The Test of Speech Sound Perception in Noise (ToSSPiN) – Effect of first language, spatial separation and reverberation on speech sound identification

By

Christian Boyle

Bachelors in Audio (BAudio), Masters in Clinical Audiology (MClinAud)

Faculty of Medicine, Health and Human Sciences

Department of Linguistics

Macquarie University, Sydney, Australia

Thesis submitted in partial fulfillment of the requirements for the degree of:

Master of Research

December 2020

MACQUARIE
University

Table of Contents

Declaration

The work in this Master of Research thesis is my own original work. I have not submitted this thesis for a higher research degree at any other university or institution. The following work reported in this thesis was undertaken during the period of enrolment as a Master of Research student at Macquarie University under the supervision of Dr Sharon Cameron and Professor Harvey Dillon. We obtained ethics approval for this project comprising human participants from Macquarie University's Human Research Ethics Committee (HREC), reference no. 52020632715958. I obtained informed consent from all participants in this study. The study conformed to the National Statement on Ethical Conduct of Human Research (NHMRC 2018).

Although the written component of this thesis is my own work, all aspects of the test development, software development and data extraction were not completed by me. Details relevant to this project that I am not responsible for are:

- Dr Sharon Cameron developed the ToSSPiN specification. Portions of the specification are contained within the methods section.
- The ToSSPiN software was developed by Dr Nicky Chong-White.
- Professor Harvey Dillon was responsible for extracting the data from the ToSSPiN software that was analysed for the results section.
- Professor Harvey Dillon was responsible for analysing the psychometric functions for each phoneme.
- Dr Sharon Cameron and Professor Harvey Dillon were responsible for recording and editing of the ToSSPiN Stimuli, albeit with some assistance from me.

Signed

Christian Boyle

6th December 2020

Acknowledgements

In the bigger picture, a 12-month Master of Research thesis is not the longest, or even the most complex project career researchers will undertake. For me, having no research or academic experience has made this process incredibly challenging and rewarding. This thesis would not be possible without the support and invaluable guidance from Dr Sharon Cameron. You have been patient and helpful. Your fingerprint will be on everything I do that is related to research for as long as I am doing it. Thank you, Sharon.

I would also like to acknowledge the support and guidance of Professor Harvey Dillon. If anyone had told me that I would have the opportunity to work closely with someone as influential and so highly regarded in audiology, I would have said they were crazy. You have been an important part of my development in 2020, you have been unbelievably generous with your time (even on weekends!) and you have been a fantastic mentor. I have many highlights from this year but the one that stands out the most is your mini stats lectures (I'm getting there…). Thank you, Harvey.

I am indebted to the academic staff in the Audiology Department at Macquarie University. To Associate Professor Mridula Sharma, I would not be here without your introduction to Professor Dillon and your advice in undertaking research. To Dr Rebecca Kim, thank you for your constant encouragement (and being a voice of reason). To Dr John Newall, although this may not be important to you, I would not be here without you accepting me into the Master of Audiology. You were kind and patient enough to allow me to be present in your clinic while reacquainting myself with audiology practice, and I am grateful for that opportunity.

To family and friends, although many of you may have questioned why I would pursue further training, I hope that when all is said and done, it will all be worth it. I owe a debt of gratitude for all your love and support.

To my nephew Harvey, I mentioned to you I was a 'scientist' and you asked me to tell you everything I know. I may not know a lot, but hopefully one day you and your brother Harrison can read this and see that I might know something, even if it is small. I dedicate this to both of you.

COVID-19 Impact Statement

Dear Examiner,
Many of our Higher Degree Research candidates have had to make changes to their research due to the impact of COVID-19. Below you will find a statement from the candidate, approved by their Supervisory Panel, that indicates how their original research plan has been affected by COVID-19 restrictions. Relevant ongoing restrictions in place caused by COVID-19 will also be detailed by the candidate.

Candidate's Statement:
"Please briefly describe (250 words or less) how your research ***project*** *direction changed as a result of COVID-19. This should not include any disruption due to changes to personal circumstances or illness as this is addressed by the Variations to Candidature processes. Please also note relevant ongoing restrictions that may impact any future work that may be requested by examiners."*

Thesis Title: The Test of Speech Sound Perception in Noise (ToSSPiN) – Effect of first language, spatial separation and reverberation on speech sound identification.

Candidate Name: Christian Boyle

Department: Linguistics

Statement:
The original plan for this Masters of Research Project required face-to-face testing for all subjects included in the study. Because of COVID-19, Macquarie University placed an indefinite hold on face-to-face data collection, including for HDR candidates. We were in a fortunate position that development of the test allowed for test delivery via an iPad and we were able to adapt the methods for data collection to test remotely. Remote testing allowed us to add a Canadian English-speaking group that would not otherwise have been possible. As teleaudiology is a relevant topic in current audiology research, it was valuable to the project, and the field, to test participants remotely. Toward the end of data collection, Macquarie University allowed face-to-face research to recommence and we could collect data from these participants. Although COVID-19 could have been a hinderance, we could expand the project to include remotely tested participants, an extra language group and a larger sample size. Our original plan was to test 46 participants, however, we recruited 83 participants. The expansion of the project allowed for significant diversification, although we could not achieve a completely balanced design as it was not possible to test the overseas participants face-to-face. Overall, however, the change in methods needed to cope with COVID-19 restrictions had a positive impact on this project.

Abstract

Aims: The first aim of the study is to investigate the effect of native language on the ToSSPiN in Australian English, Canadian English, and non-native English-speaking people. The second aim is to investigate the differences in performance on the Test of Speech Sound Perception in Noise (ToSSPiN) in face-to-face and remote delivery modes. The final aim is to determine if each phoneme is equal in difficulty and adjust them so that, on average, each are identified 71% of the time at an identical signal-to-noise ratio.

Design: ToSSPiN targets comprised consonant-vowel-consonant-vowel (CVCV) pseudo-words (e.g. /tigu/). Distractors comprised CVCVCVCV pseudo-words. Stimuli were presented using an iPad and headphones. Participants were tested face-to-face at Macquarie University with a researcher recording their responses or remotely via Zoom with a testing partner recording the responses. Scoring occurred adaptively to establish a participant's speech reception threshold (SRT) expressed as dB signal-to-noise ratio. The listening environment was simulated using reverberant and anechoic head-related transfer functions, creating ecologically valid acoustics. The listening environment also varied in whether the distractors were voiced by the same or different voices from the targets. In the baseline ToSSPiN conditions, the targets originated from 0° azimuth. The distractors originated from ±90°, ±67.5° and ±45° in the spatially separated conditions and 0° in the co-located condition. Reverberation impact (RI) was calculated as the SRT (in dB) in the anechoic condition minus the SRT (in dB) in the reverberant condition. Spatial advantage (SA) was calculated as the SRT (in dB) in the spatially separated condition minus the SRT (in dB) in the co-located condition.

Samples: SRTs were collected in young adult native Australian-English speakers (*n* = 24), native Canadian-English speakers (*n* = 25) or non-native English speakers (*n* = 34).

Results: No significant effects of language occurred for the baseline measures, RI or SA. A small but significant effect of delivery mode occurred for RI, but not for SA or baseline measures. Psychometric functions obtained for individual phonemes differed notably and phonemes required adjustments ranging from -2.0 dB for /t/ to +8.7 dB for /h/ to attain equal intelligibility.

Conclusion: The results are consistent with ToSSPiN being a language-independent test of speech sound perception. The ToSSPiN could be used in multiple language contexts to assess the auditory processing abilities of adults. The ToSSPiN could be appropriate for remote delivery.

Brief Introduction to the Master of Research Thesis

First, the literature review will discuss the different perspectives held by professional audiological organizations and researchers regarding the definition and diagnosis of auditory processing deficits. A consequence of the opposing views discussed is that clinicians face a complex task in deciding how best to assess and remediate children with listening difficulties. In the context of the thesis, poor performance on an auditory processing could be caused by an APD, cognitive deficits, and/or language deficits, or a combination of these deficits. 'Listening difficulties' will refer to the real-life consequences arising from the aforementioned causes.

Second, the literature review will discuss some commonly used types of speech-based auditory processing tests and the impact on speech-based tests of listening in a second language. The reason for discussing the impacts of native language is to show the consistent reduced performance present in non-native speakers of the test's language, compared with native speakers, when presenting target speech stimuli in a non-native language. Discussing these differences will highlight the importance of developing language-independent stimuli.

Finally, an additional aspect examined is the need for auditory processing assessments to better reflect real-world listening environments, including the impact of reverberation as well as background noise.

One of the significant components of the Master of Research is to write an essay regarding a major open question in the chosen research field. Here, the open question chosen for the essay was an examination of the impacts of cognition and language ability on auditory processing test performance. The essay's content cannot directly overlap with the thesis literature review and, therefore, any discussion on the impacts of cognition and language ability on auditory processing test performance is necessarily minimal.

1. Introduction

1.1 Background

According to the Australian Bureau of Statistics (ABS), in Australia over 27% of the population speak a primary language other than English, increasing from 23% in 2011 (ABS, 2017). Language diversity also exists in the United States with roughly 21% of the population over the age of five speaking a language other than English at home (U.S. Census Bureau, 2011). In the European Union, only 13% of inhabitants speak English as a first language, and English is the most widely spoken foreign language (TNS Opinion & Social, 2012). There are 23 official languages and 60 indigenous/minority languages in Europe (TNS Opinion & Social, 2012). Linguistic diversity is everywhere. For example, there are approximately 6500 spoken languages unintelligible from one another (Hammarström, 2016) and in India and China alone there are 427 and 241 languages spoken respectively (Gordon, 2005). While language differences can support identity and culture, these differences present a problem when developing speech-based auditory assessments. Linguistic differences potentially limit broad use of speech-based assessments. The difficulties in broadly adopting speech-based AP assessments may also extend to different dialects of the same language. Different dialects of the same language may require different normative data to compare performance on speech-based AP assessments (Dawes & Bishop, 2007).

Many studies investigating the impact of language on performance in auditory tests that use speech stimuli have established that native speakers perform better than non-native speakers when presented with target speech in background noise (Kilman et al., 2014; Krizman et al., 2017; Rogers et al., 2006; Tabri et al., 2011). Previous studies have reported that irrespective of a bilingual speaker's age of acquisition and proficiency in a non-native language, non-native bilinguals have significantly reduced performance compared with native monolingual talkers in background noise on clinically available speech-based auditory tests (discussed below) (Krizman et al., 2017; Shi, 2010; Van Engen, 2010). Differences in performance also exist in people who speak different dialects of the same language (Dawes & Bishop, 2007; Marriage et al., 2001).

Krizman et al. (2017) found that monolinguals performed better than bilinguals on the Quick Speech-In-Noise Test (QuickSIN) (Killion et al., 2001) and the Hearing in Noise

Test (HINT) (Nilsson et al., 1994), however, both groups scored comparably on the Words-in-Noise test. The QuickSIN and HINT measure the perception of English sentences in noise. Each participant repeated five target words from each sentence (30 words in total). The QuickSIN subtracts the total correct words from 25.5 to give an SNR-loss score, with a lower score representing better performance on the task (Killion et al., 2001). The Words-in-Noise test was a non-adaptive measure that required participants to repeat English single-words embedded in a carrier phrase (i.e. say the word), in background babble (55 dB HL) (Krizman et al., 2017). The number of correctly identified words gave an SNR score with a lower SNR threshold showing better performance (Krizman et al., 2017). For the QuickSIN, monolinguals had a smaller SNR loss than bilinguals (1.87 $\pm$ 1.5 dB and 3.17 $\pm$ 1.99 dB respectively; p = 0.007). For the HINT, monolinguals had a lower SNR than the bilingual group (-0.97 $\pm$ 0.82 dB SNR and -0.19 $\pm$ 1.12 dB SNR respectively; p = 0.004). The mean performance in monolingual and bilingual groups did not significantly differ on the Words-in-Noise test (5.51 $\pm$ 1.01 dB SNR and 6.22 $\pm$ 1.83 dB SNR respectively; p = 0.087). The authors concluded that performance differences between each of the language groups depended on the linguistic requirements of the speech material. Krizman et al. (2017) concluded that an increase in linguistic load for each of the sentence tests resulted in poorer performance for the bilingual group.

Hu and Lau (2020) found that native Mandarin speakers had a significantly larger right-ear advantage on dichotic listening tasks (p = 0.003) when listening to CVs, conjugated by /a/, spoken by a native Mandarin speaker as real Mandarin words (e.g. 趴 [pa]) compared to English spoken English CV syllables (e.g. /pa/). One hypothesis the authors proposed is that the enhanced lexical meaning attached to the Mandarin words increased participant’s right-ear advantage. The results suggested that even for phonetically similar words and syllables, an increase in lexical meaning increased speech intelligibility.

Based on the results of the research described above, speech-based auditory assessments may show a bias toward listeners whose native language and dialect are identical to that of the target speaker. As such, it may seem appropriate to develop speech-based assessments for each language and dialect. According to Cameron et al. (2020), however, the time required to develop and validate tests in different languages and dialects is impractical. There is, therefore, a need to develop language-independent speech-based AP tests to use across various language contexts.

A recent study has stated that speech-based audiological tests should incorporate a wide range of real-world listening conditions (Madsen et al., 2019) to better reflect an ecologically valid listening environment. Noise and reverberation occur together with speech in many typical listening environments (e.g. a classroom) (Lewis et al., 2014), yet there are no speech-based audiological tests that include a controlled degree of reverberation. Noise and reverberation together are more detrimental to speech identification than noise or reverberation alone (Lewis et al., 2014; Rogers et al., 2006) and listening skills should, therefore, be measured in a similar acoustic environment. A natural progression, therefore, is to develop AP assessments that closely resemble everyday natural listening environments such as a combination of speech, noise and reverberation.

The present study will evaluate a newly developed test of AP, The Test of Speech Sound Perception in Noise (ToSSPiN) (Cameron & Dillon, 2020c). The ToSSPiN measures AP skills in a simulated, real-world listening environment that incorporates both reverberation and spatially separated noise. The ToSSPiN stimuli constitute pseudo-words comprising consonants and vowels (CVs) common to most of the world's languages (see 'Phonemic Inventories and Cultural and Linguistic Information Across Languages' (ASHA, n.d.) and 'The Speech Accent Archive' (Weinberger, n.d.) and could, therefore, be appropriate for use around the world. Cameron et al. (2020) have previously used the CVs that comprise the ToSSPiN stimuli in the development of the Listening in Spatialized Noise – Universal test (LiSN-U) (Cameron & Dillon, 2020b). The main aim of this study is to investigate the effects of native language on various ToSSPiN measures to determine its suitability for worldwide use.

The ToSSPiN measures the auditory processing abilities of spatial processing and reverberation integration/suppression. Both spatial processing and reverberation integration/suppression utilize interaural cues (i.e. interaural timing and level differences) (Arweiler & Buchholz, 2011; Cameron & Dillon, 2007). The aforementioned AP skills fall under the definition of binaural integration as stated by the American Speech-Hearing-Language Association (ASHA, 2005) and the American Academy of Audiology (AAA, 2010). Binaural integration refers to the combination of auditory information presented binaurally to compare differences in intensity, time, and spectrum (ASHA, 2005). While these skills are important for AP, there are many other important AP skills that the ToSSPiN will not measure directly as mentioned by ASHA (2005) and AAA (2017), such as auditory discrimination, auditory temporal processing, and dichotic listening skills.

1.2 Auditory Processing Disorder

1.2.1 Definition and Presentation

APD is an umbrella term used to describe the difficulty for an individual to identify or discriminate sounds in the presence of normal hearing thresholds (AAA, 2010; ASHA, 2005; Dawes & Bishop, 2009; Jerger & Musiek, 2000; Moore, 2006; Tomlin et al., 2015). APD is considered an auditory specific disorder and the listening difficulties apparent in the presence of an APD are not caused by deficits in cognition or language (AAA, 2010; ASHA, 2005). Some listening difficulties experienced by individuals diagnosed with APDs include distraction and/or difficulty hearing in the presence of background noise (a classroom, for example) (Bamiou et al., 2001; Cameron & Dillon, 2019; Dillon et al., 2012; Sharma et al., 2009; Tomlin et al., 2015); difficulty following complex instruction (Moore, 2006; Vermiglio, 2014); difficulty lateralizing and localizing sound sources (Rowan et al., 2015); and identification of spatially separated speech (Cameron & Dillon, 2007; Dillon et al., 2012). According to Bamiou et al. (2001) children with APDs seem to present with uncertainty about what they hear.

1.2.2 Pathophysiology

The pathophysiological mechanisms that underpin APD can be considered in terms of the neurological conditions, delayed maturation of the central nervous system, or developmental disorders (Bamiou et al., 2001) that occur in the presence of an APD. Although, according to Bamiou et al. (2001), neurological conditions involving cases of APD are rare. In some instances an APD might be the only presenting symptom. Some of the neurological conditions associated with an APD are central auditory nervous system (CANS) tumors, epilepsy, metabolic disorders, or an acquired brain injury (Bamiou et al., 2001). A delay in the maturation of the auditory system, specifically the myelin sheath that surrounds auditory neurons in the higher auditory pathways, is also suspected as potentially causing an APD (AAA, 2010; ASHA, 2005; Bamiou et al., 2001; BSA, 2017). Some of the developmental abnormalities associated with APD include Dyslexia, language impairment, attention deficit hyperactivity disorder, or learning disabilities (ASHA, 2005; Bamiou et al., 2001). Developmental abnormalities associated with an APD are not without controversy. Many authors have suggested that cognitive and/or language are the causes for listening difficulties (de Wit et al., 2016; Moore et al., 2010; Neijenhuis et al., 2017). Other authors have suggested that there is a need to develop assessments that can effectively separate the causes of listening difficulties (Dawes & Bishop, 2009; Dillon et al., 2012; Sharma et al., 2009).

1.2.3 Diagnostic Criteria

Wilson and Arnott (2013) stated that the American Speech-Language-Hearing Association (ASHA, 2005), the American Academy of Audiology (AAA, 2010), and the British Society of Audiology (BSA, 2017) predominantly influence the diagnostic criteria for APD, its conceptual framework, and definition. Many other professional organizations offer opinions on APD, for example, the Canadian Interorganizational Steering Group (2012), International Bureau of Audiophonologie (2007), the Dutch Position Statement (Neijenhuis, de Wit & Luinge, 2017) and New Zealand review of APD (Esplin & Wright, 2014). Each of these organizations offers a unique perspective on the definition and diagnosis of APD.

The ASHA (2005) technical report on (Central) Auditory Processing Disorders states that AP relates to many skills including auditory discrimination, temporal processing, and binaural integration. ASHA defines APD as a deficit in one or more of these skills, as diagnosed using AP assessments. The criteria chosen for an APD diagnosis are is adopted from Chermak and Musiek (1997), in which an individual must score at least two standard deviations (SD) below the mean on two or more assessments in the APD test battery. The battery should incorporate tests which use both speech and non-speech sounds to reflect the variety of auditory processes and regions within the central auditory nervous system that underlie auditory listening skills. The ASHA report also suggests that higher-order cognition and language skills may be reliant on intact AP, due to the extensive neural interfacing between cognitive, auditory, and language areas, and the integrated processing of sensory information between these domains. While deficits in areas of cognition and language may co-exist in the presence of an AP deficit, ASHA excludes these deficits from the definition of APD (ASHA, 2005). The report suggests that while many categories of diagnostic assessments exist to assess many AP skills, clinicians should not use all types of AP tests in every APD evaluation. Instead, it is up to the clinician to determine the assessments to use based on individual listening difficulties.

The AAA (2010) Clinical Practice Guidelines on the diagnosis and management of APD agree with much of what ASHA's report suggests regarding an APD’s definition and diagnostic criteria. The guidelines do, however, state the importance of ear advantage consistency between tests, specifically, the right ear advantage that can be present in dichotic listening tasks (Hugdahl & Helland, 2014). An ear advantage shows a hemispheric dominance for speech and language (Hugdahl & Helland, 2014; Prete et al., 2018). The guidelines highlight that the importance of a consistent ear advantage is to rule out potential non-auditory confounds (AAA, 2010). For example, a right ear deficit on one task coupled with a left ear deficit on a similar task within the same individual may be an indicator of cognitive contributions, such as attention, rather than an APD.

The BSA (2017) position statement does not stipulate a specific diagnostic criterion. The statement acknowledges the need for universally accepted diagnostic criteria and the present lack of a practical diagnostic framework. Further, the BSA (2017) position statement notes an intrinsic link with neural processing systems that modulate top-down (i.e. efferent) control of the central auditory nervous system (CANS). These systems include language, reading, speech, attention, executive function, memory, emotion, vision and motor action (BSA, 2017). Therefore, the BSA have suggested that AP has inherent links to control systems operating outside of the auditory modality.

While there are three major audiological organisations with clear conceptualisations of APD, other perspectives take a cautious approach. Dutch perspective, authored by Neijenhuis et al. (2017) suggested that cases of APD in the absence of other developmental disorders are rare. Further, any diagnosis related to listening difficulties should begin with ruling out non-auditory deficits. According to the authors, a broad, multidisciplinary approach reduces the need for AP assessment as other developmental disorders more than likely cause listening difficulties (Neijenhuis et al., 2017).

Based on the aforementioned position statements on APD, several problems arise for clinicians and researchers. First, it is difficult to define APD due to the relationship between cognition, language, and AP. Second, cognition and language can impact AP test performance and may therefore lead to an inappropriate diagnosis. Third, there is not a universally agreed diagnostic protocol. Consequently, the recommended APD assessment varies depending on the organization cited.

Besides the professional organizations providing guidelines and opinions on the definition and diagnosis of APD, many authors and researchers also propose other conceptualizations (Cacace & McFarland, 2005; Cameron & Dillon, 2008; Chermak, Bamiou, et al., 2017; Chermak, Musiek, et al., 2017; Dawes & Bishop, 2009; Dillon et al., 2012; Jerger & Musiek, 2000; Miller, 2011; Moore, 2006; Sharma et al., 2009; Tomlin et al., 2015; Vermiglio, 2014; Wilson & Arnott, 2013; Wilson et al., 2004). Dawes and Bishop (2009) stated that although APD is commonly diagnosed, there is still considerable disagreement surrounding the diagnostic criteria, definition, the relationship between APD and language and cognition, and whether APD even exists at all. As discussed below, a clinician requires knowledge of many position statements and opinions surrounding APD. Because of the extensive amount of information present, there could be clinician confusion and an increased likelihood of adopting an inappropriate test battery.

1.2.4 Diagnostic Protocols for Auditory Processing: Controversies

Tomlin et al. (2015) have suggested that adopting a cut-off at -2 SDs from the mean on any two APD tests is not evidence-based or well justified. The criterion may over- or under-identify individuals with listening difficulties caused by APDs and/or other deficits (e.g. cognition), depending on the importance to speech perception of the ability being measured. Tomlin et al. (2015) correlated child performance (aged 7-12) on five AP measures that involved the AP skills suggested by ASHA (2005) and AAA (2010), which are understanding spatially separated speech in noise; lateralization; dichotic listening; auditory discrimination; auditory temporal processing; and auditory performance with competing or degraded signals. The auditory measures used were the Frequency Pattern Test (FPT) (Musiek & Pinheiro, 1987); the Gaps in Noise Test (GIN) (Paulovicks, 2008); masking level difference (MLD); the Listening in Spatialized Noise – Sentences test (LiSN-S) (Cameron & Dillon, 2007a); and the Dichotic Digits Test (DDT) (Musiek, 1983). The authors hypothesized that as each of the tests measure different auditory skills, there would be minimal correlations in performance on the tasks. The results of the study showed significant but weak to moderate (up to $r = 0.38$) correlations among FPT, DDT, and LiSN-S baseline measures. The GIN, MLD and the spatial advantage measure of the LiSN-S showed no significant correlations with other AP measures. As the correlations were only weak to moderate, each AP test was a poor predictor of performance on another test. Failing only one test can show a deficit for at least one of the auditory skills suggested by ASHA and AAA and therefore, the need to fail two tests for a diagnosis of APD could be too strict for individuals who experience real-life listening difficulties caused by a deficit in only one auditory skill. Results from Cameron and Dillon (2008) support this suggestion and showed that performance on the LiSN-S did not correlate with performance on other common AP tests, however, children who only failed the LiSN-S still reported real-life listening difficulties.

Other authors have also questioned whether -2 SDs reflects the level of deficit in a given auditory skill needed to impact on real-life listening difficulties (Ahmmed & Ahmmed, 2016; Dillon et al., 2012). Consequently, some authors have suggested that to identify real-life listening difficulties, assessment should begin with a well-validated questionnaire (BSA, 2017; Cameron et al., 2015; Moore, 2018) to compare subjective symptoms to AP test performance.

There are also suggestions that an individual falling two SDs below the mean on any two tests of AP for an APD diagnosis is potentially arbitrary, as the choice of test battery will likely influence an individual's performance (Dillon et al., 2012; Wilson & Arnott, 2013). For example, a diagnostic test battery comprising mainly speech materials with language-dependant responses will also likely identify someone with a language disorder as having APD than a test battery weighted with non-speech materials (Wilson & Arnott, 2013). Children with listening difficulties are heterogeneous as a group (Bench et al., 2020; de Wit et al., 2016; Dillon et al., 2012; Sharma et al., 2009) and there is a need to develop assessments that accurately identify the cause of listening difficulties, not just confirm their presence (Dillon et al., 2012).

The criteria used to diagnose APDs differs across the world. Wilson and Arnott (2013) investigated the pass/fail rates across nine different diagnostic criteria (AAA, 2010; ASHA, 2005; Bellis, 2011; BSA, 2017; Dawes & Bishop, 2009; McArthur, 2009) using four common APD assessments in 150 children sampled from an audiology clinic at the University of Queensland. The authors found that failure rates of the different criteria had a range of 7.3% to 96% failure depending on the criteria used. The authors concluded that until a greater consensus occurs for diagnosing APDs, the specific diagnostic criteria used should accompany a diagnosis of an APD (Wilson & Arnott, 2013).

1.2.5 Audiological Tests

Before any AP specific testing occurs, peripheral hearing assessment is recommended to rule out any possible peripheral hearing loss that could be the cause of listening difficulties (AAA, 2010; ASHA, 2005; Bamiou et al., 2001). Peripheral hearing assessment is comprised of pure tone audiometry, tympanometry, and otoacoustic emissions (ASHA, 2005; Bamiou et al., 2001). The assessments that for the AP test battery fall into four categories: monaural low-redundancy speech tests (e.g. low-pass filtered speech) (AAA, 2010; ASHA, 2005; Bamiou et al., 2001), dichotic assessments (e.g. dichotic digits test) (AAA, 2010; ASHA, 2005; Bamiou et al., 2001), temporal sequencing tasks (e.g. frequency pattern test) (AAA, 2010; ASHA, 2005; Bamiou et al., 2001), and tests of lateralization/localization and binaural integration (e.g. Listening in Spatialized Noise Sentences Test (LiSN-S) (AAA, 2010; ASHA, 2005; BSA, 2017). Each of these tests aim to measure a specific auditory skill as suggested by ASHA and AAA (these tests, their function, and the auditory processes they measure are discussed later).

According to Jerger and Musiek (2000) electrophysiological tests, such as Auditory Brainstem Response (ABR), Middle Latency Response (MLR), and P300 provides important information regarding the integrity of the auditory nerve and brainstem pathways. Although Jerger and Musiek (2000) have suggested that ABRs are appropriate for APD clinical evaluation, AAA (2010) state that the value of such tests is limited as only a small proportion of children with an APD show abnormal ABR results.

1.2.6 Management

According to many of the professional audiological organisations, treatment should encompass an individualised approach based on the pattern of results observed for individuals (AAA, 2010; ASHA, 2005; Bamiou et al., 2001; BSA, 2017). Management of an APD consists of multiple intervention strategies such as signal enhancement, auditory training, and compensatory strategies (AAA, 2010; ASHA, 2005; Bamiou et al., 2001; BSA, 2017). Signal enhancement strategies are comprised of improving the signal-to-noise ratio by using a frequency modulation (FM) system, improving room acoustics (i.e. in the classroom), or a combination approach (ASHA, 2005; Bamiou et al., 2001; BSA, 2017). Auditory training can encompass interactive training software (e.g. LiSN & Learn) (BSA, 2017; Cameron et al., 2012) or musical training (BSA, 2017). The LiSN & Learn software (now Sound Storm) is a targeted remediation tool for spatial processing disorder (SPD) which is the inability for an individual to utilize binaural cues in the presence of background noise (Cameron et al., 2012). Compensatory strategies can take the form of improving listening skills that aim to teach an individual to develop an awareness of the active processes (such as attention) that are involved when listening (BSA, 2017). Additional strategies include developing self-regulation by training metacognitive and metalinguistic skills (BSA, 2017). Some of the metalinguistic training can involve phonological awareness, semantic network expansion, and using context to understand and build vocabulary (ASHA, 2005). Metacognitive training can involve organisation skills, problem solving, and assertiveness training (ASHA, 2005; BSA, 2017). According to Bamiou et al., (2001) APD management is in itself quite controversial as if an APD is the primary cause for a language deficit, targeted auditory training may not lead to an improvement in language.

1.2.7 Outcome Measures

According to Casady et al. (2017) there is very little research regarding the efficacy of outcome measures used for the management of APD. Loo et al. (2013) have suggested that self/caregiver report, and/or questionnaires (e.g. Fisher's Auditory Checklist) can give information regarding the personal benefit of management strategies. Wilson et al. (2013) suggested to readminister selected electrophysiological tests (e.g. P300, MLR) to determine if any changes in plasticity have occurred as a result of management for APD. ASHA (2005) have suggested that audiologists should readminister selected tests within the individual's specific deficit areas to monitor improvement. In order to monitor academic successes, ASHA (2005) have also suggested that any improvement in grades is a good way to monitor academic progress post-intervention.

1.2.8 Researcher Perspectives – Comorbidity; Cognition; and Language.

In a systematic review of 53 studies (level 1 evidence) identifying characteristics of APDs, de Wit et al. (2016) found that the evidence for a purely auditory specific deficit was inadequate because of the heterogeneous nature of the participants in each study reviewed. For example, children with listening difficulties from the studies surveyed also had varied auditory, visual, cognitive, reading and language profiles. Many of the children had comorbid deficits. The authors concluded that the listening difficulties identified in children with a diagnosis of APD, based on poor performance on AP tests, may result from deficits in memory, language and/or attention, rather than purely AP deficits.

In a comparative study of children's cognitive abilities and AP skills, Moore et al. (2010) found that poor AP performance reflected increased cognitive demands rather than bottom-up processing of auditory stimuli. The authors found that children with poor AP performance also performed poorly on cognitive measures. The authors state that in the absence of any known brain lesions, APD should be redefined as primarily cognitive rather than a sensory disorder (Moore et al., 2010). The impact of cognition on test performance has led many authors to question the existence of APD as a distinct clinical disorder (de Wit et al., 2016; Kamhi, 2011; Moore et al., 2010; Vermiglio, 2014). One aspect not addressed in the Moore et al. (2010) study is the potential impact of fatigue on test performance. According to Dillon et al. (2012) increasing the number of tests in an AP battery can increase fatigue and reduce cognitive resources. There were 12 tests used in the Moore et al. (2010) study and counterbalancing of the tests did not occur (i.e. AP assessments completed first, cognitive measures completed last). Additionally, a difficult listening environment (such as noise and reverberation) leads to an increase in listening effort and cognitive resources required to decode a given message (Yang & Bradley, 2009). Poor performance on the cognitive measures could reflect fatigue in some children with poor AP rather than inherent cognitive deficits.

One potential solution to reducing the impacts of fatigue on test performance is to divide each session so that cognitive resources are not depleted. Gyldenkærne, Dillon, Sharma, and Purdy (2014) measured the relationship between attention and AP deficits in 101 children with listening difficulties and 18 children with no listening difficulties aged 7-12. The chosen AP tests aimed to assess a range of AP abilities, as recommended by

ASHA (2005) and AAA (2010) (Gyldenkærne, Dillon, Sharma, & Purdy, 2014). Each participant was given multiple breaks throughout each two hour session. If a participant scored within 2 SD of the mean on a test, it was a pass. If a participant scored below 2 SD of the mean on any test, it was a fail (Gyldenkærne et al., 2014). Each session was two hours in length, with multiple breaks throughout the session. The authors found that 19 of the children met the criteria for AP deficits without co-occurring attention issues. Nineteen children had attention issues without any AP deficits. A further 39 had both attention and AP deficits (Gyldenkærne et al., 2014). The two groups of 19 children made up 38% of the clinical sample, and the authors conclude that AP and attention deficits were separate in these children (Gyldenkærne et al., 2014). These results contrast that of Moore et al. (2010) since the 19 children with attention issues without AP deficits would be assumed to fail the AP tests. The authors also found significant correlations ($p < 0.01$) between AP test scores and sustained attention (Gyldenkærne et al., 2014). A relationship between AP and sustained attention should not be surprising as sustained attention *can* impact scores on AP assessment scores. However, attention may not be the cause of listening difficulties or poor performance on an AP assessment.

According to Dawes and Bishop (2009), rather than a complete abandonment of the APD label, audiology requires well-standardized, age-appropriate and reliable measures of AP that reduce non-auditory deficits on AP test performance. Some authors have suggested that AP tests should directly assess the processing of auditory stimuli and a diagnosis of APD should be distinguished from cognitive and language deficits (Cacace & McFarland, 2005; Miller, 2011; Moore, 2006). Other authors have stated that a complete separation of non-auditory deficits on APD test performance is impossible because of the complex interaction of auditory and non-auditory areas in the CANS (ASHA, 2005; Musiek et al., 2005).

Wilson (2018) noted that a universal concept of APD is elusive. The author stated that to move toward a single unifying framework and reconcile challenges associated with many definitions and diagnoses, APD could be considered a spectrum disorder. The spectral approach would regard APD as having multiple sites of dysfunction encompassed within bottom-up AP and the top-down cognitive processes that affect it (Wilson, 2018). Wilson's framework has received support from many authors (Bench et al., 2020; Iliadou et al., 2018), however, a universal conceptualisation of APD remains absent.

The differing views present in conceptualizing APD result in difficulty for a clinician to determine and implement an exact definition and diagnostic framework. One of the major divisions seems to be disagreement between whether APD is a unimodal, bottom-up disorder; or a disorder with influence from top-down processes that may impact APD assessment; or another label for cognitive and language deficits that are evident when testing with auditory stimuli. The debates and controversies are still ongoing in the APD research community. A universal framework remains a challenge to address for current and future researchers. An additional challenge is to develop AP assessments that can reduce or eliminate the influence of cognition and language on test performance to more accurately diagnose AP deficits.

1.3 Hierarchical Approach to Auditory Processing Assessment

According to Moore (2018), a clinical model for AP assessment should have rigorous and standardized procedures. A model suggested by Cameron and Dillon (2019) includes a deficit-specific, hierarchical diagnostic protocol. The basis of the procedure is to optimize assessment with some tests given only when an individual performs worse than a particular cut-off value for a preceding, more general test. The model aims to pair each assessment with a targeted remediation therapy and is therefore intervention-focussed. For example, remediation with LiSN & Learn software (Cameron & Dillon, 2011), or its replacement, Sound Storm (Cameron & Dillon, 2017) accompanies a diagnosis of Spatial Processing Disorder (SPD). SPD results from an inability to use interaural timing and level differences (ITDs and ILDs) to separate a target speaker from distracting signals (Cameron & Dillon, 2007a). According to Cameron and Dillon (2019) the model may be an efficient and accurate way of diagnosing specific AP deficits in individuals to provide deficit-specific remediation.

Cameron et al. (2015) tested a limited, deficit-specific, hierarchical approach for diagnosis and remediation of AP skills in school-aged children. Hearing Australia adopted the hierarchical approach for a national CAPD service. Cameron et al. (2015) analysed data collected from Hearing Australia's clients over an 18-month period. The authors measured outcomes using the Client Oriented Scale of Improvement – Children (NAL, 2016) and the Listening Inventory for Education – Teacher Scale (Anderson & Smaldino, 1999). The results showed that the hierarchical approach allowed for efficient identification of SPD and dichotic deficits with significant improvements observed post-remediation and/or management for SPD. Before completing dichotic testing, clinicians screened clients for memory deficits using the Test of Auditory Processing (TAPS-3) (Martin et al., 2018) Number Memory Forward (NMF) and Reversed (NMR) subtests. Cameron et al. (2015) aimed to not perform dichotic testing on children with memory deficits because such deficits can negatively affect performance on dichotic tests (Ahmmed et al., 2014; Cameron & Dillon, 2020a) (the impacts of cognition on dichotic tests are discussed further below). The authors offered the participants the LiSN & Learn training software (Cameron & Dillon, 2011) for remediation and/or a wireless remote microphone system for management of SPD. For dichotic deficits, children were offered an FM system. The LiSN & Learn software aims to improve an individual's spatial processing skills by asking them to repeat a target word in the presence of spatially separated background noise (Cameron & Dillon, 2011). A wireless microphone system is a microphone and wireless transmitter coupled to an in-the-ear receiver that improves the signal-to-noise ratio (SNR) received by the child (Cameron et al., 2015). The authors noted that the results were positive in showing the hierarchical model as being efficient for the diagnosis and remediation of specific AP deficits, however, these models require continuous evaluation and improvement as research and knowledge increases (Cameron et al., 2015). Moore (2018) supported the hierarchical approach and suggested that the model will be key to better clinical management if well standardized, intervention-focused, and evidence-based.

1.4 Speech Audiometry Assessment in Auditory Processing

There are many speech-based AP assessments used in the AP test battery. Each speech test aims to assess a specific auditory skill such as dichotic listening, auditory closure, and localisation/lateralisation. According to ASHA (2005) and AAA (2017) these speech-based tests fall into several categories comprising dichotic listening, low redundancy speech tests, and assessments of binaural interaction.

1.4.1 Dichotic Speech Assessments

Dichotic listening (DL) tasks have been used for over 50 years (Katz, 1962) and have been shown to be sensitive to APDs (AAA, 2010). Dichotic listening tasks assess the brain's hemispheric dominance (Hugdahl, 2009) for speech and language (Helland et al., 2018) by presenting competing signals to both ears simultaneously. Dichotic tasks can use various types of stimuli, for example, digits; words; and sentences (AAA, 2010). One of the most commonly used tests is the Dichotic Digits Test (DD) (Musiek, 1983) which uses numbers to identify an ear advantage. Dichotic assessments can form part of an overall assessment battery, such as the SCAN-C (Keith, 2000a). The SCAN-C uses both competing words and sentences to identify an ear-advantage (Dawes & Bishop, 2007; Keith, 2000a). Recently, Cameron and Dillon (2020a) showed a moderate correlation between memory and dichotic performance ($r = 0.50$). Stavrinos et al. (2018) also showed that divided auditory attention was strongly correlated with the dichotic digits test ($r = 0.68$, $p < 0.05$). Therefore, poor performance on dichotic tests could reflect AP deficits, cognitive deficits, or a combination of these deficits.

Implementing a hierarchical test design can reduce the impacts of memory deficits on AP test performance. As mentioned previously Cameron et al. (2015) used the NMF and NMR of the TAPS-3 to identify children with memory deficits before completing the DDT. Children identified with a memory deficit completed Memory Booster training, rather than completing the DDT. Memory Booster trains children in different memory strategies over six to eight weeks (St Clair-Thompson et al., 2010). Restricting dichotic testing to children with memory within normal limits reduced but did not eliminate the impacts of memory deficits on dichotic performance (Cameron et al., 2015). Cameron et al. (2015) suggested that a hierarchical test battery had the potential to increase the accuracy of identifying the true nature of an individual's deficits by reducing memory impacts of dichotic performance.

1.4.2 Low-Redundancy Speech Assessments

Low-redundancy speech tests are another type of AP test that uses speech stimuli. These tests aim to reduce the natural redundancy present in speech signals by removing a portion of the spectral information (AAA, 2010) and assess an individual's ability to understand the signal. Optimal intelligibility occurs when the entire speech spectrum is audible (Rickard et al., 2013). However, an individual with average listening skills should still be able to comprehend speech when part of the spectral information is missing (Moore, 2006; Rickard et al., 2013). Auditory closure, the skill that low-redundancy speech tests measure, is the ability of an individual to fill in the gaps when part of a speech signal is missing, degraded or to some extent unintelligible (Keith, 2000b; Rickard et al., 2013).

One of the significant concerns with degraded speech assessments is that poor performance may reflect a language impairment rather than AP deficits (Loo et al., 2013; Moore & Hunter, 2013). For example, Loo et al. (2013) found that children with language impairments and listening difficulties performed more poorly on low-redundancy speech tests than children with listening concerns and no language impairment ($p < 0.001$). Poor performance on low-redundancy speech tests as a result of language deficits could explain why, according to AAA (2010), these are less sensitive to APDs than other speech tests. Musiek et al. (2011) showed that a low-pass filtered speech test reduces the sensitivity and specificity of an APD test battery when it is included with other AP tests. Language deficits that negatively impact performance on low-redundancy AP measures may explain the reduction in specificity for low-pass filtered speech.

1.4.3 Listening in Spatialized Noise – Sentences Test (LiSN-S)

The Listening in Spatialized Noise - Sentences test (LiSN-S) developed by Cameron and Dillon (2007a) measures an individual's auditory stream segregation abilities. Auditory stream segregation is the skill involved in separating multiple sound sources and grouping them into relevant streams or objects (Paredes-Gallardo et al., 2018). An example of stream segregation is to identify a target speaker in the presence of multiple sound sources (i.e. background noise) (Cameron & Dillon, 2007a).

The LiSN-S uses a repetition-response procedure that incorporates four baseline conditions. Each baseline condition requires an individual to identify target sentences in the presence of competing noise (looped stories) that vary in spatial location and talker identity (Cameron & Dillon, 2007a). The speakers of the distractors may have the same or different voice to the target speech and are presented at either 0^{o} azimuth or $\pm 90^{o}$ azimuth. Each condition uses convolution with head-related transfer functions (HRTFs) to simulate monaural (pinna) and binaural cues (i.e. ILDs and ITDs) to produce a three-dimensional listening environment under headphones (Cameron & Dillon, 2007b).

The LiSN-S uses differential testing (i.e. difference scores) that measures the differences between two baseline conditions that vary in one auditory element only

(Cameron & Dillon, 2007a; Cameron et al., 2006; Moore et al., 2010). In each condition, cognitive and language requirements remain constant and are therefore reduced when calculating the difference scores (Cameron & Dillon, 2019; Cameron & Dillon, 2020a; Moore et al., 2010). The derived score is a more accurate measure of the specific AP skill (Cameron & Dillon, 2019; Moore et al., 2010).

The LiSN-S is a validated measure for diagnosing spatial processing disorder (SPD). SPD is a specific AP deficit and is diagnosed based on an individual's performance using the low and high-cue speech reception thresholds (SRT) and three advantage measures. The low-cue SRT is obtained when neither voice nor spatial cues are present. The high-cue SRT is obtained when both spatial and pitch cues are present. The three advantage measures are based on the benefit (in dB) an individual receives when vocal and/or spatial cues are present in the distracting noise when referenced to the low-cue condition (i.e. no cues) (Cameron et al., 2009; Cameron & Dillon, 2007a, 2007b). SPD occurs when an individual has a reduced ability to use spatial cues to identify the target speaker when the distractors are spatially separated (Cameron & Dillon, 2008).

The LiSN-S has been used to assess AP skills in several populations including individuals with neuropathic disorders, or a history of chronic middle ear disease. Rance et al. (2012) investigated spatial processing abilities in individuals with neuropathic disorders (i.e. Friedreich ataxia and Charcot-Marie-Tooth disease type 1A) and compared performance to matched controls on the LiSN-S. The results showed that both groups with neuropathic disorders required an increase in signal-to-noise ratio (SNR) for all LiSN-S conditions and had reduced spatial advantage compared with controls (Rance et al., 2012). The authors concluded that temporal processing deficits caused by neuropathic disorders disrupt the integration of binaural information (Rance et al., 2012). The LiSN-S has been used to show a high prevalence of SPD in populations with a history of chronic otitis media (COM) (Tomlin & Rance, 2014) such as Aboriginal and Torres Strait Islander populations (Cameron et al., 2014). Chronic otitis media can cause inconsistent access to ITDs and ILDs due to fluctuating hearing loss and can impact the development of binaural cues in the auditory system that are integral to spatial processing (Tomlin & Rance, 2014).

1.4.4 Listening in Spatialized Noise – Universal Test (LiSN-U)

Cameron et al. (2020) developed a language-independent version of LiSN-S, the Listening in Spatialized Noise – Universal (LiSN-U) test. The LiSN-U aims to overcome issues in establishing the LiSN-S in many languages (Cameron et al., 2020; Mealings, Cameron, Chong-White, et al., 2020; Mealings, Cameron, & Dillon, 2020; Mealings & Dillon, 2020) and was, therefore, developed for use across various cultural and linguistic contexts. The LiSN-U comprises pseudo-words that use consonants and vowels (CVs) shared between most languages (Cameron et al., 2020).

The significant differences between the LiSN-S and LiSN-U are the choice of speech material (i.e. CV pseudo-words versus sentences) and only using the same voice at 0^{o} (SV0) and same voice at $\pm 90^{o}$ (SV90) conditions to derive a spatial advantage measure. The distractors originate from the same direction as the target in SV0 condition. For the SV90 condition, the distractors are $\pm 90^{o}$ relative to the target. The phonemes used in the LiSN-U are language-independent and may allow the diagnosis of SPD to be viable across various languages.

Mealings and Dillon (2020) compared performance on the LiSN-S and LiSN-U in Aboriginal and Torres Strait Islander children in Australia, who did not have English as their first language. The authors found significant moderate correlations between the baseline scores in LiSN-U and their comparable baseline scores in LiSN-S (SV0, $r = 0.675$, $p < 0.0005$; SV90, $r = 0.606$, $p < 0.0005$). When comparing the spatial advantage measure between the LiSN-S and LiSN-U based on the SRTs (in dB), a greater degree of random measurement error weakened the (still significant) correlation ($r = 0.384$, $p = 0.012$). The authors suggested that the LiSN-U showed promise for identifying SPD in Aboriginal and Torres Strait Islander individuals and as a language-independent test of spatial processing ability (Mealings & Dillon, 2020).

Although a correlation exists between performance on the LiSN-S and LiSN-U, Mealings and Dillon (2020) showed that some Aboriginal and Torres Strait Islander children performed significantly worse on the LiSN-S than the LiSN-U. The authors suggested that the reduced performance on the LiSN-S was more than likely due to the participant's English language proficiency, and the native English language requirements needed to complete the LiSN-S. Mealings and Dillon (2020) stated that they did not have access to a standardised English proficiency assessment or, have details regarding the participants' English language acquisition. Studies may be therefore required to determine the correlation between the LiSN-S, LiSN-U and English language proficiency in non-native English speakers. The results from Mealings and Dillon (2020) may show that applying the LiSN-U stimuli to different language contexts is appropriate as the test might reduce the need for native language skills by removing vocabulary, syntactic and semantic requirements. It remains plausible that the improvements seen on the LiSN-U were due to the elimination of linguistic requirements and using high occurring phonemes. Using a speech-based AP test that minimises the need for language-specific skills (such as syntax and semantics) may reduce the impact of higher-level language deficits that can co-occur with APDs (Sharma et al., 2009).

1.5 Language-Independent Speech Material

1.5.1 Common Speech Sound Language Universals and LiSN-U

As mentioned previously, the LiSN-U comprises phonemes common to most languages (Cameron et al., 2020). Shared phonemes are an ideal speech material, as people from most languages will already be familiar with them. To take advantage of speech sound similarity shared across different languages, phonemes should be organized into syllable structures shared amongst most languages (Alqahtani Mufleh Salem, 2019). For example, a consonant-vowel (CV) structure has increased perceptual landmarks relative to a vowel-consonant (VC) and is, therefore, a preferred structure (Hyman, 2007). The increased perceptual landmarks in the CV structure have larger modulations between phonemes than in VCs, and so persist in most languages (Ohala, 1992). Ohala (1992) stated that almost all languages use the CV structure, however, the majority of existing languages do not use the VC structure. The CV is a core syllable structure and is consistent with the notion that the CV is the first syllable type learned developmentally and preferred universally (Chen, 2011). Universal speech sound stimuli should therefore have a CV structure comprised of phonemes shared across most languages.

1.5.2 Language-Independent Dichotic Test

Additional dichotic tests have been developed to reduce the impacts of language requirements on test performance. Findlen and Roup, (2011) investigated how stimulus material, lexical content of the material, and response conditions impacted performance on dichotic speech recognition in 30 normal hearing adults aged 18 to 31. The stimuli for the test comprised CVC word pairs and nonsense CVC syllable pairs. The three conditions were free recall, directed recall right, and directed recall left. The free recall condition required participants to recall the pairs in any order. The directed recall conditions, the participants were asked to recall the stimulus from the directed ear first, followed by the opposite ear. The authors found that performance for the nonsense CVC syllables was significantly worse than for CVC words (p <0.05) and concluded that in normal hearing adults, the lexical content of dichotically presented stimuli impacts speech recognition performance. Additionally, the authors suggested using nonsense CVC syllables could reduce the impacts of language on dichotic test performance.

1.5.3 Language-Independent Monaural Low-Redundancy Test

Arnott et al., (2014) examined the impacts of language confounds in 55 native English-speaking females on a filtered word test. The participants were required to repeat 80 CVC real-words and 80 CVC non-word monosyllables that had been low-pass filtered. Thirty participants had a harsher filter range of 2000 to 500 Hz, twenty-five participants had a milder frequency range of 3000 to 1500 Hz. Unfiltered nonsense words and real words were also presented to the participants. The test was scored as the percentage of phonemes correct. The authors found that the percentage of phonemes correct for CVC real words were significantly higher when no filter was applied to the stimuli and in the filter range between 3000 Hz and 1750 Hz. Conversely, the percentage of phonemes correct were significantly higher for CVC non-word monosyllables between the filter range of 500 and 1250 Hz. The authors concluded that the better performance in unfiltered and milder filter conditions support the involvement of top-down linguistic skills as aiding the repetition of real words. Further, the authors suggested that the results support using nonsense syllables in AP tests as a way of measuring AP skills without involving higher-order language abilities.

1.5.4 Accent

The accent of target speech may be problematic for individuals unfamiliar with the accent because of the differences in acoustic cues present in the speech sounds (Ferguson, et al., 2010). An unfamiliar accent in the presence of noise can cause reduced speech intelligibility and increased response times to target material (Ferguson et al., 2010; Gordon-Salant et al., 2019; Rogers et al., 2004). Although a reduction in speech intelligibility can occur when listening to an unfamiliar accent, many studies show that individuals can rapidly adapt to accented speech.

Clarke and Garrett (2004) showed that native English listeners (American) can adapt to Spanish-accented English sentences within a minute. The authors measured the average response times required for participants to recall the last word of low probability sentences across four blocks of 16 sentences. The authors found a significant reduction in response times with each subsequent block of sentences, $F(3, 45) = 13$ ($p<0.00001$). The authors also showed that by the fourth block of sentences, the reaction times were the same in Spanish-accented and native accented sentences. The authors concluded that with sufficient exposure to Spanish-accented English, processing of the accented sentences occurred just as quickly as the native accented sentences.

Xie et al. (2018) replicated the findings of Clarke and Garrett (2004) and found that native English speakers could rapidly adapt to Mandarin-accented English. The results of the study showed significantly decreased response times ($p < 0.01$) with longer exposure to accented speech. The authors concluded that sufficient exposure to accented speech attenuated average processing speeds which they interpret as evidence of accent adaptation.

Based on the results of previous research, it is plausible that there is a period in which participants require sufficient time to adapt to accented stimuli. For individuals who do not have sufficient familiarization, an increase in response time to the stimuli could occur.

1.5.5 A Universal Speech Assessment for Auditory Processing

The literature reviewed above illustrates the importance of developing diagnostic assessments of AP for use across different language contexts. Performance on a diagnostic AP assessment that uses speech stimuli in an individual's non-native language or accent may not be an accurate reflection of the individual's AP skills. As a result, the cause of poor results may be unclear. There is, therefore, a need to develop a universal AP assessment of speech sound perception not negatively affected by native language or accent. Although there are tests of dichotic perception, low redundancy, and binaural interaction that measure AP skills with language-independent stimuli, there is also a need to develop tests that incorporate ecologically valid acoustics (discussed below).

1.6 Ecologically Valid Acoustics

1.6.1 Noise and Reverberation

Noise and reverberation coincide within oral communication in many typical listening environments (Lewis et al., 2014), for example, a classroom. Noise and reverberation can have detrimental effects on speech perception in isolation, however, their simultaneous effects are often more detrimental than the sum of their component distortions (Lewis et al., 2014; Rogers et al., 2006).

Several studies show that reverberation negatively effects the intelligibility of speech sounds (Picou et al., 2016; Rennies et al., 2014; Riley & McGregor, 2012). Although reverberation has mostly adverse effects on speech perception, portions of the reverberant energy may help intelligibility. Reverberation comprises sound arriving at the listener with different delays. Early reflections (ER) arrive at the listener within 50 ms (Arweiler & Buchholz, 2011; Bradley et al., 2003) and late reflections (LR) arrive after 80 to 200 ms of the direct sound (Hidaka et al., 2007; Rennies et al., 2014). The LR portion of the reverberant signal can persist long after the sound that caused it and can mask the incoming signals arriving at the same time (Picou, Gordon, & Ricketts, 2016; Rennies, Schepker, Holube, & Kollmeier, 2014; Riley & McGregor, 2012). The auditory system, however, inhibits these later-arriving reflections to emphasize and strengthen a stimulus onset (Shinn-Cunningham & Kawakyu, 2003). The auditory system uses ERs to assist in the recognition of an immediately preceding sound by integrating ERs with the direct sound (Fang et al., 2018). Studies have shown that these ERs can help improve the intelligibility of speech in less than ideal listening conditions (Arweiler & Buchholz, 2011; Bradley, Sato, & Picard, 2003). Based on the above research, a hypothesis is that some individuals with AP deficits may have an inability to integrate ERs and/or suppress LRs and therefore, may experience greater difficulty than typically developing peers in understanding speech in the presence of noise and reverberation.

1.6.2 Spatial Separation

As previously discussed, in everyday listening environments, many sounds can arrive at the ears simultaneously and from differing spatial locations (Cameron, Glyde, & Dillon, 2011). In order to separate the components of a target sound and the competing noise sources, a listener must use binaural processing skills (i.e. using ITDs and ILDs) to both identify the target sound and to segregate it from the noise perceived to come from different directions (Cameron, Dillon, & Newall, 2006a).

Whereas there are validated measures of spatial stream segregation skills in anechoic conditions such as the LiSN-S (Cameron & Dillon, 2008), at present, an assessment tool that measures AP skills under acoustically controlled conditions in the presence of both spatially separated noise and reverberation is unavailable. Therefore, a test of AP skills using speech stimuli in the presence of spatially separated noise and reverberation is necessary to closer reflect a real-world listening environment.

1.7 Measurement Theory and Equal Intelligibility

According to Dillon (1983) to maximize sensitivity of a speech discrimination test, all test items should be equally difficult. Two significant theories regarding measurement methodology in test development are classical test theory (CTT) and item response theory (IRT). According to CTT, each item in a test or sub-test should measure the same variable, be as good an indicator of the actual score as every other item and should be scored in the same manner (Devellis, 2006). However, equal difficulty rarely occurs as each test item may require a different mixture of traits (e.g., intelligence, memory, auditory temporal resolution) (Embretson & Yang, 2006).

Different test items within the same test might require different skills or traits that change for each test item. For example, the Speech Perception in Noise (SPIN) test requires an individual to repeat the final word of a sentence in both high- and low-predictability sentences, at a fixed SNR (Kalikow et al., 1977). For high-predictability sentences, context can aid an individual to fill in the gaps as the final word of the phrase is directly related to the overall sentence (Kalikow et al., 1977). High-predictability sentences are rich in linguistic information and therefore, linguistic context and auditory skills can aid in predicting the final word. For low-predictability sentences, the final word is not related to the overall phrase and cannot be predicted from the context (Kalikow et al., 1977). Low-predictability sentences reduce linguistic cues and require utilizing acoustic-phonetic information (Kalikow et al., 1977) and can therefore be considered more difficult than high-predictability sentences.

Even if the skills required to complete a test remain the same for each item, there may be inherent differences in difficulty for each item if each item requires multiple skills to complete (e.g. attention, memory, and/or language), or the relative difficulty between items change. For example, adaptive speech tests increase or decrease the SNR between test items, such as that used in the LiSN-S or LiSN-U. As the relationship in SNR between each item changes from item to item, the difficulty between each item changes, however, the listening skills required remain constant. Another example is the Test of Auditory Processing Skills (TAPS-3) number memory for and backward subtests (Martin et al., 2018). The TAPS-3 subtests require an individual to recall sequences of digits in correct order or backward order (Martin et al., 2018). Each sequence of digits increases in length and so each subsequent test item is inherently more difficult than the previous item. As the test progresses, memory abilities remain the same, however, the memory load increases for subsequent items and therefore the difficulty increases.

IRT considers both latent traits (e.g. unobservable traits, such as intelligence and competence in various tasks) and observed item responses (such as correctly identifying a stimulus) when calculating scores (Reise & Waller, 2009). Further, in contrast to CTT, IRT does not assume that each test item is equally difficult (Embretson & Yang, 2006). Obtaining equal difficulty between each item in a speech discrimination test, therefore, may require manipulation of each item. Equal intelligibility between each test item for a speech discrimination test will achieve maximum sensitivity for a test, satisfy the framework for CTT, and hence enable the use of simple scoring methods.

1.8 The Test of Speech Sound Perception in Noise – ToSSPiN

To address the issues mentioned above, a new assessment of AP skills named the Test of Speech Sound Perception in Noise (ToSSPiN) has been developed. The ToSSPiN is a natural progression of AP assessments from the LiSN-S and LiSN-U. The ToSSPiN aims to incorporate the use of difference scores and spatialization from these assessments, in addition to incorporating reverberation. The ToSSPiN will also incorporate the language-independent stimuli used in the LiSN-U and each phoneme (i.e. consonants and vowels) will be adjusted to be equally intelligible.

The overall aim of the ToSSPiN is to improve the diagnosis of an APD by reducing the impact of cognition and language on test performance. Investigating the impacts of cognitive and language on test performance, in clinical populations, will occur in future studies. For this study, the impacts of cognition and language will be reduced by using difference scores (discussed below) and language-independent stimuli. An additional aim is to improve the diagnosis of an AP deficit in individuals who speak any language. As previously discussed, the ToSSPiN targets and distractors comprise pseudo-words using the CVs shared in many of the world's languages as used in the LiSN-U (Cameron et al, 2020).

The stimuli are convolved with HRTFs at various locations to simulate reverberant and anechoic listening environments. The ToSSPiN will also incorporate spatial separation into the distractors to allow for an ecologically valid speech assessment.

A significant addition to the assessment of AP, and incorporated into the ToSSPiN, is to measure the impact of reverberation under controlled conditions (i.e. headphones) when distracting noise varies in spatial location. The ToSSPiN comprises four conditions varying in room type (reverberant or anechoic), distractor location (separated or co-located) and distractor talkers (different from or identical to the target talker). The four conditions selected are Reverberant Separated Different voice (RSD), Anechoic Separated Different voice (ASD), Anechoic Separated Identical voice (ASI) and Anechoic Co-located Identical voice (ACI). The LiSN-S and LiSN-U incorporates noise at two locations (0^{o} and $\pm 90^{o}$ azimuth) and the ToSSPiN will extend these two locations by adding $\pm 45^{o}$ and $\pm 67.5^{o}$. Distracting noise arrives from multiple directions that change during the test, therefore, the ToSSPiN measures AP abilities in a simulated real-world listening environment.

Two difference scores that will be obtained between two pairs of baseline conditions will aim to quantify the impact of reverberation and an individual's ability to use spatial cues. These difference scores measure the impact of reverberation (RI) (RSD minus ASD) and spatial advantage (SA) (ASI minus ACI) to identify specific auditory deficits. Using difference scores, as previously discussed, will reduce the impact of cognition and language on test results. A long-term goal for the ToSSPiN is to incorporate intrinsic attention and extrinsic memory measures that will correct test scores based on changes in attention and/or an individual's memory deficits. Intrinsic and extrinsic measures will further reduce the impacts of cognition on test performance to more accurately identify AP deficits.

1.9 ToSSPiN Development

As previously stated, the ToSSPiN software was designed and developed by Dr Sharon Cameron and Professor Harvey Dillon. The following section outlining the development of the ToSSPiN software, as described in the Methods section, is unique to the project. No previously published papers describing the ToSSPiN or it's development are available to the examiners. The development of the ToSSPiN does not form part of the research project under examination for this thesis.

1.9.1 ToSSPiN Baseline Conditions

The ToSSPiN software was developed for the iOS operating system for Apple iPad hardware in Xcode programming language. The ToSSPiN consists of four baseline conditions that differ in listening environment, speaker identity and distractor spatial location. The four conditions are described as follows (shown in Figure 1.1):

- The **RSD** condition measures an individual's ability to understand speech in **R**everberation and spatially **S**eparated noise. The competing sounds are spoken by people **D**ifferent from that of the target stimuli.
- The **ASD** condition maintains the voice and spatial cues used for RSD, however, speech sound perception is assessed in an **A**nechoic environment, so there is no reverberation.
- The **ASI** condition measures an individual's ability to identify speech sounds when the distracting noise is spatially **S**eparated from the target in an **A**nechoic environment. Voice cues are reduced by using target and distractors spoken by the same talker (i.e. **I**dentical).

- The **ACI** condition is **A**nechoic with the targets and distractors delivered by the same (I.e. **I**dentical) person. Spatial cues are removed and the distractors are **C**o-located with the target speaker. There are, therefore, no spatial or voice cues available to differentiate the target consonants and vowels (CVs) from the distractors.

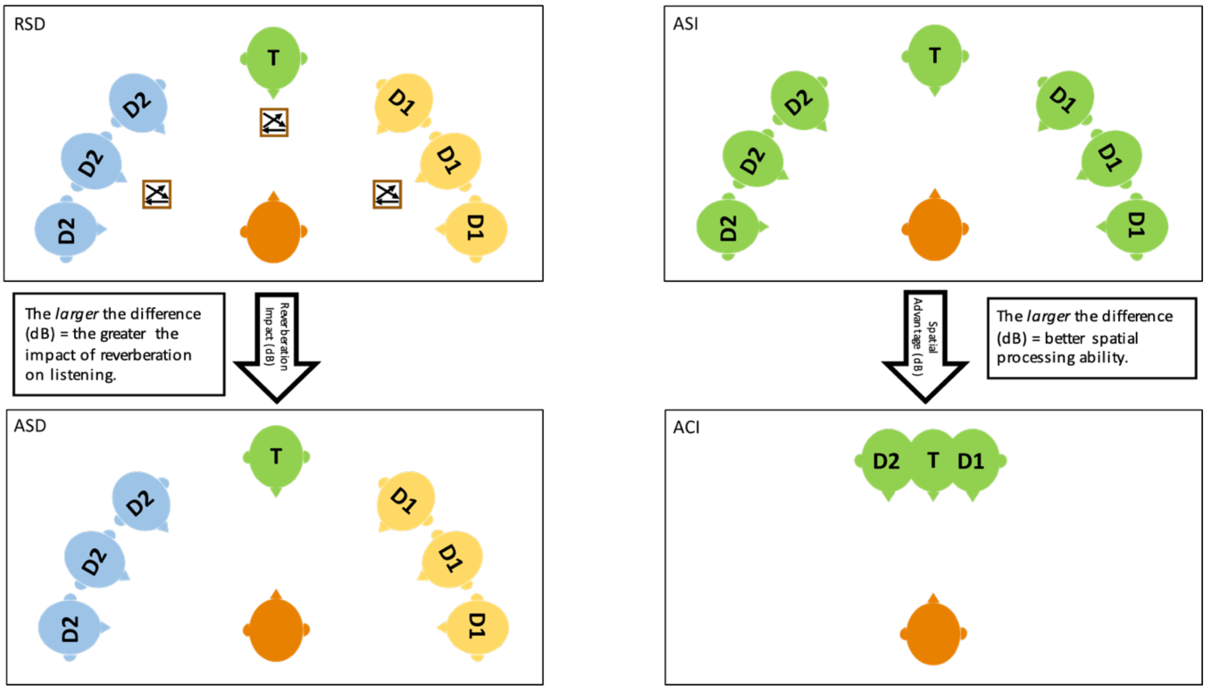


Figure 1.1 Visual graphic of each ToSSPiN baseline condition and calculation of the derived difference scores. Sources in the same colour are spoken by the same talker.

1.10 Development of the ToSSPiN Stimuli

1.10.1 CV Stimuli

The consonants chosen for this study comprised /p/ /b/ /t/ /d/ /k/ /g/ /m/ /n/ /s/ and /h/. The vowels chosen are dissimilar to each other in their formant frequencies. The vowels included were /i/ /a/ and /u/ and are a closed front vowel, an open front vowel, and a closed-back vowel respectively.

1.10.2 CV Recording, Editing and Level Normalization

Each consonant was paired with a vowel to create 30 CV pairs (e.g., /ti/ /ta/ /tu/). Three different native-Australian English female adults recorded each CV pair, in a general Australian accent, each labelled talker 1, 2 and 3 (T1, T2, and T3 respectively). Recording occurred in an anechoic chamber at Macquarie University. The CVs spoken by

T1 and T2 were recorded on a personal computer using Adobe Audition C5.6 as the digital audio workstation, a Rode NT1 cardioid condenser microphone (Silverwater, Australia) with a pop filter as the input device and an RME Babyface Pro USB as the audio interface (Haimhausen, Germany). The process used to record the CVs spoken by T3 was the same as above except that a Steinberg UR22mkII (Hamburg, Germany) was used as the audio interface. The stimuli were recorded with 24 bits and a 44,100 Hz sampling rate. Silence was removed from the start and end of each of the 30 individual CV audio files. The files have an average length of 415 ms (400 – 437 ms range).

In order to reduce the impact of headphone response variability on audibility, each CV audio file was band-pass filtered between 120 and 8000 Hz after recording. Each of the CVs was then level normalized to have a root mean square (RMS) of -20 dB FS (i.e. level relative to digital full scale).

1.10.3 Stimuli Adjustment Based on LiSN-U

Following editing and level-normalization, the output level of the individual ToSSPiN CV audio files recorded by T1, T2 and T3 were adjusted based on the psychometric functions generated from the LiSN-U stimulus intelligibility data (Cameron et al., 2020). Gains were applied to each individual consonant and vowel based on the LiSN-U data. The gains applied were equal to the difference between the relative SNR at 75% on the psychometric function minus the average of the relative SNRs at 75% across all phonemes. The adjustments ranged from +5.2 dB for /h/ to -2.4 dB for /u/. For each consonant, a constant gain was applied to each individual CV pair during the consonant-only portion of the CV. For each vowel, a constant gain was applied to each CV pair during the vowel-only portion of the CV. In between the gain varied linearly (in dB) during the transition between the consonant and vowel. Based on informal testing, an additional 0.75 dB was added to each target consonant to closer approach equal intelligibility of each phoneme.

1.10.4 Generation of HRTFs

In order to produce the ToSSPiN target stimuli and distractor tracks, the CVs recorded by T1, T2 and T3 were convolved with head-related-transfer-functions (HRTFs) recorded in both anechoic and reverberant environments. This section describes the generation of the HRTFs, while the following sections describe how the HRTFs were applied to the audio files from T1, T2 and T3 in order to produce the target stimuli and the distracter tracks.

The anechoic HRTFs were created from impulse responses recorded in an anechoic chamber at Macquarie University. The impulse responses were created from 20 Hz to 20,000 Hz logarithmic swept sine waves with a 44,100 Hz sampling frequency generated using a personal computer running MATLAB software (Mathworks Inc., 2016). Each sine-wave sweep was 15 seconds in length. The sine sweeps were presented from a single Tannoy loudspeaker model V8 BLK (United Kingdom) with a Yamaha AX-350 stereo amplifier (Japan). The centre of the loudspeaker was positioned two metres from the centre point of a Bruel & Kjaer Head and Torso Simulator (HATS) type 4128-C-001 (Denmark). An RME Fireface UC USB soundcard (Germany) was used as the audio interface between the computer, the HATS and the loudspeaker. The loudspeaker was re-located, and recordings were taken at 0°, ±45°, ±67.5° and ±90° azimuth. For each location, two sweeps were recorded and averaged.

The impulse responses generated to create the reverberant HRTFs were recorded in a 13 x 7.5 x 2.7 m room at Macquarie University. The recordings were made with the loudspeaker at three metres from the HATS. The average reverberation time (RT60) across the frequency range in all directions was 0.63 sec.

1.10.5 Target Generation

The ToSSPiN target stimuli were created from the CVs recorded by T1 and subsequently level-adjusted to approximate equal intelligibility. The 30 individual CV audio files were convolved with both the anechoic and reverberant HRTFs at 0^o azimuth using MATLAB. Note that in every ToSSPiN condition the target stimuli always emanate from 0^o azimuth. Each of the 30 target CV audio files are stored individually in the ToSSPiN program. During playback of each ToSSPiN condition, pairs of the convolved CVs are combined, generating 30 unique CVCV pseudo-words (e.g., /tigu/). The CVs used to generate the CVCV target words are selected at random by the software without replacement (i.e. CV1 cannot be the same as CV2). Each of the ten consonants occurs six times and the vowels 20 times so that each have equal occurrences. No gaps are present between the two CVs in each target word. A different, randomly selected set of 30 CVCV target tokens are generated each time the test is run.

1.10.6 Distractor Track Generation

The ToSSPiN distractor tracks were created from the 30 CV audio files recorded by T1, T2 and T3 using MATLAB and stored at an RMS level of -20 dB FS. The CVs from T1 were used to create the distractor tracks for conditions that do not include voice cues (ASI and ACI). The CV's from T2 and T3 were used for the conditions where the target speaker and distracting talkers are different (RSD and ASD). Each distracter track is comprised of 30 multi-syllable pseudo-words created from four CVs (i.e. CV1-CV2-CV3-CV4). For example, /ti gu ba da/. The CVCV pseudo-words were randomly generated using MATLAB, with each randomly selected CV occurring only once in each pseudo-word. The CVCV pseudo-words were concatenated and extended to 2200 ms by inserting silence after CV4 so that there is a gap between CV4 and CV1 of the following pseudo-word.

The tracks generated for the ACI condition (where the target and distracter speech is co-located) were convolved with HRTFs at 0^o azimuth. The tracks generated for the spatially separated conditions of the ToSSPiN (RSD, ASD and ASI) were created from a series of master tracks convolved with HRTFs recorded from + and - 45^o, 67.5^o and 90^o. In order to produce tracks featuring multiple locations. The distractor track emanating from the left side (i.e. negative azimuths) changes direction every five CVCVCVCV pseudo-words, in the order -45°, -67.5°, -90°. The track from the right side (i.e. positive azimuths) changes direction every five CVCVCVCV pseudo-words, in the order 67.5°, 90°, 45°.

The distractor tracks were looped continuously during playback. In order to ensure that gaps in each distractor track never overlap, the right-side distractor starts with a 250 ms silence and then a 5-CV pseudo-word. The left-side distractor starts immediately with a 4-CV pseudo-word. As such, the left side distractor track leads the right side track by 670 ms (on average).

The previous section has provided an overview on the development of the ToSSPiN software, including generation of the stimuli used. The administration of the

ToSSPiN is described in the following section outlining the methodology for the present study.

1.11 Summary

In order to measure an individual's AP abilities in many language contexts, there are several things to consider. First, speech assessments for AP should reflect real-world listening environments in which noise and reverberation simultaneously occur. Second, speech assessments that measure AP skills in people from various language backgrounds and accents should comprise speech sound material that is part of a listener's native language, for instance, shared CVs. The construction of pseudo-words should occur in a structure that is universally preferred (i.e. CV structure) due to better perceptual landmarks. Third, each item in the test should be as equally difficult as all other items in the test to increase test precision.

1.12 Project Aims

The primary aim of the present study is to determine the main effect of language on each ToSSPiN measure in three different native language groups. The ToSSPiN measures include the baseline SRT for each condition and the RI and SA measures. Additional measures that will be analysed to determine any effects of native language are response times, the variability of the adaptive track and the rate of learning each baseline condition. A further aim is to determine whether there is any difference in performance on the various ToSSPiN measures between face-to-face (FTF) and remotely tested participants. A final aim is to determine the relative intelligibility of each phoneme and adjust them so that, on average, each are identified 71% of the time at an identical signal-to-noise ratio.

The specific research questions addressed are:

1. Is there an effect of first language or accent on performance on any of the ToSSPiN baseline conditions? This question will determine whether language-independent normative data is appropriate. It is hypothesized there will be no effect of first language or accent on any ToSSPiN baseline condition.

2. Is there an effect of first language, or accent, on the reverberation impact and/or spatial advantage difference scores? If there is a significant effect of native language, language or accent-specific normative data may be required for the reverberation impact and spatial advantage difference scores. It is hypothesized there will be no effect of first language on either of the ToSSPiN difference scores.

3. Is there an effect of first language or accent on response times (RT msec) for ToSSPiN stimuli? Do non-native English and Canadian-English speakers have longer RTs compared to Australian-English speakers for ToSSPiN stimuli? It is hypothesized that there will be no significant difference between language groups on ToSSPiN response times.

4. Is there an effect of first language or accent on the rate of learning, measured as the gradual improvement in SRT for each ToSSPiN condition? It is hypothesized that there will be no significant difference in the rate of learning between language groups if there is no effect of language on baseline SRT.

5. Is there an effect of remote and face-to-face delivery modes on the ToSSPiN baseline SRTs and difference scores? It is hypothesized that there will be no significant difference in performance based on delivery modes.

6. Are all phonemes equally intelligible across all ToSSPiN conditions? It is hypothesized that as each phoneme has already been adjusted for equal intelligibility, minimal adjustments will be required to improve overall intelligibility.

2. Method

2.1 Participants

The participants comprised 83 adults aged 18 years, 9 months to 35 years, 6 months. There were 60 females and 23 males. A Human Research Ethics Application (HREA) was submitted and approved by the Macquarie University Human Research Ethics committee. Participants received a $30 gift card for their participation. Students at Macquarie University received a choice of a 5% mark for experiment participation toward their final grade for the semester or a $30 gift card.

This research project was initially designed to be delivered face-to-face. Due to COVID-19 and the high likelihood of transferrable viral infections, Macquarie University placed a ban on face-to-face data collection for most of the data collection phase. As a result, a convenience sampling method was used and a skewed distribution of subjects toward being remotely tested (n = 60). Participants recruited from Macquarie University were done so by advertising across different faculties to student notice boards. The researcher also contacted different universities across the world to recruit potential students. Professor Astrid van Wieringen from Ku Leuven University, Belgium, subsequently advertised the study to PhD students in the Department of Neurosciences who took part in the study. The study was also advertised on the Reddit subreddit r/audiology where almost half the sample was obtained. One user from Reddit offered to help with the recruitment process and subsequently advertised the study to friends and colleagues from her university in Canada. Non-English speakers were sampled from Australia, Belgium, Canada, Singapore, Hong Kong, and China. The limitations of the current sampling approach will be examined in the discussion.

Participants signed an information and consent form outlining the test procedure before completing the study. Participants were also asked to complete a short questionnaire as part of the information and consent form to determine their self-reported hearing status; age; native language; English proficiency (spoken and receptive); age at which English was learned; and proportion of their day spent communicating in English. The questionnaire is attached as Appendix A.

The participants were divided into three groups based on their native language: Australian English (n = 24), Canadian English (n = 25), and non-English (n = 34). The non-

English group comprised 15 Mandarin speakers, five Cantonese speakers, four Belgian Dutch (Flanders) speakers, three Malayalam speakers, three French speakers, one Russian speaker, one Korean speaker, one Vietnamese speaker, and one Filipino speaker.

2.2 Materials

For the remotely tested participants, the target and distractor tracks for the ToSSPiN were presented using an Apple iPad running at least version 10 of iOS. A variety of consumer-level circumaural or in-ear headphones were used for remote participants. Face-to-face participants were tested with an Apple iPad Pro 1st generation running iOS 13.6 and Sennheiser HD 200 Pro circumaural headphones (Hanover, Germany).

The ToSSPiN app incorporated a separate audio level calibration screen to overcome a potential problem of varying output volumes with different headphones types for remotely tested participants. The participants were asked to play a calibration track with CVs whose rms levels spanned a dynamic range of 21 dB and adjust the volume of the iPad to their most comfortable listening level. The calibration track was based on a modified version of the ASD distractor track with each CV in a CVCV pseudo-word being at a level of 0, -7, -14 and -21 dB respectively. A comfort-adjustment stimulus with CVs varying in level was used to minimise the risk of participants setting constant-level stimuli to the top or bottom of their comfort range. If the level of the stimuli was at the top or bottom of a participant's comfort range, it could create problems if the test stimuli adapted to levels higher or lower than this, respectively.

2.3 Procedures

Due to COVID-19 social distancing restrictions, 60 participants were tested remotely and 23 were tested face-to-face (FTF). The ToSSPiN is an app-based program delivered on an iPad, therefore, testing of participants could occur remotely. The processes for testing FTF and remote participants are described below.

As pure tone audiometry could not be performed for remotely tested participants, hearing screening was undertaken for all participants via self-report. Potential participants reporting hearing difficulties were excluded from the study. Additionally, as part of the questionnaire, each participant was asked to report any hearing difficulties. All 83 participants reported no hearing difficulties.

For remote participants, scoring of the ToSSPiN was performed by a 'test partner' who lived in the same residence as the participants. The researcher remotely monitored each session via Zoom. Participants and their test partner were given detailed instructions on how to operate the ToSSPiN application before participating (see Appendix B). All participants were asked to familiarise themselves with operating the software but not complete the test before the scheduled test session. In most instances, the helper and test partner would reverse roles after the first participant completed the test.

In total, two participants' data were removed from the analysis. One had completed the test the previous day unsupervised. Another participant's test partner did not speak sufficient English to understand how to score the test. Data from the remaining 83 participants are included in the Results section. Participant details and group allocation information are provided in Table 3.1.

2.4 Data Collection

2.4.1 Target CVCV Presentation

For all ToSSPiN conditions, the CVCV target stimuli were presented once the tester pressed 'start' or 'next item' on the ToSSPiN app's graphical user interface. A 200 ms, 1 kHz cue tone was presented at +5 dB relative to the distractors, followed by a 300 ms silence. The CVCV target was then presented, and a 200 ms gap followed. The target was then repeated. For example: Cue tone – 300 ms gap - /tigu/ - 200 ms gap - /tigu/.

2.4.2 Scoring

On each trial, the participants received one point per consonant correctly identified and one point per vowel correctly identified. Each phoneme had to be repeated in the correct position, or no point was given for that phoneme. A maximum of four points was given per trial. The only instance in which phonemes repeated in the wrong order were not scored incorrectly was for complete reversals of CV1 and CV2. Each pseudo-word was presented twice and there was, therefore, a possibility to hear the word in reversed order at a participant's approximate threshold. A participant might have missed the first syllable of CV1 and the second syllable of CV2. Built into the ToSSPiN app is a 'correct but reversed' button for such a reversal. When pressed, a score of three is given for the trial, so the SNR remains unchanged. Scoring occurred adaptively with step sizes targeting the SNR (in dB) at which the participant identified 71% of the targets correctly, as shown in Table 2.2.

Table 2.2 Score per trial and adaptive step size for each phase and test condition. The transition from -4 dB to -2 dB in RSD, ASD and ASI occurs when the measurement phase begins, and at least 10 trials have occurred. Only the scoring for all phonemes correct varies between conditions and phases.

Score for trial	**Step size (dB)**		
	Spatially separated (RSD, ASD, ASI)		**Co-located (ACI)**
	Practice and non-measurement phase	**Measurement phase**	**All phases**
0	+3	+3	+3
1	+2	+2	+2
2	+1	+1	+1
3	0	0	0
4	-4	-2	-2

2.4.3 Practice and Test Conditions

2.4.3.1 Familiarization Phase

Before commencing the ToSSPiN, participants were tested in a familiarisation phase. The purpose of the familiarisation phase was to ensure each participant understood and identified each of the CVs needed to complete the test. Five CVCV pseudo-words containing all ten consonants were created. Each vowel occurred at least three times. Phonemes that were incorrectly identified were repeated until the participant identified them correctly. Any participant who could not correctly repeat all CVs during this phase was to be excluded from data analysis. All participants in this study achieved a satisfactory level of performance during the familiarisation phase.

2.4.3.2 Practice Phase

At the beginning of each condition, five CVCV practice trials occurred before the test phase. The practice trials aimed to familiarise each participant with each condition and to ensure the task was understood. The presentation level of the distractor tracks was five dB below the level of the most intense CV in the comfort adjustment stimulus. The initial presentation level of the targets was +10 dB SNR relative to the distractors. Scoring occurred adaptively during practice trials, however, did not contribute to the participant's overall SRT. The targets were presented randomly during practice.

2.4.3.3 Test Phase

After the five practice trials for each condition, the test phase was initiated. The initial playback level for the ToSSPiN targets in each condition of the test phase was +10 dB SNR relative to the distractors. The test phase for each condition began with a non-measurement phase (NMP). Any changes in SNR during the NMP were not used in the calculation of a participant's speech reception threshold. A measurement phase (MP) began in the RSD, ASD and ASI conditions when at least six trials had occurred, and a participant correctly identified two phonemes or less on a single trial or three phonemes on two consecutive trials. For the ACI condition, the minimum number of trials was four instead of six. A reduced number of trials were required for the ACI condition as fewer responses typically occur, relative to the spatialized conditions, for a participant to achieve their SRT.

Each consonant and vowel occurred an equal number of times for a total of 30 CVCV pseudo-word presentations. The participants were encouraged to give a response, even if it was a guess or they were unsure. If a participant did not hear the target at all, they were instructed to say 'pass'.

2.4.4 Calculation of ToSSPiN Measures

2.4.4.1 Target Step Sizes

As discussed in the previous section on scoring, the ToSSPiN uses an adaptive scoring procedure with unequal step sizes depending on the phase of the test, and the number of phonemes correctly identified (see Table 2.2). During the practice phase and

NMP for the RSD, ASD and ASI conditions, the SNR decreased by four dB if all phonemes were correctly identified. The SNR remained constant for the next trial if three phonemes were correctly identified. The SNR increased by one dB if two phonemes were correctly identified. If one phoneme was correctly repeated, the SNR increased by two dB. If a participant did not repeat any phonemes, the SNR increased by three dB.

In the MP for spatialized conditions and for all phases for the ACI condition, the step sizes remained the same as the NMP except if all four phonemes were correctly identified, in which case the SNR decreased by two dB and not four dB. Spatial and voice cues are removed in the ACI condition and typically result in a higher SRT, therefore, a two dB decrease was used in the NMP and MP phases of the ACI condition.

2.4.4.2 Calculation of the Speech Reception Threshold

At the end of the 30 trials for each condition, the participant's SRT (in dB) was calculated. The SRT was calculated as the mean level of the targets in the MP, minus the level of the distractors. As explained later, the SRT calculation corresponds approximately to the SNR at which 71% of phonemes are correctly perceived.

2.4.4.3 Calculation of Difference Scores

Two difference scores were calculated to identify the impact of reverberation and spatial processing abilities. The difference scores are measured as the difference in dB between two baseline conditions that vary in one auditory element only. Differential testing aims to reduce the impact of cognition and language deficits on ToSSPiN performance as these two aspects remain constant between each condition.

The reverberation impact (RI) measure aimed to identify a participant's ability to suppress reverberation and improve the clarity of the target speaker. Reverberation Impact was measured as the difference in SRT (dB) when the target is presented in an anechoic versus a reverberant environment. The RI was calculated as the SRT (dB) in the RSD condition minus the SRT (dB) in the ASD condition. A smaller difference in dB between each condition indicates a smaller impact of reverberation on test performance.

The spatial advantage (SA) measure aimed to measure a participant's spatial processing abilities. The SA is a measure of a participant's ability to use interaural cues (level and timing differences) and better-ear glimpsing to separate auditory streams to identify the target speaker. The SA was measured as the difference in SRT (dB) when the target and distractors were spatially separated (i.e., ±45°, ±67.5°, ±90° azimuth) versus co-located (i.e., 0° azimuth). The SA was calculated as the SRT (dB) in the ASI condition minus the SRT (dB) in the ACI condition. A larger difference in dB between each condition indicates better spatial processing abilities.

2.4.4.4 Calculation of Response Time

Response times in milliseconds (msec) were recorded for each trial. The response times per trial were measured as the interval (msec) from the trial end to when the scorer enters the first phoneme. The iPad has an inherent limitation of recording the response times. The iPad requires a cue to begin timing (i.e. the end of the stimulus presentation) and a cue to stop recording the response time (i.e. when the clinician scores the first phoneme). The response times are therefore, a combination of the participant's response time to the stimuli, and the scorer's response time to scoring the presentation. The response times used for data analysis were the average response times in the MP. Average response times were calculated for each ToSSPiN condition.

2.4.4.5 Results and Report Generation

An initial results screen displayed each participant's SRT, response time and adjusted standard error (aSE) for each baseline condition. The aSE is based on the variability in the SNR within the adaptive track (Cameron & Dillon, 2007b). The aSE is calculated as 2*SD/√N:

- where N is the number of trials from and including the first trial of the MP;
- SD is the standard deviation of a given trial's SNR around the mean SNR, and;
- The number two is an empirical adjustment that allows for the non-independence of items in the adaptive track.

The RI and SA were also displayed on the initial results screen. By pressing 'next' on the initial results screen, a second results screen provided detailed graphs of the adaptive track for each ToSSPiN condition. On the second results screen, a button allowed the user to email each participant's report as a .pdf.

2.5 Data Analysis

Data analysis was completed using SPSS Statistics Version 27. An examination of the effects of native language on a participant's SRT and response times for RSD, ASD, ASI and ACI was completed to satisfy the project's aims. Further analyses included examining the effects of native language on a participant's reverberation impact (RI), spatial advantage (SA) and average response times for each condition.

For this study, there was a diverse range of participants from different language backgrounds who were tested both face-to-face and remotely. A further analysis investigating the effects of delivery mode on the SRT for RSD, ASD, ASI and ACI and the derived RI and SA was completed.

To support the findings for this study, an additional analysis investigating the performance variability within the adaptive track was completed. The analysis includes examining the effects of native language and delivery mode on the aSE for each baseline condition, RI and SA measures.

3. Results

Table 3.1 details the participant numbers for each language group and delivery mode. This chapter initially analyses the effect of language group and delivery mode on the baseline SRTs and difference scores. The chapter also analyses the effect of language group and delivery mode on response times, variability in the adaptive track and rate of learning.

Table 3.1 Sample sizes based on language group and delivery method.

		Delivery Mode		
		Face to Face (*n*)	**Remote (*n*)**	**Total (*n*)**
Language Group	**Australian English (*n*)**	14	10	24
	Canadian English (*n*)	0	25	25
	Non English (*n*)	9	25	34
	Total (*n*)	23	60	83

3.1 Effect of Native Language and Delivery Mode on Baseline Measures

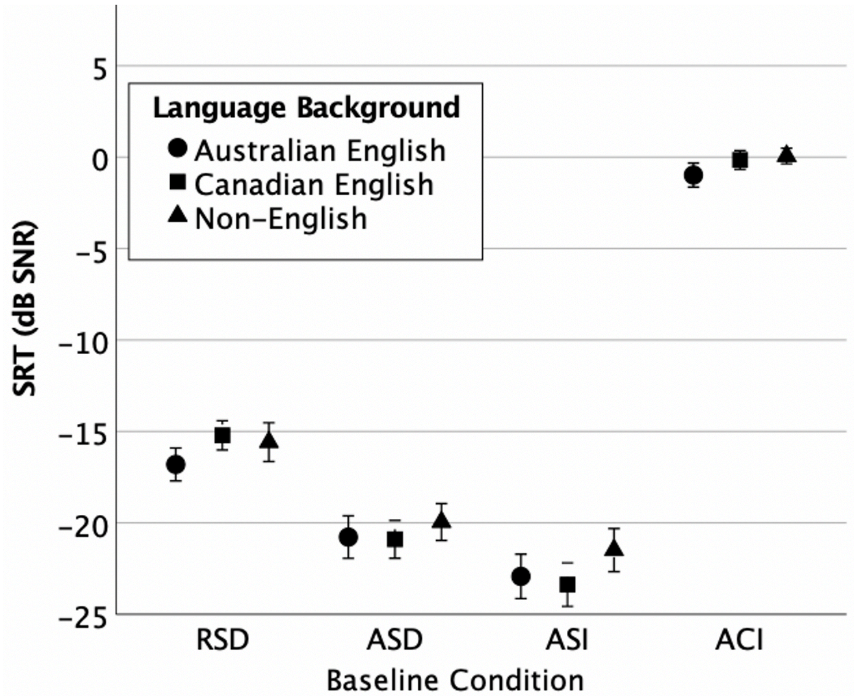
Figure 3.1 details the mean SRTs (in dB SNR) across each language group for the RSD, ASD, ASI, and ACI ToSSPiN conditions. The lowest (i.e. best) mean SRTs were observed in the ASD and ASI conditions. A lower mean SRT is to be expected in the ASD and ASI conditions due to having spatial cues available and being presented in an anechoic listening environment. Mean SRTs were slightly higher in the RSD condition indicating that the addition of reverberation reduces speech sound identification at reduced SNRs. For the ACI condition, a considerable reduction in SRT was observed.

A three-way, analysis of variance (ANOVA) with two between-groups factors (language group and delivery method) and one repeated-measures factor (baseline condition) was conducted to estimate the effect of these variables on speech reception threshold (SRT). The results of the ANOVA are shown in Table 3.2. There was no significant main effect on the baseline SRTs of either language $F(2, 79) = 2.27$, $p = 0.11$ or delivery mode, $F(1, 79) = 2.76$, $p = 0.10$. There was, as expected, a significant effect for baseline condition, $F(3, 237) = 1734$, $p<0.000001$.

Table 3.2 Language group and delivery mode results of three-way ANOVA for SRT in ToSSPiN.

	SS	*df*	*MS*	*F*	*p*
Intercept	55989.79	1	55989.79	4104.93	0.00
Delivery Mode	37.59	1	37.59	2.76	0.10
Language Group	62.00	2	31.00	2.27	0.11
Error	1077.53	79	13.64		
Baseline Condition	18957.46	3	6319.15	1734.64	0.00
Baseline Condition *Delivery Mode	27.07	3	9.02	2.48	0.06
Baseline Condition *Language Group	31.42	6	5.24	1.44	0.20
Error	863.37	237	3.64		

The effect of language group on SRT in each of the four baseline conditions is shown in Figure 3.1. The ANOVA shows that there was no significant interaction between language group and baseline condition SRT $F(6, 237) = 1.44$, $p = 0.20$.



<u>Figure 3.1</u> Mean SRT (in dB SNR) for each language group across ToSSPiN RSD, ASD, ASI and ACI conditions. Bars represent 95% confidence intervals.

The effect of delivery mode in each baseline condition is shown in Figure 3.2. A weak interaction between delivery mode and baseline can be observed, with remotely tested participants requiring a slightly poorer SNR than for FTF delivery in the RSD condition only. However, the interaction failed to reach statistical significance $F(3, 237) = 2.48$, $p = 0.06$.

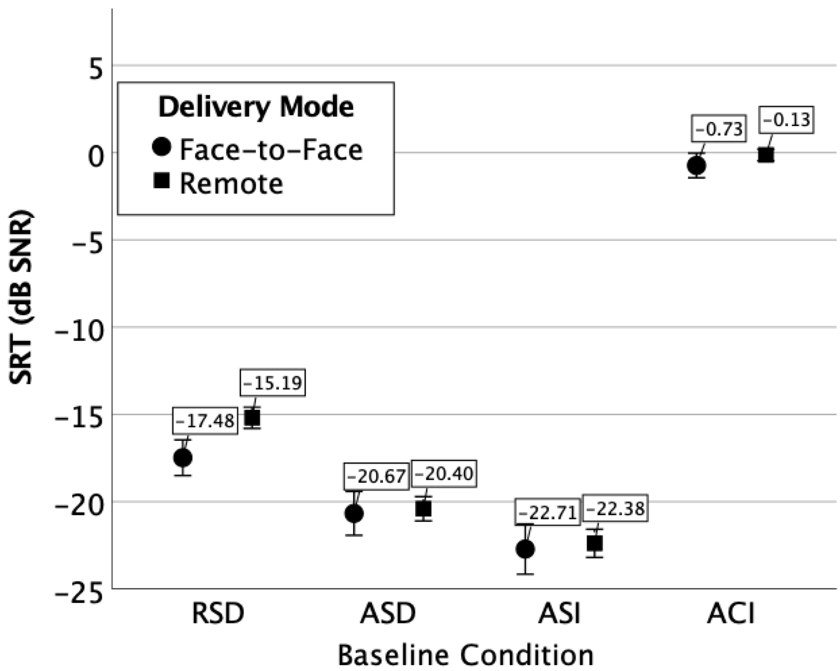


Figure 3.2 Mean SRT (in dB SNR) for delivery mode across ToSSPiN RSD, ASD, ASI and ACI conditions. Bars represent 95% confidence intervals.

3.2 Effect of Native Language and Delivery Mode on Reverberation Impact and Spatial Advantage.

Figure 3.3 and 3.5 detail the means for the two difference scores, reverberation impact (RI) and spatial advantage (SA), for each language group. Separate two-way ANOVAs, with language group and delivery method as between-groups factors, were conducted for each of RI and SA as the dependent variables expressed in dB.

The RI measure is the difference in SRT between the RSD and ASD conditions and represents the increased listening difficulty when reverberation is present. A lower RI score indicates a lower impact of reverberation. The mean RI is considerably lower than the SA. The SA measure is the difference in SRT between ASI and ACI conditions and represents the amount of benefit an individual receives when the distractors are spatially separated from the target versus co-located with the target. A larger SA score indicates an increased benefit of spatial separation.

The results of the ANOVA for RI is shown in Table 3.3. There was no significant main effect of language $F(2, 79) = 0.88$, $p = 0.42$. There was a significant main effect of delivery mode on RI $F(1, 79) = 6.5$, $p = 0.01$. On average, remotely tested participants had a significantly higher mean RI than participants tested face-to-face.

Figure 3.4 and 3.6 detail the mean RI and SA based on face-to-face and remote delivery modes. On average, the RI impact was higher for remote participants $M = 5.21$ (SD = 2.53) than for face-to-face participants $M = 3.19$ (SD = 1.93). The SA was comparable for both face-to-face and remote groups. The face-to-face group had an average SA of 21.98 dB (SD = 2.67), the remote group SA was $M = 22.25$ dB (SD = 3.04). The effects of delivery mode on RI and SA are discussed below.

Table 3.3 Language group and delivery mode results of ANOVA for RI.

	SS	***df***	***MS***	***F***	***p***
Intercept	1146.38	1	1146.38	200.87	0.00
Language Group	9.99	2	5.00	0.88	0.42
Delivery Mode	37.29	1	37.29	6.53	0.01
Language Group * Delivery Mode	1.54	1	1.54	0.27	0.61
Error	450.87	78	5.71		

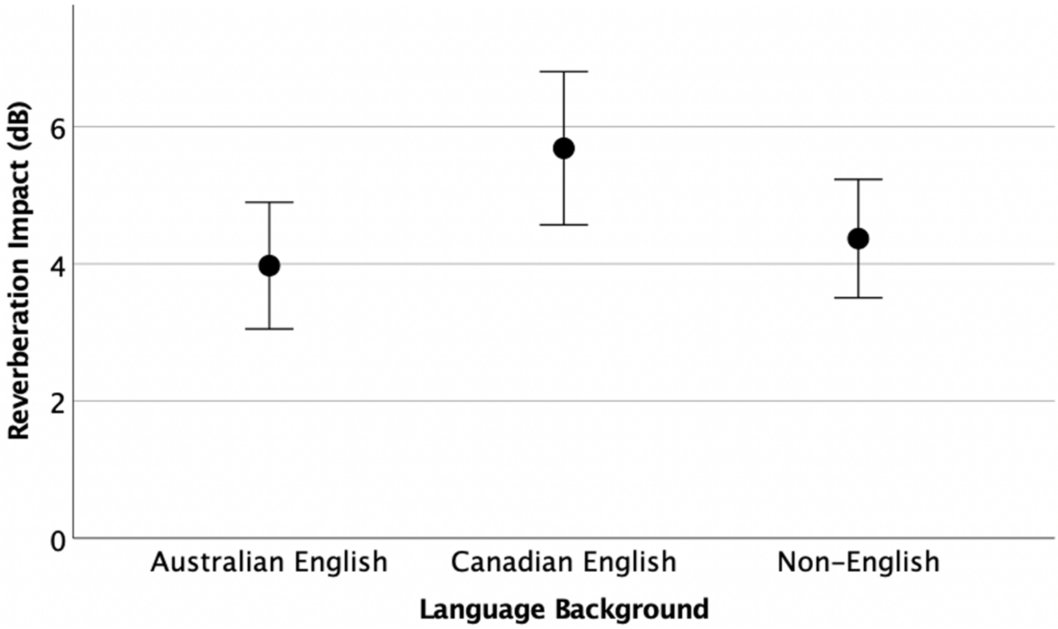


Figure 3.3 Mean RI (in dB) for non-English, Australian English and Canadian English groups. Bars represent 95% confidence intervals.

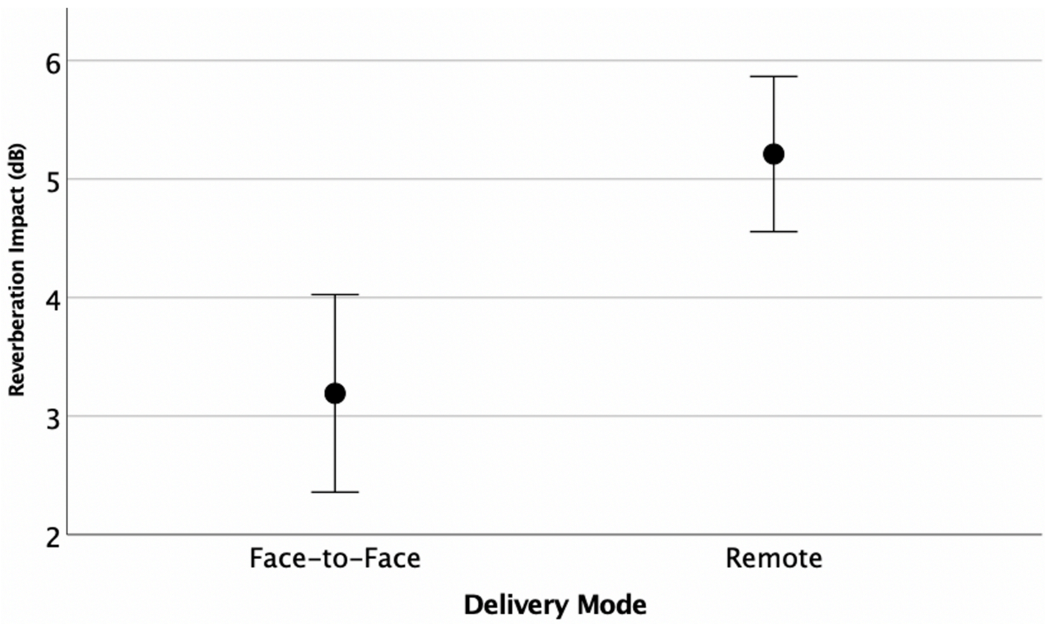


Figure 3.4 Mean RI (in dB) for FTF and Remote delivery methods. Bars represent 95% confidence intervals.

The results of the ANOVA for SA is shown in Table 3.4. There was no significant main effect of language averaged on the SA measure $F(2, 79) = 2.52$, $p = 0.09$. There was also no significant main effect of delivery mode on the SA measure $F(1, 79) = 0.17$, $p = 0.68$. Figure 3.5 shows the mean SA across the language group, averaged across delivery methods. Figure 3.6 shows mean SA for each delivery mode, averaged across the three language groups.

Table 3.4 Language group and delivery mode results of ANOVA for SA.

	SS	*df*	*MS*	*F*	*p*
Intercept	30865.55	1	30865.55	3689.96	0.00
Language Group	42.08	2	21.04	2.52	0.09
Delivery Mode	1.40	1	1.40	0.17	0.68
Language Group * Delivery Mode	17.25	1	17.25	2.09	0.15
Error	660.81	78	8.36		

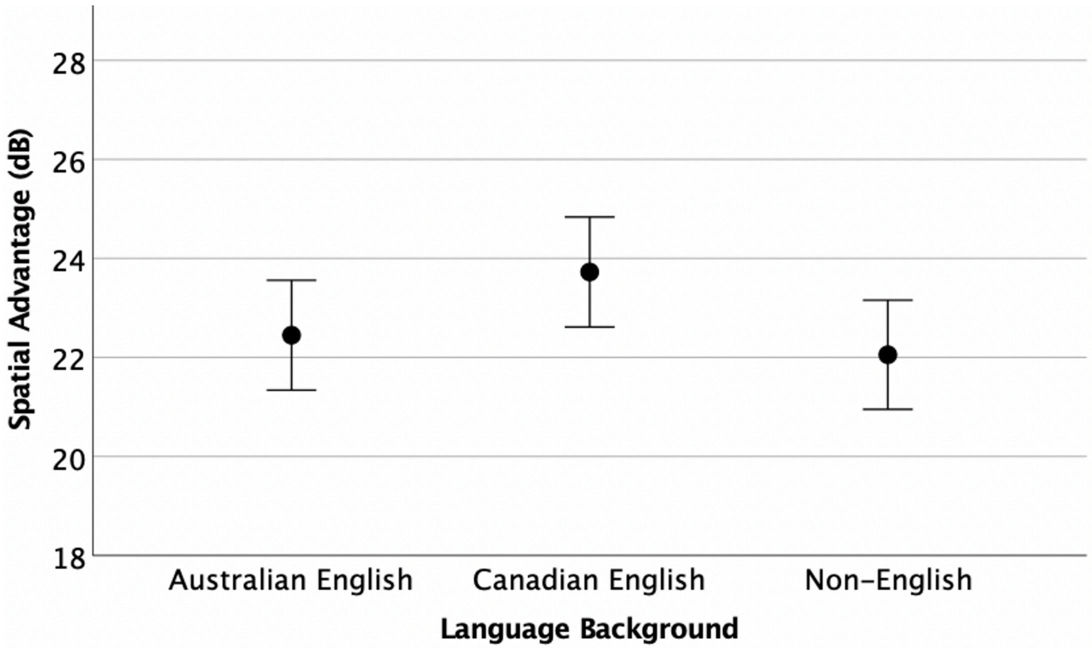


Figure 3.5 Mean SA (in dB) for non-English, Australian English and Canadian English groups. Bars represent 95% confidence intervals.

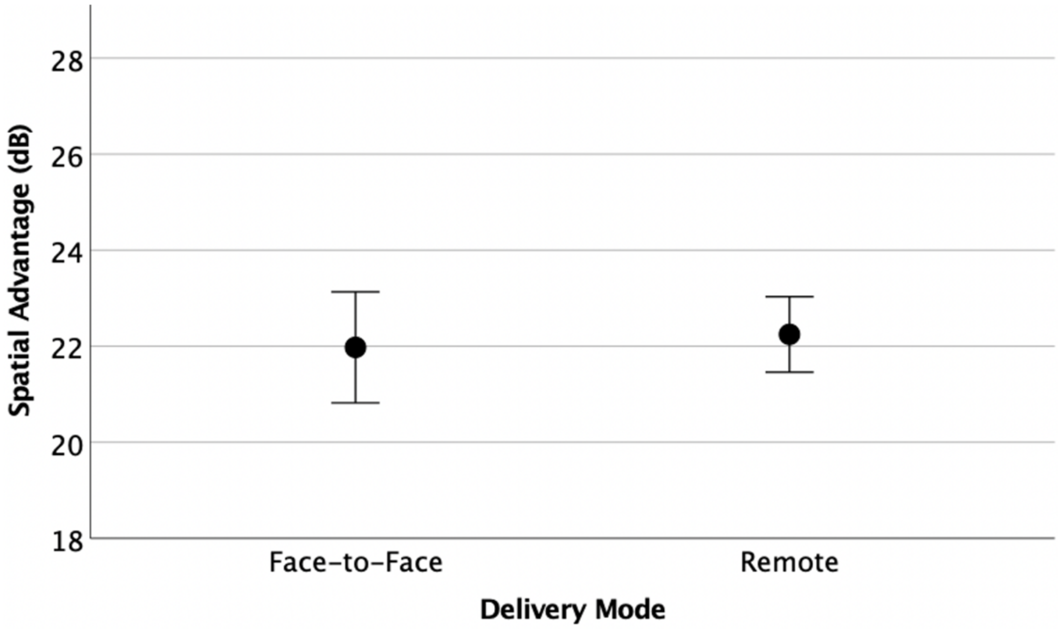


Figure 3.6 Mean SA (in dB) across face-to-face and remote delivery modes. Bars represent 95% confidence intervals.

3.3 Effect of Language Group and Delivery Mode on Response Time

Figure 3.7 provides detailed means (in sec) for the response times (RT) of each language group in the RSD, ASD, ASI and ACI ToSSPiN conditions. A three-way ANOVA with two between-groups factors (language group and delivery method) and one repeated-measures factor (baseline condition) was conducted to examine the effect of language group, delivery method and baseline condition on response time. The ANOVA results are shown in Table 3.5. Remote delivery resulted in significantly longer response times than FTF delivery. Language group did not significantly affect response time.

Table 3.5 Language group and delivery mode results of three-way ANOVA for response time (RT).

	SS	*df*	*MS*	*F*	*p*
Intercept	1082.95	1	1082.95	709.57	0.00
Language Group	4.00	2	2.00	1.31	0.28
Delivery Mode	24.64	1	24.64	16.15	0.00
Error	120.57	79	1.53		
Baseline Condition	1.04	3	0.35	2.65	0.05
Baseline Condition *Language Group	1.34	6	0.22	1.71	0.12
Baseline Condition *Delivery Mode	0.85	3	0.28	2.16	0.09
Error	31.00	237	0.13		

There was no significant interaction between baseline condition and language groups $F(6, 237) = 1.71$, $p = 0.12$. Figure 3.7 shows mean RTs of each language group across each of the baseline conditions.

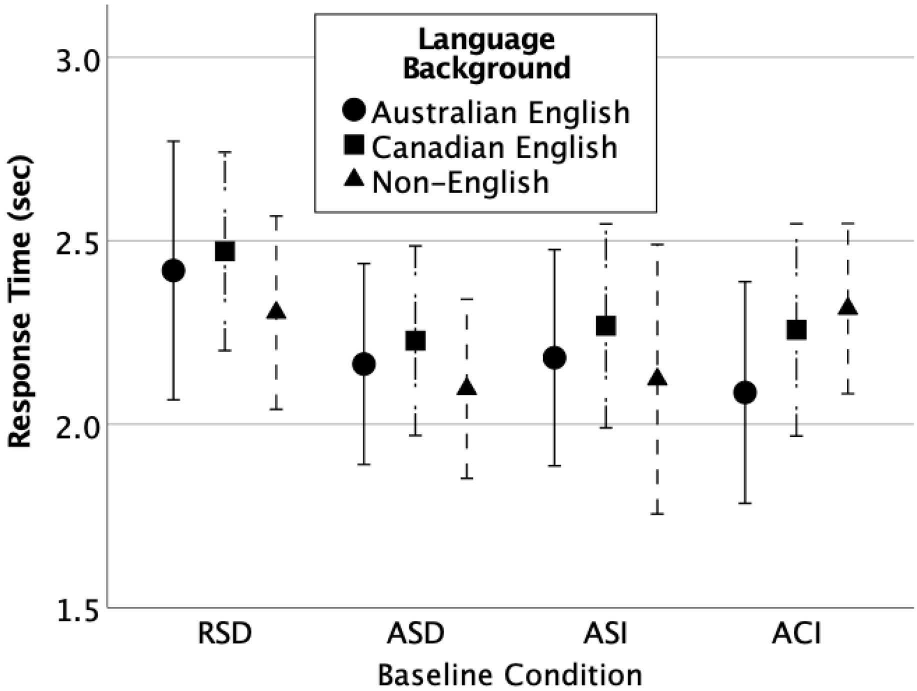


Figure 3.7 Mean RTs (in sec) for each language group across each baseline condition. Bars represent 95% confidence intervals.

3.4 Examination of Gradual Learning for Each Language Group Across Baseline Conditions

The slope of the adaptive track in the measurement phase was analysed to examine the rate of learning in each baseline condition. The slope of the adaptive track is quantified as the average change in SNR per trial (dB per trial) from the start of the measurement phase up to and including trial 30.

Figure 3.8 details the means and standard deviations for the average change in dB per trial for each language group across each of the baseline conditions. Table 3.6 shows the mean measurement phase trials for each language group across each baseline condition, and the total change in dB (slope) from the start of the measurement phase to the end of each condition.

The change in dB per trial averaged across all 83 participants was comparable for the RSD, ASD and ACI conditions M = -0.15 (SD = 0.25), M = -0.18 (SD = 0.27), and M = -0.21 (SD = 0.28) respectively. The average change in dB per trial for the ACI condition averaged across all 83 participants was considerably lower than the other three conditions M = -0.03 (SD = 0.14).

Table 3.6 Mean number of trials in the measurement phase and total change in dB (i.e. the slope times the number of trials in the measurement phase) from the start to the end of the measurement phase for each baseline condition for each language group.

Language Group	Baseline Condition	Mean (number of trials)	Total Change in dB	*n*
Australian English	RSD	22.29	-2.90	24
	ASD	21.63	-3.46	
	ASI	21.25	-4.89	
	ACI	23.75	-0.95	
Canadian English	RSD	23.00	-4.60	25
	ASD	21.56	-4.31	
	ASI	20.91	-4.81	
	ACI	24.20	-0.24	
Non-English	RSD	22.41	-2.69	34
	ASD	21.79	-4.36	
	ASI	21.44	-3.43	
	ACI	24.00	-0.48	

The results of the ANOVA are shown in Table 3.8. A three-way ANOVA with two between-group factors (language group and delivery mode) and one repeated measures factor (baseline condition) was conducted to compare the slope of the adaptive track (in dB per trial) in each baseline condition. The results of the ANOVA showed no significant effect of language or delivery mode on the slope of the adaptive track $F(2, 79) = 0.82$, $p = 0.44$. Overall, each of the language groups showed comparable slopes for each of the baseline conditions. Figure 3.8 shows a graphical representation of the mean slope of the adaptive track for each language group across each baseline condition.

Table 3.8 Language group and delivery mode results of three-way ANOVA for the slope of adaptive track in each ToSSPiN condition.

	SS	*df*	*MS*	*F*	*p*
Intercept	5.71	1	5.71	92.58	0.00
Language Group	0.10	2	0.05	0.82	0.44
Delivery Mode	0.13	1	0.13	2.10	0.15
Error	4.87	79	0.06		
Baseline Condition	1.46	3	0.49	8.13	0.00
Baseline Condition *Language Group	0.22	6	0.04	0.62	0.72
Baseline Condition *Delivery Mode	0.03	3	0.01	0.19	0.90
Error	14.17	237	0.06		

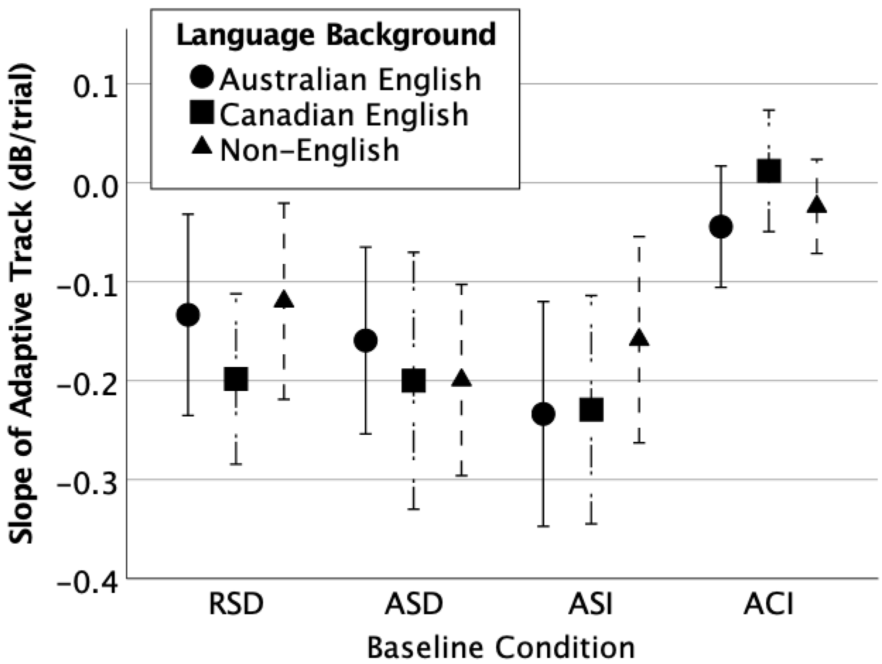


Figure 3.8 Mean change in dB per trial, representative of the slope of the adaptive track for each language group across each baseline condition. Bars represent 95% confidence interval.

3.5 Examination of the Variability in Performance for Each Language group.

An additional analysis was completed to examine the performance variability for each language group and delivery mode. The aSE characterises the trial-to-trial variability in the adaptive track for each baseline condition. Figure 3.9 details the mean aSE (in dB) of the aSE values across individuals, for each language group and baseline condition. Table 3.9 shows the mean aSE based on delivery mode across each of the baseline conditions.

Table 3.9 Mean aSE (in dB) and standard deviations for each baseline condition based on delivery mode.

Delivery Mode	Baseline Condition	Mean (aSE in dB)	Std. Dev.	*n*
Face-to-Face	RSD	1.10	0.33	23
	ASD	1.01	0.38	
	ASI	1.16	0.51	
	ACI	0.85	0.22	
Remote	RSD	1.00	0.34	60
	ASD	1.10	0.41	
	ASI	1.08	0.39	
	ACI	0.86	0.17	

Table 3.10 shows the results of the ANOVA. A three-way ANOVA with two between-groups factors (language group and delivery method) and one repeated-measures factor (baseline condition) was conducted to compare the adjusted standard errors (aSE) in each baseline condition.

Table 3.10. Language group and delivery mode results of three-way ANOVA for aSE in ToSSPiN.

	SS	*df*	*MS*	*F*	*p*
Intercept	262.41	1	262.41	2315.16	0.00
Language Group	0.22	2	0.11	0.96	0.39
Remote	0.09	1	0.09	0.81	0.37
Error	8.95	79	0.11		
Baseline Condition	2.61	3	0.87	6.83	0.00
Baseline Condition *Language Group	0.27	6	0.05	0.36	0.91
Baseline Condition *Delivery Mode	0.38	3	0.13	0.98	0.40
Error	30.22	237	0.13		

There was no significant main effect of language group on baseline aSE $F(2, 79) = 0.96$, $p = 0.39$. The was no significant main effect of delivery mode on baseline aSE $F(1,$

79) = 0.81, *p* = 0.37. There was no significant interaction between baseline condition and language group *F*(6, 237) = 0.36, *p* = 0.91 or delivery mode *F*(3, 237) = 0.98, *p* = 0.40.

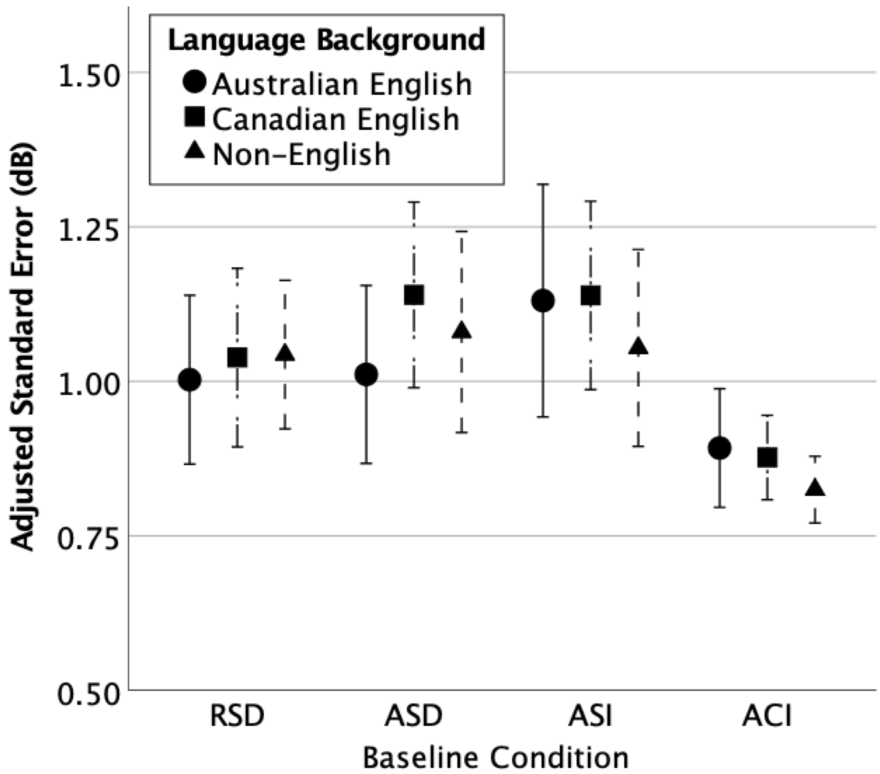


Figure 3.9 Mean aSE for each baseline condition across each language group. Bars represent 95% confidence interval.

3.6 Post-Study Target Stimuli Adjustment

By calculating percentage correct as a function of the SNR of each trial relative to the SRT found for that person in that baseline condition (i.e. the relative SNR), we can compute psychometric functions for the test overall, or for individual consonants and vowels. At the person's SRT, there were 71% of phonemes correct. Points at a relative SNR less than -5 dB were based on very few observations (i.e. the adaptive track only infrequently went more than 5 dB below the person's SRT) and were not taken into account in fitting the psychometric function. Figure 3.10 shows the percentage correct of all phonemes as a function of the relative SNR. Figure 3.11 shows the percentage correct of individual consonants as a function of the relative SNR. Figure 3.12 shows the percentage correct of individual vowels as a function of the relative SNR. The 'x' points on each figure represent the percentage correct (i.e. overall or individual phonemes) at each relative SNR, averaged across all participants. From these psychometric functions, the overall adjustment required for individual phonemes could be obtained by determining the difference between the relative SNR for 71% correct for one item and the relative SNR for 71% correct for the average of all items.

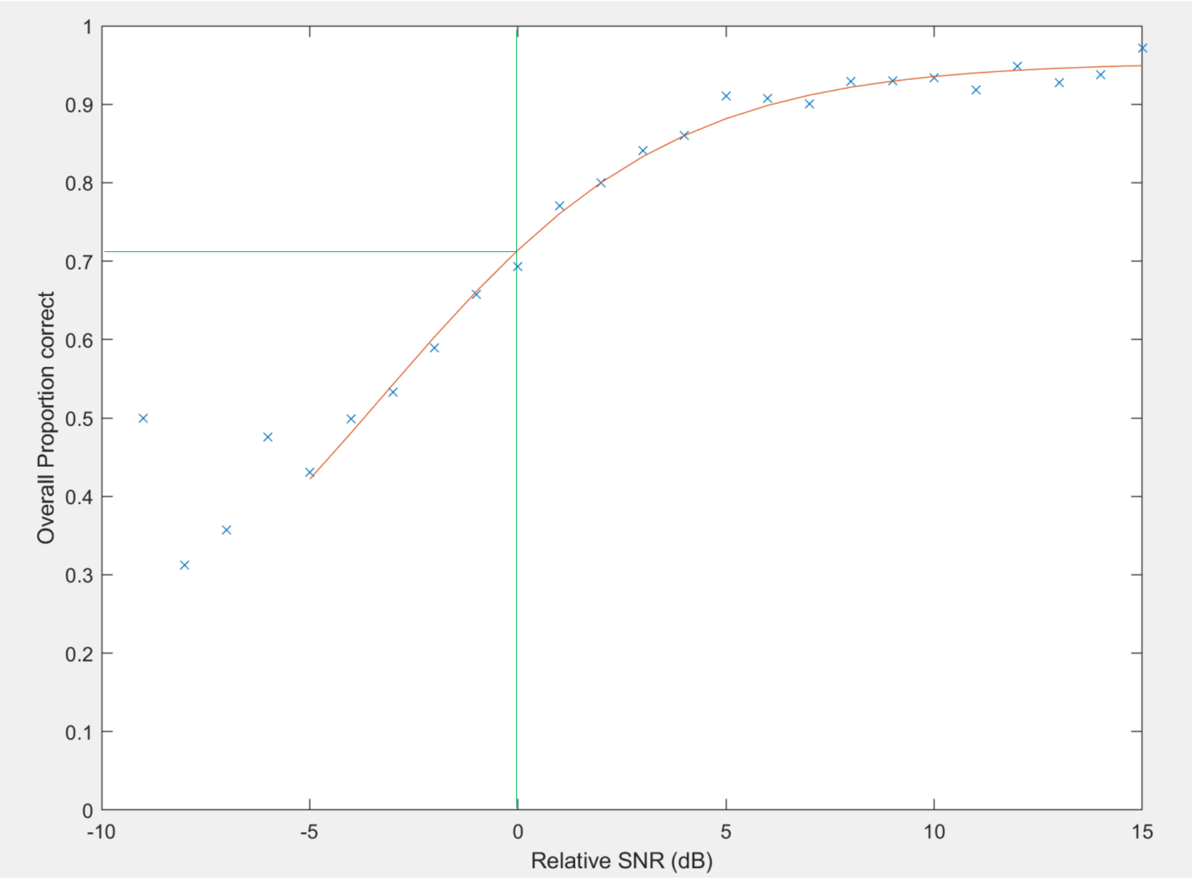


Figure 3.10 Overall psychometric function for the ToSSPiN phonemes. A relative SNR of 0 dB means that the SNR was equal to the SRT found for that person for that baseline condition.

Each graph has vertical lines showing the relative SNR that would be needed were each phoneme to give 71% correct. These values are the adjustments that must be made to achieve equal intelligibility for each phoneme. The individual adjustments for each of the consonants were; +2.3 dB for /p/; -2.0 dB for /t/; +1.5 dB for /k/; +3.3 dB for /b/; +2.1 dB for /d/; +4.0 dB for /g/; +5.4 dB for /m/; +2.3 dB for /n/; -4.8 dB for /s/ and +8.7dB for

/h/. The individual adjustments for each of the vowels were; - 4.4 dB for /a/; -1.8 dB for /i/ and -0.7 dB for /u/.

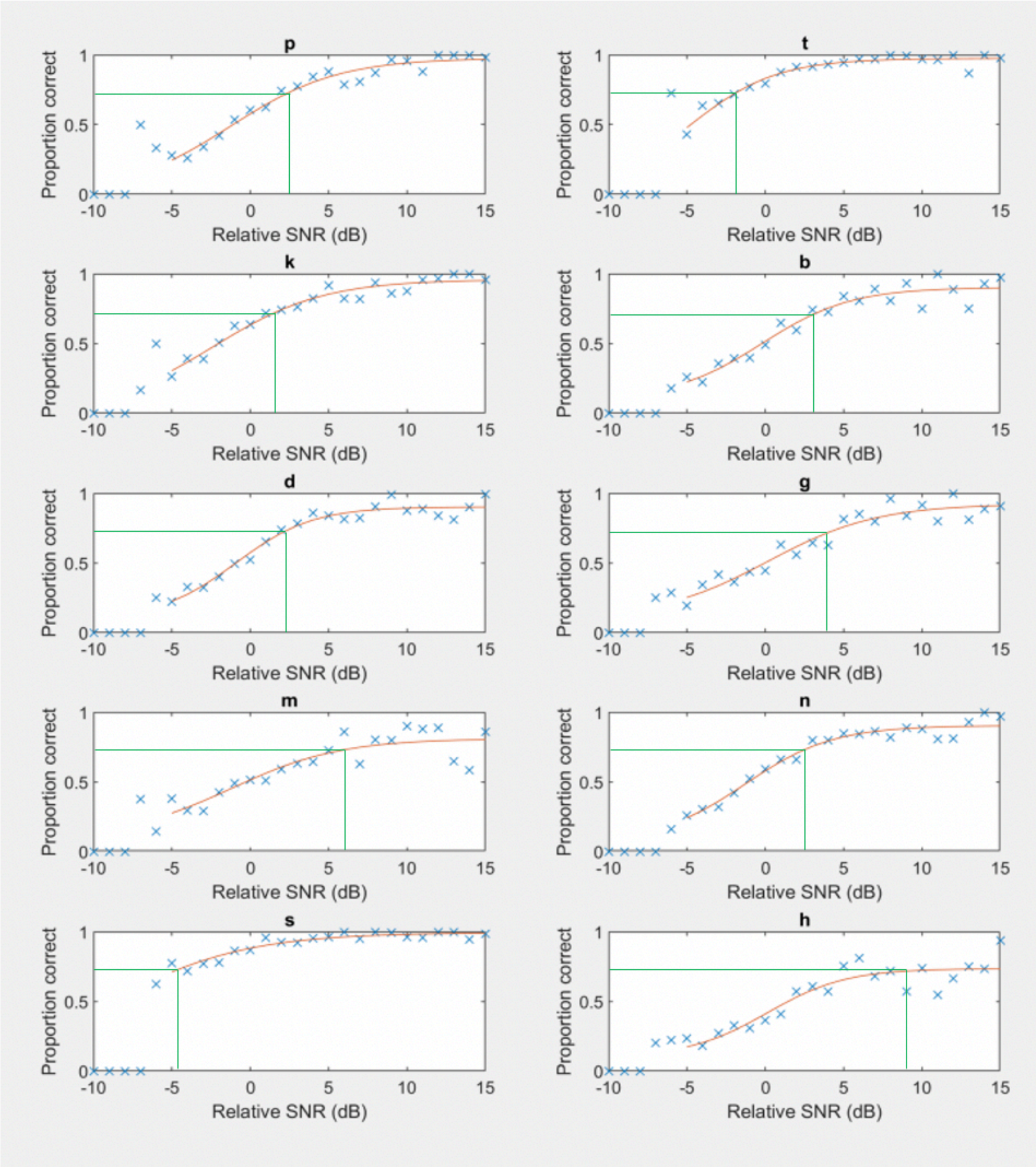


Figure 3.11 Psychometric functions for consonants at the conclusion of data collection and before post-study stimuli adjustments.

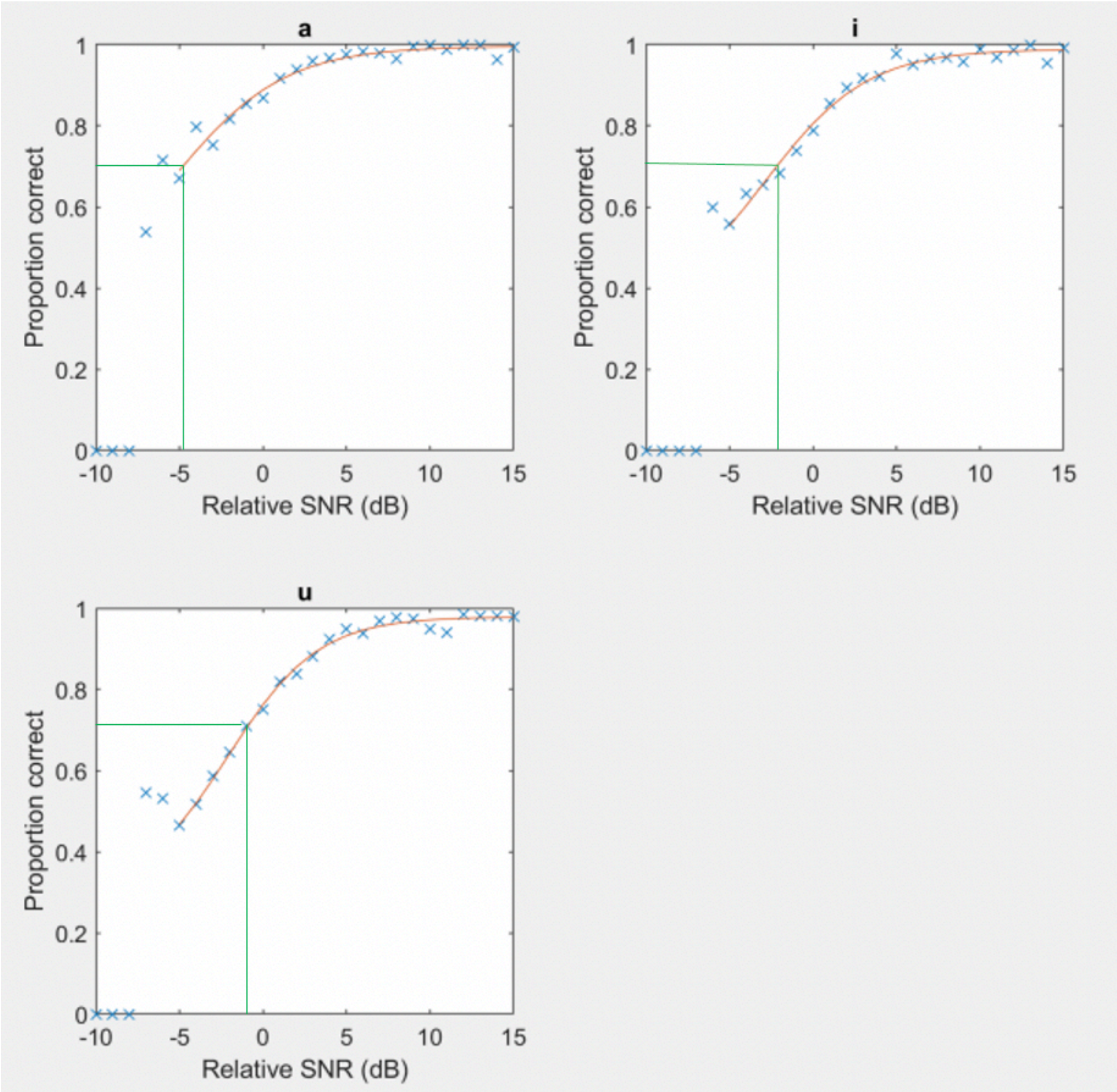


Figure 3.12 Psychometric functions for vowels at the conclusion of data collection and before post-study stimuli adjustments.

4. Discussion

The aim of this study was to evaluate a new language-independent test of speech sound perception in spatially separated noise and reverberation. Data collection and analysis included assessing spatial processing abilities and the impact of reverberation in Australian English, Canadian English, and non-Native English-speaking groups. An additional analysis included evaluating group performance on the ToSSPiN across face-to-face and remote delivery modes.

4.1 Native Language and Performance in ToSSPiN Baseline Conditions

The first research question in this study aimed to determine if there was an effect of native language or accent on each of the ToSSPiN baseline conditions. Overall, there was no significant difference in performance across each of the language groups averaged across the RSD, ASD, ASI and ACI conditions ($p = 0.20$). Intra-participant variability in performance, as measured from the aSE, was also comparable across each of the language groups ($p = 0.91$). The results correspond to the hypothesis presented in the introduction that there would be no effect of language or accent on performance. The outcomes of this study contrast with other studies that show performance differences between native and non-native speakers on speech-based auditory tests.

As mentioned in the introduction, Krizman et al. (2017) found that the signal-to-noise ratio (SNR) loss was smaller on average for native monolinguals than for bilinguals (1.87 $\pm$ 1.5 dB and 3.17 $\pm$ 1.99 dB respectively) on the QuickSIN ($p = 0.007$). Additionally, the SNR was lower for native monolinguals than for bilinguals (-0.97 $\pm$ 0.82 dB SNR and -0.19 $\pm$ 1.12 dB SNR respectively) on the HINT ($p = 0.004$). The QuickSIN and HINT required English language semantic and syntactic knowledge which explains the significant native language benefit the authors found for these tests. The native monolingual group performed only slightly better on average than the non-native bilingual group on the words in noise test (5.51 $\pm$ 1.01 dB SNR and 6.22 $\pm$ 1.83 dB SNR respectively) which was not significant ($p = 0.087$). The words in noise did not require semantic and syntactic knowledge, however, did require English vocabulary knowledge. Krizman et al. (2017) suggested that performance differences between each of the language groups depended on the linguistic requirements of the speech material. The results from Krizman et al. (2017) suggest that as linguistic cues are removed, there is a gradual decrease in performance differences between native and non-native speaking language groups. For the current study, there was no lexical meaning attached to the speech sounds. Mealings and Dillon (2020) also showed that a number of Aboriginal and Torres Strait Islander children performed much poorer on the LiSN-S (based on sentences) than the LiSN-U (based on nonsense syllables) and suggested the children's reduced English language proficiency caused the poorer performance on the LiSN-S. The results of these studies suggest that increasing linguistic information in a speech-test will increase performance in a native language group relative to performance in a non-native group.

One reason, therefore, for comparable performance between the language groups in this study could be that understanding the ToSSPiN stimuli does not place much demand on the listener's language ability. A low demand on language ability presumably occurs because the ToSSPiN targets do not contain any semantic, syntactic and vocabulary cues. The results showed no significant effects of native language on the baseline measures. Using CV pseudo-words as the ToSSPiN stimuli, therefore, has likely removed native language requirements and resulted in comparable performance for each of the language groups.

A second aspect that could have contributed to the comparable group performance was the structure of the phonemes (i.e. CV structure). Hyman (2007) suggested that world languages have a universal preference for a CV structure over a VC structure because of better perceptual landmarks. The universal preference for a CV structure, and the use of phonemes common to most languages, suggests that using ToSSPiN stimuli in the same manner used in this study, could aid in reducing performance differences on speech-based auditory processing tests in varied native language groups.

Munro and Derwing (1995) have suggested that an unfamiliar accent can cause reduced speech intelligibility for non-native listeners. Another study by Clarke and Garrett (2004) suggested that individuals can rapidly adapt to accented speech within minutes. Bradlow and Bent (2008) suggested that non-native listeners require more time to adapt to accented speech. A female with a general Australian accent presented the ToSSPiN targets and could have been unfamiliar to both groups of non-Australian English speakers. The ToSSPiN has a familiarization phase of at least five trials before the test begins. There are also practice phases of five trials before the start of each baseline condition. The results showed comparable performance between each of the language groups on each baseline measure and suggests that accent did not negatively impact performance. One possible reason for a reduction in accent effects is that there was sufficient familiarization and practice phases for those with non-Australian English language backgrounds to adapt to the accent of the stimuli. An additional possibility is that the accent of the speaker is less noticeable when delivering the stimuli as isolated CVCV nonsense words than when delivered as continuous discourse. Any accent cues that are present in continuous discourse might aid a native language listener because of the familiarity with the speaker and the context. By using nonsense words, however, the accent cues could be removed and reduce any native language benefit.

The SRT for the ACI condition was much higher (i.e. worse) than for the other baseline conditions. The higher SRT for conditions that remove spatial cues supports findings from previous studies. Cameron and Dillon (2007a) showed a significant effect of baseline condition on SRT when developing the LiSN-S. The authors found that the $\pm 0^{o}$ same-voice condition (i.e. ACI for ToSSPiN) had a significantly higher SRT (-0.5 dB, *SD* = 1.33) when compared to the $\pm 90^{o}$ same-voice condition (i.e. ASI for ToSSPiN) (-13.5 dB, *SD* = 2.52) ($p < 0.001$). Cameron et al. (2020) also showed that for LiSN-U, the mean SRT for the co-located condition (i.e. ACI for ToSSPiN) was much higher than for the spatially separated condition (i.e. ASI for ToSSPiN) (1.3 dB and -18.7 dB respectively). The ability to segregate the target from the distractors is reliant on access to ITDs and ILDs (i.e. spatial cues) (Cameron & Dillon, 2007a). When there are no spatial cues differentiating the target from the distractors, the ability to segregate the target from the distractors becomes more difficult for a listener and therefore explains the higher SRT in the ACI condition.

The findings of the current study support the use of language-independent stimuli for assessing auditory processing in varied language groups. It could be that the ToSSPiN stimuli allowed a measure of speech-sound processing without involving participants' lexical-semantic language skills, and therefore, native language requirements. The implications of the current findings are such that the removal of top-down language processing required for sentence-based auditory processing assessments, by using nonsense words, allowed performance to be equalized across each of the language groups. Broadly speaking, using language-independent stimuli could also reduce the impact of language deficits in individuals with a comorbid language disorder. If an individual's reliance on language skills is removed when completing an AP assessment, individuals who have co-morbid language impairments could complete the test without their language impairment impacting test performance. Further implications are such that the continued use of linguistically reliant speech material to assess AP skills in individuals with a language impairment could unintentionally diagnose individuals with language deficits as having an AP deficit. At present, interpreting performance on an AP assessment that requires top-down linguistic processing should be interpreted with caution, especially in circumstances where an individual also has a language impairment.

4.2 Native Language Effects on the Impact of Reverberation and Spatial Advantage

The ToSSPiN measures the SA as the difference in SRT between the ASI and ACI conditions. The difference in SRT between ASI and ACI conditions is used to identify the benefit an individual receives when the distracting noise occurs separated spatially from the target versus co-located with the target. A reduced SA is used to diagnose SPD that occurs when an individual lacks the ability to take advantage of binaural cues (i.e. ILDs and ITDs) to segregate the distracting noise from the target (Cameron & Dillon, 2019).

Given that there was no language or accent effects present on the ToSSPiN baseline measures from which the SA measure is derived, it is not surprising that there was no significant effect of language on the SA measure (p = 0.09) or the RI measure (p = 0.42). Not only are the baseline scores not affected by the native language or accent of the listeners, the SA and RI measures being difference scores, further reduce the impact of cognition and language on test performance (Cameron & Dillon, 2007a; Moore et al., 2010). The outcomes of the current study agree with the hypothesis that language would not impact these difference measures.

4.3 Potential Impact of Auditory Processing and Reverberation

The addition of reverberation in the RSD condition resulted in a higher SRT than in the ASD condition (M = -16.10 and M = -20.42 respectively) with the difference, on average, being 4.32 dB. Arweiler and Buchholz (2011) stated that auditory system integrates the early reflections (ERs) of reverberant energy to aid in speech intelligibility. The ERs occur within 50 to 80 ms of the direct sound. Auditory neurons can integrate acoustic information arriving within 200 ms (Moore, 2013) which allows the integration of ERs. An increased RI might suggest a problematic integration of ERs within the 200 ms time window.

The late reflections (LRs), which by definition arrive over 80 ms after the direct speech signal, can persist for longer periods in a listening environment (Arweiler & Buchholz, 2011) and can therefore mask the earlier arriving direct speech signal (Rennies et al., 2014). Shinn-Cunningham and Kawakyu (2003) have suggested that to strengthen the onset of a direct sound in reverberant environments, the auditory system will inhibit LRs. There could be individuals who experience a greater impact of reverberation if there is a deficit in suppression/inhibition of LRs. Future clinical studies should examine if some children with listening difficulties display an increased RI caused by reduced integration of ERs and/or suppression LRs in reverberant energy. Including the RI is a novel aspect of the ToSSPiN and could provide important information about an individual's listening abilities when reverberation is present.

4.4 Native Language Effects on Response Time

The third question in this study examined whether there was an impact of native language on response times for each baseline condition. On average, there was no effect of native language on response times averaged across the RSD, ASD, ASI and ACI conditions ($p = 0.12$). The purpose of calculating response times was to determine if it took longer for non-native Australian English listeners listening to an unfamiliar accent to respond to the stimuli. Increased response times could indicate an increase in processing the stimuli for non-native Australian English listeners even if there were no significant effects of language on baseline scores. The results of the study support the hypothesis that there would be no differences in response times between each of the language groups. The results support previous studies that have shown a rapid adaptation to accented speech occurs with ample exposure to accented stimuli in non-native listeners.

As previously mentioned, Clarke and Garrett (2004) showed that native English listeners (American) can adapt to Spanish-accented English sentences within a minute. The authors found a significant reduction in response times as exposure increase to accented stimuli $F(3, 45) = 13$ ($p<0.00001$). The authors also showed that response times were the same in Spanish-accented and native accented sentences by the conclusion of data collection (*p*-value not given). Xie et al. (2018) also found that increased exposure to accented speech decreased response times ($p<0.01$) to the stimulus and concluded that sufficient exposure to accented speech attenuated processing difficulties.

As with the lack of effect of language and accent on recognition, one reason for comparable response times could be that there was sufficient familiarization and practice given to each of the groups. The ToSSPiN begins with a familiarization phase, and each baseline condition begins with a practice phase. Each of the non-Australian English groups did not require additional familiarization or practice prior to the measurement phase to correctly repeat each of the stimuli. An additional reason for comparable response times across the language groups is, as mentioned previously, reduced accent effects might occur if presenting the stimuli as isolated monosyllabic pseudo-words versus continuous language-dependent discourse. According to Xie et al. (2018) one possible way that rapid adaptation to unfamiliar talkers occurs is by inducing new acoustic-to-category mappings. Listeners learn to generate a model for a particular talker by relating the input of a previously unfamiliar talker with the new talker-specific phonetic cue information. Generating such a model rapidly and in real-time allows a listener to increase accuracy and decrease response times when testing and comparing performance in accented versus unaccented conditions. For this current study, the comparable response times in each native language group also provides evidence of sufficient familiarization and practice. For each participant, the induction of new acoustic-to-phonetic categories more than likely occurred prior to the test phase of the ToSSPiN. A potential limitation with concluding that rapid adaptation occurred prior to the test phase is that no response times were recorded during the familiarization and practice phases. To determine if rapid adaptation has occurred, and to what extent, response times could be recorded in future studies to support rapid adaptation occurring. An additional possibility is that the Canadian and non-English participants had prior experience with Australian-English and had the acoustic-to-phonetic categories already stored in their phonetic lexicon.

4.5 Effects of Native Language on the Rate of Learning for Each Baseline Condition

The average rate of change in dB per trial during the measurement phase (i.e. slope of the adaptive track), quantifies the rate of learning for each language group. Each group, on average, showed gradual learning in each of the baseline conditions. The total change in dB from the start of the measurement phase to the end of each condition, divided by the number of trials in the measurement phase determines the slope of the adaptive track. If the slope was negative, it would reflect a gradual improvement in SRT as each condition progressed. For excessively positive slopes, it would reflect a loss of, or fluctuating attention as the SRT gradually decreases from a better SNR at the start of the measurement phase. If any group was slower than the others to learn the stimuli, the average rate of change would be more negative in that group. It was hypothesized that if no language effects exist for baseline SRT, the rate of learning should be comparable across each of the language groups. The results of the study confirm this hypothesis. There was no significant effect of language group on the slope of the adaptive track (p = 0.72).

One explanation for the learning rates being comparable is that there was sufficient practice for each of the groups to adapt to each baseline condition before the measurement phase started. An additional reason for equal learning rates is that the stimuli themselves, and the listening environments, were equally difficult for each of the language groups. Equal learning rates further support language-independent stimuli. If the non-Australian English language groups displayed delayed learning, we might also observe increased response times and increased SRTs for each of the baseline conditions. Increased baseline SRTs, response times and a gradual change in the slope of the adaptive track, taken together, could suggest that a language group required more time to adapt to the task. The data, however, shows no effect of language on any of these measures.

There was an average gradual improvement in SNR of -3.94 dB during the measurement phase averaged across the RSD, ASD, and ASI conditions. For the ACI condition, the average improvement in SRT was -0.56 dB. The results suggest that for each language group there was gradual learning in all baseline conditions. The average change in dB per trial and overall change in dB for the ACI condition was much lower than the other baseline conditions. In the ACI condition, targets presented at negative SNRs are more difficult to hear than the targets at positive SNRs due to the removal of spatial and pitch cues. Targets presented at positive SNRs are easier to hear because of the relative level compared to the distractors. Over the course of a few presentations, the SNR will more than likely change from positive to negative SNRs rapidly as reflected in the overall mean SRT for the ACI condition (-0.30 dB SRT). As a result, the adaptive track in the ACI condition deviates less from the SRT than for RSD, ASD, and ASI conditions and could explain the lower change in dB per trial relative to the other conditions.

4.6 Remote versus Face-to-Face Delivery

Another question addressed as part of this study was whether delivery mode had a significant impact on the baseline SRTs and difference scores. A non-significant main effect of delivery mode occurred averaged across the baseline SRTs, ($p = 0.10$). Baseline SRTs averaged across the ASD, ASI and ACI conditions for face-to-face delivery were therefore comparable to that of remote delivery. Although the interaction between baseline condition and delivery mode did not reach a level of significance ($p = 0.06$), due to the closeness to significance, an examination of the effect of delivery mode on the baseline measures occurred separately but with the alpha level adjusted to 0.05/4 = 0.0125. In the RSD condition, remotely tested participants required a slightly higher SNR than participants tested face-to-face.

There was no significant effect of delivery mode on the SA measure ($p = 0.68$). Participants tested face-to-face had a significantly lower RI than those tested remotely ($M = 3.19$ and $M = 5.21$ respectively, $p = 0.01$). Reverberation negatively affected remote participants more than face-to-face participants by an average of 2.02 dB. When looking at the mean performance differences based on the baseline scores from which derivation of the RI occurs, in the RSD condition remote and face-to-face participants had an SRT of $M = -15.19$ and $M = -17.48$ respectively and in the ASD condition $M = -20.40$ and $M = -20.67$ respectively. The difference in SRT in the RSD and ASD condition between face-to-face and remote participants is 2.29 dB and 0.27 dB respectively. It is apparent that the difference in SRT in RSD condition is the cause for an increased RI for remotely tested participants.

One explanation for the significant difference in RI across delivery modes could be because of learning by the tester. Although the remote participants received detailed instructions before the day of testing, there may have been an initial period in which the tester needed to adapt to the scoring protocol. The RSD condition was the first condition completed by each participant and testing partner, which might explain why the remote RSD condition and RI were the only negatively affected measures. When completing the ASD, ASI and ACI conditions, both the tester and participant should have been more comfortable with the test and scoring protocol. As the RI is a difference score taken from the RSD and ASD conditions, the non-significant increase in SRT for the remote RSD condition could have resulted in a significant RI difference between delivery modes. The likelihood of learning effects causing significant differences in RSD SRT and RI measures for remote participants are, however, reduced when looking at the aSE and the slope of the adaptive tracks. If learning effects were apparent, it is likely that there would be an increase in variability in the adaptive tracks (i.e. larger aSE because of variable scoring) and signs of gradual learning of the tasks (i.e. larger change in SNR per trial) as both the tester and participant adjusted to the RSD condition. The results of the analysis showed no significant effect of delivery mode on baseline aSE ($p = 0.37$) and there was no significant effect of delivery mode on slope of the baseline adaptive tracks ($p = 0.90$). Taken together, remote delivery did not affect the variability of the adaptive track and did not result in gradual learning of the baseline conditions.

A second hypothesis for the effects of delivery mode on RI is that ambient sound pressure level increased the cumulative effects of noise and reverberation in the RSD condition. Testing of remote participants occurred in their home with uncontrolled and untreated acoustics. Lewis et al. (2014) have suggested that the cumulative effects of noise and reverberation is greater than the sum of noise or reverberation alone. As the RSD condition had reverberation and spatially separated noise, the addition of ambient noise could have increased masking of the ToSSPiN stimuli because of the cumulative effects of noise + reverberation + ambient noise. It is impossible to determine what the average noise floor for each remotely tested participant's listening environment. Kim et al. (2017) showed that in urban environments, traffic noise could be as loud as 73 to 85 dB (reference = L_{10} or the level exceeded 10% of the time) at 10 metres from the source with 2.2 dB reduction occurring for every 30 metres of elevation away from the noise. Further, the authors found that windows were poor attenuators with a small reduction of 3 dB when a window is closed (Kim et al., 2017). A significant portion of testing occurred in urban environments and it is possible that the ambient noise floor could have contributed to the cumulative effects of noise and reverberation resulting in increased masking.

A third hypothesis to explain the effects of delivery mode on RI is that the effect occurred due to chance. There were numerous performance measures analysed as part of this study. With an increase in the number of analyses performed, the likelihood of finding a statistically significant effect increases.

There was, however, a significant main effect of delivery mode on response times (*p*<0.001) with remote participants having, on average, longer response times than face-to-face participants across all baseline conditions. The significant effect of delivery mode on response time could suggest that the testers took slightly longer to process how to administer the test. Longer response times should not be surprising for remotely tested participants, given that the researcher administering the test face-to-face was familiar with all aspects of test administration. Effects of delivery mode on RI and response times do not remove the ToSSPiN as being appropriate for remote use, it suggests that individuals familiar with the test, such as a trained audiologist, should administer the ToSSPiN. The results of the data analysis from remote and face-to-face participants suggested that the ToSSPiN was simple to conduct via a remote delivery mode and not negatively impacted by learning effects.

4.7 Intelligibility of the Phonemes

Prior to the study commencement, individual gains applied to each ToSSPiN phoneme aimed at making the phonemes equally intelligibility. These gains were based on the psychometric functions measured by Cameron et al. (2020) on a relatively small sample of participants. The psychometric functions produced from the current study, however, showed that each phoneme required additional individual gains that ranged from -2.0 dB for /t/ to +8.7 dB for /h/ in order for correct identification of each phoneme 71% of the time at a relative SNR of 0 dB. These additional adjustments should increase test precision by increasing the slope of the psychometric function for the test (because the performance intensity functions of all the phonemes will more closely coincide) and hence decreasing the width of the adaptive track.

4.8 Limitations

One potential limitation in interpreting these results is that each of the participants in the non-English language group had English experience. All participants in the sample had Australian-English, or English language experience. Based on information received from the questionnaires and instruction, each participant had sufficient English language experience to follow complex instructions from a native Australian-English speaker. Future data collected should include individuals with no English language experience to determine whether the ToSSPiN stimuli are appropriate for these populations.

The sampling procedure used could also present limitations when interpreting the results of the study. Because a convenience sample was used, the vast majority of participants were sampled from universities. Although there was a skewed distribution toward Non-native English-speaking participants, these participants for the most part were educated in English-speaking countries or in countries where English is widely spoken. For example, English is an official language in both Honk Kong and Singapore. English is also widely spoken in Belgium. Having participants that are both educated and from a context where English is widely spoken, means the results could be potentially biased toward these types of participants. Data from the general population, and from participants with little to no English experience is required to determine the generalisability of the results.

An additional limitation is having participants self-report their own hearing status. Because it was unknown whether an underlying hearing loss was present, it is impossible to determine whether an underlying hearing loss has contributed to effects seen in the RI measure. Although it is unlikely that an underlying hearing loss would impact this measure specifically (and the RSD condition from which it is derived), the impacts of self-reporting hearing status needs to be considered. Further studies should aim to collect objective information related to each participant's hearing status prior to data collection (i.e. audiogram, middle ear status, and inner ear function).

Another potential limitation is using both the 'test partner' score the test, and then complete the test directly after. A scoring bias could have presented as a result of the 'test partner' habituating to the test, leading to decreased accuracy and an increase in response time. Additionally, because the testing partner had more familiarity with the test when they swapped roles with the initial participant, there could have been comparatively better performance when the initial scorer completed the test. If the testing partner had comparatively better performance than the initial participant, it could have reduced any effect of language group, and delivery mode. Therefore, future studies should investigate native language performance under identical testing conditions to determine the true effects of language group and delivery mode on ToSSPiN performance.

To further examine the equivalence of remote and face-to-face delivery, further studies could examine remotely tested participants, and their tester under controlled acoustic environments. Having appropriate acoustic conditions could determine whether changes in ambient noise levels of the testing environment or insufficient practice from the testers caused the small increase in SRT in the RSD condition for remotely tested participants.

4.9 Conclusions and Future Work

The present study has shown that the ToSSPiN is a simple, easy-to-use assessment for measuring speech sound perception in native Australian-English, native Canadian-English and non-native English-speaking groups. The evidence obtained with these three groups does not contradict the hypothesis that test performance is not affected by the native language or accent of the listener. However, further studies should aim to determine if different language groups achieve the same performance with no English language experience. Overall, the results of this study are consistent with ToSSPiN being language-independent and therefore appropriate to use across language groups.

Future planned studies will collect normative data, with the stimuli adjusted for equal intelligibility, in native English-speaking adults and children, followed by test-retest studies. A planned test-retest reliability study will quantify the random measurement error and systematic learning effects for repeated application of the ToSSPiN. After normative data collection, additional clinical studies will examine performance on the ToSSPiN in children who display listening difficulties or have a diagnosis of Specific Language Disorder, Attention Deficit (Hyperactivity) Disorder and Autism Spectrum Disorder. The aim of the ToSSPiN is to identify AP deficits without also identifying individuals with listening difficulties primarily caused by cognitive and language deficits. Future incorporation of intrinsic attention and extrinsic memory measures will aim to control the effects of attention and memory on test performance. Correction of overall performance on the ToSSPiN will occur for any deviations in attention during the test, and the effects of any memory deficits an individual may have. A further investigation into the impact of reverberation in clinical populations will aim to identify deficits in reverberation suppression in individuals with listening difficulties.

The ToSSPiN was also simple enough to complete in remote delivery mode. Although remote participants had a slightly but significantly higher RI than those tested face-to-face, the possible causes are inconclusive and could have occurred by chance. Ravi et al. (2018) have suggested that advances in technology and connectivity have helped bridge geographic and economic barriers between individuals and healthcare providers. An AP assessment that is readily available, easy to download and easy to administer could help increase access for individuals who require specialist services.

5. Bibliography

AAA. (2010). *Clinical practice guidelines for the diagnosis, treatment, and management of children and adults with central auditory processing disorder.* Retrieved from http://www.audiology.org/publications-resources/document-library/central-auditoryprocessing-disorder.

ABS. (2017). 2016 Census: Multicultural [Press release]. Retrieved from https://www.abs.gov.au/ausstats/abs@.nsf/lookup/Media%20Release3

Ahmmed, A., & Ahmmed, A. (2016). Setting appropriate pass or fail cut-off criteria for tests to reflect real life listening difficulties in children with suspected auditory processing disorder. *International Journal of Pediatric Otorhinolaryngology, 84*, 166-173. doi:https://doi.org/10.1016/j.ijporl.2016.03.014

Ahmmed, U. A., Ahmmed, A. A., Bath, R. J., et al. (2014). Assessment of Children With Suspected Auditory Processing Disorder: A Factor Analysis Study. *Ear and Hearing, 35*(3), 295-305. doi:10.1097/01.aud.0000441034.02052.0a

Alqahtani Mufleh Salem, M. (2019). Sonority Sequencing Principle in Sabzevari Persian: A Constraint-Based Approach. *Open Linguistics, 5*(1), 434-465. doi:10.1515/opli-2019-0024

Anderson, K., & Smaldino, J. (1999). Listening Inventories for Education: A classroom measurement tool. *The Hearing Journal, 52*(10), 74-76. doi:10.1097/00025572-199910000-00009

Arweiler, I., & Buchholz, J. M. (2011). The influence of spectral characteristics of early reflections on speech intelligibility. *The Journal of the Acoustical Society of America, 130*(2), 996-1005. doi:10.1121/1.3609258

ASHA. (2005). *(Central) Auditory Processing Disorders*. Retrieved from http://www.asha.org/policy/PS2005-00114.htm.

Bamiou, D., Musiek, F., & Luxon, L. (2001). Aetiology and Clinical Presentations of Auditory Processing disorders - A Review. *Archives of Disability in Children*(85), 361-365.

Bellis, T. J. (2011). *Assessment and management of central auditory processing disorders in the educational setting : from science to practice* (2nd ed.). San Diego: Plural Pub.

Bench, J., Jacobs, K., & Furlonger, B. (2020). On differentiating auditory processing disorder (APD) from attention deficit disorder (ADD): an illustrative example using the Cattell-Horn-Carroll (CHC) model of cognitive abilities. *International Journal of Audiology, 59*(3), 224-229. doi:10.1080/14992027.2019.1682199

Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America, 113*(6), 3233-3244. doi:10.1121/1.1570439

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

BSA. (2017). *Position Statement and Practice Guidance Auditory Processing Disorder (APD)*. Retrieved from http://www.thebsa.org.uk/wp-content/uploads/2017/04/APD-Position-Statement-Practice-Guidance-APD-2017.pdf

Bureau, U. S. C. (2011). *Language Use in the United States: 2011*. Retrieved from United States: https://www2.census.gov/library/publications/2013/acs/acs-22/acs-22.pdf

Cacace, A. T., & McFarland, D. J. (2005). The importance of modality specificity in diagnosing central auditory processing disorder. *American Journal of Audiology 14*(2), 112. doi:10.1044/1059-0889(2005/012)

Cameron, S., Brown, D., Keith, R., et al. (2009). Development of the North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): sentence equivalence, normative data, and test-retest reliability studies. *Journal of the American Academy of Audiology, 20*(2), 128. doi:10.3766/jaaa.20.2.6

Cameron, S., & Dillon, H. (2007a). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear and Hearing, 28*(2), 196-211. doi:10.1097/AUD.0b013e318031267f

Cameron, S., & Dillon, H. (2007b). The listening in spatialized noise-sentences test (LISN-S): test-retest reliability study. *International Journal of Audiology, 46*(3), 145-153. doi:10.1080/14992020601164170

Cameron, S., & Dillon, H. (2008). The Listening In Spatialized Noise-Sentences Test (LISN-S): comparison to the prototype LISN and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. *Journal of the American Academy of Audiology, 19*(5), 377-391. doi:10.3766/jaaa.19.5.2

Cameron, S., & Dillon, H. (2011). Development and Evaluation of the LiSN & Learn Auditory Training Software for Deficit-Specific Remediation of Binaural Processing Deficits in Children: Preliminary Findings. *Journal of the American Academy of Audiology, 22*(10), 678-696. doi:10.3766/jaaa.22.10.6

Cameron, S., & Dillon, H. (2017). Sound Storm (Software). Sydney: Australian Hearing Services.

Cameron, S., & Dillon, H. (2019). Deficit-Specific Diagnosis and Remediation of Auditory Processing Disorders. In D. S. Geffner & D. Ross-Swain (Eds.), *Auditory processing disorders : assessment, management, and treatment* (3rd ed., pp. 95-122). San Diego, CA: Plural Publishing, Inc.

Cameron, S., & Dillon, H. (2020a). Are “Dichotic” Deficits Uniquely Dichotic? Investigating Dichotic Performance with the Dichotic Digits Difference Test (DDdT) in a Large Clinical Population of Children Referred for an Auditory Processing Assessment. *Journal of the American Academy of Audiology, 31*(3), 233-242. doi:10.3766/jaaa.19037

Cameron, S., & Dillon, H. (2020b). *The Listening in Spatialized Noise - Universal Test (LiSN-U) (Research Version)*. Macquarie University.

Cameron, S., & Dillon, H. (2020c). *Specification: Test of Speech Sound Perception in Noise*. Dr Sharon Cameron. Macquarie University.

Cameron, S., Dillon, H., Glyde, H., et al. (2014). Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia. *International Journal of Audiology, 53*(5), 326. doi:10.3109/14992027.2013.871388

Cameron, S., Dillon, H., & Newall, P. (2006). Development and Evaluation of the Listening in Spatialized Noise Test. *Ear and Hearing, 27*(1), 30-42. doi:10.1097/01.aud.0000194510.57677.03

Cameron, S., Glyde, H., Dillon, H., et al. (2015). Results from a National Central Auditory Processing Disorder Service: A Real-World Assessment of Diagnostic Practices and Remediation for Central Auditory Processing Disorder. *Seminars in Hearing 36*(4), 216-236. doi:10.1055/s-0035-1564457

Cameron, S., Mealings, K. T., Chong-White, N., et al. (2020). The development of the listening in spatialised noise - universal test (LiSN-U) and preliminary evaluation in English-speaking listeners. *International Journal of Audiology*, 1-9. doi:10.1080/14992027.2019.1689431

Canadian Interorganizational Steering Group for Speech-Language Pathology and Audiology. (2012). Canadian Guidelines on Auditory Processing Disorder in Children and Adults: Assessment & Intervention. Retrieved from http://www.ooaq.qc.ca/publications/doc-documents/Canadian_Guidelines_EN.pdf. *Accessed:* 18/6/2020

Chen, A. S.-W. (2011). Bodies and codas or core syllables plus appendices? Evidence for a developmental theory of subsyllabic division preference. *Cognition, 121*(3), 338-362. doi:10.1016/j.cognition.2011.08.006

Chermak, G. D., Bamiou, D.-E., Iliadou, V., et al. (2017). Practical guidelines to minimise language and cognitive confounds in the diagnosis of CAPD: a brief tutorial. *International Journal of Audiology, 56*(7), 499-506. doi:10.1080/14992027.2017.1284351

Chermak, G. D., Musiek, F. E., & Weihing, J. (2017). Beyond Controversies: The Science Behind Central Auditory Processing Disorder: There is general scientific agreement on the definition, diagnosis, and treatment of CAPD. *The Hearing Review, 24*(5), 20.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647-3658. doi:10.1121/1.1815131

Danish Medical Audiological Society. (2014). Retrieved from http://dmasaud.dk/onewebmedia/DSOHH-KKR-APD.pdf

Dawes, P., & Bishop, D. (2009). Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique. *International Journal of Language & Communication Disorders 44*(4), 440-465. doi:10.1080/13682820902929073

Dawes, P., & Bishop, D. V. M. (2007). The SCAN-C in testing for auditory processing disorder in a sample of British children. *International Journal of Audiology, 46*(12), 780-786. doi:10.1080/14992020701545906

de Wit, E., Visser-Bochane, M. I., Steenbergen, B., et al. (2016). Characteristics of auditory processing disorders: a systematic review. *Journal of Speech, Language, and Hearing Research, 59*(2), 384-413. doi:10.1044/2015_JSLHR-H-15-0118

Devellis, F. R. (2006). Classical Test Theory. *Medical Care, 44*(11), S50-S59. doi:10.1097/01.mlr.0000245426.10853.30

Dillon, H. (1983). The effect of test difficulty on the sensitivity of speech discrimination tests. *The Journal of the Acoustical Society of America, 73*(1), 336-344. doi:10.1121/1.388815

Dillon, H., Cameron, S., Glyde, H., et al. (2012). An Opinion on the Assessment of People Who May Have an Auditory Processing Disorder. *Journal of the American Academy of Audiology, 23*(2), 97-105. doi:10.3766/jaaa.23.2.4

Esplin, J., & Wright, C. (2014). *Auditory Processing Disorder: New Zealand Review. A Report Prepared for the Ministry of Health and Ministry of Education*. Retrieved from https://www.health.govt.nz/system/files/documents/publications/auditory_processing_disorder.pdf

Fang, Y., Feng, H., & Chen, Y. (2018). A robust interaural time differences estimation and dereverberation algorithm based on the coherence function. *Applied Acoustics, 129*, 126-134. doi:10.1016/j.apacoust.2017.07.019

Ferguson, S. H., Jongman, A., Sereno, J. A., et al. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology, 21*(3), 153. doi:10.3766/jaaa.21.3.3

Gordon, R. (2005). *Ethnologue: Languages of the World* (15th ed.). Dallas, Tx: SIL International.

Gordon-Salant, S., Yeni-Komshian Grace, H., Bieber Rebecca, E., et al. (2019). Effects of Listener Age and Native Language Experience on Recognition of Accented and Unaccented English Words. *Journal of Speech, Language, and Hearing Research, 62*(4S), 1131-1143. doi:10.1044/2018_JSLHR-H-ASCC7-18-0122

Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution, 1*(1), 19-29. doi:10.1093/jole/lzw002

Helland, T., Morken, F., Bless, J. J., et al. (2018). Auditive training effects from a dichotic listening app in children with dyslexia. *Dyslexia 24*(4), 336-356. doi:10.1002/dys.1600

Hidaka, T., Yamada, Y., & Nakagawa, T. (2007). A new definition of boundary point between early reflections and late reverberation in room impulse responses. *The Journal of the Acoustical Society of America, 122*(1), 326-332. doi:10.1121/1.2743161

Hu, X.-J., & Lau, C.-C. (2020). Dichotic listening using Mandarin CV-words of six plosives and vowel /a/. *International Journal of Audiology*. doi:10.1080/14992027.2020.1783004

Hugdahl, K. (2009). Dichotic Listening Studies of Brain Asymmetry. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 517-522). Oxford: Academic Press.

Hugdahl, K., & Helland, T. (2014). Central Auditory Processing as seen from Dichotic Listening Studies. In G. C. F. Musiek (Ed.), *Handbook of Central Auditory Processing Disorder: Auditory Neuroscience and Diagnosis* (2nd ed.). San Diego: Plural Publishing Inc.

Hyman, L. M. (2007). Universals in Phonology. *Linguistic Review, 25*, 83-137. doi:10.1515/TLIR.2008.003

Iliadou, V., Ptok, M., Grech, H., et al. (2018). European 17 countries consensus endorses more approaches to APD than reported in Wilson 2018. *International Journal of Audiology, 57*(5), 395-396. doi:10.1080/14992027.2018.1442937

International Bureau for Audiophonologie. (2007). *Recommendations: Central Auditory Processing*. Retrieved from https://www.biap.org/en/recommandations/recommendations/tc-30-central-auditory-processes-cap

Jerger, J., & Musiek, F. (2000). Report of the consensus conference on the diagnosis of suditory processing disorders in school-aged children. *Journal of the American Academy of Audiology, 11*(9), 467-474.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337-1351. doi:10.1121/1.381436

Kamhi, A. (2011). What Speech-Language Pathologists Need to Know About Auditory Processing Disorder. *Language, speech, and hearing services in schools, 42*(3), 265-272. doi:10.1044/0161-1461(2010/10-0004)

Katz, J. (1962). The use of staggered spondaic words for assessing the integrity of the central auditory nervous system. *Journal of Auditory Research, 2*, 327-337.

Keith, R. (2000a). SCAN-C: Test for Auditory Processing Disorders in Children-Revised.

Keith, R. W. (2000b). Development and standardization of SCAN-C Test for Auditory Processing Disorders in Children. *Journal of the American Academy of Audiology, 11*(8), 438-445. doi:10.1002/stem.5530140735

Killion, M. C., Niquette, P. A., Revit, L. J., et al. (2001). Quick SIN and BKB-SIN, two new speech-in-noise tests permitting SNR-50 estimates in 1 to 2 min. *The Journal of the Acoustical Society of America, 109*(5), 2502-2502. doi:10.1121/1.4744912

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

Kim, M. K., Barber, C., & Srebric, J. (2017). Traffic noise level predictions for buildings with windows opened for natural ventilation in urban environments. *Science and Technology for the Built Environment, 23*(5), 726-735. doi:10.1080/23744731.2016.1262708

Krizman, J., Bradlow, A. R., Lam, S. S.-Y., et al. (2017). How bilinguals listen in noise: linguistic and non-linguistic factors *Bilingualism: Language and Cognition, 20*(4), 834-843. doi:10.1017/S1366728916000444

Lewis, D. E., Manninen, C. M., Valente, D. L., et al. (2014). Children's understanding of instructions presented in noise and reverberation. *American Journal of Audiology 23*(3), 326. doi:10.1044/2014_AJA-14-0020

Loo, J., Bamiou, D.-E., & Rosen, S. (2013). The Impacts of Language Background and Language-Related Disorders in Auditory Processing Assessment. *Journal of Speech, Language and Hearing Research (Online), 56*(1), 1-12. doi:10.1044/1092-4388(2012/11-0068)

Madsen, S. M. K., Marschall, M., Dau, T., et al. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports, 9*(1), 10404. doi:10.1038/s41598-019-46728-1

Marriage, J., King, J., Briggs, J., et al. (2001). The reliability of the SCAN test: results from a primary school population in the UK. *British Journal of Audiology, 35*(3), 199-208. doi:10.1080/00305364.2001.11745237

Martin, N., Brownell, R., & Hamaguchi, P. (2018). *Test of Auditory Processing Skills*. Retrieved from Novato, CA: https://www.wpspublish.com/taps-4-a-language-processing-skills-assessment

McArthur, M. G. (2009). Auditory processing disorders: can they be treated? *Current Opinion in Neurology, 22*(2), 137-143. doi:10.1097/WCO.0b013e328326f6b1

Mealings, K., Cameron, S., Chong-White, N., et al. (2020). Listening in Spatialized Noise – Universal Test (LiSN-U) test-retest reliability study. *International Journal of Audiology*, 1-6. doi:10.1080/14992027.2020.1795283

Mealings, K., Cameron, S., & Dillon, H. (2020). Correlating performance on the Listening in Spatialised Noise – Sentences test (LiSN-S) with the Listening in Spatialised Noise – Universal test (LiSN-U). *International Journal of Audiology, 59*(7), 519-523. doi:10.1080/14992027.2020.1753119

Mealings, K., & Dillon, H. (2020). English language and language-free detection of spatial processing disorders in Aboriginal and Torres Strait Islander children. *International Journal of Audiology*.

Miller, C. A. (2011). Auditory processing theories of language disorders: past, present, and future. *Language, Speech, & Hearing Services in Schools, 42*, 309.
Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing*. Leiden, The Netherlands: Brill.
Moore, D. (2018). Editorial: Auditory Processing Disorder. *Ear and Hearing, 39*(4), 617-620. doi:10.1097/AUD.0000000000000582
Moore, D. R. (2006). Auditory processing disorder (APD): Definition, diagnosis, neural basis, and intervention. *Audiological Medicine, 4*(1), 4-11. doi:10.1080/16513860600568573
Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., et al. (2010). Nature of auditory processing disorder in children. *Pediatrics, 126*(2), e382-e390. doi:10.1542/peds.2009-2826
Moore, D. R., & Hunter, L. L. (2013). Auditory processing disorder (APD) in children: A marker of neurodevelopmental syndrome. *Hearing, Balance and Communication, 11*(3), 160-167. doi:10.3109/21695717.2013.821756
Munro, M., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Lang. Speech, 38*, 289-306.
Musiek, E. F. (1983). Assessment of Central Auditory Dysfunction: the Dichotic Digit Test Revisited. *Ear and Hearing, 4*(2), 79-83. doi:10.1097/00003446-198303000-00002
Musiek, F. E., Bellis, T. J., & Chermak, G. D. (2005). Nonmodularity of the central auditory nervous system: implications for (central) auditory processing disorder. *American Journal of Audiology 14*(2), 128-138. doi:10.1044/1059-0889(2005/014)
Musiek, F. E., Chermak, G. D., Weihing, J., et al. (2011). Diagnostic accuracy of established central auditory processing test batteries in patients with documented brain lesions.(Report). *Journal of the American Academy of Audiology, 22*(6), 342. doi:10.3766/jaaa.22.6.4
Musiek, F. E., & Pinheiro, M. L. (1987). Frequency Patterns in Cochlear, Brainstem, and Cerebral Lesions: Reconnaissance mélodique dans les lésions cochléaires, bulbaires et corticales. *Audiology, 26*(2), 79-88. doi:10.3109/00206098709078409
NAL. (2016). COSI-C: Client Oriented Scale of Improvement for Children. Retrieved from https://www.nal.gov.au/wp-content/uploads/sites/2/2016/11/COSI-C-Administration-Instructions.pdf. *Accessed:* 29/09/2020
Neijenhuis, K., de Wit, E., & Luinge, M. (2017). Perspectives of Dutch health professionals regarding auditory processing disorders; a focus group study. *International Journal of Audiology, 56*(12), 942-950. doi:10.1080/14992027.2017.1347290
Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085-1099. doi:10.1121/1.408469
Ohala, J. (1992). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. Retrieved from http://linguistics.berkeley.edu/~ohala/papers/altson.pdf. *Accessed:* 22/02/2020
Paredes-Gallardo, A., Innes-Brown, H., Madsen, S. M. K., et al. (2018). Auditory Stream Segregation and Selective Attention for Cochlear Implant Listeners: Evidence From Behavioral Measures and Event-Related Potentials. *Frontiers in Neuroscience, 12*(581). doi:10.3389/fnins.2018.00581
Paulovicks, J. (2008). The Gaps-in-Noise (GIN) Test and its diagnostic significance. *The Hearing Journal, 61*(3), 67-67. doi:10.1097/01.HJ.0000314723.80439.72

Picou, M. E., Gordon, A. J., & Ricketts, A. T. (2016). The Effects of Noise and Reverberation on Listening Effort in Adults With Normal Hearing. *Ear and Hearing, 37*(1), 1-13. doi:10.1097/AUD.0000000000000222

Prete, G., D'Anselmo, A., Brancucci, A., et al. (2018). Evidence of a Right Ear Advantage in the absence of auditory targets. *Scientific Reports, 8*(1), 15569. doi:10.1038/s41598-018-34086-3

Rance, G., Ryan, M. M., Carew, P., et al. (2012). Binaural speech processing in individuals with auditory neuropathy. *Neuroscience, 226*, 227-235. doi:10.1016/j.neuroscience.2012.08.054

Ravi, R., Gunjawate, D. R., Yerraguntla, K., et al. (2018). Knowledge and Perceptions of Teleaudiology Among Audiologists: A Systematic Review. *Journal of audiology & otology, 22*(3), 120. doi:10.7874/jao.2017.00353

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology, 5*(1), 27-48. doi:10.1146/annurev.clinpsy.032408.153553

Rennies, J., Schepker, H., Holube, I., et al. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America, 136*(5), 2642-2653. doi:10.1121/1.4897398

Rickard, N. A., Heidtke, U. J., & O'Beirne, G. A. (2013). Assessment of auditory processing disorder in children using an adaptive filtered speech test. *International Journal of Audiology, 52*(10), 687-697. doi:10.3109/14992027.2013.802380

Riley, K. G., & McGregor, K. K. (2012). Noise hampers children's expressive word learning.(Research Article)(Report). *Language, Speech, & Hearing Services in Schools, 43*, 325.

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language & Speech, 47*(2), 139-154.

Rogers, C. L., Lister, J. J., Febo, D. M., et al. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(3), 465-485. doi:10.1017/S014271640606036X

Rowan, D., Papadopoulos, T., Edwards, D., et al. (2015). Use of binaural and monaural cues to identify the lateral position of a virtual object using echoes. *Hearing Research, 323*, 32-39. doi:10.1016/j.heares.2015.01.012

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals - Fourth Edition*. Retrieved from San Antonio, Tx:

Sharma, M., Purdy, S. C., & Kelly, A. S. (2009). Comorbidity of auditory processing, language, and reading disorders.(Report). *Journal of Speech, Language, and Hearing Research, 52*(3), 706. doi:10.1044/1092-4388(2008/07-0226)

Shi, L.-F. (2010). Perception of Acoustically Degraded Sentences in Bilingual Listeners Who Differ in Age of English Acquisition. *Journal of Speech, Language and Hearing Research 53*(4), 821-835. doi:10.1044/1092-4388(2010/09-0081)

Shinn-Cunningham, B., & Kawakyu, K. (2003, 19-22 Oct. 2003). *Neural representation of source direction in reverberant space.* Paper presented at the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684).

Social, T. O. (2012). *Europeans and Their Languages: Special Eurobarometer 386*. Retrieved from European Union: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

St Clair-Thompson, H., Stevens, R., Hunt, A., et al. (2010). Improving children's working memory and classroom performance. *Educational Psychology, 30*(2), 203-219. doi:10.1080/01443410903509259

Stavrinos, G., Iliadou, V.-M., Edwards, L., et al. (2018). The Relationship between Types of Attention and Auditory Processing Skills: Reconsidering Auditory Processing Disorder Diagnosis. *Frontiers in Psychology 9*, 34-34. doi:10.3389/fpsyg.2018.00034

Tabri, D., Abou Chacra, K. M. S., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders 46*(4), 411-422. doi:10.3109/13682822.2010.519372

Tomlin, D., Dillon, H., Sharma, M., et al. (2015). The Impact of Auditory Processing and Cognitive Abilities in Children. *Ear and Hearing, 36*(5), 527-542. doi:10.1097/AUD.0000000000000172

Tomlin, D., & Rance, G. (2014). Long-Term Hearing Deficits After Childhood Middle Ear Disease. *Ear and Hearing, 35*(6), e233-e242. doi:10.1097/AUD.0000000000000065

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

Vermiglio, A. (2014). On the Clinical Entity in Audiology: (Central) Auditory Processing and Speech Recognition in Noise Disorders. *Journal of the American Academy of Audiology, 25*(9), 904. doi:10.3766/jaaa.25.9.11

Weinberger, S. H. (n.d.). The Speech Accent Archive. Retrieved from https://accent.gmu.edu/browse_native.php. *Accessed:* 22/02/2020

Wilson, W. J. (2018). Evolving the concept of APD. *International Journal of Audiology, 57*(4), 240-248. doi:10.1080/14992027.2017.1409438

Wilson, W. J., & Arnott, W. (2013). Using different criteria to diagnose (central) auditory processing disorder: how big a difference does it make? *Journal of Speech, Language, and Hearing Research 56*(1), 63-70. doi:10.1044/1092-4388(2012/11-0352)

Wilson, W. J., Heine, C., & Harvey, L. A. (2004). Central Auditory Processing and Central Auditory Processing Disorder: Fundamental Questions and Considerations. *Australian and New Zealand Journal of Audiology, The, 26*(2), 80-93. doi:10.1375/audi.26.2.80.58277

Xie, X., Weatherholtz, K., Bainton, L., et al. (2018). Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker. *The Journal of the Acoustical Society of America, 143*(4), 2013-2031. doi:10.1121/1.5027410

Appendix A.

Name of participant: ________________________
Date of birth: ______________________________
Email address:______________________________
Phone number:______________________________

MACQUARIE University SYDNEY·AUSTRALIA

Hearing and Language Experience/Use Questionnaire

This questionnaire forms part of a research project conducted by Christian Boyle and supervised by Dr Sharon Cameron and Professor Harvey Dillon at Macquarie University. Thank you for taking part in this project, your participation is greatly appreciated.

This basic questionnaire will allow us to better understand your early language experience and daily language use. The questionnaire will also help us to understand your self-described hearing experience. By providing this information, you will allow us to appropriately analyse any data collected from your participation in this study.

By answering the questions contained in this document, you give us permission to use this information in any internal and external publications that result from this study. In any publication resulting from this project we will keep your identity confidential and any information will be communicated honestly.

1. Do you feel you have any difficulties hearing? Yes / No (Please circle).

2. If you experience difficulties, in which environments do you typically experience these difficulties?

3. Have you ever completed a hearing test? If so, what were the results?

4. Is English your native language? Yes / No (Please circle).

   If Yes, there is no need to answer any further questions.

5. What is your native language?

6. On a scale of 1 to 10 (1 = no proficiency and 10 = fluent), how would you rate your English-speaking ability? (Please circle).

1 2 3 4 5 6 7 8 9 10

7. On a scale of 1 to 10 (1 = no proficiency and 10 = fluent), how would you rate your English understanding (i.e. when conversing in English)? (Please circle).

1 2 3 4 5 6 7 8 9 10

8. At what age did you begin learning English?

9. What language did you speak most at home as a child?

10. Of the time you spend communicating what proportion is in English? (Please circle).

0% 25% 50% 75% 100%

Appendix B.

ToSSPiN Instructions to Adult

Instructions in red will tell ***you*** how to run the test.

Read the instructions in black to the person being tested.

1. Before starting the test we need to make sure your iPad is set to stereo. Open the 'settings' menu. Scroll down to 'accessibility' and open that menu. Near the bottom of the screen will be an 'Audio/Visual' menu, open this menu. The setting for 'mono audio' should be set to the off position (grey not green) and the balance setting should be set to the middle.
2. Plug in the headphones that you have to the iPad but don't put them on the participant just yet.
3. Open the ToSSPiN app.
4. Enter the first and last name, date of birth and most proficient language of the person being tested.
5. Press the 'start' button. This will take you to the 'Comfort Adjustment' screen. This is to ensure the volume of the test is at a comfortable level for the participant.
6. Hand the person the iPad.
7. "You are going to hear some made-up words, I want you to adjust the slider  or the iPad volume control until all the words are comfortably loud".
8. Place the headphones on the participant.
9. Press the play button. 
10. Once they have finished adjusting, take back the iPad and press the stop button. 
11. Press the 

button.
12. The next screen is the 'Consonant & Vowel Familiarisation Screen'. This is to ensure the person understands the made-up words being spoken.
13. Take the headphones off the participant and explain the task like this:
14. "You are going to hear a lady saying some made-up words, like 'ti-gu' over these headphones. The lady will sound as if she is standing right in front of you. There will be a 'beep' before each word so you will know when the lady is going to speak. Your job is to repeat back the made-up word. So if you heard 'ti-gu', you would say [prompt adult to say 'ti-gu']. If you heard 'ba-ga', you would say [prompt adult to say 'ba-ga']. Even if you are a little unsure I want you to have a guess anyway, you will still get points for guesses."
15. At this point, place the headphones on the participant and sit directly in front of them without letting them see the iPad screen.
16. Once the participant is ready, press the 

button. The first word will play. If they say it is too loud, it is OK to turn the volume control down one or two clicks.
17. In the lower portion of the screen there are four boxes 

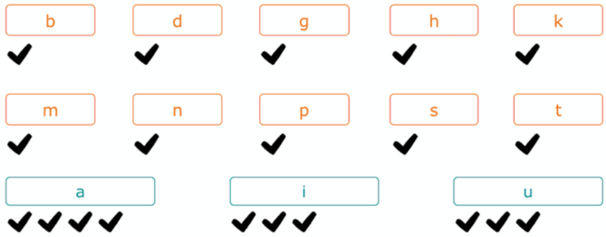
Once the participant has repeated back the first word, press each box to indicate which sounds have been correctly repeated. If all of them are correct, press 

button instead. The boxes will turn green for each sound correct.

18. Press ⏭ Next Item to continue.

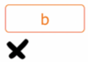
19. Continue these steps until the familiarization phase is completed.

20. When familiarization is finished, check to see if every speech sound has at least one tick:



   a. If yes, press ⏩ Go to Test

   b. If no, press ↻ More Familiarisation Then press on the letter (i.e. sound) that they need more practice on. The incorrect letter should have an ✖ underneath it, for example:



   c. When you press an incorrect letter, a new word will play.

   d. Once all letters have a tick underneath them press ⏩ Go to Test

21. The next screen will look like this:

Each box on the right of this screen represent a new condition: RSD/ASD/ASI/ACI

The ▶ Start button begins each new condition.

Each of the four blue boxes at the bottom of this screen represent each sound to be scored.

22. Take the headphones off the person and repeat below:

23. “This time, when the lady is saying the word there will be other people talking. These other tricky people also saying made-up words and will come from both sides of your head. No matter where the tricky people are, I don’t want you to listen to them. Just listen for the ‘beep’ and the made-up word said directly after it by the lady in front of you.”

24. “If you only hear a bit of the word I want you to tell me what you hear, because you get a point for each sound you get right. So if you just hear ‘ti’ repeat that back. Even if it is just a guess, have a go. The tricky people will start first and will sound like they are next to you, and their voices will be very similar to the lady’s voice. Ignore them and just listen for the beep and the word the lady says. This time the lady will say the word twice so it is a bit easier for you to hear,

but you only need to repeat it back to me once. So if you hear 'beep ti-gu ti-gu'. Then you say [prompt person to say 'ti-gu']."

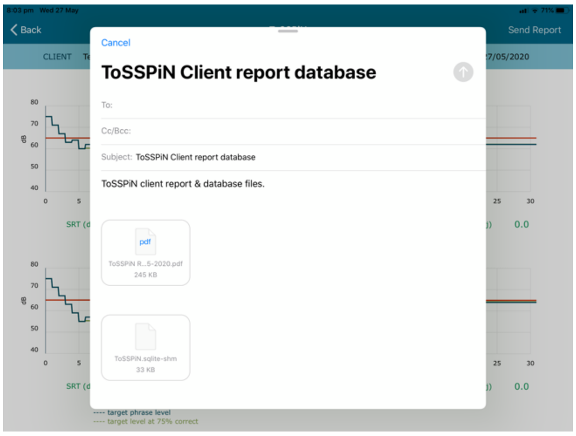
25. "Sometimes the lady's voice will get softer, just keep trying your best to hear the word. We're trying to find the softest words you can hear. If you hear something but are not sure, you can guess. If you can't hear the lady at all, just shake your head and we'll go on to the next word.
26. Although this screen looks different, the scoring method is almost the same. Once the participant has repeated back the first word, press each of the four boxes to indicate which sounds have been correctly repeated. If all of them are correct, press 'all'. If the participant swaps each half of the word, e.g. 'gu ti' instead of 'ti gu', press the correct but reversed button.
27. "We will have a practice first. Ready?"
28. Place headphones back on the participant and sit in a position so they cannot see the iPad screen. Commence and complete the RSD practice condition by pressing Start
29. Score the word and then press Next Item
30. When the practice phase has finished, a notice showing Practice Complete will appear.
31. "Great job! Now we will do the real thing."
32. Press Start Test to start the test phase.
33. Repeat these steps until Condition Complete. appears at the bottom of the screen.
34. "Well done! We have three more conditions to do. Just remember to listen to the lady, not the tricky people. We will have a practice first. Ready?"
35. Press the ASD button and then Start to begin the next practice condition.
36. If you are unsure of the condition that is currently active, look at the colour of the button. Light green such as ASD means this is not the active condition. Dark green such as ASD means this is the active condition.
37. Score as you have been previously until Practice Complete appears.
38. "Great job! Now we will do the real thing again." [Complete ASD condition].
39. Press Start Test
40. Score as you have been previously and until Condition Complete. appears at the bottom of the screen.
41. "Now we will do the third condition. We will have a practice first." Select the ASI condition button and press Start
42. When finished, say "Great, just one condition to go". Select the ACI condition.
43. "This time the tricky people will sound like they are next to the lady. You will have to listen very hard for the 'beep' and the word the lady says. This one is much harder."
44. Complete the practice first and then the test phase of the ACI condition.
45. Once the ACI condition has been completed Test Complete will be displayed at the bottom of the screen. You can take the headphones off the participant.

46. "Congratulations, you have completed the ToSSPiN! Thank you for your participation!"

47. The final step requires a report to be emailed to the researcher. Press the Results button in the bottom right corner. This will take you to the 'results' screen.

48. Press the Next button on the 'results' screen. This will take you to another screen labelled 'ToSSPiN'.

49. In the top right corner of this screen will be text labelled, Send Report press this text. A pop-up box will appear that will look like this:



In the 'Subject' line, insert: the participant name *after* the text present in the 'Subject' line, i.e:
ToSSPiN Client Report Database – Joe Bloggs

When this has been completed, press the ↑ in the top right corner of this screen to send the report to the researcher.

Tips:

- Stress to the participant that he or she is not to listen to the tricky people. Only listen for the beep and the made-up word. This is particularly important in the ACI condition.
- If the participant appears to be confused at the start of the first session, stop the test and restart the condition to ensure the participant understands the task.
- If the participant loses attention during a particular condition prompt the participant to "keep listening for the beep and the made-up word".
- If you are unsure of anything during the test, the researcher (Christian Boyle) will be monitoring the session and you may ask questions if required.
- If you are unsure whether you have started the test phase or are still in the practice phase, check the middle left-hand portion of the test screen, if you see Practice you are still in the practice phase. If you see Test you are in the test phase.

Human Sciences Subcommittee
Macquarie University, North Ryde
NSW 2109, Australia

MACQUARIE University

26/05/2020

Dear Dr Cameron,

Reference No: 52020632715958
Project ID: 6327
Title: Development of the Test of Speech Sound Perception in Noise (ToSSPiN) – Effect of First Language, Spatial Separation and Reverberation on Speech Sound Identification

Thank you for submitting the above application for ethical review. The Human Sciences Subcommittee has considered your application.

I am pleased to advise that ethical approval has been granted for this project to be conducted by Dr Sharon Cameron, and other personnel: Professor Harvey Dillon, Mr Christian John Boyle.

This research meets the requirements set out in the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the National Statement, available from the following website: https://nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018.

2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol. You will be sent an automatic reminder email one week from the due date to remind you of your reporting responsibilities.

3. All adverse events, including unforeseen events, which might affect the continued ethical acceptability of the project, must be reported to the subcommittee within 72 hours.

4. All proposed changes to the project and associated documents must be submitted to the subcommittee for review and approval before implementation. Changes can be made via the Human Research Ethics Management System.

The HREC Terms of Reference and Standard Operating Procedures are available from the Research Services website: https://www.mq.edu.au/research/ethics-integrity-and-policies/ethics/human-ethics.

It is the responsibility of the Chief Investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Faculty Ethics Officer.

The Human Sciences Subcommittee wishes you every success in your research.

Yours sincerely,

A/Prof Naomi Sweller

Chair, Human Sciences Subcommittee

The Faculty Ethics Subcommittees at Macquarie University operate in accordance with the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018), [Section 5.2.22].