Differences between listeners with early and late immersion age in spatial release from masking in various acoustic environments

C.T. Justine Hui[a], Yusuke Hioka[a,*], Hinako Masuda[b], Catherine I. Watson[c]

[a]*Acoustics Research Centre, Department of Mechanical and Mechatronics Engineering, University of Auckland, Auckland 1142 New Zealand*
[b]*Faculty of Science and Technology, Seikei University, Tokyo, 180-8633 Japan*
[c]*Department of Electrical, Computer, and Software Engineering, University of Auckland, Auckland 1142 New Zealand*

Abstract

It is well-known that we benefit from binaural hearing when listening to the speech of interest amongst noises, where spatial cues may release us from masking. However, this benefit deteriorates with external factors such as the reverberation in the room, as well as internal factors such as our familiarity with the language of interest. The current study examined spatial release from masking (SRM) experienced by listeners with different age of immersion to New Zealand English (NZE) in varying room acoustics. We conducted a speech intelligibility test using an Ambisonic-based sound reproduction system to reproduce speech and noise as if they were produced in a seminar room and a chapel at two distances between the source and the listener: 2 m and 5 m. The rooms differed in reverberation time (RT), and the distances modified the speech clarity (C50) index. The participants were split into an early immersed group (n = 20), and a late immersed group (n = 37), where the participants in the early immersed group were immersed in NZE before the age of 13, and those in the late immersed group were immersed after the age of 15. A babble noise was played from eight azimuthal angles (0, ±45°, ±90°, ±135°, 180°) while the target speech, which was sentences from the BKB corpus, was played from 0°. We found the listeners who were immersed early could identify speech better than listeners who were immersed late within most of the room acoustics tested. However, once the room acoustics cause too adverse listening conditions at a high RT and low C50, neither group could benefit from SRM. The early immersed group was also able to make use of spatial cues to benefit from SRM more than the late immersed group, even in the least reverberant room scenario in the current study. Finally, we found that while room acoustics affected how the groups benefitted from SRM, this effect was only observed when the source was located 5 m from the listener.

Keywords: speech intelligibility, age of immersion, non-native, spatial release from masking, room acoustics

1. Introduction

Speech communication in real life often occurs in noisy environments. Most of the time, we are able to process speech signals mixed with noise when perceived by our ears, picking out the speech we want to listen to and ignore parts of the signal that we do not need. Of course, the "we" here mostly refers to normal hearing native listener of the target language (Loizou, 2013; Bronkhorst, 2015). Many studies have shown how the level of exposure and familiarity to the target language affects our ability to understand speech in noise (Lecumberri et al., 2010; Scharenborg and van Os, 2019). This process of segregating speech in presence of interfering noise is coined as the *cocktail party* phenomenon by Cherry, which essentially describes "how we understand what one person is saying when others are speaking at the same time" (Cherry, 1953; Bronkhorst, 2000; Loizou, 2013). When we are able to pick out the speech of interest amongst competing noise, we are said to be *released* from the masking (of speech), where masking refers to the noise covering the speech of interest (Levitt and Rabiner, 1967; Carhart et al., 1969; Loizou, 2013). In a monoaural hearing scenario, this is mostly due to some temporal and spectrotemporal fluctuations in the masker manifested as *gaps*, allowing the listener to hear out or *glimpse* into the speech signal of interest (Miller and Licklider, 1950;

*Corresponding author
Email addresses: justine.hui@auckland.ac.nz (C.T. Justine Hui), yusuke.hioka@ieee.org (Yusuke Hioka)

Howard-Jones and Rosen, 1993; Cooke, 2006; Gnansia et al., 2008; Vestergaard et al., 2011).

In the real world, we would typically be using both ears for hearing sounds, known as binaural hearing, which gives us an additional advantage for listening to speech in noise (Pulkki and Karjalainen, 2015). As each ear receives slightly different signals due to the sound wave reaching each of the ears at different amplitudes and times characterised as interaural level difference and interaural time difference, respectively, we can acquire spatial cues of the environment by binaural hearing (Hioka et al., 2008). When speech is masked by noise, through binaural hearing we can make use of the spatial separation between the target speech and the competing noise (Middlebrooks et al., 2017). In other words, listening with both ears is what enables us to hear the target speech better when the target speech and competing noise are projected from different locations. Conversely, when target speech and competing noise are projected from the same location, we cannot benefit from binaural hearing (i.e. target speech becomes more difficult to hear). This binaural benefit is referred to as *spatial release from masking* (SRM) (Bronkhorst and Plomp, 1988; Litovsky, 2012; Bronkhorst, 2015). SRM can be attributed to two components, head shadow, where there is a *better ear*, which receives the target signal with a higher signal-to-noise ratio (SNR) due to the head blocking the noise, and the interaural level and time differences between the signals entering the two ears (Kidd et al., 1998; Bronkhorst, 2000; Freyman et al., 2001; Culling et al., 2004; Litovsky, 2005; Glyde et al., 2013).

Spatial acoustics of the environment, such as amount of reverberation and spatial arrangement of competing sound sources, have been shown to influence how well listeners can benefit from SRM and segregate competing signals (Culling et al., 2003; Litovsky, 2005; Marrone et al., 2008; Lavandier and Culling, 2008). The benefit from SRM decreases when the environment is reverberant, causing the sound image to be diffused and the noise sources are distributed around the head. Speech intelligibility deteriorates from the reduction in the contribution of head shadow and the lack of temporal fluctuations in amplitude of the noise sources which prevent the listeners from glimpsing (Brown and Bacon, 2010; Vestergaard et al., 2011), resulting in a decrease in benefit from SRM (Bronkhorst and Plomp, 1992). To examine the effect of reverberation, Bronkhorst and Plomp (1992) used multiple masker sources distributed around the listeners in an anechoic environment, while Kidd et al. (2005) manipulated reverberation time of the stimuli, where both found a reduction in benefit from SRM. However, there has been few studies that investigated SRM in terms of varying acoustic environments as a whole, likened to how communication would occur in real life, possibly due to the lack of reproducibility and feasibility of physically conducting perceptual experiments in rooms with various types of acoustics.

While the above studies on how detrimental reverberation and noise are to speech intelligibility are limited to native listeners, it is evident that non-native listeners would experience much greater disadvantages in their speech perception. Previous studies found that non-native listeners perform poorer than native listeners from low-level speech perception such as ability to discriminate and identify phonetic contrasts, to spoken word recognition and understanding unfamiliar speech (e.g. Flege, 1993; Bradlow et al., 1999; Best et al., 2001; Watson et al., 2013; Osawa et al., 2018; Hui and Arai, 2020). While non-native listeners' perception of speech in noise has been investigated thoroughly (Lecumberri et al., 2010), few has looked into how well non-native listeners can make use of spatial cues to listen to speech in adverse acoustic environments. Ezzatian et al. (2010) examined whether non-native listeners would struggle more to use spatial cues to their advantage using stimuli without noise or reverberation and they found that both native and non-native listeners benefited from SRM equally. Building from this, the current study examines whether there is any difference in how listeners of different language experiences make use of spatial cues when we introduce varying acoustic environments.

One of the most important factors that determines a listener's nuanced use of language is the age of language exposure or immersion (Munro and Mann, 2005; Ben-David et al., 2016; Gordon-Salant et al., 2019). While debates are still ongoing, theories have suggested that when the onset age of immersion to the language falls beyond a *critical period*, the learner may never achieve native-likeness in both speech perception and production (Lenneberg, 1967; Flege, 1995; Mayo et al., 1997; Abrahamsson and Hyltenstam, 2009). Even when a fluent non-native speaker, who has acquired the language at a later age, can understand speech in quiet and optimal acoustic environment similarly to a listener who learnt the language at birth, studies have found that the former would perform poorer in acoustically adverse environments such as noisy and reverberant spaces (van Wijngaarden et al., 2002; Rogers et al., 2006; Cooke et al., 2008; Lecumberri et al., 2010; Scharenborg and van Os, 2019). Lenneberg (1967)'s notion of a critical period for language acquisition marks puberty (age of immersion $\sim$ 12) as the cutoff for attaining a native degree of perceived accent (Munro and Mann, 2005). This

break down of the age of immersion is also in line with the immersion studies in Canada and Ireland, where typically late immersion programme starts at the end of primary school and start of secondary school between age 12 - 13 (MacIntyre et al., 2003; Ó Muircheartaigh and Hickey, 2008).

The aim of the current study is to investigate how varying room acoustics affect listeners of different immersion age in their use of spatial cues to understand speech in noise. There are three factors in the current paper: room acoustics, SRM, and age of immersion. Room acoustics conditions were chosen to differ in both reverberation time (RT) and the distance between source and the listeners, which in turn, changes the clarity (C50). RT and C50 are key acoustical metrics of a room for examining speech intelligibility (Pulkki and Karjalainen, 2015). RT is the time it takes for the energy of sound in a room to decay by -60 dB (Zuckerwar, 2003), and C50 is an objective quantification of speech intelligibility that is derived by the energy ratio between direct and reverberant components of a measured room impulse response (Pulkki and Karjalainen, 2015). Due to the ambiguity in the definition of native speakers within the multi-cultural demographics of Auckland, New Zealand (Ross et al., 2021; Meyerhoff et al., 2020), where the study was conducted, we have chosen to examine the differences between listeners who were immersed in New Zealand English (NZE) early in life (before age of 13) and later in life (after age of 15), where listeners were considered to be immersed in NZE at the age they moved to New Zealand (Munro and Mann, 2005).

Specifically, we aim to answer the following research questions:

- How does immersion age affect speech intelligibility in different room acoustics?
- How are listeners of different immersion age affected by room acoustics in terms of benefit from SRM?

We hypothesised: a) more adverse room acoustics would cause a detrimental effect to both groups' benefit from SRM; b) listeners who were immersed in New Zealand English (NZE) later in life would perform worse than their early immersed counterparts within the varying room acoustics; c) the late immersed group would be able to benefit from SRM, but this benefit would diminish with more adverse room acoustics; d) the early immersed group would be able to benefit from SRM and would be less affected by the room acoustics than the late immersed group.

The current study realises rooms with varying acoustics using an Ambisonics-based sound reproduction system, wherein experiment was conducted to examine listeners' benefit from SRM. Ambisonics utilises the concept of spherical harmonics to reproduce three-dimensional sound fields (Zotter and Frank, 2019). Speech intelligibility in realistic acoustic environments using Ambisonics-based sound reproduction has been studied for hearing impaired listeners and hearing aid testing (Marschall, 2014; Cubick and Dau, 2016; Mansour et al., 2019; Badajoz-Davila et al., 2020), where Ambisonic sound reproduction is viewed as a valuable tool to conduct behavioural tests in realistic acoustic environments (Marschall, 2014; Cubick et al., 2018; Ahrens et al., 2017, 2019; Dagan et al., 2019). Using spherical harmonics of different orders, Dagan et al. (2019) found that SRM can be observed under as low as first order Ambisonics. The current study utilises a higher order Ambisonics based sound reproduction system. Using higher order Ambisonics allows for a more refined sound reproduction through the use of both real and complex spherical harmonics (Poletti, 2009) compared to first order Ambisonics (Hui et al., 2020a; Au et al., 2021).

2. Methodology

A perceptual experiment was carried out to examine the benefit from spatial release from masking (SRM) in varying room acoustics in terms of speech intelligibility. This section outlines the details of the experimental design. The study has been approved by University of Auckland Human Participants Ethics Committee.

2.1. Participants

Fifty seven participants (mean age = 31.4, sd = 7.9), 31 of whom were identified as female; 26 as male; none as gender diverse) participated in the current study. All participants lived in New Zealand at the time of the experiment and were exposed to NZE daily. Those that immigrated to New Zealand before the age of 13 are considered in the current study as the *early immersed group* (n = 20) and those that arrived after the age of 15 are considered as the *late immersed group* (n = 37). While all participants spoke NZE daily as they worked and studied in NZE environments, there were participants in both groups who spoke either another language or a mixture of English and another language daily at home. Seven of the 20 participants in the early immersed group were monolingual speakers of NZE. In the early immersed group, the number of years the participants had lived in New Zealand are: 5 - 10 years

(n = 3), 10 years and more (n = 17), and languages spoken at home other than English include: Korean, Mandarin, Cantonese, Japanese, and Malay. In the late immersed group, the number of years the participants had lived in New Zealand was: 1 – 2 years (n = 3), 2 – 5 years (n = 17), 5 – 10 years (n = 9), 10 years and longer (n = 8), and their first languages were: Japanese, Mandarin, Cantonese, Russian, Greek, Italian, German, Thai, Filipino, Urdu, Hindi, Malayalam, and Marathi. Only those that were self reported to have no diagnosed hearing impairment proceeded to the experiment. Participants were given a NZ$20 monetary *koha* (a New Zealand Māori custom which can be translated as gift, present, offering, donation or contribution) to thank them for their participation.

2.2. Stimuli

The stimuli used in the test consisted of target speech and noise, which were projected simultaneously to evaluate the effect of the noise on the intelligibility of the speech. Four parameters were involved in generating the stimuli: speech utterances, noise type, speech-noise separation (i.e. spatial (angular) separation between the speech and noise sources), and room acoustics.

2.2.1. Speech utterances

The Bamford-Kowal-Bench (BKB) sentence lists (Bench et al., 1979) from the Speech Perception Assessments New Zealand (SPANZ) corpus were used as the speech stimuli (Kim and Purdy, 2015). The BKB sentences have been chosen in previous studies for assessing speech intelligibility of non-native listeners due to the corpus' limited vocabulary and simple syntactic structures (Cañete and Purdy, 2015; Engen, 2010; Calandruccio et al., 2019; Bradlow and Bent, 2002). The SPANZ corpus was created to be used for New Zealand English (NZE) speakers and was chosen for the current study due to the focus on immersion into NZE. Created for hearing assessments for NZ patients, words that are deemed unfamiliar for NZE speakers in the corpus have been modified from the original British English based sentences to better suit the New Zealand population (Cañete and Purdy, 2015; Kim and Purdy, 2015).

2.2.2. Noise type and level

The babble noise from the NOISEX-92 corpus was used in the current study, consisting of 100 people speaking in a canteen (Varga and Steeneken, 1993). A target-masker-ratio (TMR) of −3 dB was used in this study in keeping with Jiang et al. (2012) and recent studies (Hioka et al., 2016, 2020; Masuda et al., 2019; Au et al., 2021). The speech stimuli (target) was played at 50 dBA (± 1 dBA), with the masking noise played at 53 dBA, measured at where the participant was seated (see Section 2.3.2). Calibration of noise level was performed through the use of a free field microphone (GRAS 46AE) as per the procedure outlined in Au et al. (2021).

2.2.3. Spatial separation

The position of the noise source varied among eight azimuthal angles (the angles on a horizonal plane) : 0°, ±45°, ±90°, ±135° and 180° while the angle of the target speech was fixed at 0° to examine the effect of SRM. Elevation angles of both the target speech and noise were fixed at 0° (i.e. the plane level with the height of participant's ears).

2.2.4. Room acoustics

To investigate the effect of acoustical properties of rooms on speech intelligibility, two rooms (seminar room and chapel) at two distances (2 m and 5 m) between listener and sound source were tested in the experiment. The seminar room was the room # 405-430 at the University of Auckland typically used for tutorials and seminars and the chapel was the Maclaurin Chapel also at the University of Auckland. As shown in Table 1, the two rooms differ in reverberation time (RT), and the two distances in each room alters the speech clarity (C50), giving a total of four different acoustical conditions. RT and C50 are key metrics used to characterise the acoustics of rooms, both of which are known to be relevant to speech intelligibility (Pulkki and Karjalainen, 2015). The values in Table 1 were calculated using room impulse responses measured by an omnidirectional microphone (miniDSP UMIK-1) with the acoustic measurement software (REW Room Acoustics) (Mulcahy, 2021). T20 measure (Pulkki and Karjalainen, 2015) was used to calculate the RT values. As can be seen in Table 1, the seminar room has a shorter RT than the chapel. The C50 on the other hand varies depending on the distance.

Apart from the four types of room acoustics conditions, an anechoic case was also examined to collect baseline performance without the effect of room acoustics, which was realised by projecting the sound sources (target speech and noise) directly from the loudspeakers on the middle ring (i.e. 0° elevation) of the 16-channel loudspeaker array (see Section 2.3.1).

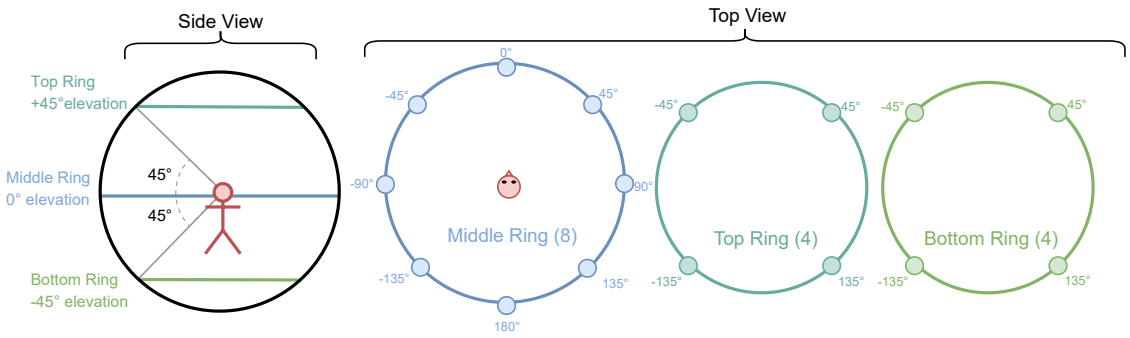

Figure 1: Side and top views of the 16-channel loudspeaker array used in the experiment. Shaded circles denote the loudspeaker placement. (Taken from Au et al. (2021))

Table 1: Acoustical properties of the measured rooms

Room	RT (s)	C50 (dB)	
		2 m	5 m
Seminar room	0.7	15.9	6.8
Chapel	1.8	10.3	3.3

2.3. *Procedure*

2.3.1. *Testing environment*

Testing was performed using a 16-channel loudspeaker array installed in the anechoic chamber at the University of Auckland where the chamber measured a negligible reverberation time of 0.04 s. The loudspeaker array specifications shown in Figure 1 can be found in Au et al. (2021). A monitor was installed below the 0° azimuth loudspeaker on the middle ring and controlled by a wireless keyboard for participants to enter their answer through a graphical user interface (GUI) similar to that used in Au et al. (2021).

2.3.2. *Pre-test preparation*

Upon arrival, participants were presented with an information sheet outlining the testing procedure. Participants were then seated in a chair placed in the middle of the loudspeaker array, with adjustments made to the height of the chair ensuring that their ears were level with the middle ring loudspeakers and parallel to the ±90° loudspeakers using a laser pointer.

2.3.3. *Test Format*

Participants were required to transcribe speech sentences that were played simultaneously with noise via the GUI using a keyboard. A practice test of five sentences that were not included in the main test was used to ensure participants were familiar with the testing procedures. Participants proceeded to the main test once they were confident with the procedures.

The test involved the stimuli discussed in Section 2.2 where each combination of the parameters included four repetitions. Thus, participants were required to transcribe a total of 160 masked sentences (5 room acoustics × 8 speech-noise separation × 4 repetitions). The participants were given a break after listening to 80 sentences. The test took roughly on average 35 - 40 minutes.

2.3.4. *Ambisonics-based sound reproduction system (SRS)*

The effect of room acoustics stated in Section 2.2.4 was collected by measuring the room impulse responses (RIR) of the actual rooms using an Ambisonics microphone array (MH Acoustics Eigenmike). The room impulse responses were measured by playing a swept sine signal at 48 kHz sampling frequency using a modified version of the ScanIR application (Boren and Roginska, 2011; Vanasse et al., 2019) and recorded using the Eigenmike. The swept sine signal was played over a loudspeaker (Genelec 8020D) connected to two audio interfaces (Eigenmike Microphone Interface Box and RME MADIface). The loudspeaker was placed at eight azimuthal angles mentioned in Section 2.2.3 to capture the 3-dimensional spatial sound effect of a single speaker facing the listener at different points in space. The cone of the loudspeaker and the centre of the Ambisonics microphone array were set at a height of 1.51 m from the floor.

The higher-order spatial impulse response rendering (HO-SIRR) toolbox (McCormack et al., 2020) in MATLAB was used to generate the stimuli. The 32-channel room impulse responses measured by the Eignemike were firstly encoded into third order spherical harmonic format and then decoded using the layout of the loudspeaker array. The encoded/decoded room impulse responses are convolved with the speech and the noise stimuli specified in Sections 2.2.1 and 2.2.2, respec-

tively.

The rendered files were played back from the 16-channel loudspeaker array connected to an audio interface (MOTU 16A) at 48 kHz sampling frequency using a digital audio workspace (Cockos Reaper). A more detailed description of the sound reproduction sound system used in the current study including its sound reproduction performance can be found in Hui et al. (2021).

2.4. Marking rubric

To quantify speech intelligibility, scoring was performed manually according to the recommendations set out in the SPANZ corpus (Kim and Purdy, 2015), where the root of the word is scored as opposed to the whole word. For example, the word "run" would be scored similarly to the word "running" or "ran". Homonyms (e.g. meat/meet, sun/son) and words that include the New Zealand English vowel merger (/iə/ and /eə/ (Maclagan and Gordon, 1996) creating homonyms such as ear/air) were not penalised. Each sentence from the BKB corpus consists of 3 - 4 keywords to be marked, and each correct keyword identified was given one mark. The marks were then normalised to proportion correct within one condition per participant.

2.5. Statistical analysis

The speech intelligibility scores in terms of the proportion correct were analysed using a linear mixed effect model with the R (R Core Team, 2015) package *lme4* (Bates et al., 2015) and model fitting was carried out using the step function from *lmerTest* (Kuznetsova et al., 2017). Interactions between two and more factors were included when it improved the fitness of the model. Post-hoc pairwise comparisons of the models were carried out using the *emmeans* package (Lenth, 2019) with *p*-values adjusted using the Tukey method.

3. Results

A model with a three-way interaction between the speech-noise separation, the room acoustics, and age of immersion (whether the listener was considered as early or late immersed NZE speaker) were used to analysed the results. For the random effect, the participant ID was included. Random slopes were excluded due to singular fit. We found a significant three-way interaction from the model analysis using a likelihood ratio comparison ($\chi^2(28) = 56.48$, $p = 0.001$).

Figure 2 shows the predicted probabilities of proportion correct for the five room acoustics conditions across the eight speech-noise separations, separated according

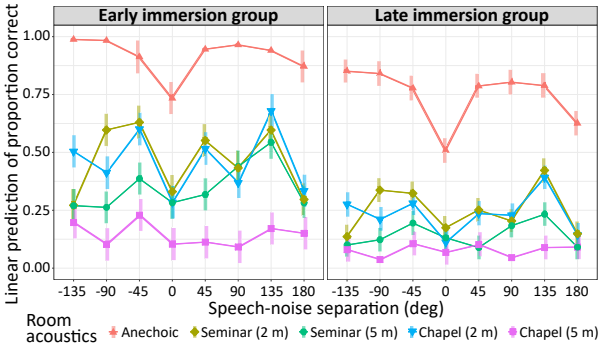

Figure 2: Linear prediction of proportion correct scores from the linear mixed model in terms of room acoustics across speech-noise separation separated by age of immersion grouping.

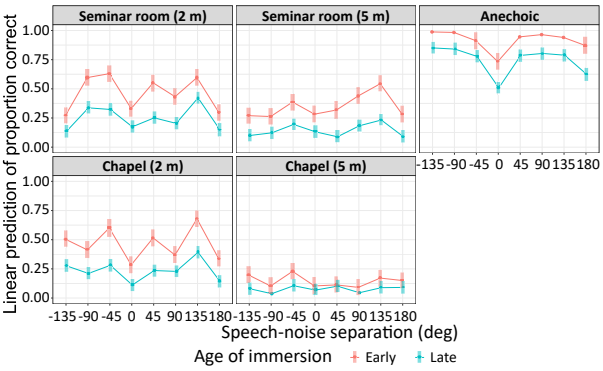

Figure 3: Linear prediction of proportion correct scores from the linear mixed model in terms of age of immersion grouping across speech-noise separation separated by the room acoustics.

to the participants' age of immersion grouping (early vs late). The colours of the plots denote the different room acoustics. Similarly, Figure 3 shows the predicted probabilities of proportion correct for the early and late immersed groups across the eight speech-noise separations, separated according to the five room acoustics conditions. The colours of the plot denote the participants' grouping in terms of their immersion age. The bars at each respective point display the 95% confidence interval. Table 2 consists of the post-hoc pairwise contrasts between early and late immersed groups in terms of the room acoustics and the speech-noise separation. Tables 3 and 4 display the post-hoc pairwise contrasts of the significant interactions between each speech-noise separation in terms of the room acoustics for the early and late immersed groups, respectively.

The anechoic case in Figures 2 and 3 shows the baseline performance of the participants when listening to speech in noise under an acoustic environment without any reverberation. The *dip* at 0° compared to the other

angles where the position of the masker source was spatially separated from that of the target speech indicates that the participants benefitted from spatial release from masking (SRM, i.e. the binaural hearing benefit where target speech can be better separated from the competing masker). The lower speech intelligibility scores at 180° for both groups could be explained by the effect of front-back confusion (Wightman and Kistler, 1999; Rychtáriková et al., 2011; Litovsky, 2012). The presence of this 'dip' and its depth will be used to indicate whether the participants could benefit from SRM and to what extent the benefit was in the varying room conditions. While both groups benefitted from SRM in the anechoic case, the late immersed group performed significantly worse than the early immersed group at all speech-noise separation.

3.1. Effect of room acoustics on speech intelligibility

Illustrated in Figure 2, adding the effect of reverberant room acoustics to the speech stimuli decreased both groups' performances overall, with significant differences at all angles between the anechoic case and the room conditions within the early and late immersed groups, separately. The groups yielded similar patterns in terms of how the room acoustics affected their speech intelligibility results within their respective groups. The seminar room (2 m) and chapel (2 m) exhibited similar performances within the groups, followed by a reduction in performance for the seminar room (5 m), and finally both groups performed the most poorly within themselves for the chapel (5 m) condition.

3.2. Difference between early and late immersed groups

The differences between the two groups with different immersion age can be observed in Figure 3. Between the early and late immersed groups, the late immersed group performed significantly worse than the early immersed group in the anechoic, seminar room (2 m, 5 m) and chapel (2 m). In the chapel (5 m), there were no significant differences between the groups apart from speech-noise separation at -45° and -135°, confirmed in Table 2. In the seminar room (2 m) and chapel (2 m), we can observe the largest differences between the two groups to be at angles where the early immersed group could benefit from SRM (e.g. ±45°), and smallest differences at angles of separation where the groups could not benefit from SRM (e.g. 0°, 180°). This trend was less distinct in the seminar room (5 m). As the environment became more adverse (reverberant) in the chapel (5 m), the differences between the two groups reduced, where they only differed at angles -135° and -45°.

3.3. Comparison between early and late immersed groups in terms of benefit from SRM

For both groups, there is a clear pattern of SRM where listeners performed higher in terms of speech intelligibility when the speech and noise sources were separated spatially compared to when the speech and noise sources were colocated at 0°. This can be observed in Figure 2 and Figure 3 where there is a dip at 0° for the anechoic, seminar room (2 m), chapel (2 m) and marginally for seminar room (5 m). While both groups could benefit from SRM at least in the conditions with higher C50 (seminar room (2 m) and chapel (2 m), seminar room (5 m)), they did not benefit from SRM in the chapel (5 m), where no speech-noise separation contrasts were significant regardless of their age of immersion as shown in Table 3 and Table 4. For the early immersed group, at seminar room (2 m) and chapel (2 m), the dip in intelligibility scores at 0° compared to the other angles exhibits similar depth to the anechoic case, but the depth diminished as the environment became more adverse in terms of clarity, e.g. seminar room (5 m), to completely no evidence of dip (i.e. SRM) at chapel (5 m). For the late immersed group, the dip in intelligibility scores compared to the other angles was shallower than the early immersed group even at the least adverse (highest clarity) environment of seminar room (2 m). Comparing the two groups in terms of the increase in intelligibility scores when the noise was separated from the speech by 45°, early immersed group had a difference range of 0.21 - 0.32 scores at the 2 m conditions, compared to the late immersed group, where they had a difference range of 0.08 - 0.17.

4. Discussion

The significant three-way interaction between the speech-noise separation, the room acoustics and the age of immersion suggests that the relationship between our language experiences and our abilities to use spatial cues when understanding speech in different room acoustics is complex. This section discusses the interactions between these three factors in detail, specifically, in terms of spatial release from masking (SRM) and how listeners could benefit from binaural listening to hear out the target speech amidst the competing masker.

4.1. Effect of room acoustics on benefit from SRM

The different room acoustics in terms of varying reverberation times (RT) and speech clarity (C50) affected both groups in how much benefit from SRM they would gain when listening to speech. While the two rooms

Table 2: Pairwise contrasts of speech intelligibility proportion correct scores between early and late immersed groups in terms of room acoustics

	Anechoic			Seminar room (2 m)			Seminar room (5 m)			Chapel (2 m)			Chapel (5 m)		
angle	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value
-135	0.137	3.153	**0.002**	0.136	3.113	**0.002**	0.170	3.904	**<.001**	0.229	5.248	**<.0001**	0.117	2.688	**0.007**
-90	0.143	3.279	**0.001**	0.260	5.968	**<.0001**	0.140	3.213	**0.001**	0.202	4.635	**<.0001**	0.066	1.511	0.131
-45	0.135	3.093	**0.002**	0.307	7.049	**<.0001**	0.192	4.408	**<.0001**	0.320	7.358	**<.0001**	0.123	2.829	**0.005**
0	0.225	5.172	**<.0001**	0.156	3.585	**<.001**	0.152	3.486	**0.001**	0.172	3.958	**<.001**	0.036	0.817	0.414
45	0.159	3.649	**<.001**	0.300	6.901	**<.0001**	0.229	5.266	**<.0001**	0.280	6.429	**<.0001**	0.011	0.249	0.803
90	0.161	3.706	**<.001**	0.229	5.254	**<.0001**	0.256	5.880	**<.0001**	0.140	3.214	**0.001**	0.046	1.055	0.292
135	0.152	3.490	**0.001**	0.174	4.000	**<.001**	0.311	7.139	**<.0001**	0.288	6.606	**<.0001**	0.082	1.886	0.060
180	0.246	5.639	**<.0001**	0.148	3.409	**0.001**	0.192	4.402	**<.0001**	0.191	4.387	**<.0001**	0.059	1.363	0.173

Table 3: Pairwise contrasts of speech intelligibility proportion correct scores between speech-noise separation angles in terms of room acoustics for the early immersed group

Early	Anechoic			Seminar room (2 m)			Seminar room (5 m)			Chapel (2 m)			Chapel (5 m)		
contrast	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value
(-135) - (-90)	0.004	0.099	1.000	-0.325	-7.183	**<.0001**	0.007	0.166	1.000	0.092	2.025	0.465	0.094	2.075	0.431
(-135) - (-45)	0.075	1.658	0.715	-0.358	-7.917	**<.0001**	-0.117	-2.575	0.165	-0.097	-2.133	0.394	-0.032	-0.701	0.997
(-135) - 0	0.254	5.617	**<.0001**	-0.059	-1.299	0.900	-0.013	-0.285	1.000	0.220	4.865	**<.0001**	0.093	2.058	0.443
(-135) - 45	0.042	0.936	0.983	-0.280	-6.185	**<.0001**	-0.048	-1.067	0.964	-0.011	-0.251	1.000	0.084	1.863	0.577
(-135) - 90	0.023	0.513	1.000	-0.161	-3.558	0.009	-0.169	-3.744	**0.005**	0.136	2.999	0.055	0.105	2.330	0.278
(-135) - 135	0.047	1.048	0.967	-0.325	-7.177	**<.0001**	-0.273	-6.042	**<.0001**	-0.175	-3.868	**0.003**	0.026	0.571	0.999
(-135) - 180	0.117	2.578	0.164	-0.025	-0.555	0.999	-0.012	-0.274	1.000	0.170	3.761	**0.004**	0.046	1.027	0.970
(-90) - (-45)	0.071	1.559	0.775	-0.033	-0.734	0.996	-0.124	-2.741	0.111	-0.188	-4.158	**0.001**	-0.126	-2.776	0.101
(-90) - 0	0.250	5.517	**<.0001**	0.266	5.884	**<.0001**	-0.020	-0.450	1.000	0.129	2.840	0.086	-0.001	-0.017	1.000
(-90) - 45	0.038	0.837	0.991	0.045	0.997	0.975	-0.056	-1.232	0.922	-0.103	-2.276	0.308	-0.010	-0.212	1.000
(-90) - 90	0.019	0.414	1.000	0.164	3.625	**0.007**	-0.177	-3.910	**0.002**	0.044	0.975	0.978	0.012	0.255	1.000
(-90) - 135	0.043	0.949	0.981	0.000	0.006	1.000	-0.281	-6.208	**<.0001**	-0.267	-5.893	**<.0001**	-0.068	-1.504	0.805
(-90) - 180	0.112	2.479	0.205	0.300	6.628	**<.0001**	-0.020	-0.439	1.000	0.079	1.737	0.663	-0.047	-1.048	0.967
(-45) - 0	0.179	3.959	**0.002**	0.299	6.618	**<.0001**	0.104	2.291	0.299	0.317	6.998	**<.0001**	0.125	2.759	0.106
(-45) - 45	-0.033	-0.722	0.996	0.078	1.731	0.667	0.068	1.509	0.803	0.085	1.882	0.564	0.116	2.564	0.170
(-45) - 90	-0.052	-1.145	0.947	0.197	4.359	**<.001**	-0.053	-1.169	0.941	0.232	5.132	**<.0001**	0.137	3.031	0.050
(-45) - 135	-0.028	-0.610	0.999	0.033	0.740	0.996	-0.157	-3.466	**0.013**	-0.079	-1.735	0.664	0.058	1.272	0.909
(-45) - 180	0.042	0.920	0.984	0.333	7.362	**<.0001**	0.104	2.302	0.293	0.267	5.894	**<.0001**	0.078	1.728	0.669
0 - 45	-0.212	-4.680	**<.001**	-0.221	-4.886	**<.0001**	-0.035	-0.782	0.994	-0.232	-5.116	**<.0001**	-0.009	-0.195	1.000
0 - 90	-0.231	-5.104	**<.0001**	-0.102	-2.259	0.317	-0.157	-3.459	**0.013**	-0.084	-1.866	0.575	0.012	0.272	1.000
0 - 135	-0.207	-4.568	**<.001**	-0.266	-5.878	**<.0001**	-0.261	-5.757	**<.0001**	-0.395	-8.733	**<.0001**	-0.067	-1.487	0.815
0 - 180	-0.138	-3.038	**0.049**	0.034	0.744	0.996	0.001	0.011	1.000	-0.050	-1.103	0.956	-0.047	-1.031	0.970
45 - 90	-0.019	-0.424	1.000	0.119	2.628	0.147	-0.121	-2.677	0.130	0.147	3.250	**0.026**	0.021	0.467	1.000
45 - 135	0.005	0.112	1.000	-0.045	-0.992	0.976	-0.225	-4.975	**<.0001**	-0.164	-3.617	**0.007**	-0.058	-1.292	0.902
45 - 180	0.074	1.642	0.725	0.255	5.631	**<.0001**	0.036	0.793	0.994	0.182	4.012	**0.002**	-0.038	-0.836	0.991
90 - 135	0.024	0.535	1.000	-0.164	-3.619	**0.007**	-0.104	-2.298	0.295	-0.311	-6.867	**<.0001**	-0.080	-1.759	0.648
90 - 180	0.093	2.065	0.438	0.136	3.003	0.055	0.157	3.471	**0.012**	0.034	0.762	0.995	-0.059	-1.303	0.898
135 - 180	0.069	1.530	0.791	0.300	6.622	**<.0001**	0.261	5.768	**<.0001**	0.345	7.630	**<.0001**	0.021	0.456	1.000

Table 4: Pairwise contrasts of speech intelligibility proportion correct scores between speech-noise separation angles in terms of room acoustics for the late immersed group

Late	Anechoic			Seminar room (2 m)			Seminar room (5 m)			Chapel (2 m)			Chapel (5 m)		
contrast	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value	estimate	t.ratio	p.value
(-135) - (-90)	0.010	0.300	1.000	-0.201	-6.035	**<.0001**	-0.023	-0.680	0.998	0.065	1.951	0.515	0.043	1.283	0.905
(-135) - (-45)	0.072	2.177	0.366	-0.187	-5.619	**<.0001**	-0.095	-2.844	0.085	-0.005	-0.141	1.000	-0.026	-0.770	0.995
(-135) - 0	0.342	10.282	**<.0001**	-0.038	-1.150	0.946	-0.031	-0.935	0.983	0.164	4.929	**<.0001**	0.012	0.352	1.000
(-135) - 45	0.064	1.922	0.536	-0.115	-3.458	**0.013**	0.011	0.330	1.000	0.040	1.203	0.931	-0.022	-0.657	0.998
(-135) - 90	0.047	1.421	0.848	-0.068	-2.038	0.456	-0.083	-2.508	0.192	0.047	1.418	0.849	0.034	1.032	0.970
(-135) - 135	0.062	1.867	0.574	-0.286	-8.601	**<.0001**	-0.133	-3.987	**0.002**	-0.116	-3.485	**0.012**	-0.009	-0.273	1.000
(-135) - 180	0.225	6.760	**<.0001**	-0.012	-0.368	1.000	0.009	0.278	1.000	0.133	3.990	**0.002**	-0.011	-0.336	1.000
(-90) - (-45)	0.062	1.877	0.567	0.014	0.417	1.000	-0.072	-2.164	0.374	-0.070	-2.093	0.420	-0.068	-2.053	0.446
(-90) - 0	0.332	9.982	**<.0001**	0.163	4.886	**<.0001**	-0.008	-0.255	1.000	0.099	2.978	0.059	-0.031	-0.931	0.983
(-90) - 45	0.054	1.622	0.737	0.086	2.578	0.165	0.034	1.010	0.973	-0.025	-0.748	0.996	-0.065	-1.941	0.523
(-90) - 90	0.037	1.121	0.952	0.133	3.997	**0.002**	-0.061	-1.828	0.601	-0.018	-0.533	1.000	-0.008	-0.251	1.000
(-90) - 135	0.052	1.567	0.770	-0.085	-2.566	0.169	-0.110	-3.307	**0.022**	-0.181	-5.436	**<.0001**	-0.052	-1.556	0.776
(-90) - 180	0.215	6.460	**<.0001**	0.189	5.668	**<.0001**	0.032	0.958	0.980	0.068	2.038	0.456	-0.054	-1.619	0.739
(-45) - 0	0.270	8.105	**<.0001**	0.149	4.469	**<.001**	0.064	1.909	0.545	0.169	5.070	**<.0001**	0.037	1.122	0.952
(-45) - 45	-0.008	-0.254	1.000	0.072	2.161	0.376	0.106	3.175	**0.033**	0.045	1.344	0.882	0.004	0.112	1.000
(-45) - 90	-0.025	-0.755	0.995	0.119	3.581	**0.008**	0.011	0.337	1.000	0.052	1.560	0.774	0.060	1.802	0.619
(-45) - 135	-0.010	-0.309	1.000	-0.099	-2.983	0.058	-0.038	-1.142	0.947	-0.111	-3.344	**0.019**	0.017	0.497	1.000
(-45) - 180	0.152	4.583	**<.001**	0.175	5.251	**<.0001**	0.104	3.123	**0.038**	0.137	4.131	**0.001**	0.014	0.434	1.000
0 - 45	-0.278	-8.359	**<.0001**	-0.077	-2.308	0.290	0.042	1.265	0.912	-0.124	-3.726	**0.005**	-0.034	-1.009	0.973
0 - 90	-0.295	-8.860	**<.0001**	-0.030	-0.888	0.987	-0.052	-1.573	0.767	-0.117	-3.511	**0.011**	0.023	0.681	0.998
0 - 135	-0.280	-8.414	**<.0001**	-0.248	-7.452	**<.0001**	-0.102	-3.052	**0.048**	-0.280	-8.414	**<.0001**	-0.021	-0.625	0.999
0 - 180	-0.117	-3.522	0.010	0.026	0.782	0.994	0.040	1.214	0.928	-0.031	-0.939	0.982	-0.023	-0.688	0.997
45 - 90	-0.017	-0.501	1.000	0.047	1.420	0.848	-0.094	-2.838	0.086	0.007	0.215	1.000	0.056	1.690	0.694
45 - 135	-0.002	-0.055	1.000	-0.171	-5.144	**<.0001**	-0.144	-4.317	**<.001**	-0.156	-4.688	**<.001**	0.013	0.384	1.000
45 - 180	0.161	4.837	**<.0001**	0.103	3.090	**0.042**	-0.002	-0.052	1.000	0.093	2.787	0.099	0.011	0.321	1.000
90 - 135	0.015	0.446	1.000	-0.218	-6.563	**<.0001**	-0.049	-1.479	0.819	-0.163	-4.903	**<.0001**	-0.043	-1.306	0.897
90 - 180	0.178	5.339	**<.0001**	0.056	1.671	0.707	0.093	2.786	0.099	0.086	2.571	0.167	-0.046	-1.369	0.871
135 - 180	0.163	4.893	**<.0001**	0.274	8.234	**<.0001**	0.142	4.265	**0.001**	0.249	7.474	**<.0001**	-0.002	-0.063	1.000

(seminar room vs chapel) differed in RT, the C50 also varied by changing the distance between the source and the listener in the respective environments. The seminar room (2 m) and chapel (2 m) showed similar effect on speech intelligibility within the groups, respectively, despite the chapel had an RT that was 2.57 times longer than than that of the seminar room. The listeners' intelligibility scores diminishing in the seminar room (5 m) to almost flooring at chapel (5 m) (3.3 dB) showed that neither clarity nor reverberation time was adequate to predict speech intelligibility by themselves. While C50 has been used as an objective measure of speech intelligibility in general (Bradley et al., 1999; Pulkki and Karjalainen, 2015), a reduction of C50 of 7.9 dB between the seminar room (2 m) and chapel (2 m) in the current study did not affect how the listeners understood the speech apart from at angles -135° and -90°. The results only corroborated our hypothesis of more adverse room acoustics causing an increased detrimental effect to the listeners' benefit from SRM at the two 5 m conditions, seminar room (5 m) and chapel (5 m). Previous studies have shown binaural cues to help in small reverberant rooms with shorter RT (Culling et al., 2003; Palomäki et al., 2004), which we could observe in the seminar room (2 m). However, evidence from previous research could not be used to explain the SRM observed in the chapel (2 m), where RT was much greater than the seminar room. As the current study only examined two distances in two rooms, further studies need to be carried out to examine the relationship between distance, room acoustics and speech intelligibility.

4.2. Effect of immersion age on speech intelligibility in varying room acoustics

The results, supporting our hypothesis, showed the late immersed group performing worse in speech intelligibility than the early immersed group at all conditions apart from the chapel (5 m). This was despite of the speech corpus having been designed with grammar, semantics and vocabulary targeted at children age 8 - 15, where a high intelligibility score of 95% and above was achieved by listeners who came to New Zealand after the age of 20 when no reverberation and noise was added (Hui et al., 2020b). Many second language theories such as PAM-L2, NLM-e, SLM have argued that second language (L2) speakers have less robust phonetic categories than first language speakers (Flege, 1995; Best and Tyler, 2007; Kuhl et al., 2008). Note that the L2 speakers usually referred to in these theories are learners who learnt the language later in life, unlike some of the participants in the current early immersed group from the immigrant diaspora who were immersed in NZE before the age of 13 while speaking another language at home. The overlapping of speech

sounds caused by masking in the reverberant rooms would make the speech contrasts less distinct for the late immersed group, corroborating previous studies on nonnative listeners' identification of speech sounds (Masuda, 2016; Osawa et al., 2018). Without noise and reverberation added, the late immersed group was able to understand the speech stimuli above 95% correct (Hui et al., 2020b). Once noise was added as shown in the anechoic results, the late immersed group was already disadvantaged compared to the early immersed group. This may be that some of the acoustic cues were lost due to the added noise, which in turn created a mismatch with the acquired prototype instances of the particular speech sound. This could have made accurate perception more difficult for listeners in the late immersion group due to insufficient ability to compensate for the mismatch, on top of already being disadvantaged by the challenge of listening to nonnative sounds. Listening to the speech further distorted by the surrounding room acoustics would have made the sound differed even more from the prototype they learnt (Osawa et al., 2021,). Future studies should look into the individual phonemes and how spectrally and temporally they were contaminated by the room acoustics and in turn, how listeners would respond perceptually.

4.3. Differences between age of immersion group in terms of effect of room acoustics on benefit of SRM

We also found that while both groups could benefit more from SRM at less reverberant environments, the early immersed group could benefit from SRM more than the late immersed group, thus providing evidence to our third and fourth hypotheses: the late immersed group would be able to benefit from SRM, but this benefit would diminish with more adverse room acoustics; and that the early immersed group would be able to benefit from SRM and would be less affected by the room acoustics than the late immersed group. This was not due to the late immersed group's inability to use SRM, as the depth of the intelligibility dip at 0° in their anechoic results illustrate that they could benefit from SRM similarly to their early immersed counterparts under an ideal acoustic environment. The anechoic case corroborates Ezzatian et al. (2010)'s study which showed that non-native listeners could also benefit from SRM similarly to the native listeners. However, our study showed that reverberation impacts on this benefit where adding a slight reverberation, e.g. seminar room (2 m), could worsen a late immersed group's ability to use spatial cues. This may be due to the less robust categories the listeners in the late immersed group may have of the language, causing them to be more affected by the reduction in temporal fluctuation of the masking noise, making it more difficult for the listeners to hear out or *glimpse* through the masker. In addition, the advantage from the spatial cues such as interaural level and time differences that the early immersed group could make use of may no longer be enough for the late immersed group to segregate the two signals.

While we had a range of listeners in the late immersed group in terms of the year they arrived in New Zealand and their language background, there was surprisingly less variation in their speech intelligibility scores compared to the early immersed group as can be observed by the confidence interval in Figure 3. This may be because while we set the boundary age between early and late immersed groups at 13, language experiences and language usage such as whether they are bilingual or monolingual, the language they used at home and the language they used at school with their peers may affect a participant's ability to listen to speech in adverse environments. Recruiting more participants in the early immersed group and dividing them into more granular age groups and their language background may provide more insights as to how the effect of age of immersion can affect listeners' benefit from SRM when listening in noisy, reverberant environments.

4.4. Limitations

Finally, limitations of the current study include the use of babble noise from the NOISEX-92 database. The babble noise was recorded in a canteen, which we presume to have non-negligible reverberation. This means that even in the lowest reverberant condition (anechoic), the noise may have artefacts from the original recordings before convolving with the room impulse response. Having said this, we could still observe a clear SRM effect in the anechoic case in both groups, suggesting that the non-negligible reverberation in the dry babble source was acceptable to be used for our study.

In addition, Ambisonic-based sound reproduction system used in the current study relies on the participant to be seated in a *sweet spot* of the loudspeaker array, which is typically not large. As we did not secure the participant's head, the participants may have moved out of the sweet spot depending on their sitting posture and thus not fully immersed in the virtual acoustics, creating possibly a trickle down effect on their speech intelligibility performance. We do not think there was a significant effect from this however due to the relatively small confidence interval in the results and the clear general trend across the participants regardless of them being in the early immersed or late immersed group.

While the speech intelligibility results of the reverberant rooms showed a clear SRM effect, there were some angles that did not follow expected patterns such as -135° for the seminar room, where we expected the proportion correct to be similar to the -90° case. While this may be a product of the specific room acoustics that the room impulse responses were taken from, more work using different room types and conditions need to be carried out to generalise our results further.

5. Conclusions

The current study examined the benefit from spatial release from masking (SRM) listeners with different age of immersion to New Zealand English have in varying room acoustics conditions. We found in general the listeners who were immersed early before the age of 13 to identify speech better than listeners who were immersed late after the age of 15. However, when the acoustics of the room was too adverse with a long reverberation time and low clarity, neither groups could benefit from SRM. The early immersed group was also able to make use of spatial cues to benefit from SRM more than the late immersed group, even when reverberation introduced to the stimuli was the lowest in the current study. While the current study only focused on the difference between listeners who were immersed in New Zealand English before and after puberty, future work can investigate how different age of immersion and language experiences may impact on listeners' ability to use spatial cues to benefit from SRM. Finally, we found that while room acoustics of varying reverberation time and clarity affected how the groups benefit from SRM, this effect was only observed when the source was far from the listener at 5 m and was not observed when the distance was 2 m. More studies should be carried out to examine how distance between the source and listener affects the acoustics of speech and in turn affects speech intelligibility.

Acknowledgement

We would like to thank our participants, Dr. Suzanne Purdy for the use of the SPANZ corpus, the Maclaurin Chapel for letting us use their space to record the room impulse responses, Gian Schmid and James Schmid for their help in building the SRS, Charlene Lo for collecting the data, Douglas Hing for his help in the acoustic measurements. This work was supported by the Faculty of Engineering Research Development Fund at the University of Auckland.

References

Abrahamsson, N., Hyltenstam, K., 2009. Age of onset and native-likeness in a second language: Listener perception versus linguistic scrutiny. Lang. Learn. 59, 928. doi:10.1111/j.1467-9922.2009.00530.x.

Ahrens, A., Marschall, M., Dau, T., 2017. Evaluating a Loudspeaker-Based Virtual Sound Environment using Speech- on-Speech Masking Evaluating a Loudspeaker-Based Virtual Sound Environment using Speech-on- Speech Masking. Fortschritte Der Akust. DAGA 2017 Kiel .

Ahrens, A., Marschall, M., Dau, T., 2019. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. Hear. Res. 377, 307–317. doi:10.1016/j.heares.2019.02.003.

Au, E., Xiao, S., Hui, C.T., Hioka, Y., Masuda, H., Watson, C.I., 2021. Speech intelligibility in noise with varying spatial acoustics under Ambisonics-based sound reproduction system. Appl. Acoust. 174, 107707. doi:10.1016/j.apacoust.2020.107707.

Badajoz-Davila, J., Buchholz, J.M., Van-Hoesel, R., 2020. Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments. J. Acoust. Soc. Am. 147, 3538–3549. doi:10.1121/10.0001259.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting Linear Mixed-Effects Models using lme4. J. Stat. Softw. 67, 1–48. doi:doi:10.18637/jss.v067.i01.

Ben-David, B.M., Avivi-Reich, M., Schneider, B.A., 2016. Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? Hear. Res. 341, 9–18. doi:10.1016/j.heares.2016.07.016.

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

Best, C.T., McRoberts, G.W., Goodell, E., 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. J. Acoust. Soc. Am. 109, 775–794. doi:10.1121/1.1332378.

Best, C.T., Tyler, M.D., 2007. Nonnative and second-language speech perception: Commonalities and complementarities, in: Munro, M.J., Bohn, O.S. (Eds.), Lang. Exp. Second Lang. speech Learn. Honor James Emil Flege. John Benjamins, Amsterdam, pp. 13–34. doi:10.1075/lllt.17.07bes.

Boren, B., Roginska, A., 2011. Multichannel impulse response measurement in Matlab, in: Audio Eng. Soc. Conv., pp. 380–385.

Bradley, J.S., Reich, R., Norcross, S.G., 1999. A just noticeable difference in C50 for speech. Appl. Acoust. 58, 99–108. doi:10.1016/S0003-682X(98)00075-9.

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., Tohkura, Y., 1999. Training Japanese listeners to identify english /r/and /l/: Long-term retention of learning in perception and production. Percept. Psychophys. 61, 977–985. doi:10.3758/BF03206911.

Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Am. 112, 272–284. doi:10.1121/1.1487837.

Bronkhorst, A.W., 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions.

Bronkhorst, A.W., 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Attention, Perception, Psychophys. 77, 1465–1487. doi:10.3758/s13414-015-0882-9.

Bronkhorst, A.W., Plomp, R., 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am. 83, 1508–1516. doi:10.1121/1.395906.

Bronkhorst, A.W., Plomp, R., 1992. Effect of multiple speechlike

maskers on binaural speech recognition in normal and impaired hearing. J. Acoust. Soc. Am. 92, 3132–3139. doi:10.1121/1.404209.

Brown, C.A., Bacon, S.P., 2010. Fundamental frequency and speech intelligibility in background noise. Hear. Res. 266, 52–59. doi:10.1016/j.heares.2009.08.011.

Calandruccio, L., Wasiuk, P.A., Buss, E., Leibold, L.J., Kong, J., Holmes, A., Oleson, J., 2019. The effect of target/masker fundamental frequency contour similarity on masked-speech recognition. J. Acoust. Soc. Am. 146, 1065–1076. doi:10.1121/1.5121314.

Cañete, O.M., Purdy, S.C., 2015. Spatial speech recognition in noise: Normative data for sound field presentation of the New Zealand recording of the Bamford-Kowal-Bench (BKB) sentences and Consonant-Nucleus-Consonant (CNC) monosyllabic words. Bull. New Zeal. Audiol. Soc. .

Carhart, R., Tillman, T.W., Greetis, E.S., 1969. Perceptual Masking in Multiple Sound Backgrounds. J. Acoust. Soc. Am. 45, 694–703. doi:10.1121/1.1911445.

Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.

Cooke, M., 2006. A glimpsing model of speech perception in noise. J. Acoust. Soc. Am. 119, 1562–1573. doi:10.1121/1.2166600.

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, 414–427. doi:10.1121/1.2804952.

Cubick, J., Buchholz, J.M., Best, V., Lavandier, M., Dau, T., 2018. Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners. J. Acoust. Soc. Am. 144, 2896–2905. doi:10.1121/1.5078582.

Cubick, J., Dau, T., 2016. Validation of a virtual sound environment system for testing hearing aids. Acta Acust. united with Acust. 102, 547–557. doi:10.3813/AAA.918972.

Culling, J.F., Hawley, M.L., Litovsky, R.Y., 2004. The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. J. Acoust. Soc. Am. 116, 1057–1065. doi:10.1121/1.1772396.

Culling, J.F., Hodder, K.I., Toh, C.Y., 2003. Effects of reverberation on perceptual segregation of competing voices. J. Acoust. Soc. Am. 114, 2871. doi:10.1121/1.1616922.

Dagan, G., Shabtai, N.R., Rafaely, B., 2019. Spatial release from masking for binaural reproduction of speech in noise with varying spherical harmonics order. Appl. Acoust. 156, 258–261. doi:10.1016/j.apacoust.2019.07.015.

Engen, K.J.V., 2010. Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. Speech Commun. 30, 943–953. doi:10.1038/mp.2011.182.doi.

Ezzatian, P., Avivi, M., Schneider, B.A., 2010. Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? Speech Commun. 52, 919–929. doi:10.1016/j.specom.2010.04.001.

Flege, J.E., 1993. Production and perception of a novel, second-language phonetic contrast. J. Acoust. Soc. Am. 93, 1589–1608. doi:10.1121/1.406818.

Flege, J.E., 1995. Second Language Speech Learning: Theory, Findings, and Problems. Speech Percept. Linguist. Exp. Issues Cross-Language Res. , 233–277doi:10.1111/j.1600-0404.1995.tb01710.x.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. 109, 2112–2122. doi:10.1121/1.1354984.

Glyde, H., Buchholz, J.M., Dillon, H., Cameron, S., Hickson, L., 2013. The importance of interaural time differences and level differences in spatial release from masking. J. Acoust. Soc. Am. 134, EL147–EL152. doi:10.1121/1.4812441.

Gnansia, D., Jourdes, V., Lorenzi, C., 2008. Effect of masker modulation depth on speech masking release. Hear. Res. 239, 60–68. doi:10.1016/j.heares.2008.01.012.

Gordon-Salant, S., Yeni-Komshian, G.H., Bieber, R.E., Jara Ureta, D.A., Freund, M.S., Fitzgibbonsa, P.J., 2019. Effects of listener age and native language experience on recognition of accented and unaccented english words. J. Speech, Lang. Hear. Res. 62, 1131–1143. doi:10.1044/2018_JSLHR-H-ASCC7-18-0122.

Hioka, Y., James, J., Watson, C.I., 2020. Masker design for real-time informational masking with mitigated annoyance. Appl. Acoust. 159, 107073. doi:10.1016/j.apacoust.2019.107073.

Hioka, Y., Okamoto, M., Kobayashi, K., Haneda, Y., Kataoka, A., 2008. A display-mounted high-quality stereo microphone array for high-definition videophone system. Dig. Tech. Pap. - IEEE Int. Conf. Consum. Electron. doi:10.1109/ICCE.2008.4587858.

Hioka, Y., Tang, J.W., Wan, J., 2016. Effect of adding artificial reverberation to speech-like masking sound. Appl. Acoust. 114, 171–178. doi:10.1016/j.apacoust.2016.07.014.

Howard-Jones, P.A., Rosen, S., 1993. Uncomodulated glimpsing in "checkerboard" noise. J. Acoust. Soc. Am. 93, 2915–2922. doi:10.1121/1.405811.

Hui, C.T., Au, E., Xiao, S., Hioka, Y., Watson, C.I., Masuda, H., 2020a. Benefit from spatial release from masking for native and non-native speakers in virtual acoustic environments, in: Internoise2020.

Hui, C.T.J., Arai, T., 2020. Pitch and duration as auditory cues to identify Japanese long vowels for Japanese learners. Acoust. Sci. Technol. 41, 797–799.

Hui, C.T.J., Hioka, Y., Watson, C.I., Masuda, H., 2021. Spatial release from masking in varying spatial acoustic under higher order Ambisonic-based sound reproduction system, in: Internoise2021, Washington DC.

Hui, C.T.J., Masuda, H., Hioka, Y., Watson, C.I., 2020b. Speech intelligibility of New Zealand English (NZE) by native Japanese listeners, in: Linguistic Society of New Zealand Language Society Conference.

Jiang, B., Liebl, A., Leistner, P., Yang, J., 2012. Sound masking performance of time-reversed masker processed from the target speech. Acta Acust. united with Acust. 98, 135–141. doi:10.3813/AAA.918499.

Kidd, G., Mason, C.R., Brughera, A., Hartmann, W.M., 2005. The role of reverberation in release from masking due to spatial separation of sources for speech identification. Acta Acust. united with Acust. 91, 526–536.

Kidd, G., Mason, C.R., Rohtla, T.L., Deliwala, P.S., 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J. Acoust. Soc. Am. 104, 422–431. doi:10.1121/1.423246.

Kim, J.h., Purdy, S.C., 2015. Speech Perception Assessments New Zealand (SPANZ). New Zeal. Audiol. Soc. Bull. 24, 9–16.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. B Biol. Sci. 363, 979–1000. doi:10.1098/rstb.2007.2154.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. 82. doi:10.18637/jss.v082.i13.

Lavandier, M., Culling, J.F., 2008. Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. J. Acoust. Soc. Am. 123, 2237–2248. doi:10.1121/1.2871943.

Lecumberri, M.L.G., Cooke, M., Cutler, A., 2010. Non-native speech

perception in adverse conditions: A review. Speech Commun. 52, 864–886. doi:10.1016/j.specom.2010.08.014.

Lenneberg, E.H., 1967. Biological foundations of language. New York.

Lenth, R., 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means. URL: https://cran.r-project.org/package=emmeans.

Levitt, H., Rabiner, L.R., 1967. Binaural Release From Masking for Speech and Gain in Intelligibility. J. Acoust. Soc. Am. 42, 601–608. doi:10.1121/1.1910629.

Litovsky, R.Y., 2005. Speech intelligibility and spatial release from masking in young children. J. Acoust. Soc. Am. 117, 3091–3099. doi:10.1121/1.1873913.

Litovsky, R.Y., 2012. Spatial release from masking in adults. Acoust. Today , 18–25.

Loizou, P.C., 2013. Speech Enhancement: Theory and Practice. 2nd ed., CRC Press, Boca Raton.

MacIntyre, P.D., Baker, S.C., Clément, R., Donovan, L.A., 2003. Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. Lang. Learn. 53, 137–166. doi:10.1111/1467-9922.00226.

Maclagan, M., Gordon, E., 1996. Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. Lang. Var. Change 8, 125–147.

Mansour, N., Marschall, M., Westermann, A., May, T., Dau, T., 2019. Speech intelligibility in a realistic virtual sound environment. Proc. 23rd Int. Congr. Acoust. Aachen, Ger. , 7623–7630.

Marrone, N., Mason, C.R., Kidd, G., 2008. Tuning in the spatial dimension: Evidence from a masked speech identification task. J. Acoust. Soc. Am. 124, 1146–1158. doi:10.1121/1.2945710.

Marschall, M., 2014. Capturing and reproducing realistic acoustic scenes for hearing research. PhD Thesis - Tech. Univ. Denmark 17.

Masuda, H., 2016. Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments. Speech Commun. 79, 74–87. URL: http://dx.doi.org/10.1016/j.specom.2016.02.007, doi:10.1016/j.specom.2016.02.007.

Masuda, H., Hioka, Y., James, J., Watson, C.I., 2019. Protecting speech privacy from native/non-native listeners - effect of masker type, in: Int. Congr. Phonetic Sci., pp. 3070–3074.

Mayo, L.H., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech, Lang. Hear. Res. 40, 686–693. doi:10.1044/jslhr.4003.686.

McCormack, L., Politis, A., Pulkki, V., Scheureggter, O., 2020. Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution. AES J. Audio Eng. Soc. 68, 248–260. doi:10.17743/JAES.2020.0026.

Meyerhoff, M., Birchfield, A., Ballard, E., Watson, C.I., 2020. Definite Change Taking Place: Determiner Realization in Multiethnic Communities in New Zealand. Univ. Pennsylvania Work. Pap. Linguist. (Selected Pap. from NWAV47) 25, 71–78.

Middlebrooks, J.C., Simon, J.Z., Popper, A.N., Editors, R.R.F., 2017. The auditory system at the cocktail party. volume 60. doi:10.1007/978-3-319-51662-2.

Miller, G.A., Licklider, J.C., 1950. The Intelligibility of Interrupted Speech. J. Acoust. Soc. Am. 22, 167–173. doi:10.1121/1.1906584.

Mulcahy, J., 2021. REW Room Acoustics Software. URL: https://www.roomeqwizard.com/.

Munro, M., Mann, V., 2005. Age of immersion as a predictor of foreign accent. volume 26. doi:10.1121/1.4779726.

Ó Muircheartaigh, J., Hickey, T., 2008. Academic Outcome, Anxiety and Attitudes in Early and Late Immersion in Ireland. Int. J. Biling. Educ. Biling. 11, 558–576. doi:10.1080/13670050802149184.

Osawa, E., Arai, T., Hodoshima, N., 2018. Perception of Japanese consonant-vowel syllables in reverberation: Comparing non-native listeners with native listeners. Acoust. Sci. Technol. 39, 369–378. doi:10.1250/ast.39.369.

Osawa, E., Hui, C.T.J., Hioka, Y., Arai, T., 2021,. Effect of prior exposure on the perception of Japanese vowel length contrast in reverberation for nonnative listeners. Speech Commun. in press.

Palomäki, K.J., Brown, G.J., Wang, D.L., 2004. A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation. Speech Commun. 43, 361–378. doi:10.1016/j.specom.2004.03.005.

Poletti, M., 2009. Unified description of ambisonics using real and complex spherical harmonics. Proc. Ambisonics Symp. , 1–10.

Pulkki, V., Karjalainen, M., 2015. Communication acoustics: an introduction to speech, audio and psychoacoustics. John Wiley and Sons.

R Core Team, 2015. R: A language and environment for statistical computer. URL: https://www.r-project.org/.

Rogers, C.L., Lister, J.J., Febo, D.M., Besing, J.M., Abrams, H.B., 2006. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Appl. Psycholinguist. 27, 465–485. doi:10.1017.S014271640606036X.

Ross, B., Ballard, E., Watson, C., 2021. New Zealand English in Auckland. Asia-Pacific Lang. Var. 7, 62–81. doi:10.1075/aplv.19014.ros.

Rychtáriková, M., Van Den Bogaert, T., Vermeir, G., Wouters, J., 2011. Perceptual validation of virtual room acoustics: Sound localisation and speech understanding. Appl. Acoust. 72, 196–204. doi:10.1016/j.apacoust.2010.11.012.

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. doi:10.1016/j.specom.2019.03.001.

Vanasse, J., Genovese, A., Roginska, A., 2019. Multichannel Impulse Response Measurements in MATLAB: An Update on ScanIR, in: Audio Eng. Soc. Conv., York.

Varga, A., Steeneken, H., 1993. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. Speech Commun. 12, 247–251.

Vestergaard, M.D., Fyson, N.R.C., Patterson, R.D., 2011. The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. J. Acoust. Soc. Am. 130, 429–439. doi:10.1121/1.3596462.

Watson, C., Liu, W., Macdonald, B., 2013. The Effect of Age and Native Speaker Status on Intelligibility, in: 8th ISCA Speech Synth. Work.

Wightman, F.L., Kistler, D.J., 1999. Resolution of front–back ambiguity in spatial hearing by listener and source movement. J. Acoust. Soc. Am. 105, 2841–2853. doi:10.1121/1.426899.

van Wijngaarden, S.J., Steeneken, H.J.M., Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native listeners. J. Acoust. Soc. Am. 111, 1906–1916. doi:10.1121/1.1456928.

Zotter, F., Frank, M., 2019. Ambisonics - A Practical {3D} Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality-. Springer International Publishing.

Zuckerwar, A.J., 2003. Acoustical Measurement, in: Meyers, R.A. (Ed.), Encycl. Phys. Sci. Technol.. Academic Press.