULD	FRIGHT/SCARED	婴儿	幼儿	不	应该	惊吓	yinger	youer	bu	yinggai	jingxia
193	1D187.wav	将这几列加起来并将总和放在这。	Add the column and put the sum here.	ADD	COLUMN	PUT	SUM	HERE	加	列	放	和	这里	jia	lie	fang	he	zheli
194	1D188.wav	我们钦佩并喜爱优秀的厨师。	We admire and love a good cook.	WE	ADMIRE	LOVE	GOOD	COOK	我们	钦佩	喜爱	优秀的	厨师	women	qinpei	xiai	youxiude	chushi
195	1D189.wav	那里的红色标记是十英寸。	The red mark there is 10 inches.	RED	MARK	THERE	TEN	INCHES	红色的	标记	那	十	英寸	hongsede	biaoji	na	shi	yingcun
196	1D190.wav	他从圆形白色大理石块上雕刻了一个头	He carved a head from the round white marble.	CARVED	HEAD	ROUND	BLOCK/WHITE	MARBLE	刻	头	圆形	白色	大理石	ke	tou	yuanxing	baise	dalishi
197	1D191.wav	农民来收获燕麦和玉米。	Farmers came to harvest the oats and corn.	FARMERS	CAME	THRESH/HARVEST	OAT	CORN	农民	来	收获	燕麦	玉米	nongmin	lai	shouhuo	yanmai	yumi
198	1D192.wav	无花果树的果实是苹果状的。	The fruit of a fig tree is apple shaped.	FRUIT	FIG	TREE	APPLE	SHAPED	水果	无花果	树	苹果	形状	shuiguo	wuhuaguo	shu	pingguo	xingzhuang
199	1D193.wav	玉米芯可以用来点火。	Corn cobs can be used to kindle a fire.	CORN	COBS	USED	KINDLE/LIGHT	FIRE	玉米	芯	用来	点	火	yumi	xin	yonglai	dian	huo
200	1D194.wav	噪音开始时他们在哪里?	Where were they when the noise started.	WHERE	THEY	WHEN	NOISE	STARTED	哪里	他们	什么时候	噪音	开始	nali	tamen	shenmeshihou	zaoyin	kaishi
201	1D195.wav	纸盒中装满了白色图钉。	The paper box is full of white pins.	PAPER	BOX	FULL	THUMB/WHITE	TACKS/PINS	纸	盒子	满	白色	图钉	zhi	hezi	man	baise	tuding
202	1D196.wav	将你的礼物卖给买家以获得高收益。	Sell your gift to a buyer to get a high income.	SELL	GIFT	BUYER	GOOD/HIGH	GAIN/INCOME	卖	礼物	买家	高	收益	mai	liwu	maijia	gao	shouyi
203	1D197.wav	刀放在冰桶旁边。	The knife is placed beside the ice bucket.	TONGS/KNIFE	LAY/PLACED	BESIDE	ICE	PAIL/BUCKET	刀	放	旁边	冰	桶	dao	fang	pangbian	bing	tong
204	1D198.wav	风把花从树上吹下来。	The wind blows the flowers from the tree.	PETALS/BLOWS	FALL/FLOWERS	NEXT/FROM	PUFF/TREE	WIND	吹	花	从	树	风	chui	hua	cong	shu	feng
205	1D199.wav	把最好的指南针带到头等舱。	Bring your best compass to the first class.	BRING	BEST	COMPASS	THIRD/FIRST	CLASS	带	最好	罗盘	头	舱	dai	zuihao	luopan	tou	cang
206	1D200.wav	他们虽然伤心仍笑了出来。	They still laugh although they are sad.	COULD/STILL	LAUGH	ALTHOUGH	WERE/THEY	SAD	仍	笑	虽然	他们	伤心	reng	xiao	suiran	tamen	shangxin

5.4 APPENDIX D: Mandarin Keyword Rater Script

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Tue Aug  4 15:38:35 2020
@author: lyndonrakusen
"""
import pandas as pd
import jieba #MUST type 'pip install jieba' in cmd line before running

def get_response_lines(df):
    drop_list = []
    for index, df_line in df.iterrows():
        # if the line is a practice trial get rid of it (?)
        if str(df_line['display']) == 'prac':
            drop_list.append(index)
        # if the line isn't a response line get rid of it
        elif str(df_line['Zone Type']) != 'response_text_entry':
            drop_list.append(index)

    df = df.drop(drop_list).reset_index()
    return df

def mandarin_keyword_rater(df_line):
    response = str(df_line['Response']).replace('他','她')
    #store all keywords as values in dictionary with new score column name as keys
    key_words = {f'score{x}':df_line[f'Mandarin_KEYWORD{x}'] for x in range(1,6)}
    #sort dictionary so longest keywords are scored first
    key_words = dict(sorted(key_words.items(), key = lambda item : len(item[1]), reverse =
True))
    total_score = 0
    for key, word in key_words.items():
        if '他' in word:
            word = word.replace('他','她')
        if word in response:
            df_line[key] = 1
            # remove scored character cluster from response
            # so individual characters can't get scored on their own
            response = response.replace(word, '')
        else:
            df_line[key] = 0
        total_score += df_line[key]

    df_line['score_all'] = total_score/5
    return df_line

def mandarin_masker_rater(df_line):
```

```
    masker_total_score = 0
    len_key_words = 0
    for masker in ['high_pitch_masker', 'low_pitch_masker']:
        # check if trial had a masker, otherwise skip rating
        if type(df_line[masker]) is not str:
            return df_line
        response = str(df_line['Response']).replace('他','她')
        key_words = df_line[masker].split(' ')
        #sort so longest keywords are scored first
        key_words = sorted(key_words, key=len, reverse = True)

        target_sentence = str(df_line['Mandarin Sentence'])
        total_score = 0
        for word in key_words:
            if '他' in word:


                word = word.replace('他','她')
            # check if masker word is in response
            for resp_word in jieba.cut(response):
                if resp_word == word:
                    #Do not score if masker word is also in target sentence
                    if word in target_sentence:
                        total_score += 0
                        target_sentence = target_sentence.replace(word, '')
                    else:
                        total_score += 1


        df_line[f'{masker}_score'] = round(total_score/len(key_words), 2)

        masker_total_score += total_score
        len_key_words += len(key_words)

    df_line['masker_score'] = round(masker_total_score/len_key_words, 2)
    return df_line


files = ['data_exp_23825-v5_task-wh8s.csv',
         'data_exp_23825-v5_task-jts5.csv',
         'data_exp_23825-v6_task-wh8s.csv',
         'data_exp_23825-v6_task-jts5.csv',
         'data_exp_23825-v7_task-wh8s.csv',
         'data_exp_23825-v7_task-jts5.csv',
         'data_exp_27323-v2_task-194c.csv', #SET1.2
         'data_exp_27323-v2_task-3zcy.csv',
         'data_exp_27323-v2_task-8vam.csv',
         'data_exp_27323-v2_task-927j.csv',
         'data_exp_27323-v2_task-b1pp.csv',
         'data_exp_27323-v2_task-ctwa.csv',
```

'data_exp_27323-v2_task-ekgb.csv',
'data_exp_27323-v2_task-etpk.csv',
'data_exp_27323-v2_task-f8yv.csv',
'data_exp_27323-v2_task-h5tp.csv',
'data_exp_27323-v2_task-imak.csv',
'data_exp_27323-v2_task-iy9h.csv',
'data_exp_27323-v2_task-pi5q.csv',
'data_exp_27323-v2_task-pmlg.csv',
'data_exp_27323-v2_task-qotz.csv',
'data_exp_27323-v2_task-sn2x.csv',
'data_exp_26986-v3_task-1svi.csv', #SET1.3
'data_exp_26986-v3_task-33lh.csv',
'data_exp_26986-v3_task-3vfl.csv',
'data_exp_26986-v3_task-7ppm.csv',
'data_exp_26986-v3_task-at3k.csv',
'data_exp_26986-v3_task-hfei.csv',
'data_exp_26986-v3_task-hh4m.csv',
'data_exp_26986-v3_task-jirf.csv',
'data_exp_26986-v3_task-noe1.csv',
'data_exp_26986-v3_task-pn4g.csv',
'data_exp_26986-v3_task-qiz8.csv',
'data_exp_26986-v3_task-vwnz.csv',
'data_exp_27535-v2_task-1tmu.csv', #SET2.1
'data_exp_27535-v2_task-3h7h.csv',
'data_exp_27535-v2_task-3oa6.csv',
'data_exp_27535-v2_task-8rdt.csv',
'data_exp_27535-v2_task-8vpc.csv',
'data_exp_27535-v2_task-a51u.csv',
'data_exp_27535-v2_task-a5zk.csv',
'data_exp_27535-v2_task-anjw.csv',
'data_exp_27535-v2_task-auut.csv',
'data_exp_27535-v2_task-b9vd.csv',
'data_exp_27535-v2_task-c6h9.csv',
'data_exp_27535-v2_task-cpnt.csv',
'data_exp_27535-v2_task-dcae.csv',
'data_exp_27535-v2_task-dmeh.csv',
'data_exp_27535-v2_task-ez76.csv',
'data_exp_27535-v2_task-f2oa.csv',
'data_exp_27535-v2_task-fico.csv',
'data_exp_27535-v2_task-g8ul.csv',
'data_exp_27535-v2_task-gazf.csv',
'data_exp_27535-v2_task-gytr.csv',
'data_exp_27535-v2_task-h33d.csv',
'data_exp_27535-v2_task-iyul.csv',
'data_exp_27535-v2_task-k88z.csv',
'data_exp_27535-v2_task-m9uo.csv',
'data_exp_27535-v2_task-murv.csv',
'data_exp_27535-v2_task-oyuf.csv',
'data_exp_27535-v2_task-pxrm.csv',
'data_exp_27535-v2_task-py79.csv',

```
        'data_exp_27535-v2_task-q3tc.csv',
        'data_exp_27535-v2_task-r7bl.csv',
        'data_exp_27535-v2_task-uqrx.csv',
        'data_exp_27535-v2_task-v9ce.csv',
        'data_exp_27535-v2_task-w6bi.csv',
        'data_exp_27535-v2_task-xy1s.csv',
        'data_exp_27535-v2_task-zvg1.csv',
        'data_exp_27535-v2_task-zxcz.csv'
        ]

for file in files:
    df = pd.read_csv(file)
    col_order = df.columns

    df = get_response_lines(df)
    df = df.apply(mandarin_keyword_rater, axis =1)
    df = df.apply(mandarin_masker_rater, axis =1)

    df = df.reindex(col_order.append(df.columns).drop_duplicates(), axis=1)
    df = df.drop('index', axis = 1)

    df.to_csv(file[:-4]+'_scored'+'.csv', index= False)
```

5.5 APPENDIX E: English Keyword Rater Script

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Tue Dec 15 10:50:13 2020

@author: lyndonrakusen
"""
import pandas as pd
from pathlib import Path
import string

def get_response_lines(df):
    drop_list = []
    for index, df_line in df.iterrows():
        # if the line is a practice trial get rid of it (?)
        if str(df_line['display']) == 'prac':
            drop_list.append(index)
        # if the line isn't a response line get rid of it
        elif str(df_line['Zone Type']) != 'response_text_entry':
            drop_list.append(index)

    df = df.drop(drop_list).reset_index()
    return df

def english_keyword_rater(df_line):
    response = str(df_line['Response']).lower()
    # strip punctuation from responses (needs testing with apostrophes)
    response = response.translate(str.maketrans('', '', string.punctuation))
    resp_words = set(response.split(' '))

    #store all keywords as values in dictionary with new score column name as keys
    key_words = {f'score{x}':df_line[f'KEYWORD{x}'].lower() for x in range(1,6)}
    total_score = 0
    for key, word in key_words.items():
        if word in resp_words:
            df_line[key] = 1
            response = response.replace(word, '')
        else:
            df_line[key] = 0
        total_score += df_line[key]
    # divide score by number of keywords
    df_line['score_all'] = total_score/5
    return df_line

def english_masker_rater(df_line):
    response = str(df_line['Response']).lower()
    # strip punctuation from responses (needs testing with apostrophes)
```

```
    response = response.translate(str.maketrans('', '', string.punctuation))
    resp_words = set(response.split(' '))

    for masker_pitch in ['high_pitch_masker', 'low_pitch_masker']:

        masker = str(df_line[masker_pitch]).lower()
        masker_words = [word for word in masker.split(' ') if '_' not in word]

        score = 0
        for word in masker_words:
            if word in resp_words:
                score+= 1
        try:
            masker_len = df_line[masker_pitch+'_keywords']
        except KeyError:
            masker_len = len(masker_words)
        df_line[masker_pitch+'_score'] = score/masker_len

    return df_line
# put working directory here
fpth = Path('E:/allfiles/study1/STUDY8/ANALYSIS/raw_files')

# list file names in directory
fnames = [file.stem for file in fpth.rglob('*.csv') if 'scored' not in file.stem]

df_master = pd.DataFrame()
for fname in fnames:
    infile = Path(fpth / fname)

    df = pd.read_csv(infile.with_suffix('.csv'))
    df = get_response_lines(df)

    df = df.apply(english_keyword_rater, axis =1)
    df = df.apply(english_masker_rater, axis =1)
    df_master = df_master.append(df)
    df = df.dropna(how='all', axis=1)
    print(fname)
    fname+='_scored'
    outfile = Path(fpth / fname)
    df.to_csv(outfile.with_suffix('.csv'), index = False)

# df_master = df_master.dropna(how='all', axis=1)
df_master.to_csv(Path(fpth / 'data_exp_35837_all_scored.csv'), index = False)
```

6. REFERENCES

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *J. Acoust. Soc. Am.*, *126*(5), 2649-2659. https://doi.org/10.1121/1.3216914

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *J. Exp. Psychol. Hum. Percept. Perf.*, *35*(2), 520-529. https://doi.org/10.1037/a0013552

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear Hear.*, *40*(5), 1084-1097. https://doi.org/10.1097/AUD.0000000000000697

Altmann, G. T. M., & Young, D. (1993, September). Factors affecting adaptation to time-compressed speech. Paper presented at *Eurospeech 9*, Berlin, Germany.

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

Arbogast, T. L., Mason, C. R., & Kidd, G. Jr. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, *117*(4), 2169-2180. https://doi.org/10.1121/1.1861598

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403-450. https://doi.org/10.1146/annurev. neuro.28.061604.135709

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *J. Mem. Lang.*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barker, J., & Cooke, M. (1999). Is the sine-wave speech cocktail party worth attending? *Speech Commun.*, *27*(3-4), 159-174. https://doi.org/10.1016/S0167-6393(98)00081-8

Barker, J., & Cooke, M. (2007). Modelling speaker intelligibility in noise. *Speech Communication*, *49*(5), 402-417. https://psycnet.apa.org/doi/10.1016/j.specom.2006.11.003

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., Baayen, R. H. (2018). Parsimonious Mixed Models. arXiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.*, *91*(2), https://doi.org/10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 142–162). New York, NY: Cambridge University Press.

Bench, J., Kowal, Å., & Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *Br. J. Audiol.*, *13*(3), 108-112. https://doi.org/10.3109/03005367909078884

Bent, T., Buchwald, A., & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *J. Acoust. Soc. Am.*, *126*(5), 2660-2669. https://doi.org/10.1121/1.3212930

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2014). Objective assessment of listening effort in the oscillatory EEG: comparison of different hearing aid configurations. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, *2014*, 2653-2656. https://doi.org/10.1109/EMBC.2014.6944168

Biçer, A., Koelewijn, T., & Başkent, D. (2023). Short implicit voice training affects listening effort during a voice cue sensitivity task with vocoder-degraded speech. *Ear Hear.*, *44*(4), 900-916. https://doi.org/10.1097/aud.0000000000001335

Bird, J., & Darwin, C. J. (1998). Effects of a difference in fundamental frequency in separating two sentences. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and Physiological Advances in Hearing*, pp. 263-269. Whurr, London.

Boersma, P. & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program]. Version 6.1, retrieved 28 September 2019 from http://www.praat.org/

Borghini, G., & Hazan, V. (2018). Listening Effort During Sentence Processing Is Increased for Non-native Listeners: A Pupillometry Study. *Front. Neurosci.*, *12*(152), 1-13. https://doi.org/10.3389/fnins.2018.00152

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012a). Perceptual Learning of Dysarthric Speech: A Review of Experimental Studies. *J. Speech Lang. Hear. Res.*, *55*(1), 290-305. https://doi.org/10.1044/1092-4388(2011/10-0349)

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012b). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Lang. Cognit. Proc.*, *27*(7-8), 1039-1055. https://doi.org/10.1080/01690965.2011.610596

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2012c). A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech. *J. Acoust. Soc. Am.*, *132*(2), EL102-EL108. https://doi.org/10.1121/1.4736952

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.*, *121*(4), 2339-2349. https://doi.org/10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729. https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.*, *101*(4), 2299-2310. https://doi.org/ 10.1121/1.418276

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun.*, *20*(3), 255-272. https://doi.org/10.1016/S0167-6393(96)00063-5

Bregman, A. (1990). *Auditory scene analysis: The perceptual organisation of sound*. Cambridge, MA: MIT Press.

Brokx, J. P., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *J. Phon.*, *10*(1), 23-36. https://doi.org/10.1016/S0095-4470(19)30909-X

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 control. *J. Acoust. Soc. Am.*, *131*(2), 1449-1464. https://doi.org/10.1121/1.3675943

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Q. J. Exp. Psychol.*, *73*(9), 1431-1443. https://doi.org/10.1177/1747021820916726

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.*, *109*(3), 1101-1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.*, *120*(6), 4007-4018. https://doi.org/10.1121/1.2363929

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.*, *110*(5), 2527-2538. https://doi.org/10.1121/1.1408946

Burke, L. A. & Naylor, G. (2020). Daily-Life Fatigue in Mild to Moderate Hearing Impairment: An Ecological Momentary Assessment Study. *Ear Hear.*, *41*(6), 1518-1532. https://doi.org/ 10.1097/AUD.0000000000000888

Buss, E., Calandruccio, L., Oleson, J., & Leibold, L. J. (2020). Contribution of Stimulus Variability to Word Recognition in Noise Versus Two-Talker Speech for School-Age Children and Adults. *Ear Hear.*, *42*(2), 313-322. https://doi.org/10.1097/AUD.0000000000000951

Calandruccio, L. & Zhou, H. (2014). Increase in Speech Recognition due to Linguistic Mismatch Between Target and Masker Speech: Monolingual and Simultaneous Bilingual Performance. *J. Speech Lang. Hear. Res.*, *57*(3), 1089-1097. https://doi.org/10.1044/2013_JSLHR-H-12-0378

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. *Am. J. Audiol.*, *22*(1), 157-164. https://doi.org/10.1044/1059-0889(2013/12-0072)

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 Hear.*, *21*, 2331216517709385. https://doi.org/ https://doi.org/10.1177%2F2331216517709385

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J. Acoust. Soc. Am.*, *128*(2), 860-869. https://doi.org/10.1121/1.3458857

Carlyon, R. P. (1996). Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J. Acoust. Soc. Am.*, *99*(1), 517-524. https://psycnet.apa.org/doi/10.1121/1.414510

Cherry, C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.*, *25*(5), 975-979. https://doi.org/10.1121/1.1907229

Ching, T. Y. C. & Dillon, H. (2013). A Brief Overview of Factors Affecting Speech Intelligibility of People With Hearing Loss: Implications for Amplification. *Am. J. Audiol.*, *22*(2), 306-309. https://doi.org/10.1044/1059-0889(2013/12-0075)

Cooke, M., & García Lecumberri, M. L. (2020). Sculpting speech from noise, music, and other sources. *J. Acoust. Soc. Am.*, *148*(1), EL20-26. https://doi.org/10.1121/10.0001474

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. *J. Acoust. Soc. Am.*, *123*(1), 414-427. https://doi.org/10.1121/1.2804952

Cooke, M., Scharenborg, O., & Meyer, B. T. (2022). The time course of adaptation to distorted speech. *J. Acoust. Soc. Am.*, *151*(4), 2636-2646. https://doi.org/10.1121/10.0010235

Cooke. M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.*, *119*(3), 1562-1573. https://doi.org/10.1121/1.2166600

Culling, J. F., & Stone, M. A. (2017). Energetic Masking and Masking Release. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. 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

Cunningham, L. L. & Tucci, D. L. (2017). Hearing Loss in Adults. *N. Engl. J. Med.*, *377*(25), 2465-2473. https://doi.org10.1056/NEJMra1616601/

da Silva Castanheira, K., LoParco, S., & Otto, A. R. (2020). Task-evoked pupillary responses track effort exertion: Evidence from task-switching. *Cogn. Affect. Behav. Neurosci.*, *21*, 592–606. https://doi.org/10.3758/s13415-020-00843-z

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. *J. Acoust. Soc. Am.*, *114*(5), 2913-2922. https://doi.org/10.1121/1.1616924

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigen, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *J. Exp. Psychol., 134*(2), 222-241. https://doi.org/10.1037/0096-3445.134.2.222

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Biling. Lang. Cog.*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Sci. Rep.*, *9*, 11278. https://doi.org/10.1038/s41598-019-47643-1

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Elsevier.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *J. Exp. Psychol. Hum. Percept. Perf.*, *23*(3), 914-927. https://doi.org/10.1037/0096-1523.23.3.914

Durlach, N. I., Mason, C. R., Kidd, G. Jr., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *J. Acoust. Soc. Am.*, *113*(6), 2984-2987. https://doi.org/10.1121/1.1570435

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychol.*, *50*(9), 2154-2164. https://doi.org/10.1016/j.neuropsychologia.2012.05.013

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The Brain Dynamics of Rapid Perceptual Adaptation to Adverse Listening Conditions. *J. Neurosci.*, *33*(26), 10688-10697. https://doi.org/10.1523/JNEUROSCI.4596-12.2013

ETS (2010). *Linking TOEFL iBT Scores to IELTS Scores: A Research Report TOEFL iBT IELTS ESL*. https://www.ets.org/research/policy_research_reports/publications/report/2010/isol.html, Retrieved 27 Nov 2023.

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

Ezzatian, P., Li, L., Pichora-Fuller, M. K., & Schneider, B. A. (2012). The effect of energetic and information masking on the time-course of stream segregation: Evidence that streaming depends on vocal fine structure cues. *Lang. Cognit. Proc.*, *27*(7-8), 1056-1088. https://doi.org/10.1080/01690965.2011.591934

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague, Netherlands: Moulton & Co.

Felty, R. A., Buchwald, A., & Pisoni, D. B. (2009). Adaptation to frozen babble in spoken word recognition. *J. Acoust. Soc. Am.*, *125*(3), EL93-EL98. https://doi.org/10.1121/1.3073733

Feng, Y.-M., Xu, L., Zhou, N., Yang, G., & Y, S.-K. (2012). Sine-wave speech recognition in a tonal language. *J. Acoust. Soc. Am.*, *131*(2), EL133-EL138. https://doi.org/10.1121/1.3670594

Festen, J. M. & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.*, *88*(4), 1725-1736. https://doi.org/10.1121/1.400247

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *J. Exp. Psychol. Hum. Percept. Perf.*, *32*(5), 1276-1293. https://doi.org/10.1037/0096-1523.32.5.1276

Fox, C. M. & Ramig, L. O. (1997). Vocal Sound Pressure Level and Self-Perception of Speech and Voice in Men and Women With Idiopathic Parkinson Disease. *Am. J. Speech Lang. Pathol.*, *6*(2), 85-94. https://doi.org/10.1044/1058-0360.0602.85

Fox, C. M., Ramig, L. O., Ciucci, M. R., Sapir, S., McFarland, D. H., & Farley, B. G. (2006). The science and practice of LSVT/LOUD: neural plasticity-principled approach to treating individuals with Parkinson disease and other neurological disorders. *Semin. Speech Lang.*, *27*(4), 283-299. https://doi.org/ 10.1055/s-2006-955118

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

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2008). Spatial release from masking with noise-vocoded speech. *J. Acoust. Soc. Am., 124*(3), 1627-1637. https://doi.org/10.1121/1.2951964

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceptive spatial separation in the unmasking of speech. *J. Acoust. Soc. Am., 106*(6), 3578-3588. https://doi.org/10.1121/1.428211

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *J. Acoust. Soc. Am., 119*(4), 2445-2454. https://doi.org/10.1121/1.2180210

Gaudrain, E., Li, S., Ban, V. S., Patterson, R. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Proceedings of *Interspeech 2009*, Brighton, United Kingdom. https://hal.archives-ouvertes.fr/hal-02144510/document

Geller. J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behav. Res. Methods, 52*(5), 2232-2255. https://doi.org/10.3758/s13428-020-01374-8

Golomb, J. D., Peelle, J. W., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *J. Acoust. Soc. Am., 121*(3), 1701-1708. https://doi.org/10.1121/1.2436635

Granholm, E. & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *Int. J. Psychophysiol., 52*(1), 1-6. https://doi.org/10.1016/j.ijpsycho.2003.12.001

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiol., 33*(4), 457-461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Händel, G. F. (ca. 1731-35). *Bourrée*. From Water Music HWV 348. Dieren: Canzona Music. https://imslp.org/wiki/Special:ReverseLookup/527960, Retrieved 18 August 2020

Haro. S., Rao, H. M., Quatieri, T. F., & Smalt, S. J. (2022). EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. *Eur. J., Neurosci., 55*(5), 1262-1277. https://doi.org/10.1111/ejn.15616

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland: Elsevier Science Publishers B.V. https://doi.org/10.1016/S0166-4115(08)62386-9

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *J. Acoust. Soc. Am., 116*(5), 3108-3118. https://doi.org/10.1121/1.1806826

Helfer, K. S., & Freyman, R. L. (2010). Aging and Speech-on-Speech Masking. *Ear Hear., 29*(1), 87-98. https://doi.org/10.1097/AUD.0b013e31815d638b

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol., 110*, 100-106. http://doi.org/10.1016/j.biopsycho.2015.06.013

Howard-Jones. P. & Rosen, S. (1993). The Perception of Speech in Fluctuating Noise. *Acustica, 78*, 258-272.

Huyck, J. J., & Johnsrude, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *J. Acoust. Soc. Am.*, *131*(3), EL236-242. https://doi.org/10.1121/1.3685511

IEEE Subcommittee on Subjective Measurements (1969). IEEE Recommended Practices for Speech Quality Measurements. *IEEE Trans. Audio Electroacoust.*, *17*, 227-246. https://doi.org/10.1109/TAU.1969.1162058

International English Language Testing System (IELTS; 2020). *How IELTS is scored*. https://www.ielts.org/about-the-test/how-ielts-is-scored, Retrieved 17 April 2020.

Irino, T. & Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *J. Acoust. Soc. Am.*, *99*(4 Pt 1), 2316-2331. https://doi.org/10.1121/1.415419

Jin, S.-H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *J. Acoust. Soc. Am.*, *132*(5), EL391-EL397. https://doi.org/10.1121/1.4757730

Kaandorp, M. W., De Groot, A. M. B., Festen, J. M., Smits, C., & Goverts, S. T. (2014). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *Int. J. Audiol.*, *55*(3), 157-167. https://doi.org/10.3109/14992027.2015.1104735

Kidd G., Mason C.R., Richards V.M., Gallun F.J., Durlach N.I. (2008) Informational Masking. In: Yost W.A., Popper A.N., Fay R.R. (eds) *Auditory Perception of Sound Sources. Springer Handbook of Auditory Research*, vol 29. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-71305-2_6

Kidd, G. Jr., & Colburn, H. S. (2017). Informational Masking in Speech Recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.) *The Auditory System at*

the Cocktail Party (pp. 75-109). Springer Nature. https://doi.org/10.1007/978-3-319-51662-2

Kidd, G. Jr., Arbogast, T., L., Mason, C. R., & Walsh, M. (2002). Informational Masking in Listeners with Sensorineural Hearing Loss. *J. Assoc. Res. Otolaryngol., 3*, 107-119. https://doi.org/10.1007/s101620010095

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Front. Psychol., 5*(651), 1-9. https://doi.org/10.3389/fpsyg.2014.00651

Klasner, E. R. & Yorkston, K. M. (2005). Speech intelligibility in ALS and HD dysarthria: The everyday listener's perspective. *J. Med. Speech Lang. Pathol., 13*(2), 127-139.

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2014a). The pupil response is sensitive to divided attention during speech processing. *Hear. Res., 312*, 114-120. https://doi.org/10.1016/j.heares.2014.03.010

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2014b). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *J. Acoust. Soc. Am., 135*(3), 1596-1606. http://doi.org/10.1121/1.4863198

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing Load Induced by Informational Masking Is Related to Linguistic Abilities. *Int. J. Otolaryngol., 2012*(865731), 1-11. https://doi.org/10.1155/2012/865731

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008) Accommodating variation: Dialects, idoletcts, and speech processing. *Cognition, 107*(1), 54-81. https://doi.org/10.1016/j.cognition.2007.07.013

Kuchinsky S. E., Ahlstrom J. B., Vaden K. I. Jr., Cute S. L., Humes L. E., Dubno J. R., & Eckert M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiol.*, *50*(1), 23–34. https://doi.org/10.1111/j.1469-8986.2012.01477.x

Ladefoged, P. (2014). *A course in phonetics*. Stamford, CT: Cengage Learning.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *J. Exp. Psychol. Gen.*, *133*(3), 339-354. https://doi.org/10.1037/0096-3445.133.3.339

Licklider, J. C. R., & Miller G. A. (1951). The perception of speech. In S. S. Stevens (Ed.) *Handbook of experimental psychology* (p. 1040-1074). Wiley.

Lie, S., Zekveld, A. A., Smits, C., Kramer, S. E., & Versfeld, N. J. (2024). Learning effects in speech-in-noise tasks: Effect of masker modulation and masking release. *J. Acoust. Soc. Am.*, *156*(1), 341-349. https://doi.org/10.1121/10.0026519

Lindblom, B. (1990) Explaining Phonetic Variation: A Sketch of the H&H Theory. In: Hardcastle W.J., Marchal A. (eds) *Speech Production and Speech Modelling*. NATO ASI Series (Series D: Behavioural and Social Sciences), vol 55. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2037-8_16

Liss, J. M. (2007). The roll of speech perception in moto speech disorders. In: Weismer, G. (Ed.) *Motor speech disorders: Essays for Ray Kent* (pp.187-219). San Diego, CA: Plural.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *J. Acoust. Soc. Am.*, *107*(6), 3415-3424. https://doi.org/10.1121/1.429412

Litovsky, R. Y. (2012). Spatial Release from Masking. *Acoustics Today*, 18-25.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.*, *94*(3 Pt 1), 1242-1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.*, *96*(4), 2076-2087. https://doi.org/10.1121/1.410149

Loewenfeld, I. E., & Lowenstein, O. (1993). *The Pupil: Anatomy, Physiology, and Clinical Applications*. Oxford: Butterworth-Heinemann.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *J. Acoust. Soc. Am.*, *89*(2), 874-886. https://doi.org/10.1121/1.1894649

Mackersie, C. L. & Calderon-Moultrie, N. (2016). Autonomic Nervous System Reactivity During Speech Repetition Tasks: Heart Rate Variability and Skin Conductance. *Ear Hear.*, 37, 118S-125S. https://doi.org/10.1097/AUD.0000000000000305

MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. *Ophthalmic. Physiol. Opt.*, *22*(3), 175-182. https://doi.org/10.1046/j.1475-1313.2002.00023.x

Marsella, P., Scorpecci, A., Cartocci, G., Giannantonio, S., Maglione, A. G., Venuti, I., Brizi, A., & Babiloni, F. (2017). EEG activity as an objective measure of cognitive load

during effortful listening: A study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss. *Int. J. Pediatr. Otorhinolaryngol.*, *99*, 1-7. https://doi.org/10.1016/j.ijporl.2017.05.006

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Commun.*, *52*(11-12), 887-899. https://doi.org/10.1016/j.specom.2010.01.005

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang. Cognitive Proc.*, *27*(7-8), 953-978. https://doi.org/10.1080/01690965.2012.705006

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cogn. Sci.*, *32*(3), 543-62. https://doi.org/10.1080/03640210802035357

Mazeres, F., Brinkmann, K., & Richter, M. (2019). Implicit achievement motive limits the impact of task difficulty on effort-related cardiovascular response. *J. Res. Pers.*, *82*, 103842. https://doi.org/10.1016/j.jrp.2019.06.012

McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., O'Beirne, G. A., & Anderson, T. (2017). Effect of Dysarthria Type, Speaking Condition, and Listener Age on Speech Intelligibility. *Am. J. Speech Lang. Pathol.*, *26*(1), 113-123. https://doi.org/10.1044/2016_AJSLP-15-0182

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *J. Exp. Child Psychol.*, *161*, 95-122. https://doi.org/10.1016/j.jecp.2017.04.006

McGarrigle, R., Knight, S., Hornsby, B. W. Y., & Mattys, S. L. (2021a). Predictors of Listening-Related Fatigue Across the Adult Life Span. *Psychol. Sci.*, *32*(12), 1937-1951. https://doi.org/10.1177/09567976211016410

McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021b). Older adults show a more sustained pattern of effortful listening than young adults. *Psychol. Aging*, *36*(4), 504–519. https://doi.org/10.1037/pag0000587

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int. J. Audiol.*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

McGarrigle, R., Rakusen, L., & Mattys, S. (2020). Effortful listening under the microscope: Examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiol.*, *58*(1), e13703. https://doi.org/10.1111/psyp.13703

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupil response for accurately recognised accented speech. *J. Acoust. Soc. Am.*, *147*(2), EL151-EL156. https://doi.org/10.1121/10.0000718

McLaughlin, D. J., Braver, T. S., & Peelle, J. E. (2021). Measuring the Subjective Cost of Listening Effort Using a Discounting Task. *J. Speech Lang. Hear. Res.*, *64*(2), 337-347. https://doi.org/10.1044/2020_JSLHR-20-00086

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

Middlebrooks, J. C., Bierer, J. A., & Snyder, R. L. (2005). Cochlear implants: the view from the brain. *Curr. Opin. Neurobio.*, *15*(4), 488-493. https://doi.org/10.1016/j.conb.2005.06.004

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & 1.L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J. L. (1987). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 309-322). Cambridge, MA: MIT Press.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later occurring information on the perception of stop consonant and semivowel. *Percept. Psychophys.*, *25*(6), 457-465.

Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *J. Speech Lang. Hear. Res.*, *61*(9), 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451

Munro, M. J., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Lang. Learn. J.*, *45*(1), 73-97. https://doi.org/ 10.1111/j.1467-1770.1995.tb00963.x

Neagu, M.-B., Kressner, A. A., Relaño-Iborra, H., Bækgaard, P., Dau, T., & Wendt, D. (2023). Investigating the Reliability of Pupillometry as a Measure of Individualised Listening Effort. *Trends Hear.*, *27*. https://doi.org/10.1177/23312165231153288

O'Leary, R. M., Neukam, J., Hansen, T. A. Kinney, A. J., Capach, N., Svirsky, M. A., & Wingfield, A. (2023). Strategic Pauses Relieve Listeners from the Effort of Listening to Fast Speech: Data Limited and Resource Limited Processes in Narrative Recall by

Adult Users of Cochlear Implants. *Trends Hear.*, 27, 1-22. https://doi.org/10.1177/23312165231203514

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lynner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear. Res.*, 365, 90-99. https://doi.org/10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear. Res., 351*, 38-79. https://doi.org/10.1016/j.heares.2017.05.012

Pals, C., Sarampalis, A., van Dijk, M., & Başkent, D. (2019). Effects of Additional Low-Pass–Filtered Speech on Listening Effort for Noise-Band–Vocoded Speech in Quiet and in Noise. *Ear Hear., 40*(1), 3-17. https://doi.org/10.1097/AUD.0000000000000587

Paulus, M., Hazan, V., & Adank, P. (2020). The relationship between talker acoustics, intelligibility, and effort in degraded listening conditions. *J. Acoust. Soc. Am., 147*(5), 3348-3359. https://doi.org/10.1121/10.0001212

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear Hear., 39*(2), 204-214. https://doi.org/ 10.1097/AUD.0000000000000494

Peelle, J. E., & Wingfield, A. (2005). Dissociations in Perceptual Learning Revealed by Adult Age Differences in Adaptation to Time-Compressed Speech. *J. Exp. Psychol. Hum. Percept. Perf., 31*(6), 1315-1330. https://doi.org/10.1037/0096-1523.31.6.1315

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peng, Z. E., & Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. *J. Speech Lang. Hear. Res.*, *62*(4), 1068–1081. https://doi.org/10.1044/2018_JSLHR-H-17-0423

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear.*, *37*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Picou, E. M. & Ricketss, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear Hear.*, *35*(6), 611-622. https://doi.org/10.1097/AUD.0000000000000055

Picou, E. M., Bean, B., Marcrum, S. C., Ricketts, T. A., & Hornsby, B. W. Y. (2019). Moderate Reverberation Does Not Increase Subjective Fatigue, Subjective Listening Effort, or Behavioural Listening Effort in School-Aged Children. *Front. Psychol.*, *10*, 1749. https://doi.org/10.3389/fpsyg.2019.01749

Pierce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pollack, I. (1975). Auditory informational masking. *J. Acoust. Soc. Am., 57*, S5. https://doi.org/10.1121/1.1995329

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

Quant, J. R., & Woo, G. C. (1992). Normal values of eye position in the Chinese population of Hong Kong. *Optom. Vis. Sci., 69*(2), 152-158. https://doi.org/10.1097/00006324-199202000-00009

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947-950. https://doi.org/10.1126/science.7233191

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time-reversal of native and non-native interfering speech. *J. Acoust. Soc. Am., 118*(3 Pt 1), 1274-1277. https://doi.org/10.1121/1.2000751

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2006). Extended speech intelligibilty index for the perception of the speech reception threshold in fluctuating noise. *J. Acoust. Soc. Am., 120*(6), 3988-3997. https://doi.org/10.1121/1.2358008

Richter, M. (2016). The Moderating Effect of Success Importance on the Relationship Between Listening Demand and Listening Effort. *Ear Hear., 37*, 111S-117S. https://doi.org/10.1097/AUD.0000000000000295

Richter, M., Buhiyan, T., Bramsløw, L., Innes-Brown, H., Fiedler, L., Hadley, L. V., Naylor, G., Saunders, G. H., Wendt, D., Whitmer, W. M., Zekveld, A. A., & Kramer, S. E. (2023). Combining Multiple Psychophysiological Measures of Listening Effort:

Challenges and Recommendations. *Semin. Hear.*, *44*(2), 95-105. https://doi.org/10.1055/s-0043-1767669

Roberts, B., Summers, R. J., & Bailey, P. J. (2010). The perceptual organisation of sine-wave speech under competitive conditions. *J. Acoust. Soc. Am.*, *128*(2), 804-817. https://doi.org/10.1121/1.3445786

Rosen, S. & Hui, S. N. C. (2015). Sine-wave and noise-vocoded sine-wave speech in a tone language: Acoustic details matter. *J. Acoust. Soc. Am.*, *138*(6), 3698-3702. https://doi.org/10.1121/1.4937605

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, *336*(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

Rule, R. R., Shimamura, A. P., & Knight, R. T. (2002). Orbitofrontal cortex and dynamic filtering of emotional stimuli. *Cogn. Affect. Behav. Neurosci.*, *2*(3), 264-270. https://doi.org/10.3758/CABN.2.3.264

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Res.*, *1286*, 155-164. https://doi.org/10.1016/j.brainres.2009.06.083

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Atten. Percept. Psychophys.*, *71*(6), 1207-1218. https://doi.org/10.3758/APP.71.6.1207

Scharenborg, O., & van Os, M. (2019). Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Commun.*, *108*, 53-64. https://doi.org/10.1016/j.specom.2019.03.001

Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *J. Acoust. Soc. Am., 109*(6), 2880-2887. https://doi.org/10.1121/1.1372913

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science, 270*(5234), 303-304. https://www.doi.org/10.1126/science.270.5234.303

Shields, C., Slade, M., Bruce, I. A., Kluk, K., & Michani, J. (2023). Exploring the Correlations Between Measures of Listening Effort in Adults and Children: A Systematic Review with Narrative Synthesis. *Trends Hear., 27*, 23312165221137116. https://doi.org/10.1177/23312165221137116

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiol., 28*(2), 207-218. https://doi.org/10.3758/BF03331979

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci., 12*(5), 182-186. https://doi.org/10.1016/j.tics.2008.02.003

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am., 125*(5), 3306-3316. https://doi.org/10.1121/1.3101452

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J. Speech Lang. Hear. Res., 57*(1), 285-296. https://doi.org/10.1044/1092-4388(2013/12-0314)

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. *J. Acoust. Soc. Am.*, *122*(6), 3628-3639. https://doi.org/10.1121/1.2799507

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. *Cogn. Res. Princ. Implic.*, *9*(35), 1-17. https://doi.org/10.1186/s41235-024-00562-w

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory lexical processing for non-native listeners and accents. *Cognition*, *179*, 163-170. https://doi.org/10.1016/j.cognition.2018.06.001

Stecker, G. C. & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *J. Acoust. Soc. Am.*, *107*(6), 3358-3368. https://doi.org/10.1121/1.429407

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int. J. Psychophysiol.*, *52*(1), 77-86. https://doi.org/10.1016/j.ijpsycho.2003.12.005

Strand., J. F., Brown, V. A., Marchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring Listening Effort: Convergent Validity, Sensitivity, and Links With Cognitive and Personality Measures. *J. Speech Lang. Hear. Res.*, *61*(6), 1463-1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257

Summers, R. J., & Roberts, B. (2020). Informational masking of speech by acoustically similar intelligible and unintelligible interferers. *J. Acoust. Soc. Am.*, *147*(2), 1113-1125. https://doi.org/10.1121/10.0000688

Summers, V., & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J. Speech Lang. Hear. Res., 41*(6), 1294-1306. https://doi.org/10.1044/jslhr.4106.1294

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *J. Acoust. Soc. Am., 117*(3 Pt 1), 1285–1298. https://doi.org/10.1121/1.1854312

Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *J. Speech Lang. Hear. Res., 60*(10), 2989-3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041

Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Cogn. Brain Res., 12*(3), 431–439. https://doi.org/10.1016/S0926-6410(01)00067-2

Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psycophysiol., 36*(1), 22–34. https://doi.org/10.1017/s0048577299971056

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R.(2002). Top-down effects can modify the initially stimulus-driven auditory organisation. *Cogn. Brain Res., 13*(3), 393–405. https://doi.org/10.1016/S0926-6410(01)00131-8

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon. Bull. Rev., 25*(6), 2005-2015. https://doi.org/10.3758/s13423-018-1432-y

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Commun., 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. *J. Acoust. Soc. Am.*, *121*(1), 519-526. https://doi.org/10.1121/1.2400666

Van Engen, K. J., & McLaughlin, D, J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hear. Res.*, *369*, 56-66. https://doi.org/10.1016/j.heares.2018.04.013

van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: FMRI and ERP support for psycholinguistic models. *Brain Res. Rev.*, *64*(1), 104–122. https://doi.org/10.1016/j.brainresrev.2010.03.002

Van Wijngaarden, S. J., Teeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *J. Acoust. Soc. Am.*, *111*(4), 1906-1916. https://doi.org/10.1121/1.1456928

Versfeld, N. J., Lie, S., Kramer, S. E., & Zekveld, A. A. (2021). Informational masking with speech-on-speech intelligibility: Pupil response and time-course of learning. *J. Acoust. Soc. Am.*, *149*(4), 2353-2366. https://doi.org/10.1121/10.0003952

von Hapsburg, D., & Peña, E. D. (2002). Understanding bilingualism and its impact on speech audiometry. *J. Speech Lang. Hear. Res.*, *45*(1), 202-213. https://doi.org/10.1044/1092-4388

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018b). Relations between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During s Speech Perception in Noise Task. *Ear Hear.*, *39*(3), 573-582. https://doi.org/10.1097/AUD.0000000000000512

Wang, Y., Zekveld, A. A., Wendt, D., Lunner, T., Naylor, G., & Kramer, S. E. (2018a). Pupil light reflex evoked by light-emitting diode and computer screen: Methodology and association with need for recovery in daily life. *PLoS One, 13*(6), e0197739. https://doi.org/10.1371/journal.pone.0197739

Warzybok, A., Brand, T., Wagener, K. C., & Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *Int. J. Audiol., 54*, 88-99. https://doi.org/10.3109/14992027.2015.1063715

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ration on the pupillary response while performing a speech-in-noise test. *Hear. Res., 369*, 67-78. https://doi.org/10.1016/j.heares.2018.05.006 0378-5955/

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen., 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Winn, M. B. & Teece, K. H. (2021). Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends Hear., 25*, 23312165211027688. https://doi.org/10.1177/23312165211027688

Winn, M. B. & Teece, K. H. (2022). Effortful Listening Despite Correct Responses: The Cost of Mental Repair in Sentence Recognition by Listeners With Cochlear Implants. *J. Speech Lang. Hear. Res., 65*(10), 3966-3980. https://doi.org/10.1044/2022_JSLHR-21-00631

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for

Those Who Want to Get Started. *Trends Hear.*, *22*, https://doi.org/10.1177/2331216518800869

Winn, W. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear. Hear.*, *36*(4), c153-c165. https://doi.org/10.1097/AUD.0000000000000145

Wisniewksi, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., & Sullivan, S. C. (2015). Frontal midline θ power as an index of listening effort. *NeuroReport*, *26*(2), 94-99. https://doi.org/ 10.1097/WNR.0000000000000306

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Atten. Percept. Psychophys.*, *79*(7), 2064-2072. https://10.3758/s13414-017-1361-2

Wu, Y. H., Stangl, E., Zhang, X., Perkings, J., & Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear Hear.*, *37*(6), 660-670. https://doi.org/10.1097/aud.0000000000000335

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends Hear.*, *22*, 1-25. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear.*, *31*(4), 480-490. https://doi.org/10.1097/aud.0b013e3181d4f251