The Role of Intergroup Attitudes in Speech Perception

A Thesis Submitted in Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

The MARCS Institute for Brain, Behaviour and Development –

– Western Sydney University

by

Nhung Nguyen

Supervised by:

Assistant Prof. Jason A. Shaw (past primary supervisor)

Associate Prof. Michael D. Tyler (current primary supervisor)

Dr. Rebecca T. Pinkus (co-supervisor)

Prof. Catherine T. Best (co-supervisor)

March, 2017

List of Tables

List of Figures

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

[redacted]

Nhung Nguyen

Acknowledgements

What I remember most about grad life depicted in PhD comics (http://phdcomics.com/comics/archive_list.php) is the grad students' struggle with the frustrating poverty as well as the notorious Prof. Smith that could not care less about the students under his supervision. My PhD candidature at the MARCS Institute, Western Sydney University, has shown that these types of grad life experience are not necessarily universal, thankfully! In fact, despite all the stress about meeting other-imposed as well as self-imposed deadlines (frantically in December 2013, 2014, and 2015), my candidature was nothing short of enjoyment, especially the first three years. Such enjoyment was abstracted over the distributions of specific evidence that I have consciously and unconsciously accumulated from two different sources over the course of these absolutely definitely entirely totally terrific, awesome-ic years. Semantic memories were the first source that provided me with interesting and useful knowledge and facts, which in turn would serve me well in the future (more or less depending on which job I will get). Episodic memories were the second but more important source. They were more important because episodic memory is a marvellous neurocognitive (mind/brain) system that stores events that happened to the self (Tulving, 2002). For all the heart-warming events that have happened to my self at MARCS, I owe the following people:

First and foremost, a big fat thank you to the MARCS Institute for feeding me for three years, disabusing my stereotypes of how impoverished a grad student life should be. MARCS also flew me to New Zealand, UK, Canada, and Thailand to discuss science.

Jason Shaw was nothing short of sweet, caring, supportive, and full of brilliant ideas for my thesis project. I cannot thank him enough for being there whenever I needed help most (especially when it was near conference submission deadlines), for spending long hours to discuss my project with me, and for giving me all the freedom I needed along the way. I learned from Jason not only about how to do research but also about how to lead a happy life. I owed Jason not only for disabusing the stereotype of Prof. Smith in supervision but also for introducing me to the MARCS scholarship that helped me avoid the typical financial struggle of grad students.

Michael Tyler was an incredible help during this final leg of my PhD journey. After all my unsuccessful attempts at submitting the thesis last year, which caused me several mental crises, Michael took me under his wing to catapult me to thesis submission. I am very thankful for the fact that he stepped up to co-ordinate the circulation of thesis drafts with the rest of the panel for me. My academic writing has also benefited greatly from working with Michael.

Cathi Best and Rebecca Pinkus contributed a lot of great ideas to the thesis. Rebecca also helped me out a lot in the first and second years to finalize the design of the online survey that aimed to collect prejudice data in such a way that would obtain the most honest response from participants. Cathi kindly offered me a research assistant position to help me financially power through my days after my scholarship ran out. It was through the job with Cathi (and the task assigned by Jason) that I wrote my first R functions, which was incredibly rewarding for a beginner in programming like me. Cathi also caringly forwarded me job applications that she deemed fit.

Of course, to this day, "the girls in my room," known as Mona Faris, Sarah Fenwick, and Valeria Peretokina, have always been the biggest source of peer support. We have shared the most frustrating as well as ecstatic moments with one another. I do appreciate you gals being there for me, when I needed help with statistics, Praat, writing, or just a couple of ears to listen to me venting. Thanks for being on my team and keeping the "grumpy flow" going in the office, Mona, we need that! Thanks for being always so sweet and positive, Valeria! And thanks for always looking out for the rest of us in the room and offering wise suggestions, Sarah!

A couple of names also activate episodes indexed with the social category "good times in Building 5" in my "PhD" distribution. Sweet Joey Ong was always around to offer help, a sympathetic ear, or a good laugh. Laurence Bruggeman was helpful and supportive, and more importantly, one of the reasons my boyfriend and I got together. Muneeb Ahmad always checked in with me if I had submitted my thesis, and if I would need help with LaTeX. Jaydene Elvin was a great help with the recruitment of Australian monolinguals. Tonya Agostini was sweet and caring during my down moments after the confirmation of candidature. Beautiful Peta Mills, always remember to say "Beautiful!" in response to "How are you today?"! And thanks as well to Anne Dwyer, who clothed me!

My "PhD" distribution also contains episodes about Yvonne Leung, Leo Chong, and Benjawan Kasisopa, indexed with the social category "good times out." I was lonely outside of the office at times at the start of the PhD. By inviting me over for their food gatherings and taking me out to restaurants, Yvonne, Leo, and

Ben showed me the fun of belonging to a group that has similar food interests. Yvonne and Leo also cared about my safety, which I very much appreciated.

Together with the admin team (Darlene Williams, Karen McConachie, Sonya O'Shanna, Andrea Kantek, Craig Purcell, and Jen Armstrong), Colin Schoknecht and the tech team (Johnson Chen, Steve Fazio, Donovan Govan, Lei Jing, Muawiyath Shujau, and Benyamin Binyamin) also helped smooth my PhD journey with their prompt help whenever I was in need. Karen Mulak was also a valuable help with ePrime programming and ideas for stimulus recording.

Very importantly, when your PhD project has the limited disclosure and deception nature, you have to rely on others' help with data collection for the integrity of your data. A very special thank is due to Mona Faris, Sarah Wright, and Mark Lathouwers for being such great associates.

The lovely neighbourhood of Milperra with the awesome landlord Terry Brown and sweet landlady Diane Brown as well as the friendly neighbours were also a big source of happiness during my time at MARCS. I was very lucky to have found this accommodation.

… But grad life is not all about PhDs, it includes Masters as well. The grad life stereotypes were partially accurate back in those days when I was doing a Masters degree in Brisbane. If it had not been for the Vietnamese couple who hired me to work for their Chinese/Vietnamese take-away restaurant and the owner of a house in a Brisbane suburb who hired me to do cleaning, I would not have survived in Australia till this day to be able to store all the heart-warming events that have happened to my self in Sydney. These casual jobs from these kind souls helped me get by during my financially hard days in Brisbane. I must also thank Ron Maskell, Caitlin Maskell, and Jan McNicol for their sweetest friendship that built me up during and after my hardship in Brisbane. Without their unconditional support, I would not be able to become who I am today. I also want to thank John Ingram and Rob Pensalfini for instilling the love of Linguistics in me through the Masters program, which has helped me persist in the field… more or less.

And I must not forget that my father Nguyen Quang Trung and grandmother Nguyen Thanh Tam are the people who saw me accumulate my very first episodic memories and semantic memories, build my first distributions, and abstracted them into several social categories like "competitiveness" and "success in studies." Though these categories have faded through time, their boundless love and unconditional support stay with me.

Last but not least, my boyfriend Ethan Sherr-Ziarko came to my life last, just a little before all the mental crises with my thesis submission attempts began. However, he has given me tremendous mental support to help me power through those days and stay strong to push through the finish line. I am grateful to him for having my back during the darkest days of my research life, cooking delicious meals for me during the last three weeks of my thesis writing, and most of all, for being the biggest and most important part of my life.

Milperra, March 2017

Abstract

This thesis pursued the overall hypothesis that listeners use abstract information whenever it is available in perceiving speech. It also attempted to evaluate the role of socio-indexical information in speech perception and, in particular, the possibility of further abstractions over abstract social categories. These abstract social categories can be abstracted over two sources of socio-indexical information: (1) knowledge and/or feelings from direct speech exposure; and (2) beliefs and/or feelings from indirect sources about speech. Abstractions of both types were predicted to have a relationship with listeners' perception of speech, sometimes interacting with each other in a complex manner. These predictions were made within the episodic approach and the hybrid approach to speech perception (including the Bayesian framework), tested over the course of three perception experiments with Australia-born listeners (reported in Chapters 2, 3, and 4), and had implications for speech perception theories.

Chapter 2 discussed abstractions over socio-indexical knowledge gained from direct speech exposure. It used known properties of vowel formant measurements to probe the relationship between mental representations of vowel categories and socio-indexical information. Specifically, it tested the predictions that listeners' tolerance of variation for Australian English vowel categories would depend on the socio-indexicality of the individual vowels: more tolerance for vowels that are not rich in socio-indexical information and less tolerance for those that are. Results were consistent with the predictions and suggested that listeners had more specific and socially relevant distributions for a certain group of Australian English vowels because, through life-long exposure, they had implicit knowledge that these vowels are likely to vary more among speakers of different generations or socio-economic groups. Findings from this first experiment demonstrated the existence of high-level abstractions over socio-indexical information. They also provided support for the overall hypothesis that listeners use abstract information whenever it is available in perceiving speech as well as the hybrid approach to speech perception.

Chapter 3 turned to abstractions over socio-indexical beliefs formed from indirect sources such as taught knowledge or hearsay. It assessed how speech categories could be represented in listeners' minds without listeners' direct speech exposure. On the introduction of the socio-indexical cue about the speaker's Vietnamese accent, episodic theories would predict no shift in perception while a shift in perception could be

predicted in the Bayesian framework. Results showed that the listener's perceptual space shrunk when they were told to expect a Vietnamese accent even when they did not have previous exposure to the accent. This suggests that, in the absence of direct observations, speech categories could be formed in accordance with listeners' abstract beliefs about them. The application of Bayesian statistics into speech perception can provide an approach that accounts for how socio-indexical abstractions formed from indirect sources are used in speech perception.

Chapter 4 moves beyond the cognitive component of listener-speaker relationship (knowledge/beliefs) into the affective component (feelings), or prejudice. It asks whether prejudice relates to speech perception, and if it does, how other listener factors such as stereotypical expectations about a speaker's speech and experience play out in the relationship between prejudice and speech perception. Results indicated that the relationship between listener expectations and prejudice varied according to whether or not participants had previous exposure to the Vietnamese accent. This provided further evidence that a purely episodic theory does not have adequate power to explain listener flexibility in this case, and therefore a more nuanced theory which makes use of high-level abstractions over socio-indexical information and incorporation of both abstractions and specifics in a probabilistic algorithm is necessary to accurately describe speech perception scenarios.

In addition to theoretical advances, the thesis also proposes several methodological advances in the areas of experimentation and data analysis, such as the attempt to quantify formant variability while taking the magnitude of formant means into account, the quantification of stereotypical expectations via Bayesian formalism, the quantification of prejudice, and the use of Australian-accented English and Vietnamese-accented English as speech stimuli in the vowel perception experiments to serve different purposes.

Overall, the results in this thesis suggest that high-level abstractions over social categories are possible, and that Bayesian principles need to be integrated into the hybrid approach to explain all possible speech perception scenarios, some of which involves highly abstract social factors.

Chapter 1

Introduction

Speech signals carry a great deal of information: linguistic information, socio-linguistic information, and the speaker's personal information, all of which interact in speech perception (e.g., Ladefoged & Broadbent, 1957). As an example, consider Cẩm, a young Vietnamese female, whose English is at an intermediate level and who is currently living in Australia. Now consider Cẩm's production of the famous garden-path sentence "The horse raced past the barn fell" /ðə hoːs ɹæɪst pɐːst ðə bɐːn fel/ (transcribed with the Australian English phoneme system – Cox & Palethorpe, 2007). A lot of information can be found in the speech she produced. First of all, instead of producing [ðə hoːs ɹæɪst pʰɐːst ðə bɐ̃ːn feɫ], she might produce [də hoːs ɹes paːs də baːn feo] instead due to transfer effects from the Vietnamese phonemic system. In her production, [d] was substituted for [ð] and [aː] for [ɐː], the diphthong in [ɹæɪst] was monophthongized and then replaced by [e], the final [t] in [ɹæɪst pɐːst] was deleted, and the dark l in [feɫ] was vocalized (see T. Nguyen & Ingram, 2004, for more examples of transfer effects and connected speech processes in Vietnamese-accented English). These transfer effects are an illustration of the first type of information that we can retrieve from Cẩm's speech: *linguistic information*. Of all these transfer effects, some are quite unique to the Vietnamese accent while some others are shared with other accents. For example, the vocalization of dark /l/ is quite common in a number of native English accents (Wells, 1982a, 1982b), but the monophthongization of diphthongs in closed syllables is quite idiosyncratic to Vietnamese-accented English. As a result, the monophthongization of diphthongs in the speech signal could prompt listeners that the speaker's accent could possibly be Vietnamese. In addition, Cẩm's relatively high fundamental frequency prompts listeners that the speaker could be a female and/or a young person. These details illustrate the second type of information carried by the speech signal: *socio-linguistic information*. In addition to linguistic information and socio-linguistic information, Cẩm's speech signal also contains some husky quality and a fast speech rate. These are called the *speaker's personal information*, which is the third type of information embedded in the speech signal.

It is logically possible for all three of these types of information to be abstracted into categories. First, consider Cẩm's linguistic information again. Native Australian-accented English listeners would store all the instances of Cẩm's [d]s in [də] and of her [aː]s in [paːs] and [baːn], together with the social details "a young Vietnamese female" and her husky voice and fast speech rate, in their memory. In the beginning, the Australian listeners might be confused as to what words her [də]s, [paːs], and [baːn] indicated. However, after adaptation to her Vietnamese-accented English, they would be able to adjust and abstract all her instances of [d]s into their native /ð/ category, perhaps called "accented /ð/ category," as well as all her tokens of [a:]s into their native [ɐ:] category, perhaps called "accented /ɐ:/ category." After some more exposure and abstraction, they would further abstract "accented /ð/ category" and some other accented categories such as /f/ and /ʃ/ into "accented categories with frication," and "accented /ɐ:/ category" with some other accented categories such as /i/ and /ʊ/ into "accented categories without change in quality." These more abstract categories would then be further abstracted into "accented consonant categories" and "accented vowel categories." Such abstraction of linguistic information is the central tenet of the abstractionist approach to speech perception.

By analogy to linguistic information, Cẩm's socio-linguistic information and personal information might also be abstracted. Australian listeners could abstract the detail "Vietnamese" into a category called "nationality" or "origin," which might include other nationalities/origins such as "Thai," "Chinese," "Lebanese," and "Egyptian." Then the listeners could abstract the detail "female" into the category "gender," which might include "female" and "male." The detail "young" could be abstracted into the category "age cohort," which might consist of "young," "middle-aged," and "old." At this point, it is still possible for the listeners to further abstract "nationality/origin," "gender," and "age cohort" into a superordinate category called "socio-indexicality" (with "indexical" meaning "marking" or "indicating")[1]. Cẩm's personal information, again, could be abstracted into categories in a similar way.[2] The episodic approach to speech perception currently accounts for socio-linguistic information and speaker's personal information at the "nationality/origin," "gender," and "age cohort" categorical level, but it is uncommitted about the possibility of further abstraction of these social categories into "socio-indexicality." However, it is possible that such further abstractions do occur and have some relationship with the way listeners perceive speech, such as the Vietnamese-accented speech of Cẩm and other Vietnamese-accented speakers. If a relationship with speech perception does exist at that level of abstraction, then it is also possible to hypothesize a relationship with speech perception at even higher levels of abstraction, such as when abstraction happens in the *absence* of direct exposure to a specific type of speech. Such evidence would have implications for an integrative theory of speech perception.

[1] The term "index" (and its derivatives: "indexical," "indexicality") is used slightly differently in this thesis than its standard meaning in speech perception. In studies about the influence of social information on speech perception such as Hay and Drager (2010), "index" is used to mean "link," as in indexing/linking information about a dialect to phonetic tokens produced by speakers from that dialect. In this thesis, it means "mark" or "indicate." Therefore, "socio-indexicality" means marking or indicating social information.
[2] These sequential abstraction processes could be either a conscious or unconscious process (see related points in Pierrehumbert, 2006).

This thesis examines the different levels of listeners' abstract knowledge in speech perception. In Section 1, I will review the three types of information carried in the speech signal that are illustrated in the example above. Then I will review speech perception theories in Section 2, which propose explanations of how listeners deal with these three types of information. The first approach – the abstractionist approach – will be introduced in Section 2.1. The abstractionist approach places focus on the first type of information in the speech signal, linguistic information. I will point out some weaknesses of the approach that led to the rising popularity of the second approach, the episodic approach, which will be introduced in Section 2.2. Again, weaknesses of the episodic approach will be discussed to arrive at a solution – the hybrid approach – in Section 2.3. In Section 3, I will show two different manifestations of abstractions over socio-indexical information. I will then test the relationship of these two manifestations with speech perception in Section 4, where I lay out the project research questions and a number of methodological details that were taken care of to ensure the robustness of the research.

1 Different types of information in speech

Linguistic information is the first type of information that a speech signal carries. It is expressed through linguistic features at the different levels of a linguistic grammar such as aspiration and frication (phonetic and phonemic features), affixation (morphemic features), tone contour and duration (tonemic/stress features), event structure (semantic features), modification (phrase features), sentence types (sentential features), and speech acts (discourse features). Speech materials such as phonemes, morphemes, tonemes/stress, words, phrases, sentences, and discourse together tell us what message the speech signal conveys. Examples of studies that examine linguistic information in speech signals include Bundgaard-Nielsen, Best, and Tyler (2011) and Faris, Best, and Tyler (2016), in which listeners were presented with a /hVbə/ non-word and identified the vowel phoneme by choosing a writing symbol or keyword on a grid. Some other examples include Babel and Russell (2015) and Bradlow and Bent (2008), in which listeners were presented with sentences embedded in noise and wrote down the words they identified on answer sheets.

Socio-linguistic information, or characterization of various socio-indexical properties (Foulkes, 2010; Foulkes & Docherty, 2006; Silverstein, 2003), is another type of information that is carried by the speech signal. It indicates social group membership and is expressed through speech features that speakers have acquired through their interactions with the members of their social groups (Ladefoged & Broadbent, 1957). Illustrations of socio-indexical information could be the non-rhotic feature of postvocalic /r/ in Australian-accented English which differentiates Australians from a number of Americans (see Cox, 2006 vs. Labov, Ash, & Boberg, 2006), and the use of "ain't" for a variety of negative constructions with "have," "be," and "do" (i.e., "have/has not," "am/is/are not," "do/does not," and "did not") in contemporary African American Vernacular English which differentiates African Americans from European Americans (Howe, 2005). Features that are linguistic (carrying meaning only) could also be socio-indexical (marking social groups). For example, tones/pitch accents carry meaning in Japanese, but they also mark Tokyo Japanese from Koshikijima Japanese via the tonal patterns of both individual words and homophones (Kubozono, 2012). In the word /makudonarudo/ meaning "McDonald's," the second, third, and fourth syllables receive high pitch in Tokyo Japanese (i.e., ma.KU.DO.NA.ru.do), whereas the first, second, third, and fifth syllables receive high pitch in Koshikijima Japanese (i.e., MA.KU.DO.na.RU.do). In the homophones /a.me/ meaning both "rain" and "candy," the tonal patterns in Koshikijima Japanese are reversed from those in Tokyo Japanese. In Tokyo Japanese, A.me means "rain" and a.ME means "candy," while A.me means "candy" and a.ME means "rain" in Koshikijima Japanese.

Personal information about a speaker is the third type of information that a speech signal exhibits. In Ladefoged and Broadbent's (1957) examination of the three types of information conveyed by speech sounds in general and vowels in particular (i.e., linguistic information, socio-linguistic information, and personal information), personal information is defined as the particular information about a speaker as an individual, not his or her social groups. It contains idiosyncrasies in a speaker's speech such as a particular voice quality (e.g., breathy, husky, or laryngealized) and a certain speaking style (e.g., casual vs. formal, see Eskénazi, 1993 for an in-depth discussion of speaking styles). These idiosyncrasies could result from learned behaviour, or anatomical and physiological configurations like a particular size and shape of a speaker's vocal tract (Ladefoged & Broadbent, 1957). Similar to the relationship between linguistic and socio-indexical features, features that are indicative of a particular speaker's information in a language could represent both idiosyncrasies of a speaker and linguistic material (carrying meaning) in another language. Click consonants serve as a good example: If present in a native English speaker's speech, they only mean a particular nonlinguistic speech habit of this person (e.g., a lateral click to encourage a horse to go faster), whereas they could both indicate a speech habit and signal linguistic information in southern and eastern Africa (e.g., Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best, McRoberts, & Sithole, 1988; Brancazio, Best, & Fowler, 2006; Traunmüller, 2003). Likewise, features that are idiosyncratic in a language could represent both idiosyncrasies of a speaker and socio-indexical information (marking social groups) in another language. A fast speech rate in Vietnamese could simply indicate a speaker's particular speech habit, or according to Jacewicz, Fox, Neill, and Salmons (2009), it could as well indicate that the speaker is a young man from Wisconsin. Lastly, some speech features could convey all the three types of information. For instance, high rising terminal intonation contours in statements, or *uptalk* (speech feature), might be the speech habit of a certain speaker (personal information), but it also signals questions (linguistic information) as well as certain social groups such as young Maori New Zealanders (socio-indexical information) (Britain & Newman, 1992; see Warren, 2016, for more details on uptalk).

In the above section, I have reviewed definitions and given examples of linguistic information, socio-indexical information, and speaker's personal information, to highlight the existence of these three types of information in the speech signal. The role of these types of information in speech perception is differently acknowledged by different theories of speech perception, which has important implications for the advancement of speech perception theories. In the next section, I will review how these types of information are accounted for by theories of speech perception.

2 Theories of Speech Perception

2.1 Abstractionist approaches to speech perception

The perception of speech was initially thought to involve the processing of linguistic information only, while neglecting or making minimal use of socio-indexical information and the speaker's personal information (e.g., Joos, 1948; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Miller, 1953; Nearey, 1989; Shankweiler, Strange, & Verbrugge, 1977; Syrdal, 1985). This speech processing principle was reflected in a number of speech perception models described in, for example, Miller (1984), Syrdal (1985), and Nearey (1992). However, even when abstractionists studied linguistic features only as cues to speech perception, they had different focuses: One school of abstractionists considered acoustic characteristics alone, such as vowel formants, whereas another school took into consideration surrounding phonological contexts such as the transition from a consonant to a vowel. The two focuses are briefly outlined below.

The first school of abstractionists looked for speech perception cues in characteristics of isolated sounds (e.g., Joos, 1948; Ladefoged & Broadbent, 1957; Miller, 1953; Nearey, 1989; Syrdal, 1985). Take vowel perception as an illustration. The perception of vowel categories was suggested to involve a normalization process, where listeners take into account the relationship between the fundamental frequency and the other frequencies within a vowel (pure intrinsic normalization) (e.g., Miller, 1953; Syrdal, 1985), the relationship between the vowel of interest and the other vowels produced by the same speaker (pure extrinsic normalization) (e.g., Joos, 1948; Ladefoged & Broadbent, 1957), or a mixture of both (e.g., Nearey, 1989). The pure intrinsic normalization approach followed in Miller's (1953) study could serve as an example. With the intention to control unwanted variability, this study explored the importance of phonetic factors such as fundamental frequency (f_0), formant amplitude, and the number of formants in vowel perception by employing synthetic vowels as stimuli. The vowels were created by a harmonic tone synthesizer as stimuli in two tests to detect the perceptual effects of the change of f_0 in the center and back vowel region (one test for a low f_0 of 140Hz and the other for a high f_0 of 288Hz), two tests in the front vowels (one for a low f_0 of 144Hz and the other for a high f_0 of 288Hz), one test to explore the influence of the addition of the third formant (F_3) on vowel perception, and one last test to put together all the vowel regions at the high f_0 of 288Hz to compare changes in perception. Nine experienced observers were asked to identify those synthetic vowels as one of 11 American monophthongs (i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, o, ʊ, u, ɚ). Results indicated a perceptual shift in the boundaries of sounds with higher f_0 towards higher frequencies, a perceptual improvement from the addition of F_3 in the front (synthetic, two-formant) vowels, and the importance of the relationship among formant amplitudes in phonetic evaluation. In short, this study approached vowel perception from the acoustic, or more generally the linguistic, angle only. Using synthetic vowels freed the researcher from having to take into account the variability in speech coming from the speaker's social groups (socio-indexical information) and the speaker's idiosyncrasies (personal information). In fact, speaker's personal information was claimed to be overlooked in the speech perception process (Joos, 1948, pp. 59–60), and socio-linguistic information was believed to signal only the need for normalization (Joos, 1948, pp. 64–65).

The second school of abstractionists considered phonetic contexts as cues to speech perception (e.g., Liberman et al., 1967; Lindblom & Studdert-Kennedy, 1967; Schatz, 1954; Shankweiler et al., 1977; Strange, Verbrugge, Shankweiler, & Edman, 1974). One illustration comes from the study by Schatz (1954), which examines the role of vowel context in the perception of English stops /p/, /t/, and /k/, in one experiment with synthetic speech and one with natural speech. The experiment with synthetic speech used a playback machine, which converted spectrograms into sounds, to generate stimuli. Eighty-four syllables were created by pairing each burst from each frequency position (out of the 12 positions: 360 Hz, 720 Hz, 1080 Hz, 1440 Hz, 1800 Hz, 2160 Hz, 2520 Hz, 2880 Hz, 3240 Hz, 3600 Hz, 3960 Hz, and 4320 Hz) and each synthetic vowel (out of the seven vowels: i, e, ɛ, a, ɔ, o, u). These synthetic stimuli were randomly presented to a group of subjects, whose task was to identify the initial burst as /p/, /t/, or /k/. Stimuli from the experiment with natural speech came from an American speaker producing "keep" [k^{h}ip], "cop" [k^{h}ɑp], "coop" [k^{h}up], "heap" [hip], "hop" [hɑp], "hoop" [hup], [ski], [skɑ], [sku], [id], [ɑr], [ul]. First off, each of the [k^h]s of [k^{h}i], [k^{h}ɑ], and [k^{h}u] in [k^{h}ip], [k^{h}ɑp], and [k^{h}up] (both the stop burst and the aspiration period) was cut to be spliced with each of the vowels of [ip], [ɑp], and [up] in [hip], [hɑp], and [hup]. However, the resulting combination sounded very unnatural as the aspiration period already contained the formants of the originally recorded vowels (the vowels of [ip], [ɑp], and [up] in [k^{h}ip], [k^{h}ɑp], and [k^{h}up]). As a result, each of the [sk]s of [ski], [skɑ], [sku] was cut to be spliced with each of the vowels of [id], [ɑr], and [ul] to serve as the natural stimuli with unaspirated [k]. Nevertheless, results from these stimuli with unaspirated [k] would not be directly comparable with the synthetic stimuli with aspirated [k^h]. Therefore, the natural stimuli with aspirated [k^h] were created with each of the [k]s in [k^{h}ip], [k^{h}ɑp], and [k^{h}up] (*only* the stop burst) cut to be spliced with [hip], [hɑp], and [hup] after the duration of [h] was shortened. Keeping the initial [h]s but shortening their durations was to give the impression of aspiration without containing the formants of the originally recorded vowels. The 45 syllables with unaspirated [k] (5 samples for each of the nine combinations) were randomly presented to 20 subjects, and so were the 45 syllables with aspirated [k^h]. Results from this experiment with natural speech stimuli corresponded closely to those from the experiment with synthetic speech stimuli: The identification of /k/ when it occurs in front of a vowel is not only cued by its acoustic characteristics, but also by the vowel context, and contextual changes may result in perceptual changes of /k/ (to be perceived as /p/ or /t/). Again, this study approached speech perception from the acoustic/phonetic, or more generally the linguistic, angle only. The study did not take into account the American speaker's social group (socio-indexical information), nor did it consider the speaker's idiosyncrasies (personal information).

The abstractionist approach is economical and capable of explaining some speech perception phenomena. It is economical because the object of storage in one's mental lexicon, such as features to describe vowels, "had better be reduced … to its barest essentials, otherwise we shall still be hampered by excess of detail" (Joos, 1948, p. 50). As a result, only abstract information that is critical to meaning and free from contexts, such as phonemes, is kept in memory. Phonetic details such as r-colouring, nasalization, and pharyngealization could be inferred from neighbouring categories, and thus, are not represented in the lexicon (Joos, 1948).[3]

The approach is also capable of explaining some speech perception phenomena such as how listeners could recognize sounds quickly and accurately despite all the variations coming from different speakers (e.g., Abramson & Cooper, 1959; Peterson & Barney, 1952; Verbrugge, Strange, & Shankweiler, 1974). The study by Peterson and Barney (1952) was the first study administered with a large scale of speaker variation. Seventy-six speakers (33 men, 28 women, and 15 children) were recorded reading two lists of 10 American English vowels in 10 words (heed, hid, head, had, hod, hawed, hood, who'd, hud, and heard) with each list of words randomized. Then 70 listeners (men and women, 32 of whom were among the 76 stimulus speakers previously) listened to eight sessions with 200 words per session and, for each word that they heard, crossed out one out of the 10 given words. Despite the differences in the formant frequencies produced by men, women, and children, six vowels were accurately perceived more than half of the time: ([i] was accurately perceived 94% of the time, [æ] 76% of the time, [ɝ] 91%, [ʌ] and [ʊ] 50%, and [u] 72% of the time). This result could be explained by the extrinsic normalization approach, which was tested out by Gerstman (1968) with success. Despite the above strengths, the abstractionist approach still falls short in its explanation of various speech perception facts surrounding linguistic, socio-indexical, and speaker's personal information (described in turn below).

First off, the rules that infer phonetic details from phonological contexts were most often formulated on fairly small numbers of observations, and such rules are most often language-specific (Pierrehumbert, 2016). For instance, nasalization is considered unimportant to abstractionists like Joos (1948) as vowels are most likely nasalized before nasal stops in a number of Indo-European languages such as Germanic, Polish, Lithuanian, and Portuguese (Joos, 1948; Schourup, 1973). However, vowels are also nasalized AFTER nasal

[3] This economy principle is also reflected in Chomsky's generative grammar (Lukasiewicz, 2012).

stops in lesser-studied languages such as Sundanese in Indonesia, Yoruba in Nigeria, Navaho in the southwestern United States, and Thai in Thailand (Schourup, 1973). In addition, even within Philadelphia English, in which nasalization is a feature, the degree of nasality varies depending on speakers' age, more so for the young generation and less so for those born in the time window of 1965–1980 (Zellou & Tamminga, 2014). As a result, it is not always possible to accurately infer phonetic details from phonological contexts in all cases, and thus, phonetic details need to be represented alongside phonological representations.

Second, if socio-indexical information truly only signaled the need for normalization and was not important for speech perception, we would not be able to explain the differences in listeners' perceptual performance when different socio-indexical cues were provided to them. In a seminal study reported in Niedzielski (1999), a group of Detroit/Michigan listeners was told to expect a Canadian speaker whereas the other group was told to expect a Detroit/Michigan speaker. The two groups, however, listened to the exact same stimuli containing raised vowels produced by the same Detroit/Michigan speaker. Although raised vowels are stereotypical of Canadian-accented English among Detroit/Michigan listeners, they are unnoticed by Detroit/Michigan listeners in Detroit/Michigan speech. The speech perception task was to match vowels in sentences to vowels in isolation. The group who expected a Canadian speaker chose raised vowels while the group who expected a Detroit/Michigan speaker chose standard American-accented English vowels.[4] Hay and colleagues replicated this finding with more subtle manipulations of listener expectations. Similar changes in New Zealand listeners' vowel perception were observed when the words "Australian" and "New Zealander" were displayed on the answer sheets in one study (Hay, Nolan, & Drager, 2006a), and when stuffed toy kangaroos, koalas, and kiwis (the first two animals are associated with Australia; the last is associated with New Zealand) were surreptitiously presented in the test room in another study (Hay & Drager, 2010). Results from a number of other studies also support the importance of socio-indexical information in speech perception (e.g., Babel & Russell, 2015; Hay, Warren, & Drager, 2006b; Johnson, Strand, & D'Imperio, 1999; King & Sumner, 2014; Lindemann, 2002; Mack & Munson, 2012; McGowan, 2015; Rubin, 1992). As a result, socio-indexical

[4] The social influences studied and reviewed in this work (and other similar works) are interpreted to be post-perceptual effects, which means they bias the listeners' decision process rather than warping the incoming sensory information. Several speech perception and social psychology models support this interpretation such as the exemplar-resonance schema in Johnson (2006) and the person construal diagram in Freeman and Ambady (2011). In these models, social factors are modelled as one of the top-down influences that directly affect the category level or the exemplars. However, they are not modelled to directly affect the cue level or the input. In addition, a brief discussion on this issue can be found in Sumner (2015), which reinforces the above interpretation.

information appears to have a more important role in speech perception than abstractionist theories allow.

Third, there is now a very large body of evidence showing speaker/voice effects in speech perception. For example, listeners identify words better from familiar native-accented speakers than from unfamiliar ones (e.g., Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993). In the study by Nygaard et al. (1994), native listeners of American-accented English were trained on the voices of 10 speakers from a database. They were then tested on the names of those speakers as well as novel words presented in noise. The experimental group was tested on the novel words produced by the same speakers they had been familiarized with from training, whereas the control group was presented with the novel words produced by novel speakers. Results showed that the experimental group that was tested on familiar speakers identified novel words in noise better than the control group that was tested on novel speakers. This voice familiarity advantage in native accents also extends to foreign accents. Listeners were found to adapt better to familiar foreign-accented speakers, compared to unfamiliar ones (e.g., Bradlow & Bent, 2008). Consequently, a speaker's personal information appears to be integral in the perception of linguistic information, not noise to be filtered out as abstractionist accounts would suggest.

In sum, there is now substantial experimental evidence that all the types of information that are conveyed in the speech signal (linguistic information, socio-indexical information, and speaker's personal information) can influence speech perception. In the next section I will review the episodic approach to speech perception, which, unlike the abstractionist approach, acknowledges the role of socio-indexical information and speaker's personal information in speech processing.

2.2 Episodic approaches to speech perception

The concept of episodic memory must firstly be clarified as episodic approaches to speech perception are built on this concept. Episodic memory was first observed by J. Nielsen (1958) in a neurological report of memory and amnesia and the term "episodic memory" was later suggested by Tulving (1972). As the originator of the term defines it, episodic memory is a neurocognitive (mind/brain) system that involves three central components: "sense of subjective time," "autonoetic awareness," and "self" (Tulving, 2002). It stores events that happened to the self, and comprises two features: (1) the "what" (happenings), the "where" (particular places), and the "when" (particular times); and (2) "re-collective experience" (the self must be consciously aware of past happenings). As a result, episodic memory allows the self to time travel, to "consciously re-experience past experiences" (Tulving, 2002, p. 6). It exists in parallel and works in tandem with (but independently from) the memory of learned knowledge and facts, which bears no relation to the self, called semantic memory.

Episodic theory, which provides a framework for representing relevant phonetic details, centers on episodic memory. Unlike the abstractionist approach, the episodic approach provides a theoretical framework for capturing all the types of information conveyed in the speech signal and experienced by the self (e.g., Goldinger, 1996, 1998; Hintzman, 1986; Jacoby, 1983; Johnson, 1997; Palmeri et al., 1993; Pierrehumbert, 2001, 2003, 2006, 2016; Pisoni, 1993). According to the episodic theory of the lexicon, each encounter with speech leaves a unique trace in the listener's memory called an episode. An episode contains not only the phonetic details of the speech but also the details about the speech context such as the speaker's social and personal information including regional origin (Goldinger, 1996), age, gender, and emotional state (Abercrombie, 1967). These details together constitute the properties of an episode. Categorization is performed by direct comparisons between the properties of a speech stimulus and a whole aggregate of properties of stored episodes (e.g., Hintzman, 1986; Jacoby, 1983). Episodes are activated when similarities are detected between the properties of the activated episodes and the current incoming stimulus. The more similar the properties, the stronger the activation. The stimulus is then classified by summing the properties of all the activated episodes. Abstract representations could be formed at this later stimulus classification stage, rather than earlier in the detail-encoding stage (e.g., Hintzman, 1986; Pierrehumbert, 2001). They could be formed through either generalizing clusters of similar experiences/distributions of episodes or extracting the parameters of those distributions, which are updated every time a new speech token is experienced and a new episode is stored (e.g., Pierrehumbert, 2002, 2006, 2016). However, abstraction is considered unnecessary in the traditional version of the episodic approach (e.g., Hintzman, 1986; Pierrehumbert, 2001). The recent version of episodic theory, however, started to acknowledge the function of abstraction alongside the importance of stored details in speech perception and word recognition (e.g., Goldinger, 2007; Pierrehumbert, 2002, 2006, 2016). In fact, the recent version of episodic theory reflects the hybrid view mentioned in the next section.

Because episodic theory takes into account all the details of speech experiences, it offers a straightforward account of how familiar voices facilitate word identification as well as adaptation to foreign accents (e.g., Bradlow & Bent, 2008; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993), and how beliefs and expectations based on episodes may have an influence on speech perception (e.g., Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). Take the studies by Hay and colleagues (Hay & Drager, 2010, Hay et al. 2006a) as an example. Even though the stimuli remained the same, a shift in perception was found in accordance with the socio-indexical cues "Australian" and "New Zealander" (Hay et al., 2006a), and stuffed toy kangaroos, koalas, and kiwis (Hay & Drager, 2010). As explained in those studies, previously encountered tokens of Australian and New Zealand variants of a sound (e.g., sound [ɪ] in "fish" is produced at the high front position in the Australian variant but near the centre of the vowel space in the New Zealand variant [ɨ]) are fully stored in listeners' memory as detailed episodes. Specified within those phonetically detailed episodes is social information about the speakers such as their nationalities (i.e., Australian, New Zealander). When listeners saw the "Australian" or "New Zealander" labels on the answer sheet, or stuffed toy kangaroos, koalas, and kiwis in the test room, the social categories Australian/New Zealander were activated, which in turn activated the corresponding Australian/New Zealand episodes. Episodes that had similar phonetic values to the experimental stimuli were also activated during the course of the experiments. As a result, episodes that were activated on both phonetic and social grounds reached full activation, compared to those which were activated only on one ground or the other.

Although episodic theory has a strong explanatory capability and its principles of speech processing have been used in a number of models such as MINERVA 2 (Goldinger, 1998; Hintzman, 1986), Johnson's (1997) model, and Pierrehumbert's (2002) models, the theory still has its own explanatory weakness. In this day and age of globalization where it is easy to meet new people with entirely different speech backgrounds, it is highly probable that we have to process speech from someone to whom we have not had previous exposure. Episodic theory on its own may not be able to explain such speech perception scenarios because pure episodic theory relies heavily on listeners' previous speech exposure. Episodic theory also relies heavily on the frequency with which a phonetic material (e.g., a rhotic vowel, a tap, an unreleased final stop) is encountered, but frequency alone cannot explain some speech perception scenarios as described in, for example, Sumner (2013), Sumner and Kataoka (2013), and Sumner, Kim, King, and McGowan (2014). In addition, by ignoring the role of abstraction in speech perception, pure episodic theory cannot explain findings from studies that show the generalization of phonetic learning (e.g., Cutler & Weber, 2007; Goldinger, 2007; McQueen, Cutler, & Norris, 2006; K. Nielsen, 2011). As mentioned earlier, the recent version of episodic theory acknowledges the importance of phonological abstraction in speech processing (e.g., Goldinger, 2007; Pierrehumbert, 2002, 2006, 2016). It discusses storage of speaker details together with clusters of similar speech episodes, which are abstracted into phonological categories and social categories (Pierrehumbert, 2006). However, although the theory acknowledges further abstractions over phonological categories such as phonological classes (e.g., lower classes: stops vs. fricatives vs. approximants, labials vs. alveolars vs. velars, monophthongs vs. diphthongs; higher classes: consonants vs. vowels), it still does not take any position regarding further abstractions over social categories (see Pierrehumbert, 2016). In real life, the specific speaker details that we are exposed to and store are not as simple as just one social category "Australian" or "New Zealander," but they could include many social categories at the same time such as young Australian female with a broad[5] accent. In that case, would social categories get abstracted further into more abstract social categories, such as "socio-indexicality," in the same way that phonological categories get abstracted further into phonological classes? An exploration of this question via an experimental study could contribute not only to broaden our current understanding of how speech is processed in the face of various social variables, but also to provide an evaluation of episodic theories in comparison to other speech perception theories and models.

In short, the episodic approach in its traditional version has significantly improved our current understanding of how a variety of speech events are processed by taking into account all the three types of information carried in the speech signal. However, there are still speech scenarios that it is not able to provide a clear explanation for, especially those that involve the role of abstract phonological categories, which the episodic approach in its recent version has begun acknowledging. By positing an abstract level into its theory, the episodic approach has brought itself closer to the hybrid approach to speech perception, which will be discussed in the next section.

[5] Broad/general/cultivated accents: These are accent types which are used to categorize Australian-accented English speakers from the 1960s to the 1990s on a continuum. In Harrington, Cox, and Evans's (1997) acoustic study of the differences between these accent types, the broad accent at one end of the continuum is defined as exhibiting features similar to London Cockney English, being stigmatized but clearly marking Australian English. The cultivated accent at the other end is defined as sounding close to Received Pronunciation of British English, and the general accent lying in-between.

2.3 Hybrid approach to speech perception

The recent landscape of speech perception research has seen the rise of the hybrid approach: an acknowledgement of the role of both abstract categories and specific episodes in speech perception and word recognition, which has also been supported recently by influential researchers in the traditional episodic framework (see Goldinger, 2007; Pierrehumbert, 2002, 2006, 2016). Evidence for the need of a hybrid approach in speech perception has begun appearing in recent studies both by episodicists and abstractionists (e.g., Cutler & Weber, 2007; Goldinger, 2007; McQueen et al., 2006; K. Nielsen, 2011). In the study conducted by McQueen et al. (2006), the question of the nature of stored knowledge in the mental lexicon is explicitly pursued, which was answered by an experiment testing for generalization of phonetic retuning. Stimuli from the experiment involved Dutch words recorded from a native Dutch speaker ending in [f] and [s] sounds, which were replaced by an ambiguous [f-s] sound. In the training phase, participants were trained on the ambiguous [f-s] sound replacing word-final [f]s in one group, and the ambiguous [f-s] sound replacing word-final [s]s in another. In the test phase, they were presented with related ambiguous primes and unrelated unambiguous primes, and asked to decide on whether the letter strings they saw on the screen after hearing the primes were real words. Results showed that participants in the two groups responded differently, depending on whether they were trained on the ambiguous [f-s] sound replacing word-final [f]s or word-final [s]s. These results have implications for the hybrid nature of stored knowledge in the mental lexicon. Phonetic learning does not occur in the pure abstractionist approach as phonetic details are unimportant, while generalization of the learning does not occur in the traditional version of the episodic approach. However, results from this study support a hybrid approach that acknowledges the importance of both abstract categories and specific episodes in speech perception and word recognition.

Together with the emergence of the hybrid approach (and the recent version of the episodic approach), Bayesian models – those that make use of Bayesian statistics to optimize behaviour as well as to update knowledge and beliefs – have started to become popular in speech perception research (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kirov & Wilson, 2013; Kleinschmidt & Jaeger, 2015; K. Nielsen & Wilson, 2008; Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2016; Wilson & Davidson, 2013). There are two reasons for this trend (e.g., Clayards et al., 2008; Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008). Firstly, a probabilistic framework is consistent with the inherent ambiguity of the speech signal and, therefore, listeners' uncertainty in perceiving speech. Secondly, Bayesian modelling provides a powerful means to optimize the speech perception and word recognition process by combining all the available sources of information (e.g., speech signals, listeners' prior knowledge), whether the information is fully specified or abstract. On the "fully specified" side, a Bayesian model takes direct observations of the speech signal as one of its inputs, called "likelihood," or "signal specificity" in this thesis. It could also take previous detailed experiences as its other input, called "prior probability." While episodic models are able to account for the effects of knowledge and corresponding expectations[6] that come from previous exposure (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999), they have been shown to implement Bayesian inference (Shi, Griffiths, Feldman, & Sanborn, 2010). Bayesian models with Bayesian inference can, therefore, account for the effects of knowledge and corresponding expectations that come from previous exposure, similarly to episodic models. On the "abstract" side, a Bayesian model could incorporate as inputs for "prior probability" listeners' abstract phonological/social categories as well as their abstract psychological processes such as beliefs that do not come from direct experiences/stereotypes (defined as "beliefs that do not come from direct experiences"[7] in the context of this thesis) and corresponding expectations. Therefore, Bayesian modelling is entirely capable of representing the processing of speech from someone we do not have previous exposure to, which is presumably more abstract than direct experiences of that person, and which is difficult – if not impossible – for a pure episodic model. The flexibility of Bayesian models lies mostly in the concept of "prior probability," which tolerates both specificity and abstraction.

2.4 Summary

Overall, while abstractionist approaches only focus on linguistic information in explaining how speech is perceived, more and more speech perception studies are suggesting that socio-indexical information and a speaker's personal information also need to be taken into account. As a result, episodic approaches, which can provide one way of accounting for how socio-indexical information and speaker's personal information might influence speech perception, have become increasingly popular. However, these approaches are currently

[6] In this thesis, expectations are considered the product of either having certain knowledge (that come from previous exposure) or holding certain beliefs (that do not come from previous exposure, but instead coming from abstract sources such as socio-cultural learning and hearsay).

[7] Stereotype can also be defined in relation to over-generalization or exaggeration (as mentioned in the theoretical work of prejudice by Allport, 1954), which will be pursued in future research.

uncommitted regarding further abstractions over socio-indexical information. Hybrid accounts allow for both abstract and episodic processing, but it is not yet clear whether or when socio-indexical information is being processed in an abstract or an episodic way. The next section outlines how abstractions might be made on the basis of direct exposure to a speaker as well as in the lack thereof (stereotypes and prejudice), and how such abstractions might provide a theoretical framework to investigate the influence of abstract socio-indexical information on speech perception.

3 Abstraction over socio-indexical information

3.1 Abstractions over speaker characteristics formed from direct speech exposure

If listeners can abstract over social details to produce social categories, just like phonetic details can be abstracted to produce phonological categories, as suggested in Pierrehumbert (2006), then the question is whether social categories can be abstracted further to produce an even more abstract social category. This is not too far-fetched a question as we have seen phonological categories abstracted further into phonological classes, which are demonstrated in numerous studies to have different relationships with speech perception (see Feldman, Griffiths, & Morgan, 2009 for a summary of those different relationships). Similarly, given several different social categories such as age, nationality, gender, and accent type (as the earlier example "young Australian female with a broad accent"), it is possible to hypothesize some abstraction over those categories. In that case, the higher level of abstract social category could be represented in memory as, for example, "socio-indexicality." It could contain two members: "more socio-indexical" (when abstraction takes place over two or more social categories) versus "less socio-indexical" (when abstraction only takes place over one single category). Just like the different levels of abstraction over linguistic information, if such a highly abstract social category exists, we would expect to see some difference in the relationship between each of its members ("more socio-indexical" vs "less socio-indexical") and speech perception. A perception experiment where speech stimuli fall into the two groups of "more socio-indexical" and "less socio-indexical," or some abstract label to that effect, would be able to test this level of abstractness.

If a high level of abstract socio-indexical information is part of what is extracted in speech perception, then we might expect stereotypes to have an influence on speech perception. Stereotypes are essentially abstractions over socio-indexical information. However, the difference between stereotypes and the high level of abstraction over socio-indexical information mentioned above is that the latter originates from the socio-indexical details stored in listeners' episodic memory in the first place, whereas stereotypes are abstracted when there is no or little direct experience to draw from in episodic memory (i.e., when there is exaggeration) (Allport, 1954). The next section will discuss abstractions over speaker characteristics formed without direct speech exposure, such as stereotypes[8] and prejudice.

3.2 Abstractions over speaker characteristics formed without direct speech exposure – Stereotypes and Prejudice

As reviewed in Section 2.1, the findings from Niedzielski's (1999) study were shown to support the importance of socio-indexical information in speech perception. Niedzielski's findings could easily be explained in the episodic framework, as suggested in Hay et al. (2006a). However, a different explanatory approach concerning language attitudes was offered in Niedzielski (1999). As a recap of the findings, despite being exposed to the same speech stimuli from a Detroit speaker, only the Detroit listeners who expected to hear a Canadian speaker accurately perceived the raised vowels exactly as they were. Meanwhile, those who expected to hear a fellow Detroiter perceived standard American-accented English vowels. This contradiction seems to point to the idea that Detroit listeners had two different beliefs, one about themselves and one about Canadians: (1) Detroiters speak standard American-accented English; and (2) Canadians speak the version of English with raised vowels. As Allport (1954) puts it, if beliefs are exaggerated and applied to another group, they become stereotypes. This idea is actually confirmed by results from a language attitude survey in Niedzielski (1996): Detroiters did hold such speech stereotypes about themselves and Canadians. In other words, from a social psychology perspective, Niedzielski's (1999) study raises the prospect that there may be a strong effect of stereotypes on speech perception. Because stereotypes are an abstract social concept, pure episodic theory would fail to account for such abstractness; however, flexible Bayesian approaches (whether hybrid or current episodic) could do so. A perception experiment that is designed to tease apart these two approaches would involve only participants without previous, direct exposure to the target speech accent, and predictions for speech perception

[8] Throughout this thesis, stereotypes are discussed in terms of how they are formed: Stereotypes about a target group are formed in the absence of exposure to that group, or in the lack of direct evidence to support certain beliefs about that group. After being formed, stereotypes can persist even after contact/exposure to the target group, but stereotype persistence is not covered in this thesis.

would then be made for each approach accordingly.

Stereotypes are just one component of attitudes between groups, the cognitive component. The other components are prejudice (the affective[9], evaluative component) and discrimination (the conative[10], behavioural component) (e.g., Fiske, 1998). Prejudice is based on feelings such as admiration, contempt, envy, and pity (Fiske, Cuddy, Glick, & Xu, 2002), which may or may not have arisen from direct exposure to the target group. Prejudice has been shown to have a clear effect on behaviour in social psychological research such as the study by Dutta, Kanungo, and Freibergs (1972). In that study, three experiments were conducted to determine the role of the intensity of perceived affect/emotions in trait retrieval behaviour. More specifically, prejudiced English Canadians (Experiments 1 and 3) and French Canadians (Experiment 2)[11] went through lists of positive and negative traits, which were initially claimed to describe themselves and the other group (i.e., French Canadians as the other group for prejudiced English Canadian participants, and English Canadians as the other group for prejudiced French Canadian participants). Participants were then told that the traits that they had been told earlier to apply to their group actually applied to the other group, and vice versa. This reversal of information was introduced before the rating phase for one participant group and after the rating phase for another participant group. In the rating phase, participants rated the perceived intensity of their emotions towards the traits. In the end, they were asked to recall the lists of traits for the groups in accordance with the reversal information. Results showed that prejudiced English Canadians recalled more negative traits for French Canadians and more positive traits for their own group, and vice versa for French Canadians. These findings clearly show that prejudice affects recall, in particular. Therefore, it is reasonable to hypothesize a relationship between prejudice and speech perception. Evidence of such a relationship would further support the central hypotheses of this thesis, that high levels of abstractions can take place over socio-indexical information and that such abstractions can have a relationship with speech perception. It would also provide

[9] In theoretical works concerning attitudes between groups such as Fiske (1998) and Lambert and Lambert (1964) as well as corresponding empirical works such as Stephan and Stephan (1993), the affective component of attitudes commonly includes feelings, emotions, moods, and the like. It is often argued to also include evaluations (Fiske, 1998). In this thesis, the affective component of attitude includes feelings, emotions, moods, and evaluations (for a discussion on what the affective component of attitudes encompasses, see Appendix 5).

[10] The equivalent but more widely used term is “behavioural” (used in, e.g., Agheyisi & Fishman, 1970). In theoretical works concerning attitudinal components such as Lambert and Lambert (1964), the phrase “tendencies to react” is used instead. As a result, the “conative” component of attitudes refers to the aspect of attitudes that deals with behaviour or behavioural tendencies. In the context of this thesis, the term “conative” is just mentioned in passing while the different attitudinal components are introduced. It is not a critical term for the purposes of this thesis.

[11] Participants were pre-selected on the basis of their prejudice scores.

support for the hybrid approach to speech perception.

In general, if a relationship between stereotypes/prejudice and speech perception does exist, it serves as clear evidence that high levels of abstraction over socio-indexical information in the speech signal does take place, perhaps in a similar manner as high levels of abstractions over linguistic information. It would suggest some modification to the episodic approach, as well as emphasize the need of broader and more flexible theoretical views in speech perception that can explain any possible abstractness in the information conveyed by the speech signal.

4 The present study

In this thesis, I evaluate the role of socio-indexical information in speech perception and, in particular, the possibility of higher levels of abstractions over social categories. The overall hypothesis pursued in this thesis is that listeners use abstract information whenever it is available in perceiving speech. This includes all the information they have been exposed to, heard about, known, or believed about speakers, sometimes even how they feel about speakers or speakers' respective groups, to sharpen their predictions about how speech varies. This could include the following three possibilities:[12]

(1) knowledge about the speaker, the speaker's speech, or even the speaker's group, which are based on *direct exposure* to the speaker, the speaker's speech, or the speaker's group. Because listeners have had direct exposure, such speaker details could be stored in memory together with specific speech details as episodes and abstracted into social categories and phonological categories. These could be abstracted further into highly abstract categories. Episode matching can definitely explain speech perception at the level of social categories and phonological categories, and, if extended, can possibly explain speech perception at the level of highly abstract categories;

(2) beliefs about the speaker, the speaker's speech, or even the speaker's group, which are based entirely on *indirect sources* such as socio-cultural learning or hearsay, and which may or may not be (factually) grounded. In this case, episodic matching is not available as a mechanism in speech perception. Nevertheless, the perception of speech in such cases may be influenced by stereotypes, resulting from

[12] The three possibilities do not map onto the three components of attitudes because (1) discrimination (the behavioural component) was not tested, and (2) the first possibility is purely about sociophonetics.

abstractions over no observations or only a small number of observations, and can be modelled as a prior probability in a Bayesian framework; or

(3) feelings towards the speaker, the speaker's speech, or even the speaker's group, which could be based on either direct exposure to speech and speakers or on indirect sources such as those listed above. Episodic theories would not be efficient as a mechanism for explaining speech perception based on prejudice, resulting from abstractions over feelings.

Specifically, the thesis seeks to answer the following research questions:

(1) The first research question, addressed in Chapter 2, concerns abstractions over social categories formed from direct speech exposure. As noted earlier, it is possible for speaker characteristics to be abstracted into narrow social categories, which can then be further abstracted into broader categories. This raises the following question: Do higher levels of abstraction exist for linguistic information only, or do they also exist for socio-indexical information? If they exist for both linguistic information and socio-indexical information, the inference is that they exist for a speaker's personal information too. Then we can make generalizations about abstractions across the three types of information in the speech signal.

(2) Stepping away from direct exposure to speech, the second research question, investigated in Chapter 3, turns to socio-indexical abstractions formed from indirect sources such as socio-cultural learning or hearsay. In such a case, the representation of speaker-specific speech categories could not have been built over relevant distributions of episodes. Without direct exposure to speech, how would a speaker or speaker's group be represented? More specifically, does the representation of speech categories reflect listeners' beliefs about speakers that they have no direct exposure to? If it does, how do these manifest in speech perception?

(3) To provide another similarly abstract scenario, the third research question, explored in Chapter 4, moves beyond the cognitive component of listener-speaker relationship into the affective component, or prejudice. As reviewed earlier, a question remains as to whether prejudice is associated with speech perception. If they are associated, how do other listener factors (both abstract and specific) such as expectations about a speaker's speech and previous experience with it play out in the relationship between prejudice and speech perception?

To address the above research questions as well as to ensure the robustness of the research, numerous decisions were made regarding methodological details such as (a) the linguistic nature of the speech stimuli, (b) the origins of the speakers, (c) the origins of the listeners, (d) collecting perceptual data, (e) modeling speech perception data, and (f) finding tools to measure prejudice. Each of these is addressed in turn below.

(a) The linguistic nature of the speech stimuli

Vowels are the object of perception in this thesis as they vary substantially across accents of English, especially outside of the British Isles (Wells, 1982b). Therefore, it is easy to find patterns of variation that a given listener group is likely to have had direct exposure to and also patterns of variation that the same group has not had exposure to (but possibly heard about). Australian-accented English and Vietnamese-accented English represent those patterns of variation to the listeners in the thesis. The different patterns of variation may evoke different perceptual mechanisms, roughly characterized by episodic theories and Bayesian inference.

(b) The origins of the speakers

Vietnamese-accented English is a foreign accent of English, which was selected for use in this thesis because foreign accents appear to be able to provide clearer evidence for the effects of interest. Speech perception is shifted in studies with cues of regional accents (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999), but the expectation effects of regional accents on accuracy are still unknown. Nevertheless, this is not the case with cues of foreign accents. In studies with cues of foreign accents, the “Chinese” cue (versus “Caucasian”) was found to either undermine American-accented English listeners’ perception of the Standard American English signals (e.g., Rubin, 1992), or enhance their perception of Chinese-accented English signals (e.g., McGowan, 2015). Consequently, even though regional accents and foreign accents both exhibit variation in speech, the effects of socio-indexical cues appear to be different: Listener flexibility is more compromised when listeners expect to listen to foreign accents (e.g., McGowan, 2015; Rubin, 1992). Thus, if the stimuli expose a foreign accent to the listeners’ ears, it may be easier to find evidence for stereotypes and prejudice. This makes foreign accents a better choice for stimulus variation than regional accents, for the goals of the present research.

The Vietnamese accent was specifically selected over any other foreign accent accessible in the Greater Sydney area where the thesis project took place for two reasons. First, it is a foreign accent to which a large number of the listeners in the thesis would have little to no previous exposure. According to the Australian Bureau of Statistics (2011), in the Greater Sydney area, only 1.6% of people were born in Vietnam (and thus could possibly speak Vietnamese-accented English). The low incidence optimizes the likelihood of finding listeners who lack exposure to the Vietnamese accent. Second, Vietnamese-accented English is considered difficult to perceive (e.g., Cunningham, 2009; Ingram & Nguyen, 2007; T. Nguyen & Ingram, 2004; Zielinski, 2003, 2006, 2008), which increases the likelihood of finding listeners who are stereotyped and/or prejudiced towards it.

(c) The origins of the listeners

The listener group in this thesis speaks Australian-accented English and was selected in consideration of the accents chosen for the speech stimuli. Australian-accented English listeners have had (large amounts of) exposure to Australian-accented English. They may not have had exposure to Vietnamese-accented English but most possibly have heard about it, and most possibly stereotypically consider it difficult to perceive (Ingram & Nguyen, 2007; T. Nguyen & Ingram, 2004; Zielinski, 2003, 2006, 2008). Moreover, some Australians (especially white Anglo-Saxons) have been reported to harbour stereotypes and prejudice towards Asians (e.g., Islam & Jahjah, 2001; Walker, 1994). This makes the Australian listener group highly appropriate for investigating the relationship between speech perception and abstractions over socio-indexical information, including stereotypes and prejudice towards Vietnamese/Asians.

(d) Collecting perceptual data

The vowel perception data were collected via a categorization task, which can reveal the perception processes at the phonological and phonetic levels. One variant of the categorization task involves listeners listening to speech stimuli, typically in the form of syllables or non-words, and then selecting keywords that contained the sounds of interest in the speech stimuli (e.g., Bundgaard-Nielsen et al., 2011; Faris et al., 2016; Tyler, Best, Faber, & Levitt, 2014). This is the variant used in this research project. Other variants of the task involve matching the sounds of interest in a sentence with synthetic sounds (e.g., Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999), ticking or crossing out the word heard on an answer sheet (e.g., Ladefoged & Broadbent, 1957; Peterson & Barney, 1952), and identifying synthetic sounds as natural sounds (e.g., Miller, 1953; Schatz, 1954). The categorization task, in general, and its keyword identification variant, in particular, were selected for three reasons. First, previous studies testing the influence of socio-indexical information in speech perception used this type of task and demonstrated clear results (e.g., Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). This research project also tests the role of socio-indexical information, although at more abstract levels. Second, this type of task has been extensively used in papers that give support to certain speech perception theories or models (e.g., Hay et al., 2006a; Tyler et al., 2014). This project aims to find evidence relevant to addressing hybrid models of speech perception. Third, the keyword identification variant of the task has been used in cross-language perception studies to investigate the assimilation of non-native phones to native phonological categories (e.g., Bundgaard-Nielsen et al., 2011; Faris et al., 2016; Tyler et al., 2014). This project involves Australian-accented English listeners and Vietnamese-accented English stimuli, which is cross-accent rather than cross-language. However, on the first exposure to an unfamiliar accent, it is possible for the initial perceptual assimilation process to also occur to phones in another accent in a similar manner as phones in another language. In brief, given the aims and methodological choices in the project, the categorization task, in general, and its keyword identification variant, in particular, are necessary to address the research question of whether social categories can be further abstracted and how such high levels of abstractions are related to speech perception.

Given the high relevance of the keyword identification variant of the categorization task in this project, a detailed review of an example study using the keyword identification variant is in order. In Tyler et al. (2014), the categorization task was used in conjunction with a discrimination task to determine if principles underlying non-native consonant perception also apply to non-native vowel perception under the Perceptual Assimilation Model (PAM) and the Natural Referent Vowel (NRV) framework (Polka & Bohn, 2003, 2011). PAM predicts that non-native contrasts could be assimilated to native contrasts in several different ways: two-category assimilation (TC) when a pair of non-native contrasts are assimilated to different native phonological categories; single-category assimilation (SC) when a pair of non-native contrasts are assimilated to the same native category as similarly good or poor; category-goodness assimilation (CG) when a pair of non-native contrasts are assimilated to the same native category, but one sounds closer to the native category than the other; uncategorized-categorized (UC) when one non-native phone in the pair is assimilated to a native category while the other is not; and uncategorized-uncategorized (UU) when both phones in the pair are not assimilated to any native category. Discrimination is predicted to be excellent for TC and UC, fair to very good for CG, poor for SC, and variable for UU. The NRV predicts that the change from more peripheral vowels to less peripheral ones is harder to perceive than the change from less to more peripheral. The combined predictions from PAM and NRV state that the change from more to less peripheral vowels is harder to perceive than the change from less to more, might only occur for SC, CG, and UU, and that TC and UC do not support this prediction. Six non-native vowel contrasts were chosen as stimuli: Norwegian /ki/-/ky/ and /ki/-/kʉ/, Thai /bɯ/-/bɤ/, and French /bo/-/bõ/, /dø/-/dœ/, and /sy/-/sø/, recorded from separate female native speakers of Norwegian, Thai, and French, respectively. The participants were American university students. They were asked to first perform a discrimination task in the form of AXB (more specifically AAB, ABB, BAA, and BBA), in which A represented the more peripheral vowel in a pair of non-native contrasts and B the less peripheral one. Participants had to make judgments on whether X, a different token of the same vowel as A or B, sounded the same as A or B. After that, they were asked to complete a keyword identification task, in which they listened to a stimulus token and chose an English keyword that contained the vowel they just heard in the token. They then rated the similarity of the vowel they heard and the vowel in the keyword they just chose on a scale from 1 (*unlike*) to 5 (*identical*). The American English vowel categories /i ɪ e ɛ æ ɑ ɔ o ʊ ʌ u ɝ/ were represented by the keywords "heed, hid, aid, ed, ad, odd, awed, hoed, hood, dud, food, heard," together with several more additional keywords to allow for the categorization of some special cases involving coarticulatory influences on the vowel, such as "end, donned, owned, dude." Results supported the combined prediction from PAM and NRV for SC versus TC and UC, and would potentially support the prediction for CG with a larger sample. As demonstrated in this study, the keyword identification of the categorization task contributed to evaluate speech perception models via means of cross-language stimuli.

(e) Modeling speech perception data

Modelling speech perception data to explain the current findings then make predictions for future findings has become an increasing trend in speech perception research. It provides a means to better understand behavioural data and put theory into practice. Analyses in this project center on mixed-effects regression and

Bayesian models, in an attempt to understand actual listener perceptual data in a neat way and to evaluate support for the hybrid approach in speech perception. The vowel categorization task that this project employs produces binary data (with 0s symbolizing incorrect categorization and 1s correct categorization), which would typically be analyzed with logistic regression in sociolinguistics (aka VARBRUL analysis – Baayen, 2008). Just like any regression methods, logistic regression can account for not only the effects of interest (main effects) but also any extraneous effects present in the data (random effects) such as the variation in stimulus items and differences in participants' baseline performance in one single model, called mixed-effects logistic regression (see Baayen, 2008, for more information on the use of this type of model for psycholinguistic data). Stimulus variation and participant variation are unavoidable variables in this type of psycholinguistic research. In addition, it is more efficient to treat both types of variation in one single model as in a mixed-effects model than in two separate analyses as in the ANOVA method. As a result, the use of mixed-effects regression models as an analytical method is necessary. However, mixed-effects regression modeling generally deals more with statistics and does not provide a conceptual framework to account for the hybrid approach in speech perception, the theoretical foundation of this research project. As discussed in Section 2.3, Bayesian inference does provide a clear conceptual framework to model both abstraction and experiences in speech perception, testified by its wide application in a range of psycholinguistic studies (e.g., Clayards et al., 2008; Kirov & Wilson, 2013; Kleinschmidt & Jaeger, 2015; K. Nielsen & Wilson, 2008; Norris & McQueen, 2008; Norris et al., 2016; Wilson & Davidson, 2013). As a result, Bayesian modeling was also used alongside with regression modeling to yield the findings in this project.

(f) Finding tools to measure prejudice

Finding the right tool to measure listener prejudice was a major challenge undertaken in this thesis project. There are relatively few studies on prejudice in Australia. The attitudes of young Australians towards three minority groups, Aboriginals, Asians, and Arabs, were investigated in Islam and Jahjah (2001). The prejudice measure used in their study was quite direct. Participants were asked to rate their emotions in relation to the target minority groups. The directness of this method risks collecting non-genuine responses from participants due to their concern about the social desirability of their responses (Orne, 1959). An indirect method such as the Scale of Anti-Asian American Stereotypes (SAAAS, developed by Lin, Kwan, Cheung, & Fiske, 2005), where participants respond to stereotype items through which prejudiced attitudes can be inferred, is preferred to reduce (although not completely eliminate) social desirability concerns. SAAAS was developed out of the Stereotype Content Model (Fisk et al., 2002). The model claims that social groups who are stereotypically considered to be high on competence but low on sociability are admired but envied. Following the Stereotype Content Model, SAAAS explores prejudice as a mixture of feelings that automatically accompany a mixture of stereotypes. As a result, although SAAAS items are about stereotypes, their total score indicates prejudice: the higher the score, the more negative the prejudice. This indirectness could help reduce bias in responses that may be attributed to participants' social desirability concerns, at least in regards to their prejudice. Appendix 5 demonstrates the robustness of SAAAS as adapted into the Australian context, with some small changes in its construction, and establishes its validity in the Australian context.

The following paragraphs describe the details of the experiments that demonstrate how the above methodological choices fit together to contribute to the progression of the thesis. The three experimental chapters will be presented as stand-alone journal articles. The conclusion of these stand-alone chapters will be discussed integratively in the Discussion chapter.

The first experiment (Chapter 2) shows how Australian-accented English listeners deal with patterns of variation in Australian-accented English vowels. This experiment uses the known properties of vowel formant measurements to probe the relationship between mental representations of vowel categories and socio-indexical information. Specifically, it tests the hypotheses that listeners' phonological categories are represented by distributions of phonetic values and that categories that carry more socio-indexical information have multiple distributions. Testing these specific hypotheses informs the broader question about abstractions over social categories formed from direct speech exposure. Results showed that listeners did form and make use of distributions over phonetic values, and that listeners may have used more specific and socially relevant distributions for a certain group of vowels because, through life-long exposure, they have implicit knowledge that these vowels are likely to vary more among speakers of different generations or socio-economic groups. Findings from this first experiment potentially demonstrate the existence of high levels of abstractions over socio-indexical information and provide some support for the hybridization of abstractionist and episodic views.

The second experiment (Chapter 3) concerns how Australian-accented English listeners deal with variation patterns in Vietnamese-accented English vowels when they do not have previous exposure to the accent. This is the experiment that assessed how speech categories are represented without direct exposure to speech. As the listeners did not have previous exposure to Vietnamese-accented English vowels, no episodes of these vowels would have been stored, and thus none would have been activated to facilitate the perception of those vowels on the introduction of the socio-indexical cue about the speaker's Vietnamese accent. Thus, episodic theories would predict no shift in perception. In contrast, a shift in perception would be possible if listeners formed abstract beliefs (grounded or ungrounded) about the Vietnamese accent, which could potentially play a role in the computation of listeners' perceptual performance on the Vietnamese-accented vowels. Results showed that listeners expected to hear Vietnamese-accented vowels at the extreme edges of the Australian-accented vowel space less often and vowels towards the center of the vowel space more often. In other words, the listener's perceptual space shrank when they were told to expect a Vietnamese accent even when they did not have previous exposure to the accent. This suggests that, in the absence of direct observations, speech categories are formed in accordance with listeners' abstract beliefs about them. As a result, there is more to the formation of beliefs about accented vowels than just abstractions over direct observations. The application of Bayesian statistics into speech perception can provide an approach that accounts for how socio-indexical abstractions formed from indirect sources are used in speech perception.

The third experiment (Chapter 4) invites similar abstractness in the design by changing the focus from the cognitive component of listener-speaker relationship (exposure/experiences, beliefs/stereotypes, expectations) to the affective component (prejudice). It has appeared in University of Pennsylvania Working Papers in Linguistics (N. Nguyen, Shaw, Pinkus, & Best, 2016). The design still involved Australian-accented English listeners and variation patterns in Vietnamese-accented English vowels. However, unlike the second experiment which controlled for listeners' lack of previous exposure, this experiment examined a more complex picture of speech perception, with listeners' expectations about a speaker's accent, their previous exposure to the accent, and their prejudice towards the speaker's group all taken into account. This complex interaction between abstract and specific speech variables is currently understudied although it is important for a broader understanding of speech perception dynamics. Results indicated that the relationship between listener expectations and prejudice varied according to whether or not participants had previous exposure to the Vietnamese accent. Listeners' prejudice towards a speaker's group apparently goes hand in hand with their productive use of their episodic memory. This experiment possibly provided further evidence that episodic theory does not have adequate power to explain listener flexibility in this case. Therefore, a more nuanced theory that makes use of abstractions over socio-indexical information and incorporation of both abstractions and specifics in a probabilistic algorithm is necessary to accurately describe speech perception scenarios.

The three experiments introduced above constitute the empirical chapters of the thesis, but they are not the only studies carried out in the project. There were preliminary studies that formed the groundwork for these main chapters that are not being reported in the main body of the thesis. The first preliminary study contributed a methodological point in Chapter 2, which observed a consistent relationship between the mean and standard deviation of formant values as well as proposing a method of quantifying formant variability through the residualization of the mean formant values against the standard deviations. This study was published in the proceedings of the 15th Australasian International Speech Science and Technology Conference and is included in Appendix 3 for reference. The second preliminary study confirmed the existence of the relationship between prejudice and speech perception, before the complex interaction reported in Chapter 4 was explored. This study was published in the proceedings of the 18th International Congress of Phonetic Sciences and is also included in Appendix 6 for reference. Although these preliminary studies provided some groundwork for the main content chapters and are cited where appropriate, they are not crucial to the thesis as it is written and are included as appendices for completeness.

Summary

In this chapter, I have reviewed the relevant concepts, theories, and studies that set the scene for the introduction of my research project. More specifically, starting from the three typical types of information embedded in the speech signal (linguistic information, socio-indexical information, speaker's personal information), I evaluated the strengths and weaknesses of the two dominant approaches in speech perception, the abstractionist approach and the episodic approach, against these information types. Weaknesses of the episodic approach gave rise to the hybrid approach, which acknowledges the role of both specific details and abstract categories. With the above theoretical approaches in mind, I then turned to proposing two ways that socio-indexical information could be highly abstracted: (1) abstractions over speaker characteristics formed from direct speech exposure, and (2) abstractions over speaker characteristics formed without direct speech exposure, or stereotypes and prejudice. Finally, I demonstrate how these two types of abstractions are related to speech perception through a series of three carefully designed experiments.

Chapter 2

How socio-indexical information modulates the relationship between formant variability and vowel categorization

A version of this chapter has been submitted for publication as:

Nguyen, N., Shaw, J. A., Best, C. T., & Tyler, M. D. (under review). How socio-indexical information modulates the relationship between formant variability and vowel categorization. *LabPhon*.

Abstract: Phonological representations are often viewed as probability distributions over phonetic parameters. In this view, phonetic variability is central to phonological representations and crucial to predicting speech perception behaviour. We predict listeners' tolerance of variation for phonological categories that have wider as compared to narrower distributions over a given phonetic parameter. One challenge to testing this prediction is that listeners could rely on multiple distributions, particularly for segments carrying rich socio-indexical information such as accent types and sound change. Here we address that challenge by incorporating socio-indexicality in our prediction: (1) for vowels that are rich in socio-indexical information, a negative relationship is predicted between formant variability and categorization accuracy; and (2) for those that are not, the predicted relationship would be positive. Based on the sociophonetics literature, we divided Australian-accented English monophthongs into subsets that carry more versus less socio-indexical information. Results showed a negative correlation between formant variability and categorization accuracy for more socio-indexical vowels, suggesting that listeners used more specific, socially relevant distributions for those vowels. A positive trend was apparent for the less socio-indexical ones, possibly indicating that the positive relationship between formant variability and categorization accuracy would surface with a more controlled set of stimuli.

Keywords: phonetic variability, vowel perception, phonological representations, phonetic parameters, Gaussian distributions, likelihood of occurrence, phonological abstraction, episodic traces

1. Introduction

1.1. Phonetic variability and phonological category membership: Some predictions

Stochastic models of phonological representations frequently take the form of distributions, often Gaussian in shape, over phonetic parameters, such as formant values for vowels (Feldman et al., 2009), VOT for voicing contrasts (Clayards et al., 2008), or temporal intervals for syllables (Shaw & Gafos, 2015). In this view, phonetic variability is central to phonological representations. Distributions over phonetic parameters can be used to make quantitative predictions about phonological category membership in speech perception (Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008). For a Gaussian distribution, the likelihood of category membership depends on the mean and standard deviation (SD) of the underlying distribution. The mean of the underlying distribution can vary little across speakers, or it can vary greatly with social variables such as accent types in variation studies (e.g., Harrington, Cox, & Evans, 1997; Mitchell & Delbridge, 1965), or age cohort in studies of sound change (e.g., Cox, 1999; Cox & Palethorpe, 2008). The SDs of the underlying distribution can be wide, suggesting that the category may tolerate a wide range of variation in that phonetic parameter. They can also be narrow, suggesting that less variability in that parameter may be tolerated in the category (e.g., Norris & McQueen, 2008, p. 363).

To illustrate predictions of distribution variability for speech perception behaviour, we consider a hypothetical scenario, schematized in Figure 2.1. The figure shows two Gaussian distributions, a wide distribution and a narrow distribution, centred on the same mean but differing in variance. The y axis indicates the likelihood of a stimulus value occurring within a distribution, and the horizontal axis indicates how stimulus values fall within the narrow and wide distributions. A stimulus value such as (1) or (3) falls outside one or two SDs from the mean of the narrow distribution but inside the same SD range of the wide distribution. The likelihood of these stimuli is therefore higher for the wide distribution than for the narrow distribution. Consequently, these stimuli stand more chance of being categorized by the wide distribution than by the narrow distribution, and are thus more likely to be correctly perceived when the underlying distribution is wide (as opposed to when it is narrow). In other words, a wide distribution tolerates a wide range of variability within the category while a narrow distribution is more selective. For stimuli like (1) and (3), the larger SD of the underlying distribution is therefore predicted to aid speech perception accuracy. In contrast, when the stimulus value falls near the mean of the distribution, such as (2) in Figure 2.1, its likelihood is higher for the narrow distribution, compared to the wide distribution. As a result, it is more likely to be categorized within the narrow distribution than within the wide distribution. Stimuli near the mean of the distribution, on the other hand, benefit from narrow distributions. The variance of the underlying distribution leads to straightforward predictions for the perception of these stimulus types.

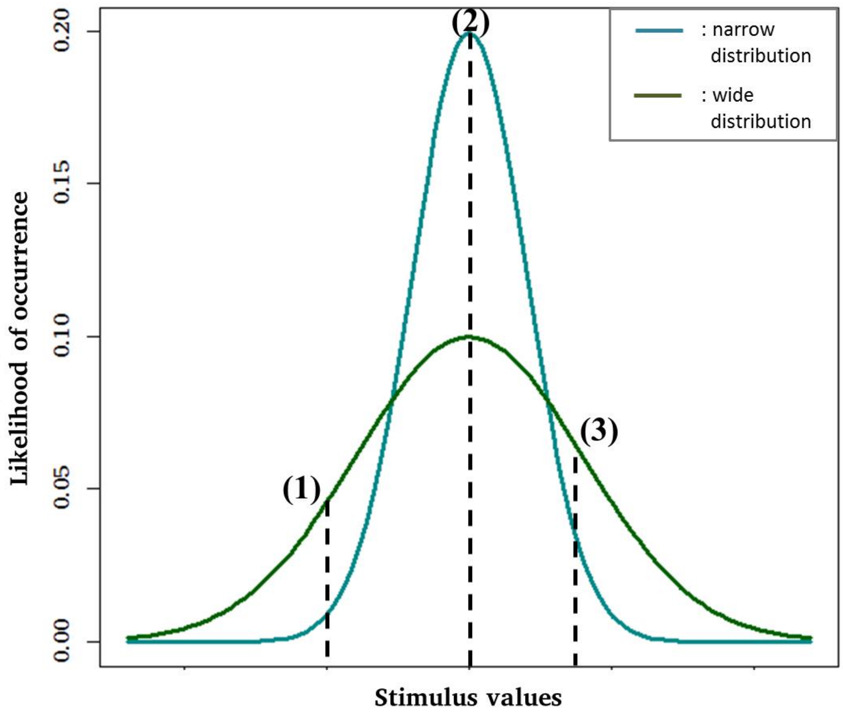


Figure 2.1: *Illustrations of the relationship between probability distributions and stimulus items* [(1) and (3) illustrate the case when the stimulus value falls on the left or right side of the mean of the two bell curves; and (2) illustrates the case when the stimulus value falls right at the mean of both distributions. The dark green curve indicates a wide distribution ($\sigma = 4$) and the dark cyan curve a narrow distribution ($\sigma = 2$). Both have the same mean ($\mu = 0$).]

Estimating the underlying distribution of phonetic values for a phonological representation is non-trivial. Nevertheless, measurements from a large corpus might offer a first approximation. Unfortunately, apart from showing the total amount of variation in the distribution of a certain phonetic parameter, a large corpus also captures subsets of such variation that occur in specific speech situations. Natural speech varies across situations such as the conditions of the speech environment (e.g., background noise, room reverberation), within-speaker factors (e.g., speaking rate, articulatory undershoot), between-speaker factors (e.g., dialect differences, speech habits), phonetic environments (e.g., coarticulation between segments, speech style), and word environments (e.g., coarticulation across word boundaries, word duration changed from syntactic requirements) (Pisoni, 1997). Considering variation in a single phonetic parameter across a large corpus with different kinds of specific speech situations will make the distribution very wide. However, that wide distribution may actually reflect the average across multiple smaller, narrower distributions due to different specific speech situations. Relatedly, there is evidence that listeners shift their perception of the relationship between phonetic parameters and phonological categories dynamically across situations (e.g., Hay & Drager, 2010; Johnson, Strand, & D'Imperio, 1999; Munson, 2011; Niedzielski, 1999; Strand, 1999; Strand & Johnson, 1996). Such shifting would be impossible if listeners used the single wide, lumped-together cross-situational distributions typically found in corpora. The above results instead suggest that listeners make use of the separate (and thus narrow) phonetic distributions that are embedded in the same phonological category in different speech situations. This intuition is formally implemented in Kleinschmidt and Jaeger (2015), in which the process of adapting categories to a new situation is proposed to involve new phonetic distributions for the same phonological category according to specific speech situations. For example, the total joint distribution of /s/ and /ʃ/ frication frequency means for all English speakers can be quite wide. However, the distribution becomes narrower when we consider all female speakers alone, and is smallest when only a particular female speaker is considered. Situation-specific distributions have been found to account well for the varied levels of adaptation effects found for speaker-specific (Bradlow & Bent, 2008; Clarke & Garrett, 2004), accent-specific (Bradlow & Bent, 2008), accent-general (Sidaras, Alexander, & Nygaard, 2009), and specific category manipulations (Eisner, Melinger, & Weber, 2013; Witteman, Weber, & McQueen, 2013).

Following Kleinschmidt and Jaeger (2015), we hypothesize that listeners can have and use multiple (narrow) distributions for a category that correspond to different social situations (e.g., accent types and age cohorts) and test this hypothesis in a speech perception experiment. Figure 2.2 is an illustration of the multiple-distribution hypothesis. For the purposes of illustration, suppose the purple dashed curve indicates the total variation of the second formant (F_2) of the GOOSE[13] vowel for speakers of all ages found in a large corpus of Australian-accented English; the individual orange, red, and blue continuous curves indicate such variation found for the three separate age groups: children, adults, and the elderly, represented in Figure 2.2 as Group 1, Group 2, and Group 3. The three groups may have the same extent of variation in their production of the GOOSE F_2. However, the mean varies according to groups. When an F_2 value falls in the distribution of Group 1, listeners are not certain that it should be recognized as GOOSE unless further information (e.g., it was produced by a member in Group 1) is provided. Otherwise, the value could well belong to a different phonological category produced by a member in Group 2, or yet another different category produced by a member in Group 3. That is, without disambiguating socio-indexical information for categories with multiple distributions, the phonological category membership is uncertain.[14] In contrast, category membership is more certain with categories that have fewer distributions (i.e., that do not vary much across socio-indexical dimensions). Therefore, degree of phonetic variability is predicted to have a positive relationship with speech perception (i.e., accuracy in identifying the phonological category), only when the mean of a phonological category is more or less the same across speakers, and thus the category has a comparatively smaller number of distributions. In other words, this type of category carries relatively little socio-indexical information, as it does little to differentiate among speakers. If the mean of a phonological category changes across speakers (i.e., the category carries relatively high levels of socio-indexical information that can differentiate among speakers), variability is predicted to have a negative relationship with speech perception.

[13] This refers to a lexical set as defined by Wells (1982) (i.e., in this case all words produced with the same vowel as in GOOSE in a given accent).
[14] The relationship between category membership and socio-indexical information has been demonstrated in a number of studies (e.g., Hay & Drager, 2010; Johnson et al., 1999; Munson, 2011; Niedzielski, 1999; Strand, 1999; Strand & Johnson, 1996).

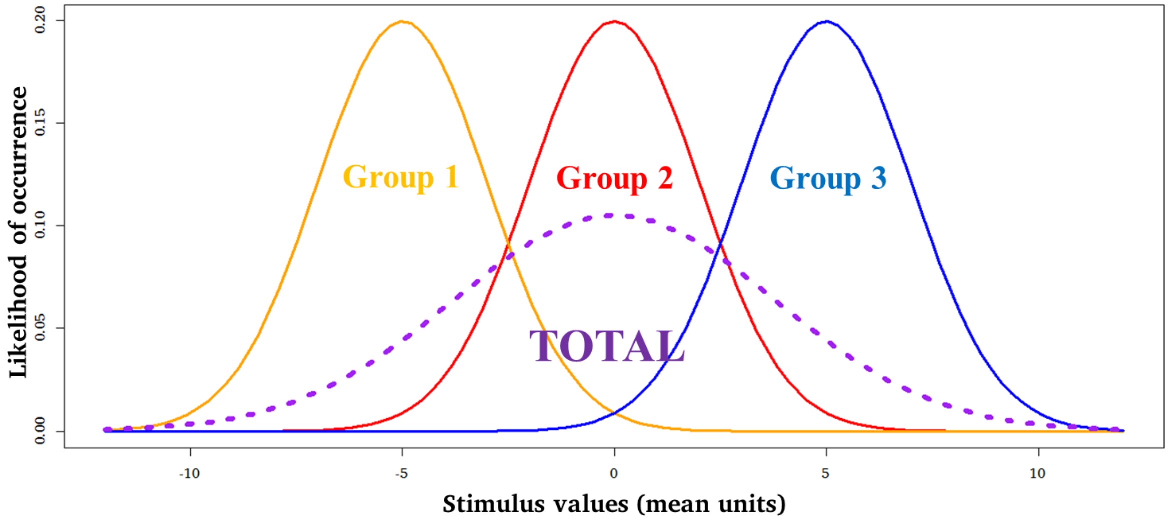


Figure 2.2: *Illustrations of the multiple-distribution hypothesis* [Purple curve indicates total variation of a category ($\mu = 0$, $\sigma = 3.8$); orange, red, and blue curves indicate the distribution within each group of a social variable that, together, makes up the total variation ($\mu = -5$, 0, and 5 accordingly; $\sigma = 2$).]

In this study, we tested the general prediction that vowel representations are distributions over formant values and the more specific prediction that vowels carrying more socio-indexical information may have a different relationship between formant variability and categorization accuracy from those carrying less socio-indexical information. We estimated formant variability from vowel distributions in a representative corpus of Australian-accented English (Cox, 2006), then explored the relationship between formant variability and listeners' performance in a vowel categorization task (similar to those used in Bundgaard-Nielsen et al., 2011; Faris et al., 2016). In discussing formant variability, for simplicity, we focus on Australian-accented English monophthongs, and the first two formant values (F_1 and F_2) only. Vowel categorization in Australian-accented English provides an appropriate empirical domain to test the predictions because some monophthongs in the relatively large Australian-accented English vowel inventory are reported in the literature to be more involved in both the classification of accent types and ongoing sound change within Australia. For these monophthongs, corpus-based estimates of phonetic parameter variability by themselves may inadequately reflect the distributions used by listeners in perception, as listeners may quickly adapt to a given speaker and apply a socially appropriate (and accordingly narrow) phonetic distribution.

1.2. Socio-indexical information in Australian-accented English monophthongs

Of all the sources of phonetic variation (e.g., social class/community/network, age/life stage, sex/gender, regional variation, ethnicity/race/bilingualism, and intra-speaker variation; Foulkes, Scobbie, & Watt, 2010), the social variables “social class,” “age,” and “sex” have received the most attention in sociophonetic studies of Australian-accented English (e.g., Bernard, 1967; Cox, 1998, 1999; Cox & Palethorpe, 2008; Harrington et al., 1997; Horvath, 1985; Mitchell & Delbridge, 1965). Socio-economic classes and sexes are often considered in studies of accent types (i.e., a “broadness” continuum[15] running from broad on one end, to general in the middle, and cultivated on the other end: Mitchell & Delbridge, 1965). Age cohorts could also be discussed in studies of accent types in a synchronic manner (e.g., Cox, 1998), but are typically discussed in studies of sound change in a diachronic manner (e.g., Cox, 1999; Cox & Palethorpe, 2008). Well-known studies of phonetic variation based on accent types include the pioneering auditory analyses of recordings from 7082 high school pupils from all over Australia (Mitchell and Delbridge, 1965), the follow-up acoustic analyses of vowel data from 171 New South Wales male speakers (Bernard, 1967), the auditory analyses of speech data from 177 Sydney speakers (Horvath, 1985), and the acoustic analyses of vowel data from 132 Sydney speakers taken from the Australian National Database of Spoken Language project (Harrington et al., 1997). Well-known studies of phonetic variation based on age cohorts include the comparison between vowel data in the 1960s and those in the 1990s (Cox, 1999), and the comparison among vowel data collected over several periods from 1885 to the 1990s (Cox & Palethorpe, 2008).

In the review below, we will use the findings from three representative studies by Mitchell and Delbridge (1965), Harrington et al. (1997), and Cox (1999) to determine a subset of monophthongs that are likely to carry more socio-indexical information than another subset that carries less socio-indexical information. The reason for the selection of these studies is that, with the reported data collected in the 1960s and the 1990s, these studies provide a diachronic picture of the phonetic variation in all 13 Australian-accented English monophthongs in relation to both accent types (the first two studies) and sound change (the third study). Given the multiple-distribution hypothesis we discussed earlier, the Australians who lived through both the 1960s

[15] The recent view of variations in Australian English is that the broadness continuum no longer applies since most younger-generation Australians now speak General Australian English. Australian English is now discussed in terms of dialect types instead, which involves Standard Australian English, Australian Aboriginal English, and ethnocultural varieties (e.g., Cox, 2006a, 2012).

and the 1990s have most likely formed multiple distributions for certain monophthongs (as well as fewer distributions for others) along the social categories of accent types and sound change. These social categories could be further abstracted into more general categories that indicate the amount of socio-indexical information carried by a vowel (i.e., more or less socio-indexical information), which facilitates the testing of our predictions.[16] We designate monophthongs as carrying more socio-indexical information when they *both* differentiate among accent types (at least one accent type contrasting with at least one other) *and* participate in sound change. Monophthongs that carry less socio-indexical information are those that differentiate among accent types only, participate in sound change only, or have no social effects.

First of all, regarding accent types, the findings from the auditory analyses in Mitchell and Delbridge (1965) and the acoustic analyses in Harrington et al. (1997) will be discussed. In Mitchell and Delbridge (1965), apart from FLEECE and GOOSE which were intended as the "chief diagnostic features" (p. 33) in the impressionistic investigation of differentiation among the broad, general, and cultivated accent types, a number of monophthongs were noted to have established different phonetic variants among the three accent types. These include the vowels in KIT, DRESS, TRAP, NORTH, NURSE, NEAR, STRUT, SQUARE, FLEECE, and GOOSE. Although the three accent types shared the variant [ɪ] for KIT and [ɛ] for DRESS, the general and broad accents had additional variants for KIT and DRESS which set them apart from the cultivated accent as well as further enabling the differentiation between them. For TRAP, NORTH, NURSE, and NEAR, the variants [æ], [ɔ], [ɜ], and [ɪː] were respectively shared among the three accent types. However, TRAP had two more variants that distinguished the broad accent from the rest, NORTH and NURSE each had one extra variant that signaled the cultivated accent, and NEAR had the additional variant [ɪə] for the cultivated and general accents and two other variants for the broad accent. The variant [ʌ] of STRUT was shared between the cultivated and general accents, but the general accent had two more variants that set it apart from the cultivated accent as well as the broad accent. Similarly, the variant [ɛ˔ː] of SQUARE was shared between the general and broad accents, but the general accent had one extra variant that distinguished it from the rest. Finally, FLEECE and GOOSE had totally different variants for each accent type. LOT, FOOT, and START were observed to

[16] Ideally only studies that included both male and female speakers in the datasets would be chosen because the phonetic variations in such datasets would better approximate the underlying distribution of phonetic values for a phonological representation. However, the only systematic study of diachronic sound change is Cox (1999) and it was based off data from male speakers only. This limitation is inevitable and will be noted for future research.

stay the same across the three accent types. In short, in the 1960s, out of 13 Australian-accented English monophthongs, only KIT, DRESS, TRAP, NORTH, NURSE, NEAR, STRUT, SQUARE, FLEECE, and GOOSE were noted to impressionistically change across the cultivated, general, and broad accents.

In Harrington et al. (1997), the cultivated, general, and broad accents were distinguished via statistical analyses of the differences between formant measurements. The formants were measured at the target of all monophthongs and the onset of FLEECE and GOOSE. The target was defined as the peak of F_2 for high front monophthongs, the trough of F_2 for high back monophthongs, the peak of F_1 for open monophthongs, the time point with maximum amplitude, or the temporal midpoint of a monophthong if all else failed. The onset was defined as the onset of voicing in the spectrogram and the onset of periodicity in the waveform. The monophthongs that showed significant accent effects were NEAR, NURSE, GOOSE, DRESS, SQUARE, and FLEECE. Apart from NEAR which showed only an overall accent effect (but unclear individual F_1/F_2 effects), these monophthongs all distinguished among the accent types along the F_2 dimension, and most of them showed the accent effects for one sex only. For example, broad NURSE F_2 target was significantly raised as opposed to cultivated and general NURSE F_2 target, and so was broad GOOSE F_2 onset; but the effects were reported for females only. Similarly, for females only, broad DRESS F_2 target was significantly raised compared with general DRESS F_2 target, and broad SQUARE F_2 target was significantly raised as opposed to cultivated SQUARE F_2 target. In contrast, for males only, broad FLEECE F_2 onset was significantly raised compared with cultivated and general FLEECE F_2 onset, and its target time was significantly delayed for the broad variety as well. The target of GOOSE was the only monophthong target that showed the accent effects for both sexes: broad GOOSE F_2 target was significantly raised as opposed to cultivated and general GOOSE F_2 target. The other monophthongs were observed to either potentially have some accent effects (but with statistical analyses not confirming the observations, such as in the cases of KIT and START), show accent effects for F_3 (which is not taken into account for the purpose of this paper, such as in the case of FOOT), or remain unchanged across accent types (NORTH, LOT, TRAP, and STRUT). In sum, in the 1990s, out of 13 Australian-accented English monophthongs, the formants for only NEAR, NURSE, GOOSE, DRESS, SQUARE, and FLEECE were shown to change across the cultivated, general, and broad accents.

Second, regarding sound change, the changes in the first two formants of Australian-accented English monophthongs from the 1960s to the 1990s were statistically documented in Cox (1999). The formants were measured at the target, onset, and offset of each monophthong. In the 1960s data, the target was defined as "the midpoint of the relatively parallel section of F_1 and F_2" (p. 6), F_2 high point for front monophthongs, and F_2 low point for high back monophthongs. In the 1990s data, the target was defined as "the point of least formant change" (p. 6): the peak of F_2 and the trough of F_1 for high front monophthongs, the trough of both F_1 and F_2 for high back monophthongs, the peak of F_1 for low monophthongs, or the point near the extreme point if all else failed. The monophthongs that showed significant change effects were DRESS, FOOT, FLEECE, NEAR, NURSE, KIT, LOT, TRAP, and GOOSE. DRESS and FOOT showed only an overall change effect (but unclear F_1/ F_2 effects). Reduced diphthongization was observed for FLEECE with the raising of F_1 at the onset. It was also observed for NEAR with the fronting of F_2 at the offset, along with the raising of F_1 and the fronting of F_2 at both the onset and target. NURSE had a fronted F_2 at the target. KIT had a raised F_1 at the target and a fronted F_2 at the offset. LOT had a raised F_1 at both the onset and target. TRAP had a lowered F_1 and a retracted F_2 at both the onset and target. GOOSE had a raised F_1 at both the onset and target, and a fronted F_2 at the onset, target, and offset. The other monophthongs were observed to either show changes for F_3 (which is not taken into account for the purpose of this paper, such as NORTH and SQUARE), or remain unchanged over the 1960s-1990s period (STRUT and START). Overall, out of 13 Australian-accented English monophthongs, only DRESS, FOOT, FLEECE, NEAR, NURSE, KIT, LOT, TRAP, and GOOSE were shown to statistically change from the 1960s to the 1990s.

As mentioned earlier, only monophthongs that *both* differentiate among accent types *and* participate in sound change would be classified into a vowel subset that is deemed to carry more socio-indexical information. The monophthongs that either differentiate among accent types or participate in sound change are listed again in Table 2.1 for easy reference.

Table 2.1: *Monophthongs that carry information about either accent types or sound change* [Monophthongs in **bold** carry **both** accent types **and** sound change. The monophthongs in the sound change list are re-ordered for easy comparison with those in the accent type list.]

Socio-indexical information	Australian-accented English monophthongs
Accent types (Harrington et al., 1997; Mitchell and Delbridge, 1965)	**KIT**, **DRESS**, **TRAP**, NORTH, **NURSE**, **NEAR**, STRUT, SQUARE, **FLEECE**, **GOOSE**
Sound change (Cox, 1999)	**KIT**, **DRESS**, **TRAP**, **NURSE**, **NEAR**, **FLEECE**, **GOOSE**, FOOT, LOT

Across accent types and sound change, we can identify the monophthongs that carry *both* social categories by picking out those that appear in *both* lists. They are the ones in bold in Table 2.1: KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE. These monophthongs satisfy our criteria to belong to a subset that carry more socio-indexical information, henceforth referred to as "more socio-indexical" vowels. The rest of the monophthongs SQUARE, STRUT, START, LOT, NORTH, and FOOT belong to another subset that carries little to no differentiating information about speaker accent type and/or age cohort, henceforth referred to as "less socio-indexical vowels". The means of the underlying distributions of the more socio-indexical vowels most likely reflect multiple internal distributions, for old speakers (speakers in the 1960s) versus young speakers (speakers in the 1990s), and for broad speakers versus general and/or cultivated speakers. Given listeners' tacit knowledge of the socio-indexical implications of these variations, as predicted earlier, they may not deploy the whole, global phonetic distribution they have experienced during the course of their lives (as reflected by corpus-based estimates) but rather would attend to a socially appropriate (and accordingly narrow) distribution instead. Therefore, we test the prediction that the relationship between formant variability and categorization accuracy is different when the monophthongs carry more socio-indexical information (i.e., multiple narrower distributions for different indexical properties of speakers: KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE) than when they carry less socio-indexical information (i.e., a single distribution that remains largely the same across different groups of speakers: SQUARE, STRUT, START, LOT, NORTH, and FOOT).

Apart from testing the above theoretical predictions, we also evaluated which method of quantifying formant variability is most effective, by evaluating the relationship between the variability yielded by each method and perceptual accuracy. Formant variability is usually estimated by how spread-out (quantified by SDs) the tokens of a vowel category are in F_1- F_2 vowel space. However, this two-dimensional spread of vowel tokens does not always serve as a reliable predictor of perceptual accuracy (Hillenbrand, Getty, Clark, & Wheeler, 1995). Another method of estimating formant variability is to take the magnitude of the formant mean into account (Eguchi & Hirsh, 1969; Kent, 1976; Lee, Potamianos, & Narayanan, 1999; N. Nguyen & Shaw, 2014). All else being equal, the degree of variability of a formant measurement is systematically related to the magnitude of the mean of that measurement. For example, vowels with high mean F_1 (or F_2) values are also more variable on F_1 (or F_2) than vowels with a low mean F_1 (or F_2) (N. Nguyen & Shaw, 2014). One way to quantify this formant variability that takes the influence of the mean measurement into account is to regress formant means on formant SDs, then consider the residuals: If a residual is positive, the vowel is variable in relation to the magnitude of the mean; if it is negative, the vowel is stable relative to its mean (N. Nguyen & Shaw, 2014). To our knowledge, the relationship between this residual method of estimating formant variability and perceptual accuracy, however, has never been evaluated. We hypothesized that the residual method indicates formant variability more effectively than standard deviation alone as it abstracts away the relationship between the variance of the formant measurements and the magnitude of their means. We reason that residuals estimate the variance of the underlying distribution in which the magnitude of the mean has already been abstracted, whereas SDs estimate the variance of the underlying distribution in which no abstraction over the magnitude of the mean has been taken into account. As a result, if the residuals are found to be better estimates of the variance in the underlying distribution than are standard deviations alone, we can infer that the underlying distribution is formed over a representation of the signal that is even more abstract than formant measurements.

In the remainder of the paper, we first present the methods we devised to test the two theoretical hypotheses (1) that vowel representations are distributions over formant values, and (2) that vowels carrying more socio-indexical information convey to listeners a different relationship between formant variability and category membership than vowels carrying less socio-indexical information. We also tested the methodological prediction that the residual method provides a better estimate of experienced variation.

2. Methods

2.1. Corpus analysis to obtain formant variability data

Measures of formant variability were obtained from the means and SDs of 13 Australian-accented English monophthongs from a large corpus of Australian-accented English (Cox, 2006). The formant values were taken from 60 teenage female speakers from Sydney's Northern Beaches producing vowels in the /hVd/ context (e.g., *hid, had, hud)*. Figure 2.3 shows the spread of the monophthongs in F_1- F_2 space expressed in Mel units.[17] In our models, and as noted earlier, we refer to KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE as the more socio-indexical vowel set and SQUARE, STRUT, START, LOT, NORTH, and FOOT as the less socio-indexical vowel set.[18]

---

[17] All the formant values used in this study are expressed in Mel scale.

[18] This vowel grouping is slightly different from the grouping presented at the first workshop on Sociophonetic Variability in the English varieties of Australia and LabPhon15, as we have taken into account the feedback received from these meetings.

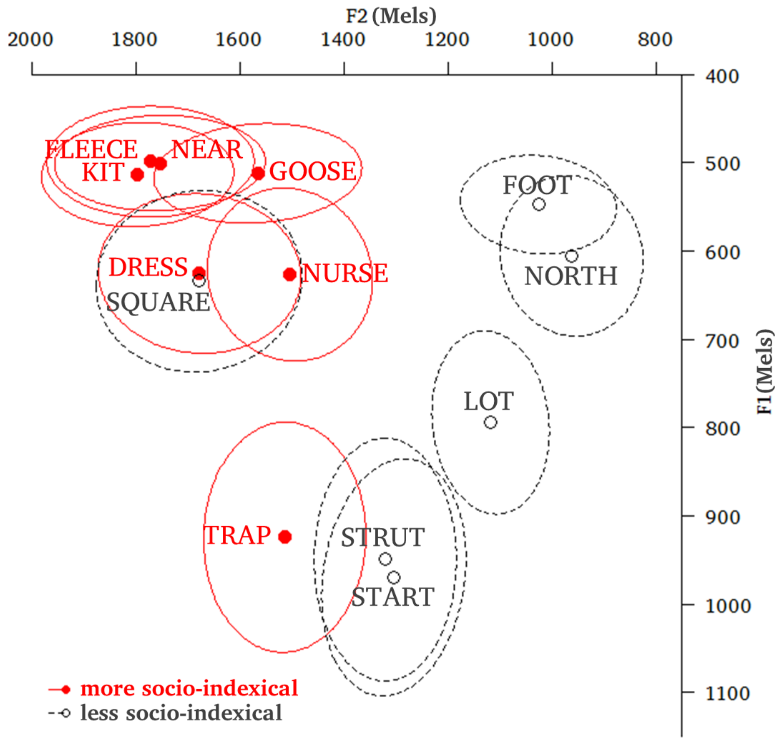


Figure 2.3: *Australian-accented English monophthongs based on Cox (2006)* [Ellipses represent 0.5 SD from the mean. Red indicates the vowels that typically carry socio-indexical information about *both* accent types *and* age cohorts in Australia.]

2.2. Vowel categorization study

Categorization data were pooled together from two experiments in a larger project designed to investigate the relationship between intergroup attitudes and vowel perception. This larger project involved two tasks: an online survey and a vowel categorization task. Australian-accented English-speaking monolingual participants listened to Australian-accented English in the training phase of the vowel categorization task and Vietnamese-accented English in the test phase. In the corresponding reports for those experiments, we reported listener perception of Vietnamese-accented English monophthongs in the test phase (N. Nguyen et al., 2016; N. Nguyen, Shaw, Tyler, Pinkus, & Best, 2015, under review). In this paper, we pooled data on perception of the Australian-accented English monophthongs from the training phase in the two experiments.

2.2.1. Participants

Data were from 64 volunteers from the Greater Sydney area, aged from 18 to 55 (*M* = 23 years, *SD* = 8.2). Participants were all born in Australia. Some were recruited from the Psychology undergraduate pool of Western Sydney University. Those participants were reimbursed by choosing either course credit or $20AUD/hr for their participation. Those who were recruited from other sources (e.g., the general public or other majors at the university where the course credit system did not apply) were reimbursed by $20AUD/hr.

2.2.2. Materials

The Australian-accented English stimuli were recorded from a female speaker who was born and raised in Western Sydney. She was in her 20s at the time of the recording, and spoke Australian-accented English only. Recordings were made with Adobe Audition software in a sound-attenuated booth at the MARCS Institute for Brain, Behaviour and Development, Western Sydney University. The speaker was recorded at 44.1 kHz sampling rate with a MOTU 896 mk3 sound card, a Shure SM10A-CN headset microphone, and an Impact core i7 tower computer. Target materials were 13 Australian-accented English monophthongs (i.e., /iː/, /ɪ/, /e/, /æ/, /ɐː/, /ɐ/, /ɔ/, /oː/, /ʊ/, /ʉː/, /ɜː/, /ɪə/, /eː/) embedded in the nonce word /ˈhVdə/ context. To ensure that the vowels in the nonce words were correctly produced, the speaker was requested to first read out a key word containing one of the vowels listed above (e.g. *book*), then the vowel itself (e.g., oo), a /hVd/ word containing the same vowel (e.g., hood), and lastly the /hVdə/ nonce word (e.g., hoodda). Each vowel was presented 10 times, each time randomized within a block of 13 to avoid list effects. For each vowel, four tokens (out of 10) which were subjectively judged to be most similar regarding speaking rate and loudness were selected as the stimuli for the vowel categorization task. Vowel /ɐ/ was an exception. Only two "hudda" tokens (out of 10) were chosen. Each "hudda" token was repeated twice so that all vowels would occur four times in the categorization task. The reason for the exception was that the other eight "hudda" /ˈhɐdə/ tokens were confused with "hadda" /ˈhædə/ in a pre-test. Sennheiser HD280 PRO Headphones were used to present all auditory stimuli to participants.

2.2.3. Reference word visual display

Participants were presented with a grid of 13 reference words (i.e., bad, bard, bead, beard, bed, bid, bird, book, bored, bud, rude, paired, and pod), which contain the monophthongs corresponding to 13 lexical set words TRAP, START, FLEECE, NEAR, DRESS, KIT, NURSE, FOOT, NORTH, STRUT, GOOSE, SQUARE, and

LOT. The relative positions of the words on the screen were randomized across participants by ePrime (version 2.0), and displayed on Acer TravelMate P645 notebook computers. The letters in light red represent the monophthong in each word, as can be seen in Figure 2.4.

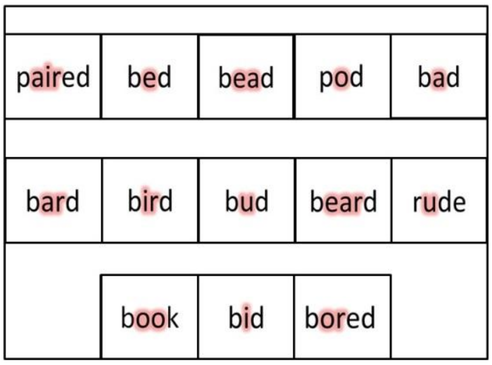


Figure 2.4: *One of the possible randomized positions of the reference words in the grid displayed to participants in the vowel categorization task*

2.2.4. Procedure

The session was run by one of three Australian-accented English associate researchers, who welcomed participants at the lab, then instructed participants to do an online survey first before moving them onto the vowel categorization task. The version of the online survey varied across participants, but both versions explored Australians' attitudes towards Asians and other ethnic groups in Australia, which was relevant to the main focus of the larger study (N. Nguyen et al., 2015; N. Nguyen et al., 2016). Because the survey most likely did not affect participants' performance in categorizing nonce words in their native accent during the training phase, which is the focus of this paper, it will not be mentioned further. Participants were informed of the speaker's Australian accent in the stimuli.

The vowel categorization task included three components: a five-trial practice, a training phase, and a test phase. A block of 52 trials (52 trials = 1 token/trial x 4 tokens/vowel x 13 vowels) were programmed for each phase and the tokens were randomized within these 52 trials. Participants completed one such block for the test phase, but they could go through up to four blocks in the training phase. After the first block of training, if they categorized three out of four tokens per vowel correctly and 10 out of 13 vowels correctly, ePrime moved them onto the test phase. Otherwise they had to complete the second 52-trial training block, the third block, or the fourth block before they could proceed to the test phase. As there may be some adaptation effect in the later blocks, for the purposes of this paper, only the first cycle of training was analyzed.

On each trial, participants listened to one token of the /hVdə/ nonce word and categorized the first vowel in that nonce word by selecting a reference word on the grid that contained the same vowel. If their categorization was correct, ePrime played the same /hVdə/ token once again and asked them to rate the goodness of fit between the vowel they heard and the vowel in the reference word they selected.[19] Low ratings indicated foreign sounding vowels while high ratings indicated native-like vowels. For our current purposes, analyses were not conducted on rating data. If participants' categorization was incorrect in the training phase, they were presented with a message on the screen advising them about their incorrect response and indicating the correct one to them.

The duration of the vowel categorization task varied greatly (i.e., from 20 min to an hour), depending on how quickly participants proceeded through the training phase. The purposes of the categorization task were debriefed to participants at the conclusion of the lab session.

3. Analyses and Results

3.1. Analyses

Mixed effects modelling was selected as an analytic approach for this study as, in the analyses of the main effects of interest (i.e., formant variability and vowel socio-indexicality), it takes into account the other factors that we did not control for such as different baseline performance by individual participants and differences among vowel tokens that potentially affect categorization performance. With the use of the lme4 package (Bates, Maechler, Bolker, & Walker, 2014), binomial mixed effects models were fitted to the accuracy data[20] (3,328 data points) in R and model comparisons were carried out. Participant and vowel token were included as random effects. Fixed effects were formant variability and the socio-indexicality of the monophthongs, which was a binary variable dividing the more socio-indexical monophthongs (KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE) from the less socio-indexical ones (SQUARE, STRUT, START, LOT,

[19] Participants could replay the stimulus in the rating phase.
[20] Our binomial mixed effects models could also be fitted to response time or goodness rating. However, because we told the participants to go with their own pace, rather than encouraging them to respond quickly as studies that collect response time as a dependent variable (DV) do, the response time data collected in our task would not be informative as a DV. Goodness rating is not good as a DV in our models either because (1) rating is subjective and (2) rating is evaluative rather than perceptual.

NORTH, and FOOT). Below are the main models:

- Data = 13 monophthongs:

  (1a) Accuracy ~ formant variability + vowel socio-indexicality + (1+formant variability|Participant) + (1+formant variability|Token)

  (1b) Accuracy ~ formant variability * vowel socio-indexicality + (1+formant variability|Participant) + (1+formant variability|Token)

- Data = less socio-indexical subset:

  (2) Accuracy ~ formant variability + (1+formant variability|Participant) + (1+formant variability|Token)

- Data = more socio-indexical subset:

  (3) Accuracy ~ formant variability + (1+formant variability|Participant) + (1+formant variability|Token)

As argued in Section 1, we predicted an interaction between formant variability and vowel socio-indexicality in which formant variability has a positive correlation with the accuracy of the less socio-indexical vowels (as long as the stimulus values fall far from the means of the distributions) and a negative correlation with the accuracy of the more socio-indexical ones. We first explored the interaction between formant variability and vowel socio-indexicality visually with a scatterplot. After that, we compared models 1a and 1b to evaluate the significance of the interaction, and then checked the significance of formant variability as a single predictor in the models with two different subsets: the less socio-indexical subset and the more socio-indexical subset.

The binomial mixed effects models above were run with two different indices of variability: the traditional SD method and the residual method which factors out the influence of the magnitude of the formant means. To compute variability according to the residual method, we fitted regression lines to the means and SDs of F_1 and F_2 values to calculate F_1 and F_2 residuals for each vowel.[21] For the SD method, we rescaled F_1 and F_2 SDs for each vowel (centering the mean on 0) to make them directly comparable to the residuals. To make a single composite measure of formant variability we summed the variability of F_1 and F_2.[22] Following Baayen

[21] This calculation of the residual method does not take into account the co-variance between F_1 and F_2. As we do not have access to individual F_1 and F_2 values but their means and SDs instead (reported in Cox, 2006), an F_1 and F_2 co-variance matrix could not be established for this paper. We would still like to thank Dave Kleinschmidt and Jonathan Harrington for this F_1-F_2 co-variance suggestion, and will consider applying it in future studies.
[22] We used the sum of F_1 and F_2 as a single index of variability because we did not have different hypotheses for F_1 variability and F_2

and Milin (2010, pp. 15-18), we also applied model criticism to identify outliers to the model fit, and no outliers influencing the modelling results were found.

3.2. Results

Figure 2.5 shows an interaction between formant variability (indexed by sums of residuals) and categorization accuracy via vowel socio-indexicality.[23] More specifically, for the less socio-indexical monophthongs there appears to be a positive relationship between formant variability and categorization accuracy, whereas for the more socio-indexical ones, the relationship between formant variability and categorization accuracy appears negative.

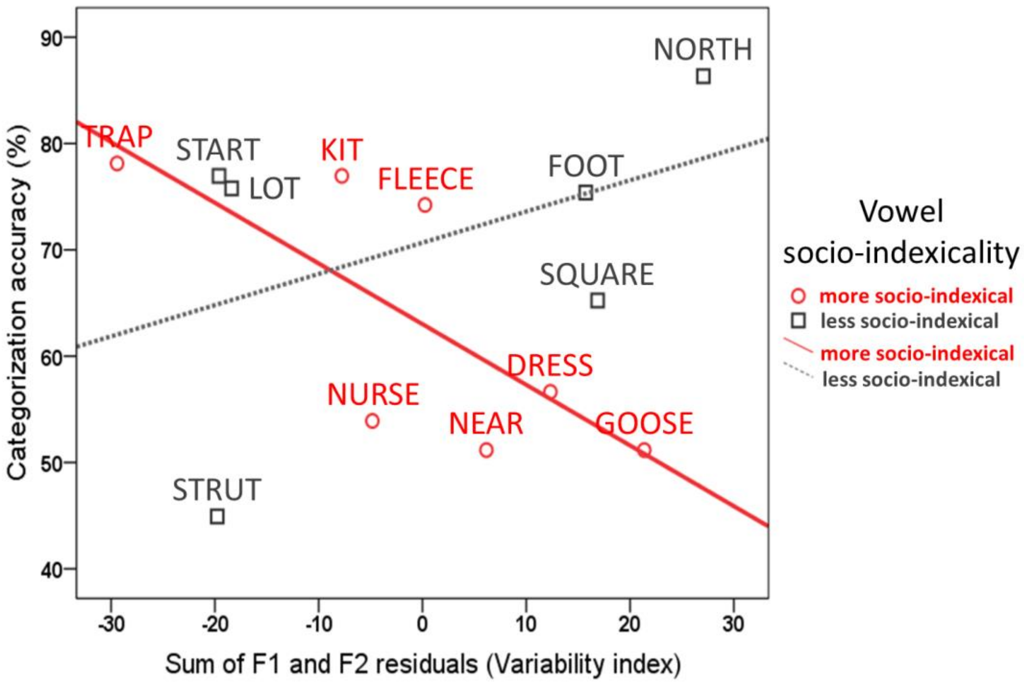


Figure 2.5: *Relationship between formant variability and categorization accuracy, depending on the socio-indexicality of the monophthongs* [Red indicates the monophthongs that carry more socio-indexical information (i.e., carrying information about *both* accent types *and* age cohorts).]

variability. In addition, the correlation between means and SDs that is applicable to F_1 and F_2 individually also applies to the sum of F_1 and F_2. We also tested using F_1 variability and F_2 variability individually to represent formant variability in the interaction model (i.e., model 1b), and compared that with the model having the variability from the sum of F_1 and F_2: Where there is an interaction between formant variability and vowel socio-indexicality, the model with the variability from the sum of F_1 and F_2 explains the most variance (i.e., it has the lowest AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), highest log likelihood, and smallest deviance).

[23] No interaction is observed between formant variability (indexed by sums of SDs) and categorization accuracy.

The interaction between formant variability (indexed by sums of residuals) and vowel socio-indexicality is significant (χ^2 = 13.27, p < .001) by means of model comparison (Table 2.2). Results from model 3 confirm that the negative correlation between formant variability and categorization accuracy for the more socio-indexical subset is significant (β= -0.04, p < .001). However, the positive correlation between formant variability and categorization accuracy for the less socio-indexical subset in model 2 is *not* significant (β = 0.01, p = .45), despite the apparent positive trend seen in Figure 2.5.

Table 2.2: *Variances explained by interaction model versus non-interaction model* [Formant variability is indexed by sum of residuals.]

Model	AIC	BIC	logLik	deviance	Pr(>χ^2)
(1a) with*out* interaction: Accuracy ~ formant variability + vowel socio-indexicality + (1+formant variability|Participant) + (1+formant variability|Token)	3692.1	3747.1	-1837.0	3674.1	
(1b) with interaction: Accuracy ~ formant variability * vowel socio-indexicality + (1+formant variability|Participant) + (1+formant variability|Token)	3680.8	3741.9	-1830.4	3660.8	<.001

Model comparison in Table 2.3 also confirms that the model with formant variability as indexed by the residual method explains more variance than the model with the variability indexed by the SD method (χ^2 = 6.71, p < .001).

Table 2.3: *Variances explained by sums of residuals versus sums of SDs*

Model	AIC	BIC	logLik	deviance	Pr(>χ^2)
(1b) with sums of SDs: Accuracy ~ (scaled) sum of SDs * vowel socio-indexicality + (1+(scaled) sum of SDs|Participant) + (1+(scaled) sum of SDs|Token)	3687.5	3748.6	-1833.8	3667.5	
(1b) with sums of residuals: Accuracy ~ sum of residuals * vowel socio-indexicality + (1+ sum of residuals|Participant) + (1+ sum of residuals|Token)	3680.8	3741.9	-1830.4	3660.8	<.001

4. Discussion

In this paper, we examined listeners' tolerance of variation for monophthongal categories over a given phonetic parameter. We predicted that this tolerance of variation would depend on the socio-indexicality of the monophthongs: more tolerance for monophthongs that are not rich in socio-indexical information and less tolerance for monophthongs that are. Our findings supported our predictions. We also found evidence for our proposal that quantifying formant variability by using the residuals of the regression between formant means and formant SDs provides a better estimate of experienced variation than quantifying formant variability by using SDs only.

At first glance, the absence of a positive relationship between formant variability and categorization accuracy for the low socio-indexicality monophthongs SQUARE, STRUT, START, LOT, NORTH, and FOOT does not seem to support our prediction. However, recall that the positive-relationship prediction was made on the assumption that the stimulus values fall far from the means of the distributions. As the prediction goes in Section 1, when the stimulus values fall near the means of the distributions, the stimulus has a better chance of being included in the category with a narrow distribution than it does with a wide distribution. In this case, formant variability would be negatively correlated with categorization accuracy. This is indeed the case for some stimulus values of SQUARE, LOT, and FOOT, most stimulus values of STRUT and START, and all stimulus values of NORTH (shown in Figure 2.6). In other words, our prediction is still supported when the stimulus values do fall close to the means of the distributions.

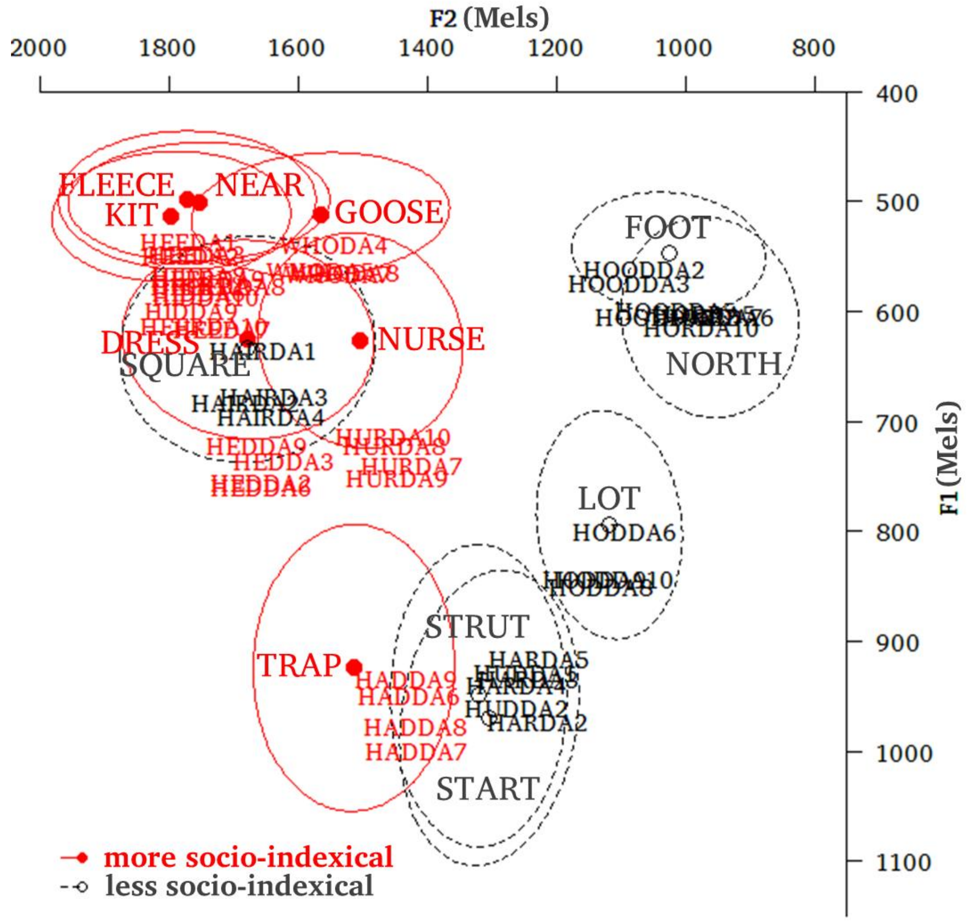


Figure 2.6: *Where stimulus values fall relative to the means of the distributions of Australian-accented English monophthongs based on Cox (2006)* [Ellipses represent half SD from the mean. Red indicates the monophthongs that carry more socio-indexical information.]

The significant negative relationship between formant variability and categorization accuracy for the more socio-indexical monophthongs KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE suggests that listeners used more specific and socially relevant distributions for these monophthongs in the perception task. Since these monophthongs carry more socio-indexical information, perhaps listeners quickly attributed certain social attributes to the speaker right after hearing the first few stimuli, and then accessed or inferred certain (narrower) distributions which they believed would be appropriate for the speech situation. These narrower distributions may or may not have served them well in the task, but they would likely disrupt the predicted positive relationship between formant variability and categorization accuracy posited for the more global, wide distributions.

The significant interaction between formant variability and vowel socio-indexicality is potentially informative in our attempts to understand the dynamics of speech perception. However, the grouping of monophthongs into socio-indexicality categories that we followed in this study is just one way of grouping 13 Australian-accented English monophthongs into two groups. It is therefore important to establish the robustness of our results by examining how likely it is for a significant interaction between formant variability and vowel groups to emerge across different ways of grouping. In the particular case of Australian-accented English monophthongs, there are 1716 possible ways to divide 13 Australian-accented English monophthongs into two groups. If vowel socio-indexicality is the only way of grouping, or one of the few groupings, which produces vowel groups which then significantly interact with formant variability, then socio-indexicality is a robust dynamic in the relationship between formant variability and speech perception. We tested this robustness prediction by re-running the model (1b) with formant variability represented by sum of residuals and each of 1716 possible groupings, resulting in 1716 corresponding models.[24] We found that a significant interaction between formant variability and vowel groups is rather likely regardless of different ways of grouping: It was found in 43.88% models (753 models out of 1716). The abundance of significant interactions between formant variability and vowel groups across diverse ways of grouping means that vowel socio-indexicality is not the only or even the primary factor in our understanding of the relationship between formant variability and speech perception. However, it does provide us some ideas for future experimentation.

Our socio-indexicality division could be questioned on the basis that the monophthongs in the more socio-indexical group (KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE) overlap with one another more than do the less socio-indexical ones (SQUARE, STRUT, START, LOT, NORTH, and FOOT). As a result, what matters in the relationship between formant variability and vowel perception could possibly be the overlap of the monophthongs rather than the socio-indexicality. Overlapping categories cause confusion and,

[24] We would like to particularly thank John Kingston for this suggestion.

therefore, hurt speech perception (Clayards et al., 2008). This could explain the negative correlation between the formant variability in KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE and their perception. However, if we were to follow out the logic of this overlap factor, then FLEECE, NEAR, KIT, DRESS, SQUARE, STRUT, and START should be grouped together (instead of KIT, DRESS, TRAP, NURSE, NEAR, FLEECE, and GOOSE). The reason is that these vowels almost overlap with one another entirely: FLEECE, NEAR, and KIT as one instance of entire overlap, DRESS and SQUARE as another, and STRUT and START as the rest. GOOSE and NURSE do not overlap that much with each other, nor do they overlap with FLEECE, NEAR, KIT, and DRESS; and TRAP is completely separate from the rest. The grouping of FLEECE, NEAR, KIT, DRESS, SQUARE, STRUT, and START is one of the 1716 groupings that were tested, but it did not result in a significant interaction with formant variability (p = .320).

Our results also show that the residuals are better estimates of the variance in the underlying distribution than the SDs. The significant interaction between formant variability and vowel socio-indexicality, which potentially contributes to our understanding of speech perception dynamics, could only be found when the sums of F_1 and F_2 residuals were used to index formant variability. The SD method did not reveal these patterns. This suggests that the residual method, compared to the SD method, provides a better differentiation between phonological variability as observed in corpora and as socially indexed in mental representations. However, as there is a large number of significant interactions with vowel groups when formant variability is indexed by sum of residuals (43.88% models or 753 models out of 1716), it is possible that there might just be something inherent in the residual method itself that makes its quantified values more susceptible to interactions with vowel groups. If this is true, then there might also be something inherent in the SD method itself that makes its quantified values less susceptible to interactions with vowel groups. In that case, re-running model (1b) with formant variability represented by sum of SDs and each of 1716 groupings would result in a much smaller number of significant interactions, compared with when formant variability is indexed by sum of residuals. We also tested out this prediction, and found that a significant interaction between vowel groups and formant variability calculated from SDs is similarly likely: It was found in 37.59% models, or 645 models out of 1716. Because the number of significant interactions for the SD method is roughly similar to that for the residual method, and because meaningful interaction patterns (such as with the socio-indexicality grouping) only emerge with the residual method, we can confidently suggest that the residual method provides a better estimate of the variance in the underlying distribution.

Although the residual method is suggested to be informative in relation to variance in the distribution, its proposal does not include elimination or preservation of variation in speech in general and phonemic variation in particular. Variation in speech was initially considered problematic and therefore needed to be discarded via a normalization process, whether the process involves taking into account the acoustic properties within a vowel (intrinsic normalization) (e.g., Miller, 1953; Syrdal & Gopal, 1986) or between the vowel of interest and the other vowels produced by the same speaker (extrinsic normalization) (e.g., Gerstman, 1968; Lobanov, 1971). The role of variation was later recognized with the rise of episodic approaches and therefore normalization processes became unnecessary (e.g., Johnson, 1997). Recently, language variation studies have used normalization processes to control for one source of variation in order to examine some other sources of variation (e.g., Adank, Smits, & van Hout, 2004; Most, Amir, & Tobin, 2000; Watson, Maclagan, & Harrington, 2000). As a result, vowel and/or speaker normalization has once again become a topic of interest in speech perception. In the context of this long-running and highly debated topic, the proposed residual method is not directly classified as either intrinsic or extrinsic normalization because it relies on regression lines that fit to the means and standard deviations of formant values from the vowels in the vowel system of a language/dialect (rather than of a single speaker). In addition, the method only seeks to propose an unbiased view of formant variability (by taking away the influence of formant means), not to take actions on that variability. However, the residual method could possibly be considered vowel-extrinsic since it involves several vowels, and currently takes a neutral stance regarding formant-intrinsic (i.e., involving only within-formant information) and formant-extrinsic (i.e., involving across-formant information).

Our findings also bear on the question of abstraction, specifically whether socio-indexical information could be stored in long-term memory as high-level abstract units in the same way as phonetic information is. According to Pierrehumbert (2006), it is possible for social details to be abstracted into social categories, similar to phonetic details being abstracted into phonological categories. However, while it is widely accepted that phonological categories could be abstracted further into phonological classes, no socio-phonetic theories have committed to the possibility of further abstraction over social categories into, for example, socio-indexicality categories. Numerous studies have demonstrated the different relationships between the high-level abstract phonological classes and speech perception (see e.g., Feldman et al., 2009). If socio-indexicality categories do exist as high-level abstract units of socio-indexical information, then different relationships between them and speech perception could be expected. This study has provided some evidence that this could be the case. Our listeners seemed to have stored two highly abstract socio-indexical categories of vowel representations: less socio-indexical monophthongs with means more or less the same across accent types and age cohorts, and more socio-indexical ones with means changing across these situations. These two highly abstract socio-indexical categories modulated the relationship between formant variability and vowel categorization differently, which suggests the existence of higher levels of abstractions over socio-indexical information.

A possible account for the categorization process of Australian-accented English monophthongs by Australian-accented English listeners, while considering the highly abstract socio-indexicality categories, could be provided based on a person construal diagram (Freeman & Ambady, 2011) and an exemplar-resonance schema (Johnson, 2006). Figure 2.7 shows the structure of categories at a linguistic and social level. The bottom of the diagram deals with concrete materials (i.e., those that can be perceived by our senses) and the top deals with abstract concepts (i.e., brain processes that cannot be seen or heard). The stereotype level in Freeman and Ambady (2011) is collapsed into the higher-order level in this thesis, which involves prejudice among others, because this thesis does not assume that stereotype is at a different level of processing than prejudice. The higher-order level in Figure 2.7, apart from a node called "task demand," also contains another node called "prior knowledge," which links the social side of the auditory input to its linguistic side. In addition, the category level in this thesis involves several sub-levels to illustrate the idea of further abstraction to social categories as well as demonstrate where the highly abstract socio-indexicality categories stand in a sound categorization process. The diagram takes external input from the world (both bottom-up and top-down sources such as sounds needed to be perceived and task demand, respectively) and internal input from within the listener (top-down sources such as expectations about what to hear). Therefore, it allows for both activation from the bottom-up sources and feedback from the top-down sources. It also allows for excitation and inhibition from neighbouring nodes. The exemplars in Figure 2.7 are treated as individual sounds. This is a simplified treatment in speech perception to illustrate the categorization process and does not conflict with the treatment of exemplars as whole words depicted in Johnson's (2006) schema.

Take the categorization of the GOOSE vowel /ʉː/ in "who'da" produced by our Australian-accented English speaker in 2014 as an example. When the network receives an auditory input, or when the Australian-accented English listeners in this thesis heard "who'da," a number of cues can be found. Those cues could indicate social attributes of the speaker of the stimulus as well as prompt listeners of the linguistic identity /ʉː/ of the stimulus (for simplification, the personal information conveyed by the stimulus is not mentioned in this diagram). Socially, cues could reveal which specific accent type the speaker of the stimulus has (i.e., Broad, General, or Cultivated) or which specific age cohort the speaker belongs to (i.e., whether the speaker sounds like someone from the 1960s or the 1990s). Linguistically, cues help listeners identify the vowel /ʉː/ in the input. The task from the top-down source (i.e., the experimenter or prompts on the experimental computer) demands listeners to categorize vowels. As a result, the task demand node excites the vowel node as well as the individual vowel categories (e.g., /ʉː/, /iː/, and /ɐː/), and inhibits the social side of the input. When both top-down and bottom-up sources are considered, the individual vowel categories receive a lot of excitation. When the auditory input of "who'da" comes into the network, the network would encode it as a spectrogram, then try to match the portion of the spectrogram corresponding to the vowel in "who" with a number of exemplars stored in memory, also in the form of spectrograms. Exemplars close enough to the GOOSE vowel would all be activated, while those different than the speech input would stay inactive. Depending on how much the close-enough exemplars are weighted in relation to a specific vowel category, for example the GOOSE vowel as opposed to the FLEECE vowel or the START vowel, that category is also activated from the sum of all those activated exemplars. Only the social side of the input is activated by the bottom-up source alone. If the input contains cues to a specific accent, when it unfolds and enough evidence is gathered, it would activate the node to that particular accent strongly and the rest only weakly. It would also activate strongly the node to a specific time range when those speech features in the input were often heard. These activations, in turn, would excite the corresponding, more abstract category nodes called "accent type" and "sound change." These nodes do not inhibit each other, but in fact interact with each other to excite the highly abstract category node called "socio-indexicality" and also get feedback from this node. This socio-indexicality node on the social side of the input again interacts with the individual vowel category nodes on the linguistic side of the input to modify a node at the higher-order level called "prior knowledge." These nodes also receive feedback from the prior knowledge node. If the Australian listeners in the thesis were those who lived through both the 1960s and the 1990s, their prior knowledge node went through both the modification and feedback processes with the socio-indexicality and vowel category nodes. However, the listeners in the thesis most likely did not live through the 1960s, and thus their prior knowledge about the connection between socio-indexicality and vowel categories most likely came from indirect sources. Consequently, their prior knowledge node mostly sends feedback to the socio-indexicality and vowel category nodes to facilitate the categorization of our Australian-accented English speaker's vowels produced in 2014 rather than receiving information from these two nodes.

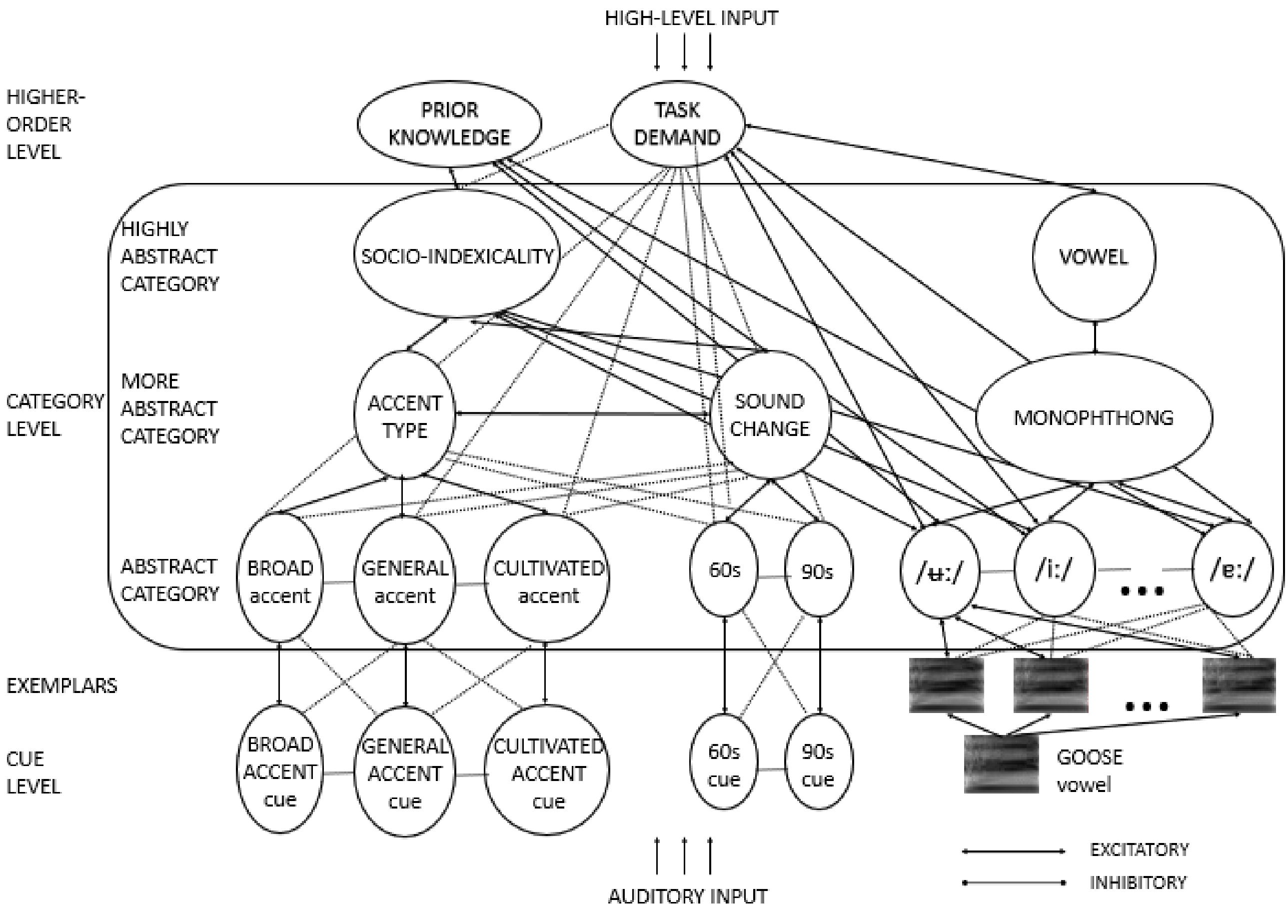


Figure 2.7: *A diagram of the structure of categories at a linguistic and social level based on Freeman and Ambady (2011) and Johnson (2006), illustrating the categorization process of Australian-accented English monophthongs by Australian-accented English listeners* [Note that the social categories in the diagram are not limited to the linguistic system, and that node sizes are only cosmetic and have no theoretical implications.]

In terms of the abstraction-versus-specificity debate in speech perception, our findings lend support to hybrid accounts of speech perception (e.g., McQueen et al., 2006; Pierrehumbert, 2016), which allow for both abstract and episodic processing. Above we suggested the existence of high-level abstract socio-indexical units, which listeners have formed by having been directly exposed to different groups of speakers (e.g., speakers with broad accent versus cultivated and general accents, old speakers versus young ones) during the course of their lives. This could be considered abstract and episodic processing of socio-indexical information. In addition, we have just argued earlier in this section in favour of the residual method, which suggests some abstraction in the underlying distributions away from the phonetic values that listeners experience in everyday speech scenario. This is evidence for abstract and episodic processing of phonetic information.

Future studies could implement a number of changes to improve the reliability of the results. First, while there were only three participants born in the 1960s in our cohort (most of our participants were born in the 1990s), our participant cohort should have involved those who had lived through both the 1960s and the 1990s. Our grouping of the monophthongs into socio-indexicality categories was based on the assumption that Australians abstracting the social categories of accent types and age cohorts into socio-indexicality categories by virtue of having lived through both the 1960s and the 1990s. As a result, having such a cohort of listeners would provide stronger and more direct evidence of the modulation of abstract socio-indexical information in the relationship between formant variability and vowel categorization. However, such a participant cohort was not readily available to us. Second, the socio-indexicality selection should be based on findings from studies with data collected in the 1990s and recent times, rather than the 1960s and the 1990s. This takes into account our actual participant cohort with most being born in the 1990s as well as our earlier assumption of high-level abstraction of socio-indexical information through experience. However, socio-phonetic data collected in recent times such as AusTalk (Estival, Cassidy, Cox, & Burnham, 2014) have not been analyzed and reported. Once that is done, not only can we select socio-indexicality vowels that are better represented in the minds of our more available participant pool, but we can also have updated estimates of the variability in the underlying distributions of vowel categories. Third, the manipulation of vowel stimuli would yield more informative findings. The manipulation of vowels in terms of their socio-indexicality[25] would be necessary for more definite conclusions on its modulation effects on the relationship between formant variability and vowel

[25] We would like to particularly thank Dave Kleinschmidt and Eleanor Chodroff for this suggestion.

categorization. In addition, the manipulation of stimulus values (i.e., close to or far from the mean of phonetic distributions) could provide direct evidence for our prediction of listeners' tolerance of variation for phonological categories that have wide phonetic distributions.

5. Conclusion

Our results support the views that phonological representations can be represented by Gaussian distributions over phonetic values and also that the distributions probably involve some level of socio-indexical abstraction, possibly along the lines proposed by Kleinschmidt and Jaeger (2015). Revealing these associations requires taking into account the relationship between mean formant values and their associated SDs. This suggests a greater degree of abstraction in the representation of socio-indexical information than is typically assumed in exemplar-based models of episodic memory for experienced speech events.

Chapter 3

Expectations about a foreign accent influence speech perception even without previous exposure

A version of this chapter has been submitted for publication as:

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (under review). Expectations about a foreign accent influence speech perception even without previous exposure. *Language and Speech*.

Abstract: When listeners have had exposure to a particular English accent and are told to expect that accent, exemplar theories predict a shift in perception in the direction of experience. Here we show that expectations about a speaker's accent induce perceptual shifts even when listeners do not have previous exposure to that accent. We manipulated listener expectations through the absence or presence of explicit instruction that the speaker in a vowel categorization task had a Vietnamese accent. Listener expectations about the accented vowels were modelled as priors in a Bayesian framework. Results indicated that Bayesian priors over vowels at the extreme corners of the vowel space decreased while priors over more central vowels increased (i.e., the perceptual space for vowels shrank) when listeners were told that the speaker had a Vietnamese accent. Despite having no previous exposure to Vietnamese-accented English, cueing speaker background shifted perception towards listener beliefs about the accent.

Keywords: exemplar theory, Bayesian priors, listener expectations, foreign accent, vowel categorization

1 Introduction

Listener expectations have been shown to shift vowel perception according to information about the speaker's regional origin, even when that information is incorrect (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). For example, listeners from Detroit, Michigan, which shares a border with Canada, were asked to match vowels heard in sentences to vowels played in isolation (Niedzielski, 1999). The vowels in the sentences were produced by a Detroit speaker, but they had the "Canadian Raising" feature that Detroiters often attribute to Canadian-accented English. The Detroit listeners who were told to expect a Canadian speaker matched the vowels in the sentences to vowels that exhibited Canadian Raising, whereas those who were told to expect a Detroit speaker selected standard American-accented English vowels instead.

Exemplar theories can straightforwardly explain how expectations may influence speech perception (Hay & Drager, 2010; Hay et al., 2006a). In exemplar theories, phonetic details of words are represented in the lexicon along with detailed information about the speech context as an episodic memory, or exemplar, including the speaker's information (e.g., Goldinger, 1998; Pierrehumbert, 2006). The Detroit listeners in Niedzielski's (1999) study were most likely familiar with Canadian-accented English because of their close proximity to Canada. Therefore, the cues "Detroit" (or "Michigan") and "Canadian" carried socio-indexical information associated with Detroit listeners' phonetic representations of Detroit and Canadian speech. When those listeners were told to expect either a Detroit or Canadian speaker, exemplars of Detroit or Canadian speech were activated. The selective activation of exemplars associated with specific regional accents can explain the observed shifts in vowel perception across conditions.

The Niedzielski (1999) result has since been replicated with other speaker/listener groups and with subtler manipulations of listener expectations (Hay & Drager, 2010; Hay et al., 2006a). These studies all involved listeners likely to have had previous exposure to the accent of the stimuli. In the *absence* of such exposure, exemplar theories would not predict a shift in perception, as there are no relevant exemplars to be activated. Alternatively, a shift remains a possibility in Bayesian approaches (described below). To evaluate these alternatives, we investigated the role of expectations in vowel perception in the *absence* of previous exposure. We asked Australian-accented English listeners in two conditions to make decisions about the same speech stimuli while varying the information that listeners received about the speaker. In the treatment condition, participants were told to expect an Australian-accented English speaker in the training phase and a Vietnamese-accented English speaker in the test phase. The control participants were told to expect two different speakers in the two phases and given no additional information. We chose the Vietnamese accent for the test phase because it is an accent to which many Australian listeners would have little to no previous exposure.

We hypothesized that beliefs about the Vietnamese accent would influence behaviour in the treatment participants, relative to control. A difference in perception could be induced by expectations arising from beliefs about the speaker's accent. For cases in which beliefs are developed from direct experiences, like beliefs about the Vietnamese accent based on hearing Vietnamese-accented English speakers, exemplar theories predict similar behaviour as Bayesian formalizations of beliefs (Shi et al., 2010). Bayesian models, however, can also formalize the influence of prior beliefs that do not come from direct experiences. These beliefs can be true or false and can come from a wide range of sources (e.g., hearsay and stereotypes).

To quantify the different expectations for speech categories as a function of our experimental manipulation, we used Bayesian prior probability over vowel categories. Bayesian models have been applied in a range of psycholinguistic studies, including spoken word recognition and phoneme categorization, the task employed in this paper (e.g., Clayards et al., 2008; Kirov & Wilson, 2013; Kleinschmidt & Jaeger, 2015; K. Nielsen & Wilson, 2008; Norris & McQueen, 2008; Norris et al., 2016; Wilson & Davidson, 2013). Bayesian priors are highly appropriate for quantifying implicit beliefs about categories, regardless of the source of those beliefs. In Kleinschmidt and Jaeger's (2015) and Norris and McQueen's (2008) models, priors are estimated based on real-world observations such as frequency counts, linguistic contexts, visual contexts, or cues to co-articulation, which are sometimes limited in number (Kleinschmidt & Jaeger, 2015). In principle, however, the Bayesian formalism is not committed to beliefs based purely on observations, and this is a point of departure from exemplar-based categorization models. In this study, we deliberately minimized relevant observations by selecting participants with no previous exposure to the Vietnamese accent. This allows us to test whether beliefs about the Vietnamese accent arising in the *absence* of direct exposure to the accent influence expectations about speech.

2 Method

2.1 Participants

Sixty psychology undergraduates from Western Sydney University participated in the vowel categorization task for course credit. Data from two participants were excluded due to not finishing the task. Of the remaining 58 participants (age range = 18-45 years), 43 evaluated themselves as having no previous exposure to the Vietnamese accent[26] (although they claimed to have had exposure to other Asian-accented Englishes such as Chinese-, Thai-, Korean-, Japanese-, and Indian-accented English). Analyses were conducted on the data from these 43 participants. There were 21 participants in the control condition and 22 in the treatment condition. All 43 participants were born in Australia. Table 3.1 lists the participants' ethnic origins.

[26] In the survey the word "experience" was used rather than "exposure." Examples of experience provided in the survey included watching a TV show in which the characters are from a different country and speak with that accent, and having a good friend, family member, or co-worker who speaks with an accent.

Table 3.1: *Details of participants' ethnic backgrounds*

Participant	European	Indigenous Australian	South Asian	Other – please specify (participants wrote in "____")			
				Australian	Fijian/ Indian	Egyptian	mixed
control condition (21)	16	0	0	3	1	0	1 (England-born father and India -born mother)
treatment condition (22)	16	1	2 (1: India -born parents, 1: Australia-born parents)	2	0	1	0

2.2 Vowel categorization task

2.2.1 Speakers

Two female speakers were recorded for auditory stimuli: one speaker of Australian-accented English and one speaker of Vietnamese-accented English. Born and raised in Western Sydney, the Australian-accented

English speaker was a monolingual in her 20s. The Vietnamese-accented English speaker was a native speaker of Vietnamese in her 30s. She learned English in Vietnam with Vietnamese teachers, and then moved to Australia when she was 19. At the time of the recording, she considered her English to be at an intermediate level.

2.2.2 Nonce word auditory stimuli

Auditory stimuli were recorded with Adobe Audition software on an Impact core i7 tower computer at a sampling rate of 44.1 kHz, with a Shure SM10A-CN headset microphone and a MOTU 896 mk3 sound card. The recordings were made in a sound-attenuated booth. The target items were 13 Australian English monophthongs embedded in the /hVdə/ context (*heeda* /hiːdə/, *hidda* /hɪdə/, *hedda* /hedə/, *hadda* /hædə/, *harda* /hɐːdə/, *hudda* /hɐdə/, *hodda* /hɔdə/, *horda* /hoːdə/, *hoodda* /hʊdə/, *who'da* /hʉːdə/, *hurda* /hɜːdə/, *heerda* /hɪədə/, and *hairda* /heːdə/). Each was repeated 10 times, four of which were chosen to be the stimuli for the vowel categorization task, based on the first author's subjective judgments of similarity in speaking rate and loudness. Australian-accented *hudda* served as an exception. In a pre-test, which involved four native speakers of Australian-accented English rating *hadda* and *hudda*, eight of the 10 Australian-accented *hudda* tokens were often judged to sound like *hadda*. For this reason, only the two tokens that were consistently judged to be *hudda* were included for this vowel. The two clear *hudda* tokens were each repeated so that all vowels were heard four times in training. The auditory stimuli were played to participants over Sennheiser HD280 PRO Headphones.

2.2.3 Reference word visual display

The task was programmed in ePrime (version 2.0). Thirteen reference words (*bad*, *bard*, *bead*, *beard*, *bed*, *bid*, *bird*, *book*, *bored*, *bud*, *food*, *paired*, and *pod*) were presented to the participants on the screen of a notebook computer. Their positions on the screen were randomized by participant. To maximize clarity, the letters representing the vowel in each word were highlighted in light red. Figure 3.1 shows an example of a reference word grid.

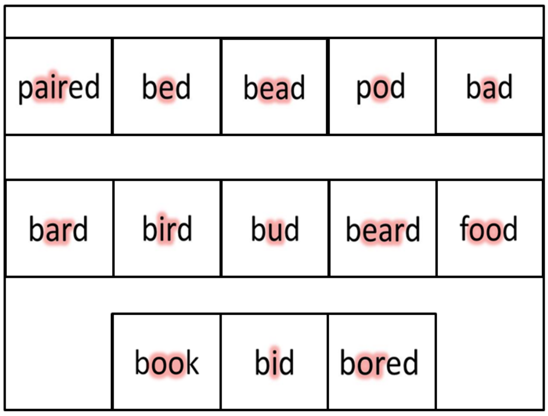


Figure 3.1: *An example of a reference word grid that participants saw in the vowel categorization task*

2.3 Procedure

An associate researcher who spoke Australian-accented English greeted participants at the lab. He asked participants to complete an online survey that investigated their previous exposure to several Asian accents (including the Vietnamese accent) and their attitudes towards the Asian group and other groups in Australia. The attitude data were collected as part of a larger project for another purpose and therefore will not be mentioned further in this paper. On each trial of the vowel categorization task, participants listened to a nonce word and selected the English word from the grid containing the vowel that best matched the first vowel sound in the nonce word. Participants also rated their choice, indicating on a scale from 1 to 7 how well the stimulus vowel matched the chosen category. Before commencing the task, participants were asked to produce the grid words, and to complete five practice trials using Australian-accented English. When the practice ended, participants in the control condition were informed that the two experimental blocks would each have a different speaker. Participants in the treatment condition were told to expect an Australian-accented English speaker in the training phase and a Vietnamese-accented English speaker in the test phase. That information was also repeated in the on-screen instruction at the start of each phase for both conditions. The associate researcher left the room before participants continued with the two phases of the vowel categorization task.

The nonce words described earlier were played to participants, one token per trial. In the training phase, participants listened to a block of 52 nonce word tokens in Australian-accented English (four tokens per vowel, 13 vowels). The stimuli were randomized across 52 trials in the block. Participants were given feedback on incorrect responses only. After participants selected an incorrect reference word, they would see the following message "Your response '[selected word]' is incorrect. The correct response is '[correct word].'" If the selection was correct, they would see another screen with their correctly selected word and the question "How good is the match? (1 = *foreign*, 7 = *native-like*)" and rated the match between the first vowel in the nonce word and the vowel in the reference word. The rating data were not analyzed for the purposes of this paper. The training phase ended after one block if at least three out of four tokens per vowel were correctly responded to for at least 10 out of the 13 vowels. Those who did not pass this training criterion after the first block (*n* = 29) repeated it by completing another block of 52 trials. The training terminated after four blocks, regardless of whether or not participants passed the training criterion (*n* = 15 participants who did not pass by four training blocks). In the test phase, all participants listened to just one block of 52 trials presenting nonce word tokens in Vietnamese-accented English (four tokens per vowel, 13 vowels), randomized across 52 trials. There was no feedback in the test phase. After selecting a reference word, participants rated the match regardless of whether or not a correct reference word was selected. The task lasted anywhere from 20 min to an hour (depending mostly on time spent in training), and this was followed by a debriefing.

3 Results

3.1 Analytical approach

Participants' performance on vowel categorization tasks has been documented to be variable and substantially lower than consonant categorization tasks. Accuracy tends to be below ceiling even when the stimuli are in the native accent of the listeners (Best et al., 2015; Faris et al., 2016). Therefore, all data from 43 participants who lacked previous exposure to the Vietnamese accent were included in the analyses, regardless of whether or not these participants passed the training phase after the fourth block. The results reported below are for the test phase of the experiment only.

We considered two possible indicators of perceptual shifts: (1) difference across conditions in overall accuracy; and (2) difference across conditions in priors over vowels. To assess the first indicator, we visually compared the mean accuracy across conditions, fitted binomial mixed effects models to the data, and evaluated the significance of ***condition*** as a fixed factor via model comparison. Since the Australian participants selected for analysis had no previous exposure to the Vietnamese accent, exemplar theories do not predict improvement.

Information about the speaker's Vietnamese origin could not have facilitated performance by activating exemplars of Vietnamese-accented English. However, it is possible from an exemplar theory perspective that information about the speaker degrades performance (cf., N. Nguyen et al., 2016; N. Nguyen et al., 2015), possibly by priming misleading (e.g., Asian but non-Vietnamese) exemplars (cf., McGowan, 2015). To assess the second indicator, we applied a Bayesian approach to detect the influence of expectations on perceptual space by exploring patterns of differences across conditions in priors over vowels. We attribute differences in vowel categorization across conditions to variation in expectations about the speaker's accent. Again, we evaluated statistical significance via model comparison.

Our calculation of priors derives from Bayes's rule (see Perfors, Tenenbaum, Griffiths, & Xu, 2011 for a recent tutorial)

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{\Sigma_{h_j \in H} P(d|h_j)P(h_j)} \qquad (1)$$

The term $P(h_i)$, or prior probability, is the expected probability of hypothesis h_i prior to observing data d. $P(h_i|d)$, or posterior probability, stands for the probability of hypothesis h_i being true given data d. $P(d|h_i)$ expresses the probability with which data d represent hypothesis h_i. $\Sigma_{h_j \in H} P(d|h_j)P(h_j)$ is the overall probability of data d across all the hypotheses. As can be seen in Equation (1), $P(h_i|d)$ depends on $P(h_i)$ and $P(d|h_i)$.

We solved Bayes's rule for the prior, as in Equation (2), and calculated the difference in priors across conditions according to the method described below

$$P(h_i) = \frac{P(h_i|d)\Sigma_{h_j \in H} P(d|h_j)P(h_j)}{P(d|h_i)} \qquad (2)$$

To simplify exposition, we refer to the term $P(d|h_i)/\Sigma_{h_j \in H} P(d|h_j)P(h_j)$ in Equation (1) as "signal specificity" because it captures the degree to which the signal picks out a particular category. We also use the familiar terms "posterior" for $P(h_i|d)$ and "prior" for $P(h_i)$. We solved for the signal specificity of each vowel category according to Equation (3). Because listeners heard each vowel with equal frequency in the training phase (and were given feedback), we assumed that the control participants assigned equal priors to the 13 vowels in the test phase that followed.[27] Accordingly, we set the prior in the control condition to 1/13. We estimated the posterior from categorization accuracy in the control condition. Estimates of the posteriors and priors in the control condition allowed us to solve for signal specificity.

$$Signal\ specificity\ = \frac{Posterior}{Prior} \qquad (3)$$

Since the stimuli were the same across conditions, we kept signal specificity constant across conditions. The signal specificity values computed according to Equation (3) for the control condition were used to solve for the prior in the treatment condition according to Equation (2), which is repeated in compact form as Equation (4). Again, we used categorization accuracy—this time in the treatment condition—to estimate the posterior. In the results, we report values for each of the terms in compact Equations (3) and (4).

$$Prior\ = \frac{Posterior}{Signal\ specificity} \qquad (4)$$

Lastly, we examined whether priors differed across conditions (control vs. treatment). To express the difference, we subtracted the value of the prior in the control condition from the value of the prior in treatment:

$\Delta\ prior = prior_{Treatment}$ - $prior_{Control}$

3.2 The effect of accent expectations on accuracy

Figure 3.2 shows that treatment participants, who were informed to expect the Vietnamese accent, performed at approximately the same mean level of accuracy as control participants. Using the lme4 package (Bates et al., 2014), we fitted two binomial mixed effects models in R (version 3.3.0 Patched) to the accuracy data. The models are summarized in Table 3.2. The baseline model contained random effects of participant and token but no fixed effects. The second model added ***condition*** as a fixed factor. The models confirmed the pattern shown in Figure 3.2 that the effect of ***condition*** on categorization accuracy is negligible.

[27] Given that listeners' accuracy varied, it is fair to say that all listeners “heard” each vowel with each frequency because they were given feedback in the training phase.

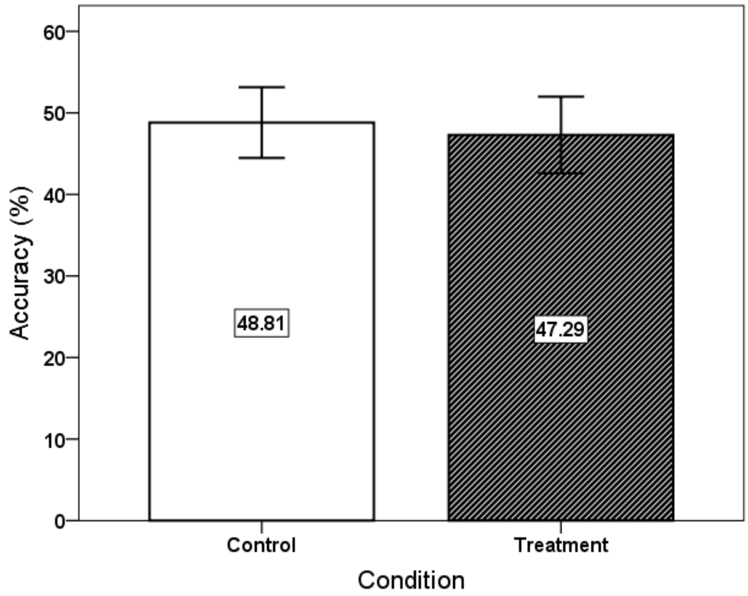


Figure 3.2: *Mean categorization accuracy across the two conditions* [Error bars indicate 95% confidence interval.]

Table 3.2: *Model comparison to check if* ***condition*** *is a significant predictor of accuracy*

Model	AIC[a]	BIC[b]	logLik[c]	deviance	χ^{2d}
(1) without any fixed factor: Accuracy ~ (1|Participant) + (1|Token)	2489	2506.2	-1241.5	2483	
(2) with ***condition*** as the only fixed factor: Accuracy ~ ***condition*** + (1|Participant) + (1|Token)	2491	2513.9	-1241.5	2483	0.00

Note: ***Condition*** is not a significant predictor of accuracy by way of model comparison, $p > .250$. A better model has smaller AIC, smaller BIC, higher logLik, and smaller deviance. [a]AIC = Akaike Information Criterion. [b]BIC = Bayesian Information Criterion. [c]logLik = log likelihood. [d]χ^2 = chi-square test based on log likelihood ratio (e.g., Busemeyer & Wang, 2000).

3.3 The effect of accent expectations on perceptual space

To investigate the difference in perceptual space across conditions, we quantified listener expectations by solving for the Bayesian priors over vowels, and mapping the differences in priors onto the vowel space of the Australian-accented English listeners. The differences in priors across conditions are reported by vowel in Table 3.3. The vowels *bored*, *book*, *bead*, *beard*, and *bard* have negative values, indicating that the priors of choosing these vowels decreased in the treatment condition (when listeners were informed of the speaker's Vietnamese accent), relative to the control condition. The priors for the vowels *bud* and *bid* did not change at all (diff = .000 for *bud*) or barely at all (diff = .003 for *bid*), indicating no perceptual shift. The other vowels have positive values, indicating that the priors of choosing those other vowels increased. Figure 3.3 expresses the results in Table 3.3 visually.

Table 3.3: *Differences in priors between control and treatment* [Red indicates the priors that decreased or hardly changed in treatment (i.e., vowels that were expected less often, or similarly often, in treatment than they were in control).]

Reference word	Posterior (control)	Prior (control)	Signal specificity	Posterior (treatment)	Prior (treatment)	Difference in priors
bored	.012	.077	0.155	.000	.000	-.077
book	.631	.077	8.202	.466	.057	-.020
bead	.738	.077	9.595	.614	.064	-.013
beard	.464	.077	6.036	.398	.066	-.011
bard	.714	.077	9.286	.659	.071	-.006
bud	.440	.077	5.726	.443	.077	.000
bid	.881	.077	11.452	.920	.080	.003
bed	.667	.077	8.667	.716	.083	.006
bird	.440	.077	5.726	.500	.087	.010
food	.655	.077	8.512	.761	.089	.013
paired	.417	.077	5.417	.500	.092	.015

pod	.214	.077	2.786	.295	.106	.029
bad	.001	.077	0.013	.011	.874	.797

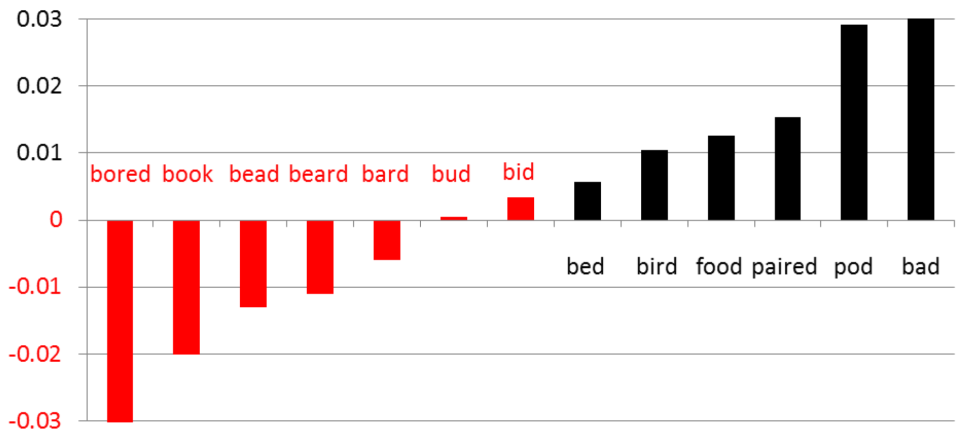


Figure 3.3: *Differences in priors across conditions* [Red bars indicate vowels that were expected less often (or similarly often) in treatment than in control; black bars indicate vowels that were expected more often in treatment than in control.]

In studies of phonetic variation across varieties of English, the relative articulatory (and auditory) positions of vowels are conventionally represented by a two-dimensional vowel space with F_1 on the vertical axis indicating the height of a particular vowel and F_2 on the horizontal axis indicating backness (e.g., Labov et al., 2006). Figure 3.4 shows the Australian-accented English vowel space that was generated from sampling the means and standard deviations reported for 60 female speakers in Cox (2006). Positioning the results in Figure 3.3 within the Australian-accented English vowel space in Figure 3.4 reveals a clear pattern: The vowels with the most negative difference in priors (*bored*/*book* (high, back), *bead*/*beard* (high, front), and *bard* (low, back)) are in the corners of the vowel space (henceforth, *corner vowels*); those with almost zero difference (*bud* (low, back) and *bid* (high, front)) are also corner vowels; and those with the most positive difference (*bed*, *bird*, *food*, *paired*, *pod*, and *bad*) are non-corner vowels. As a result, the difference in priors changes from negative to positive values for the non-corner vowels. This indicates that when our listeners, who reported having no previous exposure to the Vietnamese accent, were told to expect the Vietnamese accent, they expected to hear vowels towards the center of the vowel space more often. Vowels at the extreme edges of the vowel space were expected either less often or similarly often. In other words, the listener's perceptual space shrank just when they were told to expect a Vietnamese accent.

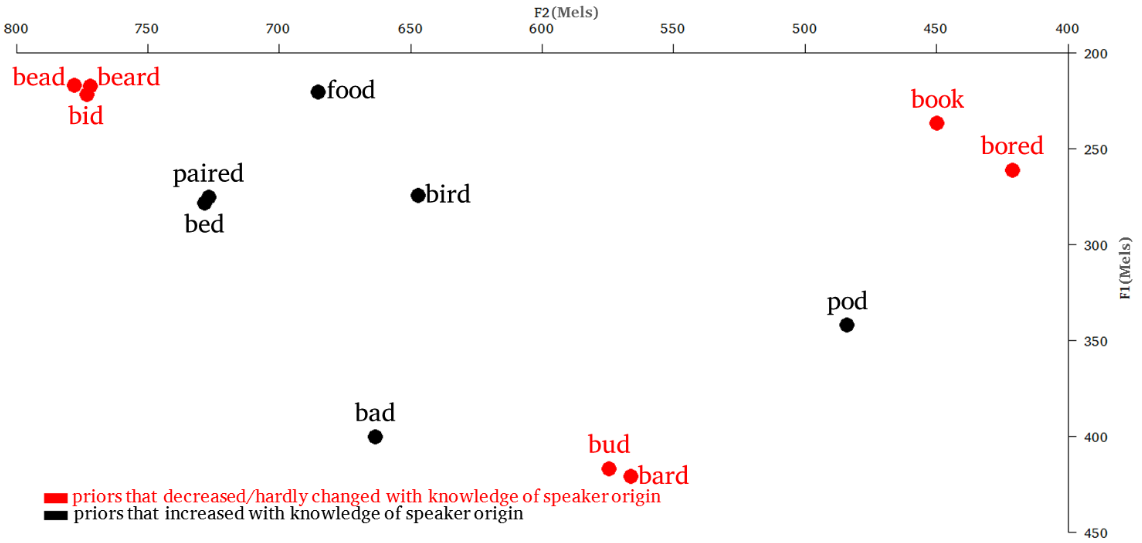


Figure 3.4: *Formant values (in Mels) for Australian-accented English vowels* [Vowels plotted in red had a decrease or no change in priors from the control to treatment conditions. Those plotted in black had an increase in priors.]

In light of the pattern in Figure 3.4, we divided the 13 vowels into two groups: vowels with decreased priors/no change in priors across conditions (those on the corners of the vowel space) and vowels with increased priors (those more towards the center of the vowel space). Figure 3.5 shows categorization accuracy by ***condition*** and by ***vowel location***. In the control condition without any cue about the Vietnamese accent, where participants may have expected the vowels to occur with equal probability, they were more accurate for the corner vowels than the non-corner vowels. In the treatment condition with cues about the Vietnamese accent, where participants may have had different expectations about the probability of the vowel categories, their categorization accuracy was basically the same for both corner and non-corner vowels (i.e., 50% vs. 46%). Therefore, compared with the control group, the treatment group was less accurate on the corner vowels and more accurate on the non-corner vowels. With ***condition*** and ***vowel location*** as two fixed factors, we fitted two binomial mixed effects models to the accuracy data in R: one with an interaction term between the two fixed factors and one without. Table 3.4 confirms that the interaction is significant by way of model comparison.

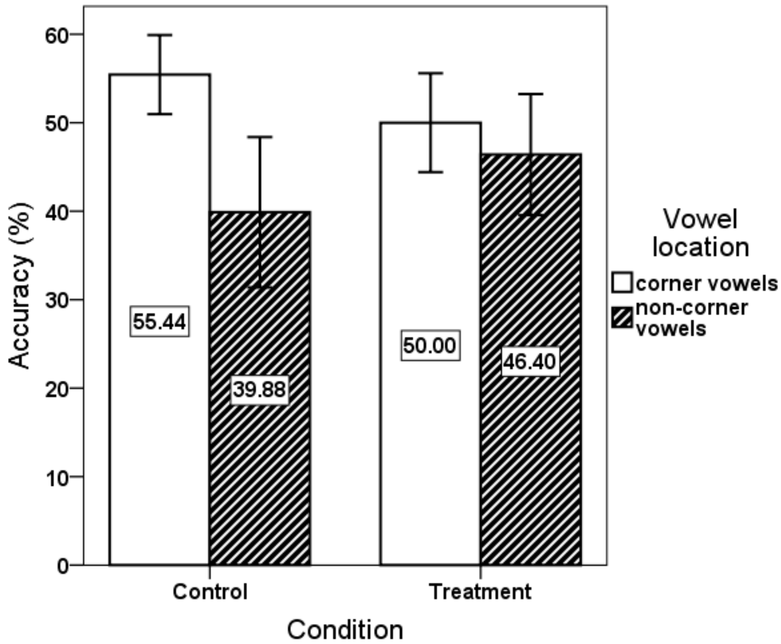


Figure 3.5: *Mean categorization accuracy as a function of* ***condition*** *and* ***vowel location*** [Error bars indicate 95% confidence interval.]

Table 3.4: *Model comparison to check the significance of the interaction between* ***condition*** *and* ***vowel location***

Model	AIC[a]	BIC[b]	logLik[c]	deviance	χ^{2d}
(1) with<u>out</u> interaction: Accuracy ~ ***condition*** + ***vowel location*** + (1+***condition***|Participant) + (1+***condition***|Token)	2496.4	2547.8	-1239.2	2478.4	
(2) with interaction: Accuracy ~ ***condition*** * ***vowel location*** + (1+***condition***|Participant) + (1+***condition***|Token)	2489.5	2546.7	-1234.8	2469.5	8.86

Note: The interaction between ***condition*** and ***vowel location*** is significant by way of model comparison, $p<.01$. A better model has smaller AIC, smaller BIC, higher logLik, and smaller deviance. [a]AIC = Akaike Information Criterion. [b]BIC = Bayesian Information Criterion. [c]logLik = log likelihood. [d]χ^2 = chi-square test based on log likelihood ratio (e.g., Busemeyer & Wang, 2000).

4 Discussion

We investigated whether expectations shifted vowel perception in the *absence* of previous exposure to a speaker's accent. Our findings suggest that Australian-accented English listeners, upon being told of the speaker's Vietnamese origin, expected to hear vowels that occur towards the center of the vowel space (vowels in *bed*, *bird*, *food*, *paired*, *pod*, and *bad*) more often. Vowels that occur at the extremes of the vowel space were expected either less often (*bored*, *book*, *bead*, *beard*, and *bard*) or similarly often (*bud* and *bid*). This pattern demonstrates a shrinkage of the perceptual space occurring when the speaker's accent is revealed, despite listeners having no previous exposure.

Our analyses demonstrated how Bayesian reasoning can be deployed to study listener expectations. Exemplar theories have been proposed as the cognitive mechanism for Bayesian computation (Shi et al., 2010).

Under such views, stored exemplars correspond to Bayesian priors, or exemplar-based beliefs. Speech exemplars are stored in memory, get weighted by the speech context, and become activated when they best explain the signal. Our study raises an interesting question about this link between exemplar theories and Bayesian computation, as our participants did not have any stored exemplars of the Vietnamese accent over which they could compute priors. Nevertheless, the group who were told to expect a Vietnamese accent demonstrated different expectations about vowel categories than the control group. This suggests that there is more to the formation of beliefs about accented vowels than just direct observations, or abstraction over exemplars. Although beliefs could certainly be formed from direct, specific experiences, they could also be formed through other more general experiences that listeners have been exposed to in the course of their lives, including hearsay and stereotypes.

In this particular case we argue that the differences in listener responses across conditions derived from the combination of two beliefs. One belief is that vowels produced in the Vietnamese accent are difficult to categorize. The other is that the non-corner vowels are harder to categorize than the corner vowels. We elaborate on each of these beliefs in turn.

The Greater Sydney area, from which our listeners were recruited, is culturally diverse, and the Australian-accented English listeners in our study have most likely had exposure to a range of foreign accents, including Arabic-, Mandarin-, and Cantonese-accented English.[28] Foreign accents reflect speakers' native language vowel systems, which both vary across languages and also differ from Australian-accented English vowels in ways that Australian-accented English listeners are unlikely to be aware of (see Yi, 2010; Thompson, 1987; and Cox, 2006 for differences in the Chinese, Vietnamese, and Australian-accented English vowel systems). This situation, we argue, is likely to result in the general belief that foreign accents are difficult to understand. In addition, there is a literature documenting specific difficulties that Australian-accented English listeners have perceiving Vietnamese-accented English in particular (Ingram & Nguyen, 2007; T. Nguyen & Ingram, 2004; Zielinski, 2003, 2006, 2008). Even without direct exposure to the Vietnamese accent, our listeners may have had the belief that the Vietnamese accent would be difficult to understand or even that any foreign or

[28] In 2011, 4.1% of the people in the Greater Sydney area spoke Arabic at home, which was the language with the largest number of speakers in non-English-speaking households. Mandarin and Cantonese tied for second place with 3% of Greater Sydney-siders speaking them at home (Australian Bureau of Statistics, 2011).

unfamiliar accent would be difficult to understand.

The second factor contributing to the shrinkage of the perceptual vowel space is, we argue, the tacit belief that corner vowels are easier to understand than non-corner vowels. That listeners have tacit knowledge of which vowel categories are more similar is readily apparent from language use as in, for example, the formation of puns and half-rimes (Kawahara, 2007; Kawahara & Shinohara, 2009), and is proposed to be a basic component of phonological cognition (Steriade, 2001/2008). The corner vowels are the most common across languages, presumably because they are best-suited for the human speech production-perception loop (Stevens, 1989). Listeners in the treatment condition reduced their expectations for these easy-to-understand vowels.

To put the two beliefs together, we surmise that the listeners who were told to expect a Vietnamese-accented English speaker thought that the task would be hard and, therefore, expected to hear vowels that are more difficult to categorize. Although this strategy did not lead to increased accuracy in this task, it may be an effective heuristic in more naturalistic listening situations and a basic component of human speech perception. More broadly, the results indicate that expectations developed outside of direct exposure can influence perception. Listeners seem to rely on preconceived ideas about unfamiliar foreign accents to guide their perception. Importantly, beliefs about a speaker's accent systematically change the way speech is perceived, even without previous exposure to the accent.

Chapter 4

Intergroup Dynamics in Speech Perception: Interaction among Experience, Attitudes, and Expectations

A version of this chapter is published as:

Nguyen, N., Shaw, J. A., Pinkus, R. T., & Best, C. T. (2016). Intergroup dynamics in speech perception: Interaction among experience, attitudes, and expectations. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 141–150. Retrieved http://repository.upenn.edu/pwpl/vol22/iss2/16/

Abstract: Experience, attitudes, and expectations have been identified as separate influences on speech perception and comprehension across groups. In this study, we investigate the interaction among these three variables. 58 Australia-born participants completed an online survey and a vowel categorization task. The survey examined participants' experience with Vietnamese-accented English and their attitudes towards Asians. The vowel categorization task examined participants' recovery of a Vietnamese-accented speaker's intended vowels. Half of the participants were told to expect a Vietnamese accent whereas the other half were not. Results indicated that the relationship between listener expectations and group attitudes varied according to whether or not participants had experience with the Vietnamese accent. This interaction was most clearly reflected on the "book" vowel. Compared to participants who did not expect a Vietnamese accent, had no experience with the Vietnamese accent, but positive attitudes towards the Vietnamese group, those who expected a Vietnamese accent showed a decrease in accuracy on "book" categorization. A decrease in "book" categorization accuracy was also found for those having experience with the accent but negative attitudes. In contrast, an increase in accuracy was found for those having no experience with the Vietnamese accent and negative attitudes towards the Vietnamese group, and those having experience with the accent and positive attitudes. We concluded that expectations, experience, and attitudes interact in their relationship with speech

perception.

1 Introduction

Since Rubin's (1992) study on the effect of perceived speaker ethnicity on speech perception, listener factors have received increasing attention in speech perception and comprehension research (e.g., Babel & Russell, 2015; Hay & Drager, 2010; Hay et al., 2006a; Lindemann, 2002; McGowan, 2015; N. Nguyen et al., 2015; Niedzielski, 1999). These studies demonstrate the importance of three listener factors: attitudes, expectations, and experience. Listener *attitudes* towards a foreign-accented speaker's group have been shown to relate to the accented speech perception and comprehension in several ways. First, listeners with negative attitudes towards Koreans reported unsuccessful communication with the accented speakers whereas those with positive attitudes reported successful communication (Lindemann, 2002). Second, when listeners had negative attitudes towards Koreans and used avoidance strategies (i.e., not giving feedback to clarify information to their Korean-accented conversational partners), on top of their perceived unsuccessful communication with their Korean partners, their interactions were also genuinely unsuccessful (Lindemann, 2002). Third, listener attitudes towards Asians have also been found to negatively correlate with categorization accuracy of Vietnamese-accented vowels (N. Nguyen et al. 2015). Listener *expectations* about a speaker's accent, in turn, have been demonstrated to shift perception of vowels in regional accents in the direction of the expected accents (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). Finally, *experience* with a speaker's accent has also been found to improve accuracy of foreign-accented speech comprehension (McGowan, 2015).

To date, however, the effects of attitudes, expectations, and experience on speech perception have been researched separately. In Niedzielski's (1999) study on *expectations*, for example, although some participants had experience with Canadian vowels, others did not; unfortunately, the study did not quantify the relationship between such *experience* and listeners' vowel perception. Research quantifying *experience* with a speaker's accent, for example McGowan's (2015) study, did not take listener *attitudes* into account. N. Nguyen and colleagues (2015) examined the relationship between affective *attitudes* towards Asians and Vietnamese-accented vowel perception, but did not take listener *experience* with the accent into consideration.

The current study, therefore, was designed to explore how attitudes, expectations, and experience interact in speech perception. Specifically, we manipulated listener *expectations* about a speaker's accent, examined which vowels were affected by this information, then explored how the perceptual effects of the experimental manipulation interact with the effects of the other two factors: listeners' *experience* with the accent and their *attitudes* towards the speaker's group. To achieve that goal, firstly, we administered a survey to our Australian English participants to examine their experience with Vietnamese-accented English. We then assessed their attitudes towards Asians via the Scale of Anti-Asian American Stereotypes (SAAAS), modified for the Australian context (N. Nguyen et al. 2015). We then revealed the speaker's Vietnamese accent to the participants in the treatment condition prior to our speech perception test, to create expectations about the speaker's accent as well as to elicit effects of group attitudes. The participants in the control condition, by contrast, did not receive such information about the accent, and thus should have had neither specific expectations nor attitudes towards the speaker's group. Expectations created in the treatment condition were predicted to have an effect on particular vowels (as seen in Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). Attitudes evoked in the treatment condition were predicted to negatively relate to participants' performance in a vowel categorization task (Bundgaard-Nielsen et al., 2011; Faris et al., 2016) (as seen in N. Nguyen et al. 2015). We also predicted a positive relationship between experience with the Vietnamese accent and participants' categorization performance (similar to McGowan, 2015).

2 Method

2.1 Participants

60 first-year Psychology students from the Western Sydney University (WSU) participant pool participated in the study for course credit. Two participants were excluded prior to data analysis because they did not complete the vowel categorization task. Data analyses were conducted on the remaining 58 participants (32 control, 26 treatment) who were between the ages of 18 and 45 (M = 21, SD = 4.7). Although Australia-born, participants had a range of self-reported family backgrounds (European = 41, Indigenous Australian = 2, South American = 1, African = 1, European and South Asian = 1 (England-born father and India-born mother), Fijian = 1 (Fiji-born parents of Indian heritage), South Asian = 3 (1 Afghanistan-born parents, 1 India-born parents, and 1 Australia-born parents), Southeast Asian = 1 (Thailand-born parents)). Seven participants chose the 'other – please specify' option and wrote in 'Australian' in the blank. Of these 58 participants, 15 reported having experience with Vietnamese-accented English while the rest reported having none (n = 43).

2.2 Survey

Our survey explored participants' experience with the Vietnamese accent and attitudes towards Asians. The question on experience was just a simple yes/no question, asking "Do you have experience with the following accent?" and a list of 10 accents (i.e., Vietnamese accent and nine filler accents: Chinese, Mexican, Italian, Thai, Lebanese, Korean, French, Japanese, and Indian). Attitudes towards Asians were quantified by the SAAAS scale (Lin et al., 2005), adapted to the Asian Australian group and three filler groups in the Australian context: Aboriginal Australians, Anglo Australians, and Arab Australians. Built on the Stereotype Content Model (Fiske et al., 2002), the SAAAS scale comprises 25 items: 12 indicating Competence and 13 indicating Sociability. The scale items are about cognitive attitudes or stereotypes, but they were designed in such a way that they can indirectly quantify affective attitudes or prejudice (i.e., positive and negative prejudice; Fiske et al., 2002, Lin et al., 2005): SAAAS prejudice comes from the combination of the Competence and Sociability dimensions, which can indicate mixed evaluations about a group. For example, Asian Americans are respected for their high Competence but disliked for their low Sociability (Fiske et al., 2002, Lin et al., 2005). Participants' responses were coded from 0 ("strongly disagree") to 5 ("strongly agree") for 19 normal items and vice versa for 6 reverse-scored items. The higher the SAAAS scores, the stronger the negative prejudice towards a group. The SAAAS scores for the control condition were negatively skewed and ranged from 18 to 105 (M = 73.28, SD = 23.17). The SAAAS scores for the treatment condition were normally distributed and ranged from 33 to 122 (M = 70.31, SD = 20.40).

It was important to distract participants from the true purposes of the survey. If participants figured out those purposes, they would be likely to respond to the survey items in a certain way to present themselves in a positive light, a bias that is called a "demand characteristic" (Orne, 1959). Therefore, the accent experience question and SAAAS scale above were interspersed with other filler questions and scales such as questions on personal details and language backgrounds, 17 emotion items (Fiske et al., 2002), a liking item (adapted from Stephan, Ybarra, Martínez, Schwarzwald, & Tur-Kaspa, 1998), Ten-Item Personality Inventory (TIPI) (Gosling, Rentfrow, & Swann Jr., 2003), Balanced Inventory of Desirable Responding (BIDR) (Paulhus, 1984), and emotional responses scales. For comparison purposes, all participants experienced the same order of questions and scales in the survey. However, to avoid order effects, groups within a scale (e.g., Aboriginal Australians, Anglo Australians, Arab Australians, and Asian Australians) were randomized, and items within a group (e.g., Asian group) were also randomized. Qualtrics Survey Software on the WSU server was used to host the survey online.

2.3 Vowel categorization task

2.3.1 Speakers

Auditory stimuli were recorded from two female speakers. One speaker was born and raised in Western Sydney, was in her 20s, and spoke only Australian-accented English. The other speaker was born in Vietnam, learned English in Vietnam with Vietnamese teachers, and immigrated to Australia at 19 years of age. At the time of the recording, she was in her 30s, self-evaluated her English to be at an intermediate level, and spoke it with a Vietnamese accent. The stimuli from the Australian-accented speaker were used in the training phase of the experiment, and those from the Vietnamese-accented speaker were used in the test phase.

2.3.2 Nonce word auditory stimuli

Auditory stimuli were recorded in a sound-attenuated booth at The MARCS Institute for Brain, Behaviour and Development, Western Sydney University. Adobe Audition software was used to record auditory stimuli on an Impact core i7 tower computer. The sampling rate was 44.1 kHz and the sound card was MOTU 896 mk3.

The speakers were recorded with a Shure SM10A-CN headset microphone. They were instructed to look at PowerPoint slides and, on each slide, read out a key word containing one of 13 Australian English monophthongs (i.e., /iː/, /ɪ/, /e/, /æ/, /ɐː/, /ɐ/, /ɔ/, /oː/, /ʊ/, /ʉː/, /ɜː/, /ɪə/, /eː/), then that monophthong on its own, then that monophthong embedded in the /hVd/ and then /hVdə/ contexts (e.g., ban, æ, had, hadda). The production steps were put in place to guide the speakers to produce the correct vowels for the /hVdə/ nonce words. For the Australian-accented speaker, the vowels were presented randomly within a block of 13, and repeated 10 times. For the Vietnamese-accented speaker, since she had difficulty producing the vowels consistently across the 10 repetitions when they were randomized, the vowels were each repeated 10 times in a row to ensure consistent productions for stimulus selection purposes. In addition, for the Vietnamese-accented speaker to correctly produce the schwa, the nonce words were presented to her as a mixture of English and Vietnamese orthography (e.g., "hadda" was written as "hadđờ").

For each set of 10 tokens belonging to the same vowel, we subjectively judged their similarity in terms of speaking rate and loudness, and selected four of them to be the stimuli for the experiment. However, for the

Australian-accented "hudda" tokens, only two were chosen as the other eight were judged by native Australian English listeners to sound closer to "hadda" in a pre-test. We repeated each of these two clear "hudda" tokens twice to ensure that the vowel would appear four times in the training phase.

2.3.3 Reference word visual display

Participants were presented with a grid of 13 reference words (i.e., bad, bard, bead, beard, bed, bid, bird, book, bored, bud, food, paired, and pod). The presentation of those words on the screen was programmed via ePrime (version 2.0), with the positions randomized by participant. For each word, light red was used to highlight the letters indicating the vowel. Figure 4.1 illustrates what a participants' screen looked like in the task.

2.4 Procedure

At the lab, participants were greeted by an associate researcher who was a Caucasian Australian and spoke Australian-accented English. They were then instructed to do the online survey first. After finishing the survey, they were asked to do the vowel categorization task, starting with a five-trial practice, then the training phase and after that the test phase. Before the training phase started, participants in the treatment condition were told to expect an Australian accent in the training phase and a Vietnamese accent in the test phase whereas those in the control condition were told to expect two different speakers only. In the training phase, participants categorized Australian-accented English vowel tokens in a block of 52 trials (one token per trial × four trials per vowel × 13 vowels). The 52 trials were randomized. Feedback was given to participants on incorrect responses only. When participants had an incorrect response, the following message appeared on the screen "Your response ''[selected word]'' is incorrect. The correct response is [correct word]." When they responded correctly, the experimental program asked them to rate the match between the highlighted vowel in the word they selected and the first vowel sound in the nonsense word they heard: 1 = "foreign"; 4 = "okay"; and 7 = "native-like." After participants finished rating, the next trial began. After one block, if participants correctly responded to at least three out of four tokens of a vowel and at least 10 out of the 13 vowels, their training ended and the experiment moved on to the test phase. If participants did not pass the above criterion, another 52-trial block was presented to them. When they reached the end of the fourth training block, irrespective of whether or not they satisfied the criterion, the test phase started. The test phase was identical to the training phase, except that the stimuli were in Vietnamese-accented English, that participants went through only one

52-trial block, and that they did not get feedback on incorrect responses.

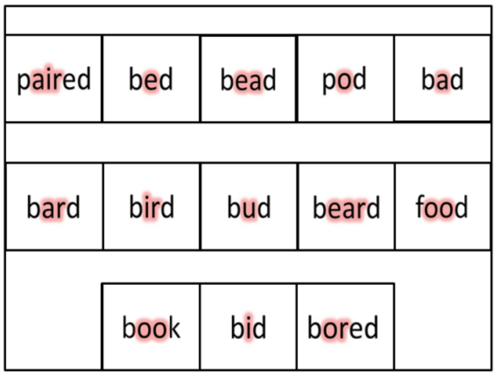


Figure 4.1: *One of the possible orders of reference words that was displayed on participants' screen in the vowel categorization task*

Participants listened to the auditory stimuli via Sennheiser HD280 PRO Headphones (once per trial) and saw the reference words on Acer TravelMate P645 notebook computers. The duration of the task was from 20 min to an hour (depending mainly on how long participants took in training). At the end of the experiment, the associate researcher debriefed the participants on the purposes of the vowel categorization task. Interested participants received a full debriefing about the connection between the survey and the vowel categorization task and a summary of results at the end of the project.

3 Results

3.1 Expectation effects

We began by fitting four binomial mixed effects models to the accuracy data in R (version 3.1.2) to examine the expectation effects using lme4 (Bates et al., 2014). We checked the main effects of expectations by comparing a model without any fixed factor and another model with expectations as the only fixed factor. We found no main effect of listener expectations on overall vowel categorization accuracy. Since previous findings establish expectation effects on individual vowels (Hay & Drager, 2010, Hay et al., 2006a, Niedzielski 1999), we then checked the interaction between expectations and vowels by comparing two other models containing vowel as a fixed factor, one with and one without the interaction term between expectations and vowels. Random effects of participants (intercept only) and tokens (intercepts and slopes varying with expectations) were included for all models. In Table 4.1, the results of model comparison show the significance of the interaction between expectations and vowels, with AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, logLik = log likelihood, Pr(>Chisq) = *p*-value of the Likelihood Ratio Test (LRT) applied for models (1) and (2), which follows a Chi-Square distribution. Compared to model (1), smaller deviance in model (2) means that model (2) fits the data better and explains more variance. In addition, the *p*-value of the LRT shows that the difference between models (1) and (2) (i.e., the interaction term) is significant. However, the higher AIC and BIC in model (2) mean that the variance explained does not justify model complexity (i.e., there is a chance that the interaction term is over-fitting the data).

Table 4.1: *Significant interaction between expectations and vowels by model comparison*

Model	AIC	BIC	logLik	deviance	Pr(>Chisq)
(1) without interaction term: Accuracy ~ expectations + vowels + (1|Participant) + (1+expectations|Token)	3213.3	3321.5	-1588.7	3177.3	
(2) with interaction term: Accuracy ~ expectations * vowels + (1|Participant) + (1+expectations|Token)	3214.0	3394.4	-1577.0	3154.0	<0.05

Figure 4.2 illustrates expectation effects on individual vowels. The vowels are arranged in the ascending order of accuracy. Error bars indicate one standard error. Expectations appear to have an effect on the categorization accuracy of "book," "bead," "food," and "bard." In the treatment condition, when listeners were told to expect the Vietnamese accent, accuracy decreased on vowels that fall around the periphery of the Australian English vowel space: "book," "bead," and "bard." Accuracy on "food," which is fronted to the degree that it is a central vowel /ʉː/ in Australian English, goes in the other direction. Its accuracy increased when listeners expected the speaker to have a Vietnamese accent. Table 4.2 shows the significant predictors in model (2). Among "book," "bead," "food," and "bard," which are observed to apparently be under

expectation effects, only "bead" and "book" are (marginally) statistically significant, with β = log odds of accuracy for the corresponding predictor, and Pr(>Chisq) = *p*-value of the LRT (as mentioned for Table 4.1).

Table 4.2: *Significant predictors in model (2)*

Predictor	β	Pr(>|z|)
(Intercept)	0.83	<0.001
bad	-5.10	<0.001
beard	-0.87	<0.01
bid	1.13	<0.001
bird	-1.20	<0.001
book	-0.53	<0.05
bored	-5.80	<0.001
bud	-1.00	<0.001
paired	-1.10	<0.001
pod	-1.91	<0.001
treatment * bead	-0.81	0.055
treatment * book	-0.78	0.059

The formant plot of the Australian English vowels in Figure 4.3 was regenerated with means in Mel units reported for 60 female speakers in Cox (2006). It highlights the locations of the vowels with decreased categorization accuracy in the treatment condition (bead, beard, bad, bud, bard, and book), as opposed to the locations of those with increased categorization accuracy (bid, paired, bed, food, bird, pod, and bored), within the Australian English vowel space. We will focus the following discussion on "bead" and "book" since they showed the most reliable effects.

3.2 Interaction between expectations and experience

Figure 4.4 shows categorization accuracy on the Vietnamese-accented "bead" and "book" vowels across conditions for listeners with or without experience with the Vietnamese accent. Error bars indicate one standard error. We ended up with unequal numbers of experienced and inexperienced participants: 11 experienced participants out of 32 in the control condition, but just four out of 26 in the treatment condition. Although the numbers of inexperienced and experienced participants were different, the effect of expectations on vowel categorization was the same for both experienced and inexperienced listeners. For both experience groups, knowledge that the speaker had a Vietnamese accent reduced accuracy on “bead” and “book” vowels, relative to lack of such knowledge. In the next section we add prejudice data to the current factors to explore why experience does not seem to help vowel perception, even when listeners know the speaker’s accent.

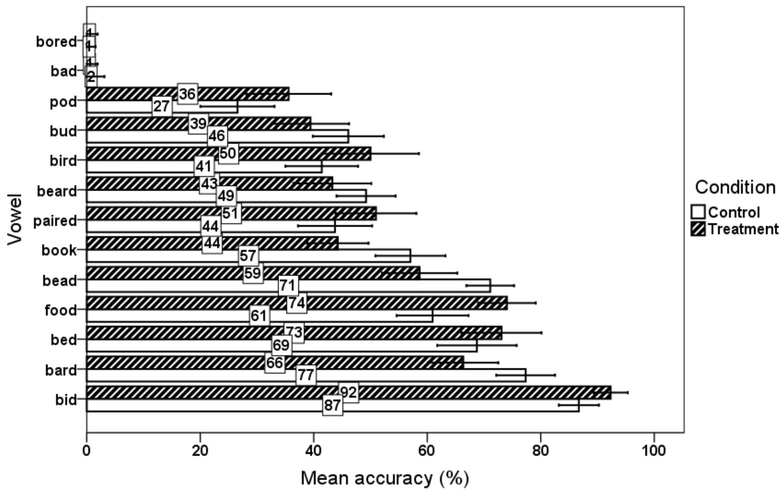


Figure 4.2: *Mean categorization accuracy by vowel in test* [Error bars indicate one standard error of the mean.]

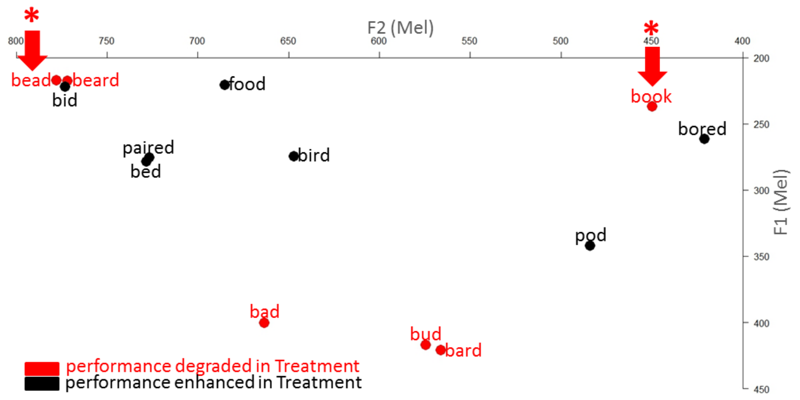


Figure 4.3: *Degraded (red) versus enhanced (black) performance in the treatment condition, relative to the control condition, as located within the Australian English vowel space* [Arrows mean (marginally) statistically significant differences between treatment and control participants.]

3.3 Interaction among expectations, experience, and prejudice

Recall that prejudice was estimated using SAAAS scores, with negative prejudice inferred from high SAAAS scores. Figure 4.5 plots participants' centered SAAAS scores against the percentages of their categorization accuracy according to the conditions they were in. The curves reflect quadratic functions fit to the data points (control: R^2 = 0.076, treatment: R^2 = 0.057). We observe that the correlation between categorization accuracy and SAAAS scores is curvilinear for both conditions, with an increase in accuracy for both low and high SAAAS scores (i.e., positive prejudice and negative prejudice) and a dip in accuracy for mid SAAAS scores (i.e., no particular prejudice). Since the relationship between categorization accuracy and SAAAS scores is not linear, SAAAS scores were standardized and divided up into three bins: mid bin consists of zScores between -0.5 and +0.5, expressing no particular prejudice (24 participants: n_{control} = 12, $n_{\text{treatment}}$ = 12); low bin consists of the scores below -0.5, expressing positive prejudice towards Asians (15 participants: n_{control} = 8, $n_{\text{treatment}}$ = 7); and high bin with the scores above +0.5, expressing negative prejudice towards Asians (19 participants: n_{control} = 12, $n_{\text{treatment}}$ = 7) (Figure 4.6).

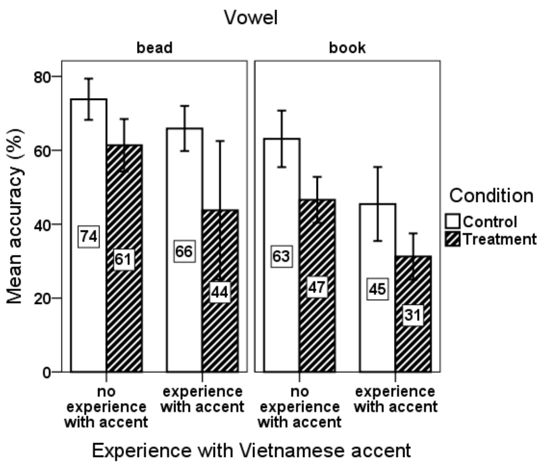


Figure 4.4: *Mean categorization accuracy by "bead" and "book" in test across conditions and across experience levels* [Error bars indicate one standard error of the mean.]

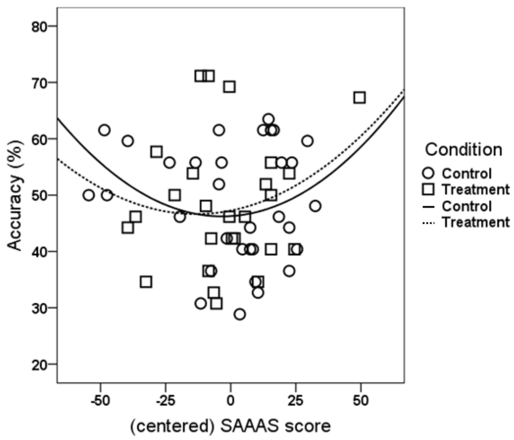


Figure 4.5: *Non-linear relationship between categorization accuracy and SAAAS scores*

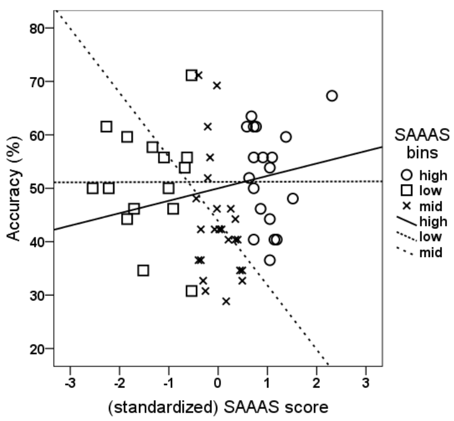


Figure 4.6: *Three prejudice bins based on standardized SAAAS scores*

Figure 4.7 plots participants' experience (+/-) with the Vietnamese accent against their accuracy percentages, according to the respective conditions they were in (i.e., no fill for control and pattern fill for treatment) and the types of prejudice they held towards Asians (i.e., bottom row for those with positive prejudice, middle row for those with no particular prejudice, and top row for those with negative prejudice) for the Vietnamese-accented "bead" vowel on the left column and "book" on the right. The rows on the left column show that there is no relationship between prejudice and the accuracy on the "bead" vowel. The pattern is the same across SAAAS bins. However, for "book," the rows on the right column show a clear relationship between prejudice and accuracy that interacts with experience and expectations. Specifically, in the experienced group, participants' performance for "book" was the same whether or not they expected a Vietnamese accent, if they did not have a particular prejudice towards Asians. However, their performance was enhanced when they expected a Vietnamese accent and had positive prejudice towards Asians. In contrast, participants' "book" vowel perception degraded with negative prejudice towards Asians. Participants without experience show the opposite pattern. Their performance degraded with positive prejudice while it was enhanced with negative prejudice towards Asians.

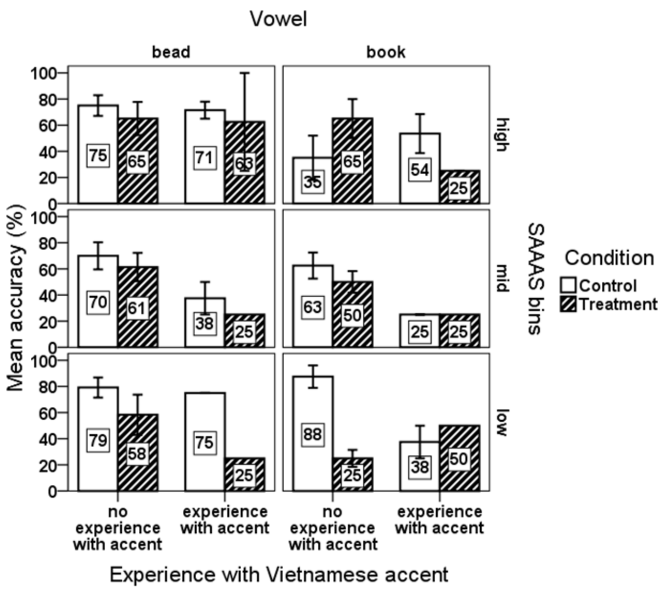


Figure 4.7: *Mean categorization accuracy by "bead" and "book" in test across conditions, across experience levels, and across prejudice levels* [Error bars indicate one standard error of the mean.]

4 Discussion

Expectations, experience, and attitudes were researched separately in previous speech perception studies. The current study was designed to put these three listener factors under scrutiny together. We predicted an *expectation* effect on particular vowels (Hay & Drager 2010; Hay et al., 2006a; Niedzielski, 1999), and we found the effect on "bead" and "book" (although only marginally significant). Since *experience* with an accent has been found to aid speech perception (McGowan, 2015), we predicted that, in our study, experience would enhance speech perception regardless of expectations and attitudes. We therefore did not predict that experience would aid accuracy only for low-prejudice listeners who were told about the speaker's accent, as was the case for the "book" vowel. We did predict the observed interaction between expectations and *prejudice*. In N. Nguyen and colleagues (2015), there was no control condition in which listeners were uninformed about the speaker's accent. All listeners were told to expect a Vietnamese accent (similar to our treatment condition). Although the data is the thinnest for the listeners reporting experience with the Vietnamese accent in our treatment condition (n = 4), negative prejudice in this group (n = 2) seems to relate to the decrease in accuracy of "book" categorization, similar to the finding by N. Nguyen and colleagues (2015). However, the relationship was not found for the control condition, suggesting an interaction between expectations and prejudice. Besides, as noted earlier, experience with the Vietnamese accent also contributes to the relationship among expectations, prejudice, and accuracy: Listeners who were told to expect the Vietnamese accent (and had no experience with it) showed a positive relationship between prejudice and accuracy, compared to the negative relationship found for those who had experience with the Vietnamese accent. In short, findings from past work motivating this study did not really prepare us to expect all aspects of the three-way interaction among expectations, experience, and prejudice in our data. The issue of how prejudice towards a group relates to the perception of the speech produced by members of that group is a complex one, which seems to be influenced by many factors.

In the remainder of the discussion, we speculate on some possible connections to the Social Psychology literature on mood effects (e.g., Isen, Means, Patrick, & Nowicki, 1982; Sinclair, 1988) that could be developed to explain some aspects of the pattern, particularly why positive prejudice could lead to decreased categorization accuracy and negative prejudice could lead to increased categorization accuracy in certain situations. According to Isen et al. (1982), individuals in a happy mood have an "intuitive" and simplified approach to problems. They generally avoid exerting cognitive effort to find optimal solutions to problems, especially when they deem the problems to be unimportant. As a result, they may make decision errors. Similar to the happy mood effects, perhaps listeners who had positive prejudice towards the speaker group may not have processed the speech signals carefully enough when they were aware of the speaker's accent, resulting in lower accuracy. In contrast, individuals in a depressed mood have been reported to be more careful and controlled in their manner, differentiate more categories, and process information in an algorithmic way, resulting in more accuracy in their performance (Sinclair, 1988). As with the depressed mood effects, for listeners with negative prejudice, disliking the speaker group may have boosted their expectations for exotic vowels. They may have attended to the speech signals more closely and followed algorithmic processing, resulting in greater accuracy.

5 Conclusions

Our results demonstrate that expectations, experience, and prejudice interact in their relationship with vowel perception. The relationship between listener expectations and group prejudice is different for experienced and inexperienced listeners. In our experiment, this result comes out most clearly for the "book" vowel. For inexperienced listeners, accuracy on the "book" vowel decreased for those with positive prejudice towards Asians but increased for those with negative prejudice. Those with negative prejudice towards Asians (and no experience with the Vietnamese accent) seemed to concentrate harder when they were told about the speaker's accent. In sharp contrast, among experienced listeners, accuracy on the "book" vowel decreased for those with negative prejudice towards Asians but increased for those with positive prejudice. Listeners' prejudice towards the speaker group goes hand in hand with whether they can make productive use of their experience. Although this study has some limitations, including uneven numbers of participants across conditions, we think the general approach can fruitfully contribute to understand the dynamics of intergroup factors in speech perception and comprehension.

Chapter 5

Discussion

1 Summary and evaluation

This thesis investigated the overall hypothesis that listeners use abstract information whenever it is available in perceiving speech. It also attempted to evaluate the role of socio-indexical information in speech perception and, in particular, the possibility of further abstraction over social categories. More specifically, it asked the following questions:

(1) Do high-level abstractions exist for linguistic information only, or do they also exist for socio-indexical information?

(2) Does the representation of speech categories reflect listeners' beliefs about speakers that they have no direct exposure to? If it does, how do these manifest in speech perception?

(3) Is prejudice associated with speech perception? If it is, how do other listener factors (both abstract and specific) such as expectations about a speaker's speech and previous experience with it play out in the relationship between prejudice and speech perception?

The answers to these three questions were pursued in the three experimental chapters, summarized in turn below.

Chapter 2 sought to answer the first research question. It hypothesized the existence of a high level of abstraction over socio-indexical information called socio-indexicality categories ("more socio-indexical" vs. "less socio-indexical"), and Australian English vowels were classified into these two categories. The relationship between phonetic variability and speech perception behaviour was then examined, which was predicted to be different across the socio-indexicality categories. Categories that carry more socio-indexical information should have multiple narrow distributions over a given phonetic parameter corresponding to different speech situations (e.g., accent types and age cohorts). Therefore, the total amount of variation in the distribution over that same phonetic parameter (as found in a large corpus) was predicted to negatively correlate with perceptual accuracy. In contrast, for categories that carry less socio-indexical information (i.e., few narrow distributions over a given phonetic parameter corresponding to different speech situations), the total amount of variation was predicted to positively correlate with perceptual accuracy. These predictions were tested in a categorization paradigm, where Australia-born participants listened to Australian-accented English nonce words categorizing them as one of the 13 Australian English monophthongs. Participants' accuracy data from the vowel categorization task corresponded to the perceptual accuracy mentioned in the predictions. Formant measurements reported from a large corpus of Australian-accented English (Cox, 2006) were used to quantify the total amount of variation over the first two formants of a monophthong in two ways: (a) formant standard deviations (SD), and (b) residuals of the regression between formant means and formant SDs. Previous findings in the Australian socio-phonetic literature regarding accent types and sound change were used to inform the classification of 13 Australian-accented English monophthongs into socio-indexicality categories (i.e., monophthongs that *both* differentiate among accent types *and* participate in sound change were classified into the "more socio-indexical" vowel subset, and the rest belonged to the "less socio-indexical" subset). Results from mixed-effects modeling showed a significant negative correlation between formant variability and categorization accuracy for more socio-indexical vowels as well as an apparent positive trend for the less socio-indexical ones. These results provided some support for the predictions about the different relationship between phonetic variability and perceptual accuracy for "more socio-indexical" vowels versus "less socio-indexical" ones. Results also showed that method (b) of quantifying formant variability, the residual method which abstracts away the magnitude of the formant means, could provide a better estimate of vowel formant variability. Together, these results possibly lend some support to the hybrid approach to speech perception (e.g., McQueen et al., 2006; Pierrehumbert, 2016) as well as the overall hypothesis of the thesis that listeners use abstract information in perceiving speech, including abstraction over social categories.

While the experiment in Chapter 2 sought evidence of a high-level of abstraction over the socio-indexical information stored in listeners' episodic memory, Chapter 3 pursued the second research question about the representation of speech categories in the *absence* of episodic memories for a particular type of speech. Without knowledge of a non-native speaker's speech categories accumulated from direct speech exposure, listeners could only use their (grounded or ungrounded) beliefs about those categories to recognize speech. Beliefs about a speaker or speaker's group are called stereotypes, and they are formed from indirect sources such as socio-cultural learning or hearsay. Chapter 3 examined the effects of expectations (as a product of stereotypes) in speech perception, in a similar manner as the study conducted in Niedzielski (1999). However, as listeners' previous exposure to the target speech was not controlled for in Niedzielski (1999), those findings are inconclusive and compatible with either episode matching (as argued in Hay & Drager, 2010, and Hay et al., 2006a) or stereotypes as the source of the effects (as argued in Niedzielski, 1999). Chapter 3 focused on stereotypes only, by involving listeners who lacked previous exposure to the speech of the target speaker's group. Within the episodic approach to speech perception, no change in the perception of speech categories was predicted in response to different types of cues about the speaker (i.e., with cues about the speaker's accent vs. without cues). However, such a change in perception could be a possibility in Bayesian approaches (e.g., Clayards et al., 2008; Kirov & Wilson, 2013; Kleinschmidt & Jaeger, 2015; K. Nielsen & Wilson, 2008; Norris & McQueen, 2008; Norris et al., 2016; Wilson & Davidson, 2013) as they are generally uncommitted about the specific source (i.e., knowledge/belief). Again, the vowel categorization paradigm was used to collect data for these predictions. Australia-born participants, who reported having no previous exposure to Vietnamese-accented English, listened to Vietnamese-accented English /hVdə/ tokens and categorized the monophthong /V/ in /hVdə/ by selecting a reference word on the grid containing the same monophthong as /V/. Participants in the treatment condition were told about the Vietnamese accent of the speaker whereas the control participants were not. Bayesian inference was used to discover patterns in the data, with participants' accuracy data from the vowel categorization task serving as the input for the posterior probability term in the Bayesian formula; mixed-effects modeling was then used to confirm the significance of the patterns found. Results indicated that a shift in perception did occur in response to cues about a speaker's accent even when listeners did not have previous exposure to the accent. More specifically, Vietnamese-accented English vowels were expected less often at the extreme edges of the Australian-accented English vowel space, and more often towards the center. This finding supported the prediction made within the Bayesian framework, as well as providing clear evidence for the claim that socio-indexical information could also be abstracted in spite of a lack of direct speech evidence. Overall, the results added support for the overarching hypothesis that listeners make use of abstract information in perceiving speech even when the abstraction draws upon beliefs rather than on direct experience.

Built on the theoretical issues in Chapter 2 (listeners' knowledge about social categories, developed via direct experience) and Chapter 3 (listeners' stereotypical expectations about social categories, developed via indirect sources such as socio-cultural learning and hearsay), Chapter 4 responded to the third research question about how these common listener factors work in tandem in speech perception. It also incorporated one further listener factor: listeners' feelings of like/dislike towards a speaker's group, or prejudice, which could be developed via direct or indirect sources. Findings in the literature to date did not lead to a prediction regarding a three-way interaction among experience, expectations, and prejudice. However, they did support a prediction of a two-way interaction between expectations and prejudice (as inferred from N. Nguyen et al., 2015), as well as the predictions about the main effects of experience (similar to McGowan, 2015), expectations (similar to Hay & Drager, 2010; Hay et al., 2006a; and Niedzielski, 1999), and prejudice (similar to N. Nguyen et al., 2015) in speech perception. As in Chapters 2 and 3, the categorization paradigm was also used in Chapter 4 to collect perceptual data. Australia-born participants (74% having no previous exposure to Vietnamese-accented English) listened to /hVdə/ tokens in Vietnamese-accented English and categorized /V/ in /hVdə/ by selecting a grid word with the same monophthong. Data for listener factors were either collected via a survey question, a scale, or manipulated in the categorization task. Experience/exposure data, for example, were collected via a simple survey question "Do you have experience with the following accent?" Expectations were manipulated via cues about the Vietnamese accent: Treatment participants were given those cues while control participants were not. Lastly, prejudice data were collected via the Scale of Anti-Asian American Stereotypes (SAAAS; Lin et al., 2005), modified for the Australian context (N. Nguyen et al. 2015). Data analyses built from the main effects of expectations, to the interaction between expectations and experience, and then to the interaction among expectations, experience, and prejudice. Results revealed a complex picture of speech perception, with listeners' prejudice towards a speaker's group correlating with their perception of the speaker's vowel categories in ways that interact with their expectations about and previous exposure to the speaker's accent. More specifically, listeners' accuracy in perceiving the Vietnamese-accented English vowels increased when listeners expected to hear the Vietnamese accent but (a) lacked previous exposure to it and had negative prejudice towards Asians, or (b) had previous exposure to the accent and positive prejudice towards Asians. This finding of the three-way interaction among experience, expectations, and prejudice was not predicted, but the relationship between listeners' feelings of like/dislike Asians and their perception of Vietnamese-accented English vowel categories was predicted, and potentially provides further evidence for the abstraction of socio-indexical information (in this case, listeners' prejudice towards Asians). The findings in Chapter 4, therefore, provided some additional support for the hypothesis that listeners use available abstract information of socio-indexical nature in perceiving speech.

To summarize, across these three chapters, evidence for the relevance of abstract social information was provided and the overall hypothesis of the thesis about listeners' ready use of abstract socio-indexical information was supported. Chapter 2 possibly provided preliminary evidence for listeners' abstraction of social categories of accent types and age cohorts into superordinate socio-indexicality categories, which in this thesis are called "more socio-indexical" versus "less socio-indexical." It also suggested that listeners used these highly abstract socio-indexicality categories, consciously or unconsciously, in working out their tolerance of phonetic variability in speech perception. Chapter 3 suggested that listeners used their (grounded or ungrounded) beliefs (i.e., stereotypes) about speech categories, formed in the lack of direct speech evidence, to build representations of those speech categories. Lastly, Chapter 4 brought abstract social categories formed from direct speech experience and those formed from indirect sources together, and potentially suggested that they all play a role in speech perception in a complex manner. In the next section, the findings from these chapters will be evaluated against findings and ideas reported in the literature.

2 Evaluations of findings from the present study: Theoretical advances

2.1 Abstractionist approaches to speech perception

Similar to the findings of previous studies that demonstrated the role of speaker's socio-indexical information in speech perception (e.g., Babel & Russell, 2015; Johnson et al., 1999; Hay & Drager, 2010; Hay et al., 2006a; Hay et al., 2006b; Lindemann, 2002; Mack & Munson, 2012; McGowan, 2015; Niedzielski, 1999; Rubin, 1992), findings in this thesis, which readily capitalized on listeners' use of socio-indexical information, warrant reconsideration of findings from studies within the abstractionist approach. For example, Miller (1953) controlled for the variability in the speech signal, including the variability indexed by social categories, by using only synthetic sounds as stimuli to investigate the importance of phonetic factors such as f_0, formant amplitude, and the number of formants in the perception of vowels. Findings suggested that listeners' phonetic evaluation was affected by high f_0 (most prominent in the center region of the American English vowel space), the addition of F_3 (at least for the front synthetic, two-formant vowels), and formant amplitudes. However, the applicability of these findings is very limited as this study disregards phonetic variability coming from different social sources; in real-life speech situations, listeners cannot disregard such variability (e.g., Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). In fact, this thesis provided strong evidence that listeners not only use socio-indexical information in speech perception but they readily use abstractions over such information whenever it is available. In everyday speech situations, socio-indexical information, together with some level of abstraction of socio-indexical information such as abstract social categories, is always tied to linguistic information (see Pierrehumbert, 2006). As a result, findings from studies stripping off socially meaningful phonetic variability, as in Miller (1953), appear not to be very informative about real-world perception of natural speech variability. Schatz (1954) used natural speech stimuli in the design, but the study did not take into account the American speaker's social group in relation to the listeners. Nor did it control for factors related to the listeners such as their experience with the speech of the American speaker's social group, expectations/stereotypes about that particular type of speech, or feelings towards the speech, all factors that could potentially affect how Schatz's listeners would process the American speaker's speech. As the experiments in this thesis (especially Chapter 4) showed, the listener-speaker dynamic in speech perception is complex and requires thorough consideration. As a result, again, Schatz's findings are not very informative when it comes to speech situations with real talkers.

The above arguments have some implications for the abstractionist approach. As a speech perception theory, the abstractionist approach advocated a sole focus on linguistic information in speech processing, while socio-indexical information and the speaker's personal information were considered unimportant (e.g., Joos, 1948; Liberman et al., 1967; Miller, 1953; Nearey, 1989; Shankweiler et al., 1977). A number of speech perception models duly reflected these speech processing principles, as described in, for example, Miller (1984), Syrdal (1985), and Nearey (1992). The approach has had its own merits by being economical and capable of explaining some speech perception phenomena such as how listeners could recognize sounds quickly and accurately in the face of between-speaker variations (e.g., Abramson & Cooper, 1959; Peterson & Barney, 1952; Verbrugge et al., 1974). However, more and more empirical evidence for the role of socio-indexical information in speech perception has accumulated (e.g., Babel & Russell, 2015; Johnson et al., 1999; Hay & Drager, 2010; Hay et al., 2006a; Hay et al., 2006b; Lindemann, 2002; Mack & Munson, 2012; McGowan, 2015; Niedzielski, 1999; Rubin, 1992). The abstractionist approach cannot explain perception across the many varied speech situations in real life, which heavily involves the consideration of social factors. Furthermore, the abstractionist approach does not have a mechanism to explain listeners' use of socio-indexical information observed in the findings of this thesis.

2.2 Episodic approaches to speech perception

This thesis was informed by previous findings that listeners make use of socio-indexical information and social categories in perceiving speech (e.g., Babel & Russell, 2015; Johnson et al., 1999; Hay & Drager, 2010; Hay et al., 2006a; Hay et al., 2006b; Lindemann, 2002; Mack & Munson, 2012; McGowan, 2015; Niedzielski, 1999; Rubin, 1992). However, it was mostly inspired by the findings in Niedzielski (1999) and Hay and colleagues (Hay & Drager, 2010; Hay et al., 2006a), as evaluated against the thesis findings in turn below.

To recap, in Niedzielski's (1999) study two groups of listeners from Detroit, Michigan listened to the same Detroit speaker but perceived the vowel stimuli differently in a vowel matching task, depending on the cues about the speaker that they were told to expect: The group that expected a Detroit speaker perceived standard American-accented English vowels whereas the group that expected a Canadian speaker perceived raised vowels (as the vowels truly sounded). The explanation of the perceptual shift proposed in Niedzielski (1999) speech stereotypes. In Niedzielski's (1996) language attitude survey, Detroit listeners were confident to claim that Detroiters spoke standard American-accented English, whereas Canadians spoke a version of English with raised vowels. The results presented in Chapter 3 are consistent with that account, as they provided clear evidence that listeners' stereotypes about a speaker's speech do shift the way listeners perceive that speech, although they also acknowledges the role of speech episodes in speech perception. In fact, this thesis has now provided a compelling test case for when stereotypes have the clearest impact in speech perception, without any interfering role from speech episodes: the *absence* of previous speech exposure. The listeners in the present research were carefully selected in terms of their lack of previous exposure to the accent in the speech stimuli. In Niedzielski's (1999) case, however, the shift in perception could possibly have come from both sources: stereotypes and speech episodes, as Niedzielski's listeners' previous experience with "Canadian English" had not been controlled for. The close proximity between Detroit/Michigan and Canada makes the explanation about speech episodes convincing, but the explanation involving stereotypes is also very likely. As laid out in the thesis, stereotypes are essentially abstractions over socio-indexical information accumulated not through direct evidence (at least not initially), but through indirect sources such as socio-cultural learning or hearsay. Consequently, the two explanations for the findings in Niedzielski (1999) ultimately boil down to listeners' use of socio-indexical information: one is abstraction over speech experiences (speech episodes), and one is abstraction over no or few speech experiences but rather over information from indirect sources instead (speech stereotypes).

Similarly, Hay and colleagues' findings (Hay & Drager, 2010; Hay et al., 2006a) could also be discussed in relation to listeners' use of socio-indexical information. The studies by Hay and colleagues obtained similar results as those in Niedzielski (1999). New Zealand listeners responded to the same vowel matching task with more Australian-like vowels when they received test sheets with the word "Australian" (vs. "New Zealander") written on them (Hay et al., 2006a), or when they were in test rooms where they were surreptitiously shown stuffed toy kangaroos and koalas (vs. kiwi birds) (Hay & Drager, 2010). Like Detroit/Michigan and Canada in Niedzielski's (1999) study, New Zealand is in close proximity to Australia. As a result, it is most likely that New Zealanders have had some level of exposure to Australian speech, and stored speech episodes indexed with the label "Australia" accordingly. The possibility of stereotype effects was not convincing to Hay and colleagues initially in the more explicit design replicating Niedzielski's (1999) study (Hay et al., 2006a) with the word Australian/New Zealand written on the test sheets. However, the effects became clearer in the more subtle design in their follow-up study (Hay & Drager, 2010) with the stuffed toy kangaroos and koalas referring to Australia, and kiwis referring to New Zealand, surreptitiously shown in the test rooms. Stereotype effects in speech perception were then duly acknowledged in Hay and Drager (2010), together with a discussion of the role of speech episodes. What was also new in Hay and Drager (2010), as opposed to Hay et al. (2006a), is the New Zealand versus Australia sporting rivalry, which was briefly discussed as the source of possible prejudice effects which could differentiate performance of male from female New Zealand participants. Similar to stereotypes, which are cognitive abstractions over socio-indexical information, prejudice is also laid out in this thesis as abstractions over socio-indexical information, but in this case they are affective abstractions. In other words, there are three ways in which listeners could use socio-indexical information that Hay and Drager (2010) mentioned: speech episodes – cognitive abstraction over speech experiences; speech stereotypes – cognitive abstraction over no or few speech experiences; and group prejudice – affective abstraction over either group experiences or no/few group experiences. This thesis further demonstrated these three uses of socio-indexical information in speech perception, particularly focusing on speech stereotypes and group prejudice, suggesting a complex picture to consider in future exploration of speech perception dynamics.

The thesis findings, as compared against the above evaluations of the findings in Niedzielski (1999) and Hay and colleagues (Hay & Drager, 2010; Hay et al., 2006a), carry some consequences for the episodic approach to speech perception. The episodic theory of the lexicon has speech exposure/experience/direct observations as its central tenet and revolves around speech episodes (e.g., Goldinger, 1996; Hintzman, 1986;

Jacoby, 1983). Speech is perceived by comparing the properties of the speech signal with those of stored episodes, activating the episodes with similar properties, and summing the properties of all the activated episodes to classify the stimulus. Although abstract representations were considered unnecessary in the earlier versions of the episodic approach (e.g., Hintzman, 1986; Pierrehumbert, 2001), they were not discarded but were acknowledged to possibly form at the later stimulus classification stage, rather than earlier. Models that reflect these episodic speech processing principles include MINERVA 2 (Goldinger, 1998; Hintzman, 1986), Johnson's (1997) model, and Pierrehumbert's (2002) exemplar models. Episodic theory, with its unique framework for representing all the types of information conveyed across the speech signals a listener has encountered, especially socio-indexical information, can account for a large number of speech scenarios in real life. However, it can only account for some, but not all, of the perceptual findings in this thesis. The theory acknowledges the existence of abstract social categories (Pierrehumbert, 2006), and therefore can account for how listeners in this thesis abstracted socio-indexical information into social categories. However, it is currently uncommitted about the possibility of higher levels of abstractions over social categories such as socio-indexicality categories, which was demonstrated in Chapter 2 to modulate the relationship between phonetic variability and speech perception behaviour. It would also fail to explain how socio-indexical information acquired through indirect sources, in the *absence* of speech exposure/experience/direct observations, could be abstracted into cognitively abstract social categories (stereotypes) and affectively abstract social categories (prejudice). Stereotypes and prejudice were shown in Chapters 3 and 4 to play a role in speech perception. The central tenet of the episodic approach cannot be changed, but the theory could be further expanded from its current claims to explicitly address the existence of more levels of abstractions over social categories. That way, the explanatory capacity of the approach would be enhanced, and more speech perception scenarios could be tested and accounted for. Just as abstraction over linguistic categories facilitates generalization to new words, abstraction over social categories may guide our expectations about new speakers.

2.3 Hybrid approach to speech perception

The thesis findings also supplement findings from studies that show evidence for both abstract categories and specific episodes in speech perception and word recognition (e.g., Cutler & Weber, 2007; Goldinger, 2007;

McQueen et al., 2006; K. Nielsen, 2011). One example in support of the hybrid approach is the study by McQueen and colleagues (2006), which was designed to demonstrate the function of phonological abstraction alongside the somewhat accepted role of speech episodes, by showing the generalization of listeners' phonetic retuning of [f]/[s] to new words. Socio-indexical cues were given in this study, with Dutch listeners being told (via written instructions) to expect either real Dutch words or non-words on hearing the stimulus list. The study results showed that listeners retuned their [f]/[s] perception and generalized the learning to new words, which suggested that there must at least be abstractions over phonetic details in speech episodes prior to lexical processing for phonetic retuning to be generalized. This suggestion acknowledged the importance of both abstraction and specificity in speech processing. This thesis did not concern phonetic details or their abstractions per se, but it did reach similar conclusions about hybrid accounts with respect to socio-indexical information, supporting the importance of socio-indexical details as well as (highly) abstract social categories in guiding speech behaviour. As a result, the essential difference between the findings of McQueen et al. (2006) and the thesis findings is the type of information in the speech signal that served as the object of the study: linguistic information in their study and socio-indexical information in this thesis. This essential difference has some implications for speech perception theories, as outlined below.

In light of the thesis findings, theoretical suggestions can be made within and beyond the hybrid approach to speech perception. Results from the thesis lend additional support to the hybrid approach, as argued above. However, the hybrid approach has been discussed in previous works solely in relation to linguistic information, or more specifically phonetic information, whereas evidence for the hybrid approach in this thesis relates to socio-indexical information. Therefore, an expansion of the approach to cover socio-indexical information (or better yet both socio-indexical information and speaker's personal information) would become necessary. In addition, although the hybrid approach is more flexible than the episodic approach, by virtue of officially acknowledging the role of abstraction as well as specificity in speech perception, by definition it is a combination of the abstractionist and episodic approaches. As discussed earlier, the abstractionist approach does not concern storage of social categories in the mind and the episodic approach only abstracts social categories over speech exposure/experience/direct evidence. Consequently, the combination of these two approaches, or the hybrid approach, still cannot explain the effects of socio-indexical information abstracted over no or few speech episodes such as stereotypes and prejudice, as demonstrated in Chapter 3. This is another theoretical area in which the hybrid approach could be modified to be capable of explaining speech perception scenarios without limits.

Another alternative is to contemplate an integrative theory of speech perception, which is built on the hybrid approach but which allows for abstract categories to be built from indirect sources. This theory would then be capable of explaining how listeners deploy (highly) abstract knowledge of all the information carried in the speech signal, regardless of whether this abstract knowledge is built on previous exposure to those specific speech characteristics or indirect sources such as socio-cultural learning or hearsay. Such an integrative theory of speech perception could be developed in a Bayesian framework, which makes use of all the factors known to belong to listeners, irrespective of how these factors arose in the first place. Such a theory would also have adequate power to explain the role of both abstraction and specificity in speech perception. As a result, although quite similar to the hybrid approach, that integrative theory of speech perception would be much more powerful and, therefore, would be able to explain more natural speech perception scenarios that happen in everyday life.

At this point, it is necessary to re-evaluate the appropriateness of Bayesian inference as a tool in the investigation of listeners' use of socio-indexical information. In this thesis, Bayesian inference was used to model listeners' expectations (Chapter 3). Modelling expectations as Bayesian priors revealed patterns that account for how socio-indexical abstractions formed from indirect sources are used in speech perception. The approach combines all the sources of information (e.g., speech signals, listeners' prior knowledge/beliefs about the speech) available to listeners, whether the information is fully specified or abstract, to optimize listeners' speech perception behaviour, described by the "posterior probability" term. Depending on particular speech contexts, either the "prior probability" term or the "likelihood"/"signal specificity" term would have more influence on the "posterior probability" term (e.g., Norris & McQueen, 2008). In the case of Vietnamese-accented English, the speech signals were ambiguous enough to the listeners, who were selected on the basis of having no previous exposure to the Vietnamese accent. As a result of this ambiguity, listeners would have had to rely more on their prior knowledge/beliefs about Vietnamese-accented English vowel categories to help them perform the categorization task. However, due to the lack of direct experiences (or so-called knowledge in the context of this thesis), listeners' prior probabilities would only be calculated from their (most likely ungrounded) beliefs about the accent, which were built on indirect sources such as socio-cultural learning or hearsay. These (possibly inaccurate) beliefs, or stereotypes, are examples of abstract information about speaker characteristics formed without direct speech exposure. Prior probabilities calculated from stereotypes would most likely not reflect the real probabilities of these categories. Indeed, the analysis in this thesis revealed a shrinkage of listeners' perceptual space. This is a very clear demonstration of the appropriateness of using Bayesian inference to model listeners' use of socio-indexical information in perception.

Bayesian inference was not necessary for data analyses in Chapters 2 and 4, but the relevant details that could facilitate Bayesian inference were all provided in the set-up of those chapters. The predictions in Chapter 2 were built on an intuition about listeners forming different phonetic distributions, corresponding to different speech situations, for the same phonological category, called the multiple-distribution hypothesis. This multiple-distribution hypothesis was formally implemented in Kleinschmidt and Jaeger's (2015) ideal adapter framework, which operates on Bayes rule. In addition, the Bayesian approach can model the modulation of the fully specified prior knowledge in Chapter 2, which was probably initially abstracted from individual talkers and then into social categories of accent types and age cohorts, and then further abstracted into socio-indexicality categories, as prior probabilities. Given the speech scenario where Australian-accented English listeners were presented with Australian-accented English stimuli, the speech signals may not have been that ambiguous. As a result, listeners may have weighted such prior probabilities and cues from the stimulus signals equally in categorizing the Australian-accented English vowels. The application of Bayesian inference in Chapter 4 would be a mixture of its use in Chapters 3 and 2 as Chapter 4 deals with both fully specified prior knowledge (experienced with the Vietnamese accent) and abstract beliefs (inexperienced with the Vietnamese accent). In summary, Bayesian inference appears to be very appropriate as a statistical modeling method to investigate listeners' use of socio-indexical information.

Bayesian inference could even be applied to explain the role of speech exposure/experience/direct evidence in the formation of speech stereotypes, on the assumption of the multiple-distribution hypothesis formally implemented in Kleinschmidt and Jaeger (2015). Take the stereotype that Vietnamese-accented English is difficult to understand (suggested in Chapter 3) as an example. In Kleinschmidt and Jaeger's (2015) proposal, listeners form new (and accordingly narrow) phonetic cue distributions for a phonological category according to specific speech situations (e.g., accents, age, and social class), compared to the wide distribution that would result from lumping phonetic experiences into a single category. Then, depending on the specific speech situation, listeners access the situationally appropriate distribution to help them perceive the correct category. However, if the distribution they deem "situationally appropriate" turns out to be the wrong one, then that would hurt their category perception. Similar to when listening to their native Australian-accented English vowels, the listeners who listened to Vietnamese-accented English vowels also tried to access appropriate distributions for these accented vowels. However, if those listeners had not encountered the accent before, they had not yet formed such distributions for the Vietnamese-accented vowels. There were two possible solutions for these listeners at this point. The first solution could be that they chose to perform the much more demanding task of inferring both the category and the parameters of the underlying distributions (i.e., the mean and variance). This inference task is captured in formula (5) by Kleinschmidt and Jaeger (2015), where x represents an acoustic cue, c represents a category, μ represents the mean of the underlying distribution, and σ the variance:

$$p(c, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 | x) \propto p\left(x \middle| \mu_c, \boldsymbol{\sigma}_c^2\right) p(c) p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$$

As the formula states, the accuracy of listeners' categorization of the acoustic cue x ($p(c, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 | x)$) relies on (1) the likelihood that the acoustic cue x belongs to the category c and falls within the distribution of c formed by the distribution parameters μ_c and σ_c^2 ($p\left(x \middle| \mu_c, \boldsymbol{\sigma}_c^2\right)$), and (2) the prior beliefs about the probability of the category and of the distribution parameters ($p(c)p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$). The task of inferring $p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ may be further complicated by the fact that the distributions of the Vietnamese-accented English vowels were shifted compared to the native norms (Ingram & Nguyen, 2007; T. Nguyen & Ingram, 2004; Zielinski, 2003, 2006, 2008), making the task of tracking (1) much harder for listeners. That would explain listeners' degraded categorization performance on the Vietnamese-accented vowels compared to their performance on their native vowels, which might have led to the belief that Vietnamese-accented English is difficult to understand. The second solution could be that, although these listeners had not yet formed appropriate distributions for the Vietnamese-accented English vowels, they might have had appropriate distributions for the vowels in the other foreign accent/s that they had encountered before (see also Bent & Holt, 2017, and McGowan, 2015), or some distributions that they mistakenly assumed to be appropriate for Asian-accented English vowels in general. These distributions would be inappropriate and inadequate for the task of categorizing Vietnamese-accented English vowels, as the phonetic distributions of the vowel system in every language are different (see Yi, 2010 for the Chinese vowel system; Thompson, 1987 for Vietnamese; and Cox, 2006 for Australian English). Assessing these distributions in the hope of accurately perceiving Vietnamese-accented English vowels would only hurt listeners' perception. This could possibly be another explanation regarding the genesis of our listeners' apparent belief that Vietnamese-accented English is difficult to understand. Whichever of the two solutions is correct, the belief that Vietnamese-accented English is difficult to understand would be spread from the hypothesized listeners above to their fellows who are also naïve about the accent, and thus, would also be willing to perpetuate such a bias and communication difficulty to more others.

To sum up, in this section, the hybrid approach has been demonstrated to be a promisingly powerful approach in explaining a variety of speech perception events. However, to explain the results of this thesis, it needs to be expanded to explicitly include all types of information in the speech signal as study objects, not just linguistic information. It is also necessary for the approach to be more flexible so that it can accept not only speech episodes but also indirect sources from which to form abstract categories. In order to achieve these improvements, it is suggested here that a modified hybrid account should be developed according to Bayesian inference, as it is an appropriate tool in the investigation of listeners' use of socio-indexical information in speech perception.

2.4 Other theoretical issues

Having established that listeners do abstract over socio-indexical information, it is interesting to consider whether the abstraction is conscious or unconscious. One answer to this question could be inferred from Pierrehumbert's discussion about whether social categories are "external (representing observations of the scientist about the population)" or "internal (imputed to the minds of individual speakers)" (Pierrehumbert, 2006, p. 527). Social categories could be argued to be internal, that is, represented in the cognitive system of individual speakers; or they could be external, that is, available for notice and awareness. Pierrehumbert's concepts of "external" and "internal" representations correspond in many ways to the concepts of conscious and unconscious processes of abstracting socio-indexical information herein discussed: Speakers are conscious of social categories when these categories are represented externally, or when speakers notice that they have formed these social categories; and vice versa, speakers are unconscious of social categories when these categories are represented internally, or when speakers are not aware that they have formed these social categories. Speakers in one speech context are also listeners in another context, and it could be reasonable to expect them to make use of these same processes as listeners. Therefore, it can be inferred that abstractions over socio-indexical information may happen consciously when listeners notice that they are abstracting properties of their observations of social groups into social categories and more highly abstract categories. Such abstractions would happen unconsciously when listeners are not aware of the abstracting process. As abstractions over phonetic information in a naïve speaker happen unconsciously, and abstractions over socio-indexical information are analogous to abstractions over phonetic information (Pierrehumbert, 2006), it would be logical to infer that abstractions over socio-indexical information also happen unconsciously. This idea is indeed suggested in Pierrehumbert (2006), with some caution that there are still cases where abstractions over socio-indexical information could happen consciously. Whether or not abstractions over socio-indexical information happen consciously is still not yet resolved, and this interesting question should be taken up by future research.

It is also interesting to compare the relationship between prejudice and speech perception against the relationship between stereotypes and speech perception. It can be seen from the thesis results that the relationship between listeners' prejudice and speech perception depends on other factors (e.g., listeners' expectations about the speaker's accent and their previous exposure to it) in a complex fashion. Meanwhile, a straightforward relationship between listeners' stereotypical expectations and speech perception was observed. This might reveal something about the role of the affective component of the listener-speaker relationship relative to the role of the cognitive component in speech perception. As noted in Islam and Jahjah (2001), the attitudes of young Australians (18-29 years old) in New South Wales towards Asians were best predicted by perceived threat, stereotypes (or the cognitive component of attitudes between groups), and prejudice (or the affective component). The studies in this thesis were also conducted in New South Wales and also on young Australians (90% of the participants taking part in the studies in Chapters 2, 3, and 4 are within 18-29 years old). As a result, it is no surprise to find a more stable effect of stereotypical expectations, compared to the rather fragile role of prejudice, as these expectations more or less indicate the cognitive component of the listener-speaker relationship in speech perception, and the cognitive component has been shown to express attitudes better than the affective component (prejudice) (Islam & Jahjah, 2001).

2.5 Summary

Overall, there is now plenty of empirical evidence in speech perception research that shows the effects of socio-indexical information on speech perception. Such effects cannot be explained by the abstractionist approach, in which socio-indexical information was believed to only signal the need for normalization and, therefore, was unimportant for speech. The episodic approach, which is built on episodic memory, provides a framework for representing phonetic details, stored in episodic memory as episodes and indexed with all types of information conveyed in the speech signal. As a result, it has the representational ability to explain the differences in listeners' perceptual performance when different socio-indexical cues were provided to them. However, due to its reliance on speech exposure and its current acknowledgement of only one level of abstraction over socio-indexical information, it can only explain some results obtained in this thesis and would need to be expanded to be able to account for more speech scenarios. The hybrid approach acknowledges both abstract categories and specific episodes in speech perception and word recognition and, thus, can also account for the effects of socio-indexical information found in speech perception studies. However, it also faces similar issues as the episodic approach and cannot explain all the findings in this thesis, unless (1) it is explicitly stated to cover socio-indexical information and speakers' personal information as its research object, and (2) it allows for not only speech episodes but also indirect sources to build abstract categories from. The process of building abstract categories could be either conscious or unconscious, which might have implications for future research. And finally, in the relationship with speech perception, cognitive abstractions over socio-indexical information appear to have clearer effects than affective abstractions over the same information.

3 Evaluations of findings from the present study: Methodological advances

In addition to theoretical advances, the thesis also provides several methodological advances in the areas of experimentation and data analysis.

The first development is the attempt to quantify formant variability while taking the magnitude of formant means into account. Formant variability is traditionally estimated by how spread-out the tokens of a vowel category are in F_1-F_2 vowel space -- the so-called standard deviation method (Lobanov, 1971). Based on observed correlations between formant means and formant standard deviations (Eguchi & Hirsh, 1969; Kent, 1976; Lee et al., 1999; N. Nguyen & Shaw, 2014), Chapter 2 (and Appendix 3) proposes a method of quantifying formant variability by calculating the residuals from regression analysis between formant means and formant standard deviations. Positive residuals indicate that the vowel is variable in relation to the magnitude of the mean, and negative residuals indicate that the vowel is stable. This method of quantifying formant variability, referred to as the residual method in the chapter, was shown to provide a better estimate of perceptual behaviour, and to be informative of meaningful social variation. Future research should take this residual method into account in developing more sophisticated methods for quantifying formant variability.

The second development concerns the quantification of expectations via Bayesian formalism. Chapter 3 employs Bayes' rule to compute listener expectations via the "prior probability" term. In the context of Chapter 3, this term expresses expectations about a Vietnamese-accented speaker's English vowels before observing the vowel stimuli (i.e., their prior probabilities), probably resulting from listeners' stereotypes towards the speaker's accent. These prior probabilities were calculated from listeners' probabilities of perceiving the vowels accurately ("posterior probability") after observing the vowel stimuli ("likelihood", or "signal specificity" in the thesis). This method of quantifying expectations was sensitive enough to detect changes in listeners' perceptual space, which would not have been detected otherwise, and to find patterns of perceptual shift for statistical analyses. As a result, it is recommended that the quantification of expectations via Bayesian formalism should be integrated in upcoming studies regarding expectation effects for deeper understanding of perceptual data.

Another development brought in by Chapter 4 is the quantification of prejudice. Chapter 4 uses an established scale from social psychology called the Scale of Anti-Asian American Stereotypes (SAAAS) modified for the Australian context to quantify listeners' prejudice towards speaker groups. On the basis of a contemporary attitudinal model in Social Psychology called the Stereotype Content Model, SAAAS quantifies prejudice (indirectly) via stereotype items. As mentioned in Appendix 5, social norms have changed such that participants are no longer willing to endorse a blatant attitude (as in Katz & Braly, 1933) but opt for a more subtle expression instead (as in McConahay, 1986). As a result, the investigation of prejudice is a very sensitive type of investigation that requires indirectness or implicitness in the method to be able to collect honest responses. SAAAS is in keeping with contemporary approaches to assess prejudice, which favor indirect or implicit measures (over direct or explicit measures) in order to minimize participants' desire to look good socially (social desirability bias) for such sensitive topics (Fazio & Olson, 2003; Uhlmann, Leavitt, Menges, Koopman, Howe, & Johnson, 2012). Although not eliminating social desirability responses, the indirectness of SAAAS helps minimize it to a large extent, and should be considered in future prejudice investigation.

The thesis also highlights the importance of careful selection of listener and speaker groups in accordance with the research goal. The listener group was selected to be Australian-accented English and the speaker groups to be Australian-accented English and Vietnamese-accented English for the following reasons: (1) To the Australian-accented English listeners in the thesis, Australian-accented English stimuli contain patterns of variation that they have had a large amount of exposure to, which was appropriate for the investigation of abstractions over speaker characteristics formed from direct speech exposure. (2) Vietnamese-accented English stimuli contain variation patterns that they may not have had exposure to (but possibly had heard about), which was appropriate to explore abstractions over speaker characteristics formed from indirect sources such as beliefs coming from stereotypes and prejudice. Episode matching can be efficient at explaining the patterns of variation that listeners have had exposure to, but the variation patterns that listeners have not yet had exposure to require some abstraction and generalization modelled in the Bayesian framework. These different variation patterns, however, were not separate in previous studies (e.g., Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). By zooming in on these different variation patterns in separate chapters, this thesis has demonstrated that it is more important than previously thought to be careful about selecting listener and speaker groups.

To summarize, several methodological contributions resulted from the thesis such as the quantification of formant variability, expectations, and prejudice. The different groups of listeners and speakers in the thesis also contribute to a representation of different patterns of phonetic and social variation, which suits the experimental purposes. In the last section, ideas for follow-up research will be suggested.

4 Limitations and Future directions

In this section, several design changes are suggested in light of the findings in this thesis, with each experimental chapter discussed in turn.

In Chapter 2, vowels classified as carrying more socio-indexical information were found to have a different relationship between formant variability and vowel categorization than those classified as carrying less socio-indexical information. This meaningful pattern of interaction between socio-indexicality as a way of grouping vowels and formant variability surfaced strongly, even with a correlational design. Therefore, the role of socio-indexicality in speech perception would be expected to show up more clearly and definitively with the manipulation of socio-indexicality. Listeners could be exposed to vowel stimuli of different manipulated distributions, following the methodology in the study by Clayards and colleagues (2008), with or without changes in speaker identity. Second, the current formant variability index quantified by the residual method was shown in this thesis to provide a good estimate of the phonetic variance in the underlying distribution. This method of quantifying formant variability worked despite the fact that it does not take into account the F_1-F_2 covariance due to lack of access to individual F_1 and F_2 values in the corpus used. Future research should use a corpus or published data without such limitations and refine the method by taking the F_1-F_2 covariance into account. The method could possibly be expanded into a vowel variability index, not just formant variability, by incorporating the variability from other vowel properties (e.g., duration). These two changes alone would considerably improve our understanding of listeners' use of socio-indexical information in speech perception.

Chapter 3 provided strong results demonstrating stereotype effects in speech perception (i.e., listeners who reported having no previous exposure to the Vietnamese accent expected to hear vowels towards the center of the vowel space more often when told to expect the Vietnamese accent). The content of the speech stereotypes should be further explored. Chapter 3 extrapolates that listeners formed speech stereotypes from indirect sources, but did not specifically survey those stereotypes or their sources. What are those speech stereotypes about the Vietnamese accent specifically, and where exactly do they come from? Chapter 3 addressed one such stereotype -- listeners believed vowels produced in the Vietnamese accent would be difficult to perceive – and suggested that this belief could come from socio-cultural learning or hearsay. Follow-up studies could administer qualitative pre-tests to explore speech stereotypes about accents in more detail in the form of open-ended questions. Then, based on the answers from those qualitative pre-tests, certain stereotypes could be selected along with different accents for experimental manipulation, in order to explore whether the shrinkage of the perceived vowel space is just about difficulty and whether the pattern of shrinkage would remain the same for cues to other accents.

The informative results in Chapter 4 revealed a complex picture of the different ways listeners use socio-indexical information in speech perception that encourages further investigation. In Chapter 4, a three-way interaction was found among stereotypical expectations about a speaker's accent, experience with that accent, and prejudice towards the speaker group. However, only a tentative conclusion was reached because only expectations were manipulated in that study, and also, the study may have been underpowered. To more directly probe into that three-way interaction and examine the modulation effects of each socio-indexical use, future studies should consider manipulating listeners' experience with the stimulus speech as well, and the information about listeners' experience could perhaps be collected earlier in a separate study to avoid demand characteristics that arise when listeners become suspicious of the experimental purpose (and therefore act accordingly). SAAAS (or other indirect measures of prejudice) could possibly be administered in this separate study also. Then participants could be divided into two groups: with and without experience with the accent, and invited back for the expectation manipulation for each of the experience groups. If the three-way interaction persists when both expectations and experience are manipulated, conclusions about causality could be made on the effects of experience in speech perception in the context of other listeners' factors. It would be ideal if prejudice could be manipulated as well. However, as mentioned in the thesis introduction, manipulating feelings is not simple, and it poses ethical issues to manipulate prejudice towards existing groups. An easier experimental manipulation would be the manipulation of listeners' experience with the accent.

Another methodological point that could use some changes in future research is the quantification of prejudice using SAAAS in the context of a speech perception experiment. SAAAS was used to quantify prejudice indirectly through items with stereotype content. The practice of using stereotypes as a proxy to prejudice, as SAAAS does, is not novel. Stereotype endorsement has been used earlier to assist the quantification of prejudice (Dutta, Norman, & Kanungo, 1969) as stereotypes and prejudice are highly integrated in the perception of groups (e.g., Hamilton & Mackie, 1993; Rosenberg, 1960). This could help the investigation of prejudice as participants are not aware that their prejudice is being examined. Indirectness could help reduce bias in participants' responses that emerges from a desire to appear politically correct and/or socially desirable. However, straightforward interpretation of the results in terms of prejudice can only be accomplished by taking on the theoretical assumptions of the Stereotype Content Model (SCM), and concerns have been voiced regarding the validity of the scale, as to whether or not SAAAS really measures prejudice. Nevertheless, the SCM model that SAAAS was constructed from is widely accepted in social psychological research as a means of understanding attitudes between groups (e.g., Cuddy et al., 2009). In addition, the inference of prejudice from stereotype items following SCM tenets has been scientifically validated through a correlational study in Fiske et al. (2002), which SAAAS was built on. SAAAS has also contributed to providing meaningful results in the search for more in-depth understanding of complex speech perception dynamics (N. Nguyen et al., 2015; N. Nguyen et al., 2016). Follow-up studies are recommended to validate the indirect approach of SAAAS with different groups and converging evidence from different prejudice measures.

5 Concluding remarks

Overall, this thesis provided evidence for highly abstract social categories and support for the overall hypothesis that listeners use abstract social information in perceiving speech. These findings lend support to the episodic approach and the hybrid approach to speech perception, but suggest an expansion of these two approaches in order to allow for more levels of abstraction over socio-indexical information (for the episodic approach) and speech episodes as well as indirect sources for category abstraction (for the hybrid approach). The thesis has also contributed several methodological advances such as the quantification of formant variability, of stereotypical expectations, and of prejudice. Future work in this area should (a) manipulate experimental variables such as socio-indexicality categories and listeners' previous exposure to a speaker's speech, (b) use corpuses or published data with individual F_1 and F_2 values reported to be able to take into account the F_1-F_2 covariance in calculating the variability index, (c) explore speech stereotypes in more detail, and (d) administer a batch of prejudice measures at the same time to validate SAAAS as an indirect measure of prejudice. As mentioned earlier, in real-world speech comprehension, socio-indexical information is always tied to linguistic information. A hybrid-Bayesian, or integrative, theory of speech perception, as suggested from the findings in this thesis, would be able to take into account not only the influences of socio-indexical information on speech perception, but also the abstractions of such information and their potential interactions with the abstractions of linguistic information. These detailed and abstract forces can all be modeled as Bayesian priors, most likely in a complex fashion.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: University of Chicago Press.

Abramson, A. S., & Cooper, F. S. (1959). Perception of American English vowels in terms of a reference system. *Haskins Laboratories Quarterly Progress Report QPR–32*, Appendix l.

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, *116*(5), 3099–3107.

Agheyisi, R., & Fishman, J. A. (1970). Language attitude studies: A brief survey of methodological approaches. *Anthropological Linguistics*, *12*(5), 137–157.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Australian Bureau of Statistics. (2011, August). 2011 Census QuickStats – Greater Sydney (code 1GSYD). Retrieved from http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/1GSYD?opendocument&navpos=220

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. doi: 10.21500/20112084.807

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of the Acoustical Society of America*, *137*(5), 2823–2833.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 [R package version 1.0-6]. Retrieved from http://CRAN.R-project.org/package=lme4.

Bent, T., & Holt, R. F. (2017). Representation of speech variability. *Wiley Interdisciplinary Reviews: Cognitive Science*. Retrieved from Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com

Bernard, J. (1967*). Some measurements of some sounds of Australian-accented English* (Unpublished doctoral thesis). Sydney University, Australia.

Best, C. T., McRoberts, G. W., LaFleur, R. & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behaviour and Development*, *18*(3), 339–350.

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 345–360.

Best, C. T., Shaw, J. A., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., Mulak, K. E., & Wood, S. (2015). From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels. In *Interspeech 2015*. Dresden Germany, September 8–10.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 1–22. doi: 10.1016/j.cognition.2007.04.005

Brancazio, L., Best, C.T. & Fowler, C.A. (2006). Visual influences on perception of speech and nonspeech vocal-tract events. *Language and Speech, 49*(1)**,** 21–53.

Britain, D., & Newman, J. (1992). High rising terminals in New Zealand English. *Journal of the International Phonetic Association*, *22*(1–2), 1–11.

Bundgaard-Nielsen, R., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian-accented English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics, 32*, 51–67. doi: 10.1017/S0142716410000287

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*(1), 171–189.

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

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809. doi: 10.1016/j.cognition.2008.04.004

Cox, F. (1998). The Bernard data revisited. *Australian Journal of Linguistics*, *18*(1), 29–55. doi: 10.1080/07268609808599557

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*, 1–27. doi: 10.1159/000028438

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179. doi: 10.1080/07268600600885494

Cox, F. (2006a). Australian English pronunciation into the 21st century. *Prospect*, *21*(1), 3–21. Retrieved from

Macquarie University. Retrieved from clas.mq.edu.au/felicity/Papers/Prospect_Erratum_v1.pdf

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350. doi: 10.1017/S0025100307003192

Cox, F., & Palethorpe, S. (2008). Reversal of short front vowel raising in Australian-accented English. *Proceedings of Interspeech 2008*. Brisbane, Australia. 342–345. Retrieved from Macquarie University. Retrieved from clas.mq.edu.au/felicity/Papers/CoxPalethorpe_Interspeech08new.pdf

Cox, F. (2012). *Australian English Pronunciation and transcription*. Cambridge University Press.

Cuddy, A. J. C., Fiske, S. T.,Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., Palacios, M., Smith, V., Perez, R., Vala, J., & Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33.

Cunningham, U. (2009). Phonetic correlates of unintellibility in Vietnamese-accented English. Paper presented at the FONETIK 2009 The XXIIth Swedish Phonetics Conference. Retrieved from http://www.ling.su.se/fon/fonetik_2009/108%20cunningham_fonetik2009.pdf.

Cutler, A., & Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 43–48). Dudweiler, Germany: Pirrot.

Dutta, S., Kanungo, R. N., & Freibergs, V. (1972). Retention of affective material: Effects of intensity of affection retrieval. *Journal of Personality and Social Psychology*, *23*(1), 64–80.

Dutta, S., Norman, L., & Kanungo, R. N. (1969). A scale for the measurement of attitudes towards French Canadians. *Canadian Journal of Behavioural Sciences*, *1*(3), 156–166.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngol. Suppl. 257*, 1–51. Retrieved from National Center for Biotechnology Information, NCBI. Retrieved from https://www.ncbi.nlm.nih.gov

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, *4,* 1–9. doi: 10.3389/fpsyg.2013.00148

Eskénazi, M. (1993). Trends in speaking styles research. *Proceedings of EuroSpeech93* (pp. 501–509). Berlin.

Estival, D., Cassidy, S., Cox, F. and Burnham, D. (2014). AusTalk: An audio-visual corpus of Australian-accented English. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. Reykjavik, Iceland. Retrieved https://www.academia.edu/6546894/AusTalk_an_audiovisual_corpus_of_Australian_English.

Faris, M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *JASA-Express Letters*, *139*, EL1. doi: 10.1121/1.4939608

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. doi: 10.1037/a0017196

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 357–414). New York: McGraw-Hill.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu. J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Journal of Laboratory Phonology*, *1*(1), 5–39.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438.

Foulkes, P., Scobbie, J. M., & Watt, D. J. L. (2010) Sociophonetics. In W. Hardcastle, J. Laver, & F. Gibbon (Eds.), *Handbook of Phonetic Sciences* (2nd ed., pp. 703–754). Oxford: Blackwell.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279.

Gerstman, L. H. (1968). Classification of self-normalized vowels. *IEEE Trans. Audio Electroacoust. AU–16*, 78–80.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain & W. J.Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 49–54). Dudweiler, Germany: Pirrot.

Gosling, S. D., Rentfrow, P., & Swann Jr., W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528.

Hamilton, D. L., & Mackie, D. M. (1993). Cognitive and affective processes in intergroup perception: The developing interface. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, Cognition, and Stereotyping Interactive Processes in Group Perception* (pp. 1–11). Academic Press.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian-accented English. *Australian Journal of Linguistics*, *17*(2), 155–84. doi: 10.1080/07268609708599550

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865–892. doi: 10.1515/LING.2010.027

Hay, J., Nolan, A., & Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379.

Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111. doi: 10.1121/1.411872

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428.

Horvath, B. M. (1985). *Variation in Australian-accented English The sociolects of Sydney*. Cambridge University Press.

Howe, D. (2005). Negation in African American Vernacular English. In Y. Iyeiri (Ed.), *Aspects of English Negation* (pp. 173-203). Amsterdam: John Benjamins.

Ingram, J. C. L., & Nguyen, T. T. A. (2007). Vietnamese accented English: Foreign accent and intelligibility judgement by listeners of different language backgrounds. Paper presented at *TESOL in the Internationalization of Higher Education in Vietnam Conference* (pp. 1–15).

Isen, A. M., Means, B., Patrick, R., & Nowicki, G. (1982). Some factors influencing decision making strategy and risk-taking. In *Affect and Cognition: The 17th annual Carnegie Mellon Symposium on Cognition*, ed. by Margaret S. Clark and Susan T. Fiske, 243–261. Hillsdale, NJ: Erlbaum.

Islam, M. R., & Jahjah, M. (2001). Predictors of young Australians' attitudes towards Aboriginals, Asians and Arabs. *Social Behaviour and Personality*, *29*(6), 569–580.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 21–38.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds), *Talker variability in speech processing* (pp. 145-165). San Diego, CA: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. doi: 10.1006/jpho.1999.0100

Joos, M. (1948). Acoustic phonetics. *Language*, *24*(2), 5–136.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, *28*(3), 280–290.

Kawahara, S. (2007). Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics*, *16*(2), 113–144.

Kawahara, S., & Shinohara, K. (2009). The role of psychoacoustic similarity in Japanese puns: A corpus study. *Journal of Linguistics*, *45*(1), 111–138.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19*(3), 421–447. doi: 10.1044/jshr.1903.421

King, S., & Sumner, M. (2014). Voices and variants: Effects of voice on the form-based processing of words with different phonological variants. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2913–2918). Austin, TX: Cognitive Science Society.

Kirov, C., & Wilson, C. (2013). Bayesian speech production: Evidence from latency and hyperarticulation. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 788-793). Cognitive Science Society, Austin, TX.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148–203. doi: 10.1037/a0038695

Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua, 122*(13), 1395–1414.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English Phonetics, phonology and sound change: A multimedia reference tool*. Mouton de Gruyter.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*(1), 98–104.

Lambert, W. W., & Lambert, W. E. (1964). The social significance of attitudes. In R. S. Lazarus (Ed.), *Foundations of Modern Psychology Series Social Psychology* (pp. 49–69). New Jersey: Prentice-Hall.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*(3), 1455–1468. doi: 10.1121/1.426686

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype Content Model explains prejudice for an envied outgroup: Scale of Anti-Asian American Stereotypes. *Personality and Social Psychology Bulletin*, *31*(1), 34–47.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, *42*(4), 830-843.

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, *31*(3), 419–441.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, *49*(2), 606–608.

Lukasiewicz, E. (2012). Perception, processing and storage of subphonemic and extralinguistic features in spoken word recognition – An argument from language variation and change. *Psychology of Language and Communication*, *16*(1), 1–19.

Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198–212.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 91–125). San Diego, CA: Academic Press.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, *58*(4), 502–521.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science: A Multidisciplinary Journal*, *30*(6), 1113–1126. doi: 10.1207/s15516709cog0000_79

Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, *25*(1), 114–121.

Miller, J. D. (1984). Auditory processing of the acoustic patterns of speech. *Archives of Otolaryngology, 110*(3), 154–159.

Mitchell, A., & Delbridge, A. (1965). *The speech of Australian adolescents*. Sydney: Angus & Robertson.

Most, T., Amir, O., & Tobin, Y. (2000). The Hebrew vowel system: Raw and normalized acoustic data. *Language & Speech, 43*, 295–308.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (L). *Journal of the Acoustical Society of America*, *130*(5), 2631–2634. doi: 10.1121/1.3641410

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, *85*(5), 2088–2113.

Nearey, T. M. (1992). Applications of generalized linear modelling to vowel data. In *Proceedings of the 1992 International Conference on Spoken Language Processing* (pp. 583–586).Banff, Canada: University of Alberta.

Nguyen, N. & Shaw, J. A. (2014). Why the SQUARE vowel is the most variable in Sydney. In J. Hay & E.

Parnell (Eds.), *Proceedings of the 15th Australasian International Speech Science and Technology Conference*. Christchurch, New Zealand: University of Canterbury. ISSN 1039-0227. Paper number 26. 37-40 retrieved http://www.nzilbb.canterbury.ac.nz/graphics/SST%202014_Proceedings%20optimized.pdf

Nguyen, N., Shaw, J. A., Pinkus, R. T., & Best, C. T. (2016). Intergroup dynamics in speech perception: Interaction among experience, attitudes, and expectations. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 141–150. Retrieved from http://repository.upenn.edu/pwpl/vol22/iss2/16/

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (2015). Affective attitudes towards Asians influence perception of Asian-accented vowels. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0561. Retrieved https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0561.pdf.

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (under review). *Expectations about a foreign accent influence speech perception even without previous exposure*. Manuscript submitted for publication.

Nguyen, T., & Ingram, J. (2004). A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology* (pp. 516–521). Sydney, Australia: Macquarie University.

Niedzielski, N. (1996). Acoustic analysis and language attitudes in Detroit, *University of Pennsylvania Working Papers in Linguistics*, *3*(1), 73–85.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. doi: 10.1177/0261927X99018001005

Nielsen, J. M. (1958). *Memory and amnesia*. Los Angeles: San Lucas.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

Nielsen, K., & Wilson, C. (2008). A hierarchical Bayesian model of multi-level phonetic imitation. In N. Abner, & J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 335–343). Somerville, MA: Cascadilla Proceedings Project.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition.

Psychological review, *115*(2), 357–395. doi: 10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Orne, M. T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal Psychology*, *58*(3), 277–299.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N.Warner (Eds.), *Laboratory phonology 7* (pp. 101–139). Berlin: de Gruyter.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*(6), 516–530. doi: 10.1016/j.wocn.2006.06.003

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of*

Linguistics, *2*, 33–52. doi: 10.1146/annurev-linguistics-030514-125050

Pisoni, D. B. (1993). Long term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, *13*(1), 109–125.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson, & J. W. Mullennix (Eds.), *Talker Variablity in Speech Processing* (pp. 9–32). San Diego, CA: Academic Press.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231.

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478.

Rosenberg, M. J. (1960). Cognitive reorganization in response to hypnotic reversal of attitudinal affect. *Journal of Personality*, *28*(1), 39–63.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531.

Schatz, C. (1954). The role of context in the perception of stops. *Language*, *30*(1), 47–56.

Schourup, A. (1973). A cross-language study of vowel nasalization. *Ohio State University Working Papers in Linguistics*, *15*, 190–221.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. Shaw, & J. Bransford (Eds.) *Perceiving, acting, and knowing: Towards an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Shaw, J. A., & Gafos, A. I. (2015). Stochastic time models of syllable structure. *PloS one*, *10*(5), e0124714. doi: 10.1371/journal.pone.0124714

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*(4), 443–464.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316. doi: 10.1121/1.3101452

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, *23*(3–4), 193–229.

Sinclair, R. C. (1988). Mood, categorization breadth, and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behaviour and Human Decision Processes*, 42, 22–46.

Stephan, W. G., & Stephan, C. W. (1993). Cognition and affect in stereotyping: Parallel interactive networks. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping interactive processes in group perception* (pp. 111–136). San Diego, CA: Academic Press.

Stephan, W. G., Ybarra, O., Martínez, C. M., Schwarzwald, J., & Tur-Kaspa, M. (1998). Prejudice towards immigrants to Spain and Israel An integrated threat theory analysis. *Journal of Cross-Cultural Psychology*, 29, 559–576.

Steriade, D. (2001/2008). The phonology of perceptibility effect: The p-map and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), *The nature of the word* (pp. 151–179). Cambridge: MIT Press.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. doi: 10.1177/0261927X99018001006

Strand, E. A. & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbons (Ed.), *Natural language processing and speech technology: Results of the 3rd konvens conference, bielfelt* (pp. 14–26). Berlin: Mouton de Gruyter.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1974). Consonant environment specifies vowel identity. *Haskins Laboratories Status Report on Speech Research SR-37/38*, 209–216.

Sumner, M. (2013). A phonetic explanation of phonological variant effects. *Journal of the Acoustical Society of America*, *134*, EL26–EL32.

Sumner, M. (2015). The social weight of spoken words. *Trends in Cognitive Sciences*, *19*(5), 238–239.

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *Journal of the Acoustical Society of America*, *134*, EL485–EL491.

Sumner, M., Kim, S. K., King, E., & McGowan, K. (2014). The socially-weighted encoding of spoken words: A dual route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13.

Syrdal, A. K. (1985). Aspects of a model of the auditory representation of American English vowels. *Speech Communication*, *4*(1–3), 121–135.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*(4), 1086–1100.

Thompson, L. C. (1987). *A Vietnamese reference grammar* (pp. 78–104). Honolulu, University of Hawaii Press.

Traunmüller, H. (2003). Clicks and the idea of a human protolanguage. In *Proceedings of Fonetik 2003 Phonum 9* (pp. 1–4). Umeå University.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21.

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods, 5*(4), 553–600.

Verbrugge, R., Strange, W., & Shankweiler, D. (1974). What information enables a listener to map a talker's vowel space? *Haskins Laboratories Status Report on Speech Research SR–37/38*, 199–208.

Walker, I. (1994). Attitudes to minorities: Survey evidence of Western Australians' attitudes to Aborigines, Asians, and Women. *Australian Journal of Psychology*, *46*(3), 137–143.

Warren, P. (2016). *Uptalk: The phenomenon of rising intonation*. Cambridge, United Kingdom: Cambridge University Press.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change*, *12*(1), 51–68.

Wells, J. C. (1982). *Accents of English: Vol. 1. An Introduction*. Cambridge: Cambridge University Press.

Wells, J. C. (1982a). *Accents of English: Vol. 2. The British Isles.* Cambridge, United Kingdom: Cambridge University Press.

Wells, J. C. (1982b). *Accents of English: Vol. 3. Beyond the British Isles.* Cambridge, United Kingdom: Cambridge University Press.

Wilson, C., & Davidson, L. (2013). Bayesian analysis of non-native cluster production. In S. Kan, C. Moore-Cantwell, & R. Staubs (Eds.), *Proceedings of the 40th Annual Meeting of the North East Linguistic Society* (pp. 265–278). Amherst, MA: Graduate Linguistic Student Association.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception & Psychophysics*, *75*(3), 537–556. doi: 10.3758/s13414-012-0404-y

Yi, Z. (2010). *Lectures on Chinese phonetics*. Taipei: Taiwan Commercial Press.

Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, *47,* 18–35.

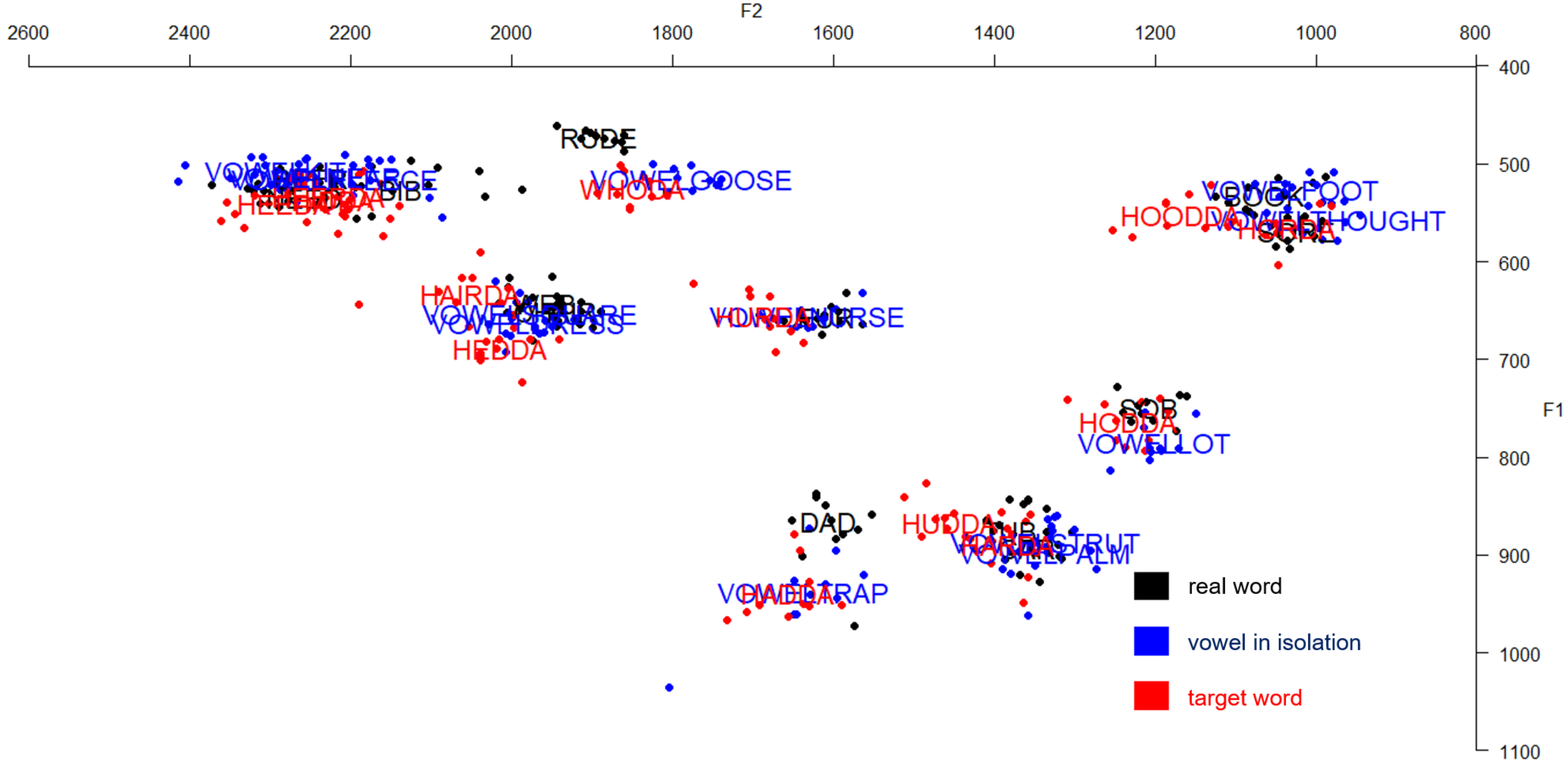
Zielinski, B. W. (2003). Intelligibility in Speakers of English as a Second Language Previous research. In *16th Educational Conference Melbourne 2003* (pp. 1–6).

Zielinski, B. W. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, *21*(1), 22–45.

Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*(1), 69–84.

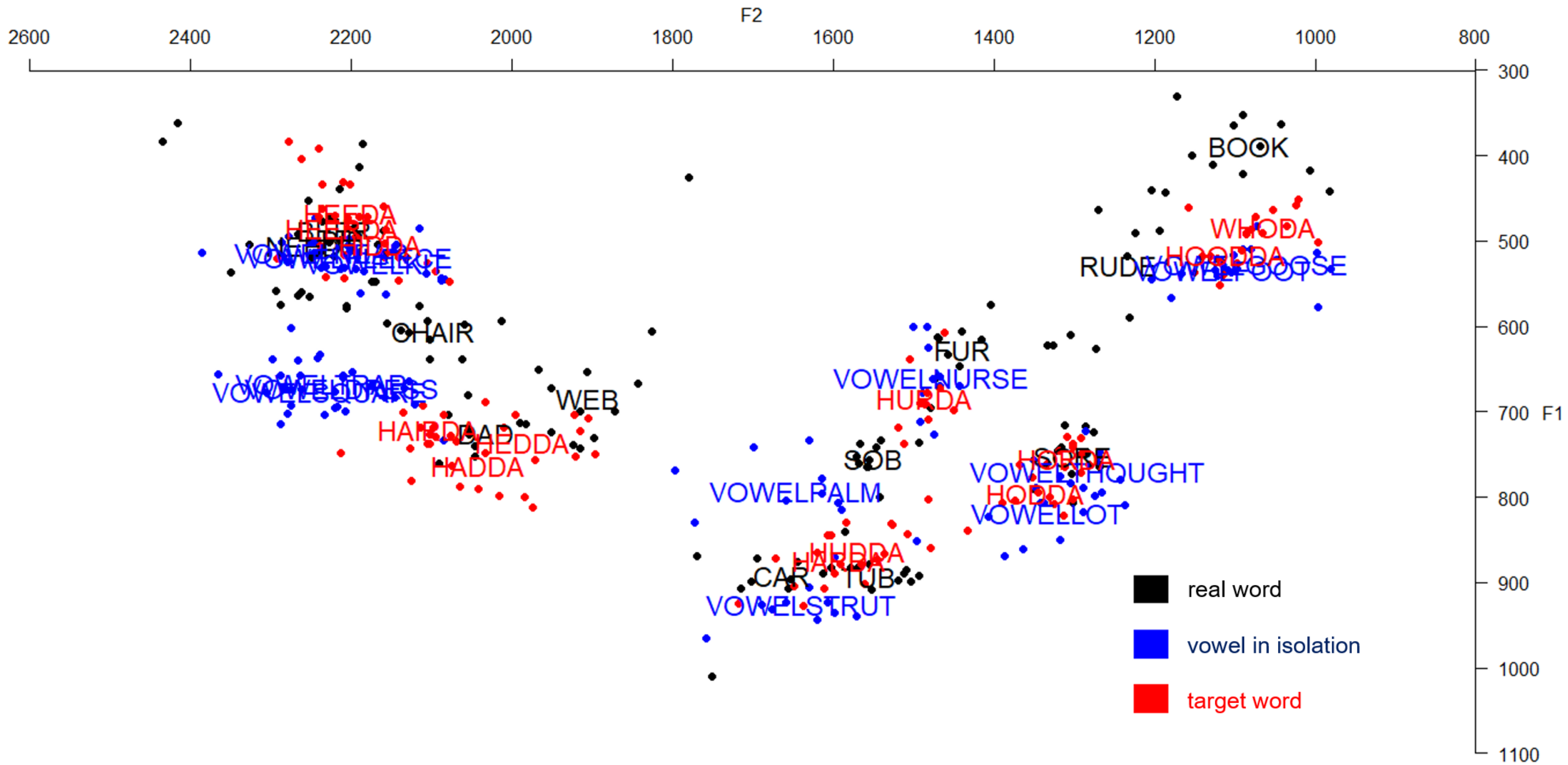
Appendices

Appendix 1a



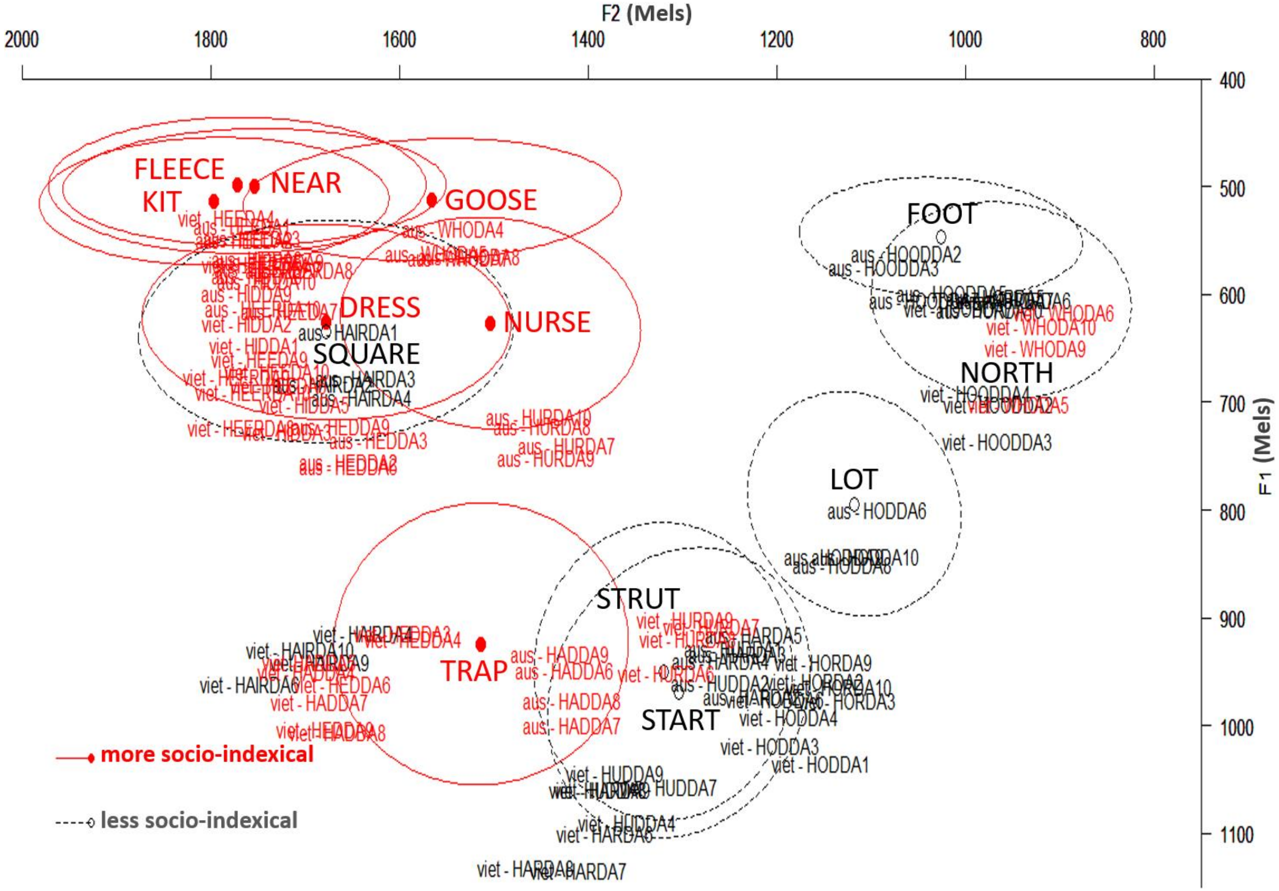
Vowel space showing how accurately the Australian-accented English speaker produced the real word items containing the target vowel, the vowel in isolation, and the vowel in the target nonword frame

Appendix 1b



Vowel space showing how accurately the Vietnamese-accented English speaker produced the real word items containing the target vowel, the vowel in isolation, and the vowel in the target nonword frame

Appendix 2



Vowel space showing *where (both Australian-accented English and Vietnamese-accented English) stimulus values fall relative to the means of the distributions of Australian-accented English monophthongs based on Cox (2006)* [Ellipses represent half SD from the mean. Red indicates the monophthongs that carry more socio-indexical information.]

Appendix 3:

Why the SQUARE vowel is the most variable in Sydney

This appendix has been published as:

Nguyen, N. & Shaw, J. A. (2014). Why the SQUARE vowel is the most variable in Sydney. In J. Hay & E. Parnell (Eds.), *Proceedings of the 15th Australasian International Speech Science and Technology Conference*. Christchurch, New Zealand: University of Canterbury. ISSN 1039-0227. Paper number 26. 37-40 retrieved http://www.nzilbb.canterbury.ac.nz/graphics/SST%202014_Proceedings%20optimized.pdf

Why the SQUARE vowel is the most variable in Sydney

Nhung Nguyen, Jason A. Shaw

The MARCS Institute, School of Humanities and Communication Arts, University of Western Sydney

nhung.nguyen@uws.edu.au, j.shaw@uws.edu.au

Abstract

Vowel variability is often explained in terms of linguistic and social factors. We have observed another factor that predicts vowel variability. Within four different corpora of Australian English vowels, we find a consistent relationship between the mean and standard deviation of formant values. For both F1 and F2, increases in mean formant values go hand in hand with increased variability. Given this observation, we propose that inferences about vowel variability take the mean formant values into account. Doing so changes conclusions about which vowels are most variable, undergoing change, or likely to reflect meaningful social variation.

Index Terms: vowel variability, acoustic phonetics, Australian English, forced alignment

1. Introduction

A typical method of visualizing vowel variability is to plot tokens in F1-F2 vowel space. The spread of tokens within this space, or sometimes ellipses summarizing the variance, indicate vowel variability [1, 2]. We exemplify this approach with a dataset from Cox's seminal paper on Australian vowels [1]. The vowel formant data from monophthongs (with the centring diphthongs /ɪə/ and /eː/ also classified as monophthongs [1]) produced by female speakers from that study (n= 60) are reproduced in Figure 1. Ellipses show two standard deviations from formant means. The shape of the ellipses indicates the spread of variation. Ellipses for FLEECE, NEAR, KIT, and GOOSE vowels [3], vowels with high F2 and low F1, are short and fat. Ellipses for TRAP, STRUT, START, and LOT vowels, vowels with high F1 and low F2, are tall and skinny. On the basis of these patterns, we might conclude that the TRAP, STRUT, START, and LOT vowels are most variable with regards to F1, and FLEECE, NEAR, KIT, and GOOSE vowels are most variable in relation to F2. This is a satisfactory conclusion. However, we observe that, on this conclusion, vowel variability is closely linked to the magnitude of mean formant values. The vowels with the highest mean values of F1 are the most variable in F1. The vowels that have the highest formant values of F2 are the most variable in F2. To highlight this point, Figure 2 plots the mean formant values from the same set of monophthongs shown in Figure 1 against the standard deviation. The result is clear. As mean formant values increase, so too does the standard deviation.

The strong correlation between the mean and standard deviation of formant values raises a number of questions about the proper characterization of vowel variation. This correlation may be crucial for interpreting vowel variability for sociolinguistic purposes such as differences across accents [2, 4], genders [1, 5], ages [6], and other social factors. In addition, knowledge of how much variability should be expected for a vowel given its mean formant values could inform the extent to

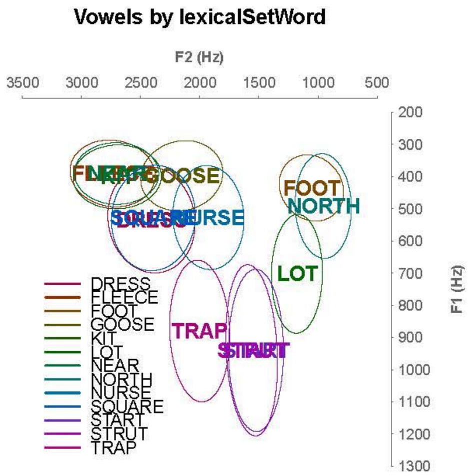


Figure 1: *Plots of Australian English monophthongs regenerated with random values sampled from means and standard deviations reported for female speakers in Cox [1]. Ellipses represent two standard deviations from the mean.*

which variability is tolerated in perception of accented speech [7] and in selective adaptation [8].

In the remainder of this paper, we seek to assess the robustness of the correlation between the mean and standard deviation of Australian English vowel formants. We report the mean-standard deviation (mean-sd henceforth) relationship for vowel formants drawn from published sources [5], publically available corpora [9], and newly collected data. We find significant correlations in all of the datasets. Given this result, we suggest methods of assessing vowel variability that take into account the magnitude of mean vowel formants. In light of this suggestion, we reassess which Australian vowels are most variable.

2. Method

2.1. Descriptions of datasets

In addition to Cox [1], we looked for a mean-sd correlation in three additional corpora. We wanted first to replicate the result in a similar dataset. For this purpose, we searched through the AusTalk corpus [9] to find speakers from the Sydney region. We also wanted to know if the relationship emerges only across inter-speaker variation or if it can be found as well for

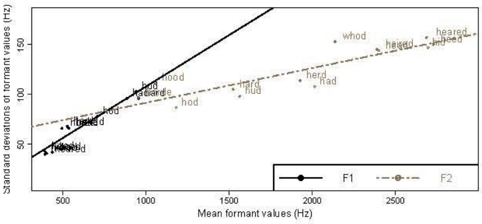


Figure 2: *Correlation between mean and standard deviation of 13 Australian English monophthongs in 13 /hVd/ words based on 60 female speakers from Sydney as reported in Cox [1].*

within-speaker variation. For this reason, we recorded several repetitions of each monophthong from another Sydney speaker. Lastly, we wanted to investigate whether the relationship is peculiar to the Sydney accent, so we explored a set of published results from Melbourne speakers [5]. We describe each of these corpora in greater detail below. For simplicity in presentation, we focus on just female speakers and on monophthongs produced in the /hVd/ environment. However, the main conclusions that we present generalize to male participants and diphthongs as well.

First, we sought to replicate the Cox result using speakers drawn from the publically available AusTalk corpus [9]. The data used for analyses were 13 monophthongs in the /hVd/ environment. We choose these words because they are in the same environment as the vowels reported in Cox [1]. We selected speakers from AusTalk based on the following criteria: female, aged 18 to 30, have Australian-born parents, live in the Sydney area, have low to no proficiency in other languages. We found five speakers that met this criteria (IDs: NSW10 Red-Squatter Pigeon, NSW28 Green-Cook's Petrel, NSW30 Gold-Terek Sandpiper, NSW19 Blue-Grey Grasswren, NSW25 Gold-Cape York Rock-wallaby). An exception to our criteria was made for ID NSW13 Blue-Laysan Albatross, who has a New Zealand-born father but typical Sydney vowels. A total of 78 observations from these six speakers were analyzed.

To evaluate whether the mean-sd correlation emerges from within-speaker variation, we recorded another female speaker from Sydney. The recording included the same 13 Australian English monophthongs embedded in the /hVd/ context. Data was collected in a sound-attenuated booth using a Shure SM10A-CN headset microphone. The speaker produced each /hVd/ words 10 times in random order, giving us 130 /hVd/ observations.

Lastly, the Melbourne dataset was taken from Billington's females [5]. Data collection approximated Cox's [1] with 13 female participants, each producing the 13 monophthongs twice in /hVd/ context. The means and standard deviations for F1 and F2 of each vowel are provided in [5]. The data are drawn from 338 total observations.

Neither of the three corpora analyzed here are as impressive as Cox's [1] in size, but they allow us to see if the trend observed in that data persists across a smaller sample of Sydney speakers, within a speaker, and in another region of Australia.

2.2. Vowel segmentation and formant extraction

Vowels in Cox's and Billington's datasets were hand-segmented and hand-measured. This work is labour-intensive, particularly given the size of those respective corpora. In addition to hand-segmenting the portion of the AusTalk data and single speaker data we report here, we also explored automatic segmentation using the Forced Alignment and Vowel Extraction (FAVE) program suite, web-based programs freely available through the University of Pennsylvania [10].

After segmentation, formants were extracted at 20%, 25%, 35%, 50%, 65%, and 80% of total vowel duration using the Mahalanobis method [11], which optimizes formant-tracking settings on a speaker-by-speaker basis. Initial settings were for five formants to be tracked with a maximum formant value of 5500Hz, a window size of 0.025s, and 12-point smoothing. Throughout this paper, we report the values of formants obtained at 25%.

To assess segmentation based on FAVE forced alignment, we correlated measurements of F1, F2, and duration based on hand- and machine-segmented values. For the AusTalk data, correlation coefficients were very high for F1 ($r = .97, p < .001$) and F2 ($r = 1.00, p < .001$). The correlation for duration was not as good ($r = .89, p < .001$). This indicates that, although there was some variation in where the boundary was placed by human versus machine segmentation, the variation in boundary placement had a negligable effect on formant extraction. Results for the single-speaker data were not quite as good: for F1 ($r = .86, p < .001$), for F2 ($r = .99, p < .001$), and for duration ($r = .76, p < .001$). Where the forced aligner deviated most from hand segmentation was for back rounded vowels (i.e., 'hod', 'hood', and 'horde'). For these vowels, the forced alignment occasionally placed the onset of the vowel boundary too early, including in the vowel some of the aperiodic energy in the /h/ of our /hVd/ context.

We think that the forced alignment results are reasonably promising, but we also note that it is subject to errors which seem to be systematic and could potentially influence the variability of vowel measurements. For this reason, we focus our discussion on the hand-segmented data and return to the human versus machine comparison in the discussion.

3. Results

For each of the four datasets under consideration, we plotted the mean formant values (F1 and F2) against the standard deviation and fit regression lines to F1 and F2. Figure 2 shows Cox's Sydney data (females only). Figure 3 shows the female Sydney speakers from AusTalk. Figure 4 shows our single Sydney speaker dataset. Figure 5 shows Billington's Melbourne data. For each of the datasets, the trend is the same. As the central tendency increases so too does the variance. Regression lines fit to the various corpora all indicate a significant positive relationship between the mean and standard deviation for both F1 and F2. The R^2 values are summarized in Table 1.

4. Discussion

4.1. The robustness of mean-sd correlation

We found a positive correlation between the mean and standard deviation of vowel formants for each of four different sets of vowel measurements. The linear relationship is stronger in Cox's data for both F1 and F2 than any of the other datasets. We take this to be an indication of the size and quality of Cox's

Table 1: *Mean-sd relationship across four datasets.*

dataset	formant	R^2	p
Cox's (Sydney)	F1	$R^2 = .93$	$p < .001$
	F2	$R^2 = .77$	$p < .001$
AusTalk (Sydney)	F1	$R^2 = .35$	$p < .05$
	F2	$R^2 = .70$	$p < .001$
Single-speaker (Sydney)	F1	$R^2 = .80$	$p < .001$
	F2	$R^2 = .40$	$p < .05$
Billington's (Melbourne)	F1	$R^2 = .78$	$p < .001$
	F2	$R^2 = .70$	$p < .001$

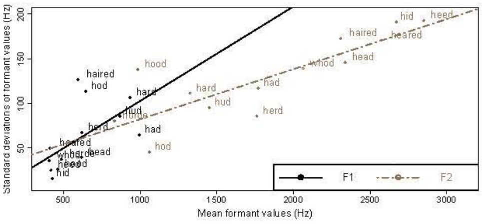


Figure 3: *Correlation between means and standard deviations of 13 Australian English monophthongs in 13 /hVd/ words, 6 female Sydney speakers, AusTalk dataset (hand-segmented).*

data. It is the largest dataset, composed of 60 female speakers. Moreover, the entire corpus was hand-segmented. The relationship is next strongest in Billington's data. Also hand-segmented and comparably large, the measurements in Billington are probably the next best estimation of population variance.

Formant values for the other datasets were extracted at a fixed percentage of vowel duration. This method may be less reliable for capturing the vowel target. Nevertheless, we find it intriguing that the mean-sd relationship persists despite this and other differences across corpora. Notably, we found the correlation regardless of whether variability was calculated within speakers or across speakers; whether in Sydney or in Melbourne; whether in adolescents (as in Cox's study and in Billington's study) or in young adults (as in the AusTalk speakers and our single speaker). Although we reported results here on just female speakers and on just monophthongs, the mean-sd

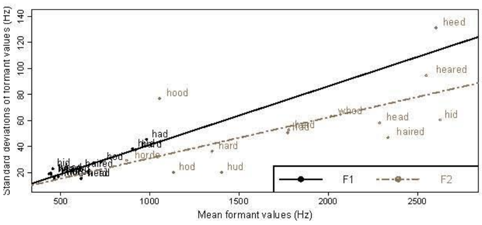


Figure 4: *Correlation between means and standard deviations of 13 Australian English monophthongs in 13 /hVd/ words, 1 female Sydney speaker, single-speaker dataset (hand-segmented).*

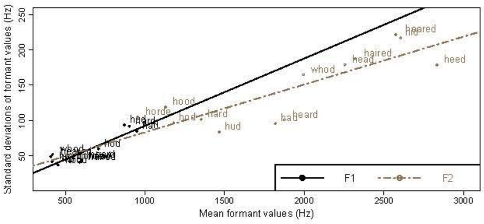


Figure 5: *Correlation between means and standard deviations of 13 Australian English monophthongs in 13 /hVd/ words, 13 female Melbourne speakers, Billington's dataset.*

relationship can also be found in male speakers in both Cox's and Billington's datasets, and it persists as well when diphthongs are included.

One thing that the four datasets reported on here have in common is that they all include vowels in the /hVd/ context. We would like to explore the mean-sd relationship in more diverse speech styles and contexts. Forced alignment and other modern corpus analysis tools could expedite such an analysis substantially. However, if our interpretation of the differences in goodness of fit between Cox's corpus and the AusTalk corpus is on the right track, then segmentation method might be very important to revealing mean-sd correlations. To underscore this point, we return to the discussion of forced alignment. We demonstrated reasonably strong correlations between measurements of vowels based on hand and machine segmentation. However, the correlation between the mean and standard deviation of vowel formants was weaker for datasets segmented using forced alignment. For the single-speaker dataset, mean-sd relationship weakened from $R^2_{F1} = .80$ (hand-segmented) to $R^2_{F1} = .59$ (machine-segmented) and $R^2_{F2} = .40$ (hand-segmented) to $R^2_{F2} = .34$ (machine-segmented). The mean-sd correlation for the AusTalk data also weakened from $R^2_{F1} = .35$ (hand-segmented) to $R^2_{F1} = .15$ (machine-segmented) and $R^2_{F2} = .70$ (hand-segmented) to $R^2_{F2} = .69$ (machine-segmented). It therefore appears that the strength of the mean-sd relationship reflects the quality of the measurements.

To sum up, the mean-sd correlation is robust across corpora. It surfaces in all of the datasets irrespective of vowel segmentation methods, formant measurement methods, regional varieties, speaker age, and gender; however, it is strongest for large datasets measured carefully. This suggests the possibility of using the relationship to evaluate the quality of (automatic) vowel segmentation and formant measurement.

4.2. So which Australian vowel is most variable?

Early descriptions of Australian English vowels claimed that the FLEECE, GOOSE, FACE, PRICE, MOUTH, and GOAT vowels were the most variable, as these vowels marked the three Australian accent varieties (i.e., Broad, General, and Cultivated) in the 60s [12]. These vowels are amongst those that we noted in the introduction have a large standard deviation (i.e., large ellipses in Figure 1). FLEECE and GOOSE show more variation in F2 than do TRAP, STRUT, START, or LOT. However, if we take the central tendency of the formants into account, we would note that these vowels have just the amount of variation that is expected, based upon their mean values. Given the mean-sd correlation, how then should we assess vowel variability? We conclude by suggesting a method that takes the mean formant value into account and apply the method to the four corpora of female Australian English vowels.

We suggest using the regression lines fit to the mean and standard deviation of formant values (e.g., Figures 2, 3, 4, 5) to establish a baseline for vowel variability. Datapoints above the regression lines (positive residuals) are variable relative to the magnitude of the mean. Values below the regression line (negative residuals) are stable relative to the mean. Table 2 shows lexical sets with the highest positive residuals.

Table 2: *F1 and F2 residuals for each dataset.*

dataset	F1 residual	F2 residual
Cox's (Sydney)	NORTH (10Hz)	GOOSE (22Hz)
	NURSE (9Hz)	FOOT (19Hz)
	SQUARE (8Hz)	NORTH (7Hz)
AusTalk (Sydney)	SQUARE (67Hz)	FOOT (57Hz)
	LOT (48Hz)	SQUARE (17Hz)
	START (11Hz)	KIT (15Hz)
Single-speaker (Sydney)	KIT (6Hz)	FLEECE (50Hz)
	TRAP (5Hz)	FOOT (44Hz)
	NEAR (3Hz)	NEAR (15Hz)
Billington's (Melbourne)	GOOSE (15Hz)	NEAR (32Hz)
	STRUT (14Hz)	FOOT (27Hz)
	NEAR (13Hz)	KIT (25Hz)

As can be seen from table 2, SQUARE has high F1 and F2 residuals in the AusTalk dataset. It also has a positive F1 residual in Cox's data. This is not a surprising result given the unclear extent of its monophthongization [2]. In addition, the Australian vowel system is undergoing change [13]. Due to the descent of the TRAP vowel, vowels surrounding TRAP, such as DRESS [13, 14], have also lowered. Since DRESS and SQUARE have been shown to be acoustically similar (except for duration) at least in the /hVd/ context [1], it is reasonable to speculate that SQUARE is also undergoing change. The residuals indicate that SQUARE is more variable than predicted by its mean in the inter-speaker Sydney datasets but not in the single-speaker dataset or in the Melbourne data. This indicates that there are comparably large differences in how the SQUARE vowel is produced across speakers in the Sydney region.

5. Conclusions

We observed a significant positive correlation between the variance of Australian vowel formant measurements and the central tendency of these measurements. The correlation emerged across corpora despite differences in measurement technique and speaker properties including age, gender, and region. The correlation also has consequences for interpreting vowel variability. When formant variability is considered relative to the mean, the SQUARE vowel emerges as the most variable in the Sydney region. This generalization is otherwise masked by the strong influence that mean formant values exert on formant variance.

6. Acknowledgements

The AusTalk corpus was collected as part of the Big ASC project (Burnham et al. 2009; Wagner et al. 2010; Burnham et al. 2011), funded by the Australian Research Council (LE100100211). See: https://austalk.edu.au/ for details. We would like to thank Jaydene Elvin for downloading the AusTalk data, Jaydene Elvin, Karen Mulak, Michael Tyler for help with recording material, Mona Faris for a PRAAT observation, and Jia Ong and Sarah Fenwick for stats discussion.

7. References

[1] Cox, F., "The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers", Australian Journal of Linguistics, 26(2):147–179, 2006.

[2] Harrington, J., Cox, F., and Evans, Z., "An acoustic phonetic study of broad, general, and cultivated Australian English vowels", Australian Journal of Linguistics, 17(2):155–184, 1997.

[3] Wells, J.C., Accents of English, Cambridge, 1982.

[4] Clermont, F., "Multi-speaker formant data on the Australian English vowels: A tribute to J.R.L. Bernard's (1967) pioneering research", Proc. 6th Australian Int. Conf. Speech Science and Technology, Adelaide, Australia, 145–150, 1996.

[5] Billington, R., "Location, Location, Location! Regional Characteristics and National Patterns of Change in the Vowels of Melbourne Adolescents", Australian Journal of Linguistics, 31(3):275–303, 2011.

[6] Millar, J., Vonwiller, J., Harrington, J., and Dermody, P., "The Australian National Database of Spoken Language", Proceedings of the International Conference on Acoustics Speech and Signal Processing, Adelaide, 2, 1994.

[7] Clarke, C.M., and Garrett, M.F., "Rapid adaptation to foreign-accented English", Journal of Acoustical Society of America, 116(6):3647–3658, 2004.

[8] Kleinschmidt, D., and Jaeger, T.F., "A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation", Proceedings of the 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan, 2012.

[9] Estival, D., Cassidy, S., Cox, F. and Burnham, D., "AusTalk: an audio-visual corpus of Australian English". Online: https://www.academia.edu/6546894/AusTalk_an_audio-visual_corpus_of_Australian_English, accessed on 25 May 2014.

[10] Rosenfelder, I., Fruehwald, J., Evanini, K. and Jiahong, Y., "FAVE (Forced Alignment and Vowel Extraction) Program Suite". http://fave.ling.upenn.edu/index.html, 2011.

[11] Evanini, K., "Automatic vowel analysis", in K., Evanini, The permeability of dialect boundaries: a case study of the region surrounding Erie Pennsylvania, 50-94, PhD dissertation, 2009.

[12] Mitchell, A.G., and Delbridge, A., "The speech of Australian adolescents: A survey", Sydney: Angus and Robertson, 1965.

[13] Cox, F., "Australian English pronunciation into the 21st century", Prospect, 21(1):3–21, 2006.

[14] Cox, F. and Palethorpe, S., "A question of broadness", Paper presented at Australian Linguistics Society Conference, Newcastle, Australia, 2004.

Appendix 4a

(This is the version of the online survey

that was used to collect the prejudice data reported in Appendices 3 and 4)

https://uwsssap.co1.qualtrics.com/jfe/form/SV_0HEF3xCCFZpNCER

Q1.1 We are conducting an online survey about student perspectives. You will be asked to mainly rate items on given scales. Occasionally, you will be asked to answer some short-answer or multiple-choice questions. By participating, you will gain first-hand experience in psychological research.   All aspects of the survey, including results, will be confidential and only the researchers will have access to information on participants, and the identity of participants will not be disclosed. The findings from this survey will constitute part of a thesis and may be submitted for publication to a journal article, and presented at conferences. This survey should not provide any discomfort or harm to you, and participation is entirely voluntary: you are not obliged to be involved and - if you do participate - you can withdraw at any time without giving any reason and without any consequences.   This study has been approved by the University of Western Sydney Human Research Ethics Committee (H10467). If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Research Ethics Office (Tel: +61 2 4736 0229 or Email: humanethics@uws.edu.au). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.   This survey is optimized for you to do on your desktop or laptop. It will take you approximately half an hour to finish, please do not begin the survey until you have this amount of time available. Partially completed surveys will not be saved, and the survey page may automatically expire after some time of inactivity. Your consent to participate is given once you click on the option 'Yes, I agree to participate' below. Participation in this survey will bring you 3 credit points. NB1: The `back' button is not available for this survey, so please choose your answers carefully before clicking on the double arrow button to proceed to the next page. NB2: If, by any chance, you have already taken this survey outside the SONA system, please do not take it again.   Do you agree to participate in this survey?

❍ Yes, I agree to participate.

❍ No, I do not agree to participate.

If No, I do not agree to parti... Is Selected, Then Skip To End of Survey If Yes, I agree to participate. Is Selected, Then Skip To Full name

Q1.2 Full name

Q1.3 Date of birth (dd/mm/yyyy)

Q1.4 Email address (please provide the email address you use regularly)

Q1.5 Do/did you have a hearing impairment, or reading difficulties (e.g., difficulties learning to read), or language development or speaking difficulties (e.g., delayed language development, stuttering, lisping, etc.)?

❍ No

❍ Yes

Display This Question:

If Do/did you have a hearing impairment, or reading difficulties (e.g., difficulties learning to read), or language development or speaking difficulties (e.g., delayed language development, stuttering... Yes Is Selected

Q1.6 Please click any/all that apply (hold the Ctrl key as you click if you want to select multiple options):

❑ hearing

❑ reading

❑ speaking

❑ language

Q2.1 What is your father's native language?

Q2.2 What is your mother's native language?

Q2.3 Where were you born? Please list the city, state, and country. If the country is Australia, specifically list the suburb. For example: Bankstown - Sydney, New South Wales, Australia

Q2.4 List cities (and countries) you've lived since you were born. If the country is Australia, specifically list the suburbs. For example: Madrid (Spain), West End - Brisbane (Australia), Bangkok (Thailand), Vancouver (Canada), Surry Hills - Sydney (Australia)

Q2.5 Spoken language: Please tell us what other languages you can speak besides English, at what age you started speaking the language and at what age you stopped (if you still speak the language, then put in the current age and the word 'now'), and how often and how well you spoke/speak it. If you don't speak any language besides English, please put 'n/a' into the cells. If you've already filled in 'n/a' into a row but the page keeps asking you to answer the question, please just click on 'Continue without answering'.

	Language	Age (start)	Age (end)	How often did/do you speak it? (1=rarely, 5=always)	How well do you speak it? (1=hardly at all, 5=highly fluent)
Example					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Q2.6 Have you ever tried the following cuisine? If yes, please tick accordingly.

	yes
American cuisine	❍
Chinese cuisine	❍
Aboriginal cuisine (in Australia)	❍
British cuisine	❍
Jordanian cuisine	❍
Vietnamese cuisine	❍
Ethiopian cuisine	❍
Korean cuisine	❍
Egyptian cuisine	❍
Sudanese cuisine	❍

Q2.7 The instructions below are for the items in the next page, please read them carefully (because there's no 'back' button for you to go back and read them again): Thinking back over the past year (2013) up to now, what is the highest frequency of interactions that you have had with people that grew up in these countries (states/territories) but are currently residing short-term or long-term in Australia? (for example, first generation immigrants, international students, etc.) For example: Over 2013 in general, Joanne has rarely had an interaction with people from Korea, except that there was a Korean friend of family coming to Australia for a 3-month English course in the beginning of the year. During these 3 months, she had conversations (in English) with this Korean guest everyday. Therefore, to answer this question, Joanne would choose 'Daily' for the column 'Korea'.

Q2.8

	Never	Once a year or less	Several times a year	Once a month	2-3 times a month	Once a week	2-3 times a week	Daily
United States of America	❍	❍	❍	❍	❍	❍	❍	❍
China	❍	❍	❍	❍	❍	❍	❍	❍
Northern Territory (Australia)	❍	❍	❍	❍	❍	❍	❍	❍
United Kingdom	❍	❍	❍	❍	❍	❍	❍	❍
Jordan	❍	❍	❍	❍	❍	❍	❍	❍
Vietnam	❍	❍	❍	❍	❍	❍	❍	❍
Ethiopia	❍	❍	❍	❍	❍	❍	❍	❍
Korea	❍	❍	❍	❍	❍	❍	❍	❍
Egypt	❍	❍	❍	❍	❍	❍	❍	❍
Sudan	❍	❍	❍	❍	❍	❍	❍	❍

Q2.9 The instructions below are for the items in the next page, please read them carefully (because there's no 'back' button for you to go back and read them again): Thinking back across your lifetime (excluding 2013), what is the highest frequency of interactions that you had with people that grew up in these countries (states/territories) but are currently residing short-term or long-term in Australia? (for example, first generation immigrants, international students, etc.) For example: Joanne is 18 years old this year. When she was 12, a Jordanian family moved to her neighborhood. The Jordanian wife came to talk to her 2 or 3 times a week. The Jordanian family then moved away when Joanne turned 15. Apart from these interactions, Joanne hardly spoke to a Jordanian. Therefore, to answer this question, Joanne would choose '2-3 times a week' for the column 'Jordan'.

Q2.10

	Never	Once a year or less	Several times a year	Once a month	2-3 times a month	Once a week	2-3 times a week	Daily
United States of America	❍	❍	❍	❍	❍	❍	❍	❍
China	❍	❍	❍	❍	❍	❍	❍	❍
Northern Territory (Australia)	❍	❍	❍	❍	❍	❍	❍	❍
United Kingdom	❍	❍	❍	❍	❍	❍	❍	❍
Jordan	❍	❍	❍	❍	❍	❍	❍	❍
Vietnam	❍	❍	❍	❍	❍	❍	❍	❍
Ethiopia	❍	❍	❍	❍	❍	❍	❍	❍
Korea	❍	❍	❍	❍	❍	❍	❍	❍
Egypt	❍	❍	❍	❍	❍	❍	❍	❍
Sudan	❍	❍	❍	❍	❍	❍	❍	❍

Q2.11 The instructions below are for the items in the next page, please read them carefully (because there's no 'back' button for you to go back and read them again): Thinking back over the past year (2013) up to now, what is the highest frequency of interactions that you have had with Australians whose parents grew up in these countries (states/territories)? For example: Over 2013 in general, Joanne has rarely had an interaction with Korean Australians, except that she stayed with Korean Australians in the same hostel room during her 15-day trip to Queensland in the beginning of the year. During these 15 days, she had conversations with these Korean Australian roommates everyday. Therefore, to answer this question, Joanne would choose 'Daily' for the column 'Korea'.

Q2.12

	Never	Once a year or less	Several times a year	Once a month	2-3 times a month	Once a week	2-3 times a week	Daily
United States of America	❍	❍	❍	❍	❍	❍	❍	❍
China	❍	❍	❍	❍	❍	❍	❍	❍
Northern Territory (Australia)	❍	❍	❍	❍	❍	❍	❍	❍
United Kingdom	❍	❍	❍	❍	❍	❍	❍	❍
Jordan	❍	❍	❍	❍	❍	❍	❍	❍
Vietnam	❍	❍	❍	❍	❍	❍	❍	❍
Ethiopia	❍	❍	❍	❍	❍	❍	❍	❍
Korea	❍	❍	❍	❍	❍	❍	❍	❍
Egypt	❍	❍	❍	❍	❍	❍	❍	❍
Sudan	❍	❍	❍	❍	❍	❍	❍	❍

Q2.13 The instructions below are for the items in the next page, please read them carefully (because there's no 'back' button for you to go back and read them again): Thinking back across your lifetime (excluding 2013), what is the highest frequency of interactions that you had with Australians whose parents grew up in these countries (states/territories)? For example: Joanne is 18 years old this year. When she was 12, a Jordanian Australian family moved to her neighborhood. She talked to their daughter, who was also 12, and played with her everyday. The Jordanian Australian family then moved away when Joanne and their daughter turned 14. Joanne still spoke to the friend 2 years after, but only once every few months. Therefore, to answer this question, Joanne would choose 'Daily' for the column 'Jordan'.

Q2.14

	Never	Once a year or less	Several times a year	Once a month	2-3 times a month	Once a week	2-3 times a week	Daily
United States of America	❍	❍	❍	❍	❍	❍	❍	❍
China	❍	❍	❍	❍	❍	❍	❍	❍
Northern Territory (Australia)	❍	❍	❍	❍	❍	❍	❍	❍
United Kingdom	❍	❍	❍	❍	❍	❍	❍	❍
Jordan	❍	❍	❍	❍	❍	❍	❍	❍
Vietnam	❍	❍	❍	❍	❍	❍	❍	❍
Ethiopia	❍	❍	❍	❍	❍	❍	❍	❍
Korea	❍	❍	❍	❍	❍	❍	❍	❍
Egypt	❍	❍	❍	❍	❍	❍	❍	❍
Sudan	❍	❍	❍	❍	❍	❍	❍	❍

Q2.15 Have you ever travelled to these countries/states/territories? If yes, please tick accordingly.

	yes
United States of America	❍
China	❍
Northern Territory (Australia)	❍
United Kingdom	❍
Jordan	❍
Vietnam	❍
Ethiopia	❍
Korea	❍
Egypt	❍
Sudan	❍

Q3.1 Please indicate your preferences in the use of hands in the following activities:

	Always left hand	Mostly left hand	Both hands	Mostly right hand	Always right hand
Writing	❑	❑	❑	❑	❑
Throwing	❑	❑	❑	❑	❑
Scissors	❑	❑	❑	❑	❑
Knife (without fork)	❑	❑	❑	❑	❑
Opening box (lid)	❑	❑	❑	❑	❑

Q4.1 Please respond to the items below according to the corresponding scale:

	strongly disagree	moderately disagree	slightly disagree	no answer/ don't know	slightly agree	moderately agree	strongly agree
I believe that there is a physical Hell where people are punished after death for the sins of their lives.	❍	❍	❍	❍	❍	❍	❍
To me the most important work of the church is the saving of souls.	❍	❍	❍	❍	❍	❍	❍
I have a duty to help those who are confused about religion.	❍	❍	❍	❍	❍	❍	❍
Even though it may create some unpleasant situations, it is important to help people become enlightened about religion.	❍	❍	❍	❍	❍	❍	❍
I find that my ideas on religion have a considerable influence on my views in other areas.	❍	❍	❍	❍	❍	❍	❍

Religion is a subject in which I am not particularly interested.	❍	❍	❍	❍	❍	❍	❍

Q5.1 Please choose true or false for the following statements:

	true	false
Seeing a cockroach in someone else's house doesn't bother me.	❍	❍
It bothers me to hear someone clear a throat full of phlegm.	❍	❍

Q5.2 Please rate how disgusting you would find the following experiences: not disgusting at all, slightly disgusting, very disgusting. If you think something is bad or unpleasant, but not disgusting, you should choose 'not disgusting at all':

	not disgusting at all	slightly disgusting	very disgusting
You see maggots on a piece of meat in an outdoor garbage pail.	❍	❍	❍
While you are walking through a tunnel under a railroad track, you smell urine.	❍	❍	❍

Q6.1 Please think about each statement that follows and rate the degree to which you agree or disagree with each one on the following scale:

	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
I feel that I have a number of good qualities.	❍	❍	❍	❍	❍	❍	❍
On the whole I am satisfied with myself.	❍	❍	❍	❍	❍	❍	❍
All in all I am inclined to feel that I am a failure.	❍	❍	❍	❍	❍	❍	❍
I certainly feel useless at times.	❍	❍	❍	❍	❍	❍	❍

Q7.1 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q7.2

	strongly disagree	disagree	neutral	agree	strongly agree
Aborigines have jobs that the Anglos should have.	❍	❍	❍	❍	❍
Most Aborigines living here who receive support from welfare could get along without it if they tried.	❍	❍	❍	❍	❍
Anglo people and Aborigines can never really be comfortable with each other, even if they are close friends.	❍	❍	❍	❍	❍
Most politicians in Australia care too much about Aborigines and not enough about the average Anglo person.	❍	❍	❍	❍	❍
Aborigines come from less able races and this explains why they are not as well off as most Anglo people.	❍	❍	❍	❍	❍

Q7.3

	very similar	similar	neutral	different	very different
How different or similar do you think Aborigines living here are to Anglo people in how honest they are?	❍	❍	❍	❍	❍

Q7.4 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q7.5

	strongly disagree	disagree	neutral	agree	strongly agree
I would not mind having a serious romantic relationship with an Aborigine.	❍	❍	❍	❍	❍
It is just a matter of some people not trying hard enough. If Aboriginal Australians would only try harder they could be as well off as Anglo people.	❍	❍	❍	❍	❍

Q7.6

	not at all bothered	not bothered	neutral	bothered	very bothered
How bothered would you be if a child of yours had children with a person of very different colour and physical characteristics than your own, and your grandchildren did not physically resemble the people on your side of the family?	❍	❍	❍	❍	❍

Q7.7 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q7.8

	very similar	similar	neutral	different	very different
How different or similar are Aborigines living here to Anglo people in the values that they teach their children?	❍	❍	❍	❍	❍
How different or similar are Aborigines living here to Anglo people in their religious beliefs and practices?	❍	❍	❍	❍	❍

Q7.9

	never	rarely	neutral	often	very often
How often have you felt sympathy for Aborigines living here?	❍	❍	❍	❍	❍
How often have you felt admiration for Aborigines living here?	❍	❍	❍	❍	❍

Q8.1 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q8.2

	strongly disagree	disagree	neutral	agree	strongly agree
Africans have jobs that the Anglos should have.	❍	❍	❍	❍	❍
Most Africans living here who receive support from welfare could get along without it if they tried.	❍	❍	❍	❍	❍
Anglo people and Africans can never really be comfortable with each other, even if they are close friends.	❍	❍	❍	❍	❍
Most politicians in Australia care too much about Africans and not enough about the average Anglo person.	❍	❍	❍	❍	❍
Africans come from less able races and this explains why they are not as well off as most Anglo people.	❍	❍	❍	❍	❍

Q8.3

	very similar	similar	neutral	different	very different
How different or similar do you think Africans living here are to Anglo people in how honest they are?	❍	❍	❍	❍	❍

Q8.4 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q8.5

	strongly disagree	disagree	neutral	agree	strongly agree
I would not mind having a serious romantic relationship with an African.	❍	❍	❍	❍	❍
It is just a matter of some people not trying hard enough. If African Australians would only try harder they could be as well off as Anglo people.	❍	❍	❍	❍	❍

Q8.6

	not at all bothered	not bothered	neutral	bothered	very bothered
How bothered would you be if a child of yours had children with a person of very different colour and physical characteristics than your own, and your grandchildren did not physically resemble the people on your side of the family?	❍	❍	❍	❍	❍

Q8.7 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q8.8

	very similar	similar	neutral	different	very different
How different or similar are Africans living here to Anglo people in the values that they teach their children?	❍	❍	❍	❍	❍
How different or similar are Africans living here to Anglo people in their religious beliefs and practices?	❍	❍	❍	❍	❍

Q8.9

	never	rarely	neutral	often	very often
How often have you felt sympathy for Africans living here?	❍	❍	❍	❍	❍
How often have you felt admiration for Africans living here?	❍	❍	❍	❍	❍

Q9.1 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q9.2

	strongly disagree	disagree	neutral	agree	strongly agree
Arabs have jobs that the Anglos should have.	❍	❍	❍	❍	❍
Most Arabs living here who receive support from welfare could get along without it if they tried.	❍	❍	❍	❍	❍
Anglo people and Arabs can never really be comfortable with each other, even if they are close friends.	❍	❍	❍	❍	❍
Most politicians in Australia care too much about Arabs and not enough about the average Anglo person.	❍	❍	❍	❍	❍
Arabs come from less able races and this explains why they are not as well off as most Anglo people.	❍	❍	❍	❍	❍

Q9.3

	very similar	similar	neutral	different	very different
How different or similar do you think Arabs living here are to Anglo people in how honest they are?	❍	❍	❍	❍	❍

Q9.4 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q9.5

	strongly disagree	disagree	neutral	agree	strongly agree
I would not mind having a serious romantic relationship with an Arab.	❍	❍	❍	❍	❍
It is just a matter of some people not trying hard enough. If Arab Australians would only try harder they could be as well off as Anglo people.	❍	❍	❍	❍	❍

Q9.6

	not at all bothered	not bothered	neutral	bothered	very bothered
How bothered would you be if a child of yours had children with a person of very different colour and physical characteristics than your own, and your grandchildren did not physically resemble the people on your side of the family?	❍	❍	❍	❍	❍

Q9.7 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q9.8

	very similar	similar	neutral	different	very different
How different or similar are Arabs living here to Anglo people in the values that they teach their children?	❍	❍	❍	❍	❍
How different or similar are Arabs living here to Anglo people in their religious beliefs and practices?	❍	❍	❍	❍	❍

Q9.9

	never	rarely	neutral	often	very often
How often have you felt sympathy for Arabs living here?	❍	❍	❍	❍	❍
How often have you felt admiration for Arabs living here?	❍	❍	❍	❍	❍

Q10.1 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q10.2

	strongly disagree	disagree	neutral	agree	strongly agree
Asians have jobs that the Anglos should have.	❍	❍	❍	❍	❍
Most Asians living here who receive support from welfare could get along without it if they tried.	❍	❍	❍	❍	❍
Anglo people and Asians can never really be comfortable with each other, even if they are close friends.	❍	❍	❍	❍	❍
Most politicians in Australia care too much about Asians and not enough about the average Anglo person.	❍	❍	❍	❍	❍
Asians come from less able races and this explains why they are not as well off as most Anglo people.	❍	❍	❍	❍	❍

Q10.3

	very similar	similar	neutral	different	very different
How different or similar do you think Asians living here are to Anglo people in how honest they are?	❍	❍	❍	❍	❍

Q10.4 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q10.5

	strongly disagree	disagree	neutral	agree	strongly agree
I would not mind having a serious romantic relationship with an Asian.	❍	❍	❍	❍	❍
I would not mind if a suitably qualified Asian was appointed as my boss.	❍	❍	❍	❍	❍
I would not mind if an Asian who had a similar economic background as mine joined my close family by marriage.	❍	❍	❍	❍	❍
Asians living here should not push themselves where they are not wanted.	❍	❍	❍	❍	❍

Q10.6 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q10.7

	not at all bothered	not bothered	neutral	bothered	very bothered
How bothered would you be if a child of yours had children with a person of very different colour and physical characteristics than your own, and your grandchildren did not physically resemble the people on your side of the family?	❍	❍	❍	❍	❍

Q10.8

	strongly disagree	disagree	neutral	agree	strongly agree
Many other groups have come to Australia and overcome prejudice and worked their way up. Asians should do the same without special favour.	❍	❍	❍	❍	❍
It is just a matter of some people not trying hard enough. If Asian Australians would only try harder they could be as well off as Anglo people.	❍	❍	❍	❍	❍

Q10.9 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q10.10

	very similar	similar	neutral	different	very different
How different or similar are Asians living here to Anglo people in the values that they teach their children?	❍	❍	❍	❍	❍
How different or similar are Asians living here to Anglo people in their religious beliefs and practices?	❍	❍	❍	❍	❍
How different or similar are Asians living here to Anglo people in their sexual values or sexual practices?	❍	❍	❍	❍	❍
How different or similar are Asians living here to Anglo people in the language that they speak?	❍	❍	❍	❍	❍

Q10.11 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement.   PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus. PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc. PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc. PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q10.12

	strongly disagree	disagree	neutral	agree	strongly agree
Asians living here teach their children values and skills different from those required to be successful in Australia.	❍	❍	❍	❍	❍

Q10.13

	never	rarely	neutral	often	very often
How often have you felt sympathy for Asians living here?	❍	❍	❍	❍	❍
How often have you felt admiration for Asians living here?	❍	❍	❍	❍	❍

Q11.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q11.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, females aim to achieve too much.	❍	❍	❍	❍	❍
Females tend to have less fun compared to males.	❍	❍	❍	❍	❍
A lot of females can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of females tend to be active and chatty.	❍	❍	❍	❍	❍
Females are not very 'street smart.'	❍	❍	❍	❍	❍
Females know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most females are not very vocal.	❍	❍	❍	❍	❍
Females are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q11.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q11.4

	strongly disagree	disagree	neutral	agree	strongly agree
Females spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, females think they are smarter than males.	❍	❍	❍	❍	❍
Females enjoy a disproportionate amount of career success.	❍	❍	❍	❍	❍
Females are not as social as males.	❍	❍	❍	❍	❍
Females are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most females function well in social situations.	❍	❍	❍	❍	❍
Many females always seem to compare their own achievements to other females'.	❍	❍	❍	❍	❍
Females rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q12.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q12.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, young people aim to achieve too much.	❍	❍	❍	❍	❍
Young people tend to have less fun compared to old people.	❍	❍	❍	❍	❍
A lot of young people can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of young people tend to be active and chatty.	❍	❍	❍	❍	❍
Young people are not very 'street smart.'	❍	❍	❍	❍	❍
Young people know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most young people are not very vocal.	❍	❍	❍	❍	❍
Young people are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q12.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q12.4

	strongly disagree	disagree	neutral	agree	strongly agree
Young people spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, young people think they are smarter than old people.	❍	❍	❍	❍	❍
Young people enjoy a disproportionate amount of career success.	❍	❍	❍	❍	❍
Young people are not as social as old people.	❍	❍	❍	❍	❍
Young people are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most young people function well in social situations.	❍	❍	❍	❍	❍
Many young people always seem to compare their own achievements to other young people's.	❍	❍	❍	❍	❍
Young people rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q13.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians,

Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q13.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, Aboriginal Australians aim to achieve too much.	❍	❍	❍	❍	❍
Aboriginal Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍
A lot of Aboriginal Australians can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of Aboriginal Australians tend to be shy and quiet.	❍	❍	❍	❍	❍
Aboriginal Australians are not very 'street smart.'	❍	❍	❍	❍	❍
Aboriginal Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most Aboriginal Australians are not very vocal.	❍	❍	❍	❍	❍
Aboriginal Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q13.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians,

Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q13.4

	strongly disagree	disagree	neutral	agree	strongly agree
Aboriginal Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, Aboriginal Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍
Aboriginal Australians enjoy a disproportionate amount of career success.	❍	❍	❍	❍	❍
Aboriginal Australians are not as social as other groups of people.	❍	❍	❍	❍	❍
Aboriginal Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most Aboriginal Australians function well in social situations.	❍	❍	❍	❍	❍
Many Aboriginal Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍
Aboriginal Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q14.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q14.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, African Australians aim to achieve too much.	❍	❍	❍	❍	❍
African Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍
A lot of African Australians can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of African Australians tend to be shy and quiet.	❍	❍	❍	❍	❍
African Australians are not very 'street smart.'	❍	❍	❍	❍	❍
African Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most African Australians are not very vocal.	❍	❍	❍	❍	❍
African Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q14.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians,

Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q14.4

	strongly disagree	disagree	neutral	agree	strongly agree
African Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, African Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍
African Australians enjoy a disproportionate amount of career success.	❍	❍	❍	❍	❍
African Australians are not as social as other groups of people.	❍	❍	❍	❍	❍
African Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most African Australians function well in social situations.	❍	❍	❍	❍	❍
Many African Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍
African Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q15.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q15.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, Arab Australians aim to achieve too much.	❍	❍	❍	❍	❍
Arab Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍
A lot of Arab Australians can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of Arab Australians tend to be shy and quiet.	❍	❍	❍	❍	❍
Arab Australians are not very 'street smart.'	❍	❍	❍	❍	❍
Arab Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most Arab Australians are not very vocal.	❍	❍	❍	❍	❍
Arab Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q15.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.
Q15.4

	strongly disagree	disagree	neutral	agree	strongly agree
Arab Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, Arab Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍
Arab Australians enjoy a disproportionate amount of career success.	❍	❍	❍	❍	❍
Arab Australians are not as social as other groups of people.	❍	❍	❍	❍	❍
Arab Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most Arab Australians function well in social situations.	❍	❍	❍	❍	❍
Many Arab Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍
Arab Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q16.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q16.2

	strongly disagree	disagree	neutral	agree	strongly agree
When it comes to education, Asian Australians aim to achieve too much.	❍	❍	❍	❍	❍
Asian Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍
A lot of Asian Australians can be described as working all of the time.	❍	❍	❍	❍	❍
The majority of Asian Australians tend to be shy and quiet.	❍	❍	❍	❍	❍
Asian Australians are not very 'street smart.'	❍	❍	❍	❍	❍
Asian Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍
Most Asian Australians are not very vocal.	❍	❍	❍	❍	❍
Asian Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍

Q16.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians, Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q16.4

	strongly disagree	disagree	neutral	agree	strongly agree
Asian Australians seem to be striving to become number one.	❍	❍	❍	❍	❍
Asian Australians commit less time to socializing than others do.	❍	❍	❍	❍	❍
In order to get ahead of others, Asian Australians can be overly competitive.	❍	❍	❍	❍	❍
Asian Australians do not usually like to be the centre of attention at social gatherings.	❍	❍	❍	❍	❍
Most Asian Australians have a mentality that stresses gain of economic power.	❍	❍	❍	❍	❍
Asian Australians can sometimes be regarded as acting too smart.	❍	❍	❍	❍	❍
Asian Australians put high priority on their social lives.	❍	❍	❍	❍	❍

Asian Australians do not interact with others smoothly in social situations.	❍	❍	❍	❍	❍
As a group, Asian Australians are not constantly in pursuit of more power.	❍	❍	❍	❍	❍

Q16.5 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement. PS1: The term 'young' in this survey refers to 18 to 24 years old, and 'old' 65 plus.PS2: The term 'Africans' in this survey refers to, for example, Ethiopians, Sudanese, etc.PS3: The term 'Arabs' in this survey refers to, for example, Egyptians,

Jordanians, etc.PS4: The term 'Asians' in this survey refers to, for example, Chinese, Vietnamese, etc.

Q16.6

	strongly disagree	disagree	neutral	agree	strongly agree
Asian Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍
Oftentimes, Asian Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍
Asian Australians enjoy a disproportionate amount of economic success.	❍	❍	❍	❍	❍
Asian Australians are not as social as other groups of people.	❍	❍	❍	❍	❍
Asian Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍
Most Asian Australians function well in social situations.	❍	❍	❍	❍	❍
Many Asian Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍
Asian Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍

Q17.1 Please respond to the following statements according to the accompanying scale. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement.

Q17.2

	strongly disagree	disagree	neutral	agree	strongly agree
Our country needs a powerful leader, in order to destroy the radical and immoral currents prevailing in society today.	❍	❍	❍	❍	❍
The 'old-fashioned ways' and 'old-fashioned values' still show the best way to live.	❍	❍	❍	❍	❍
God's law about abortion, pornography and marriage must be strictly followed before it is too late, violations must be punished.	❍	❍	❍	❍	❍
It would be best if newspapers were censored so that people would not be able to get hold of destructive and disgusting material.	❍	❍	❍	❍	❍

Our forefathers ought to be honoured more for the way they have built our society, at the same time we ought to put an end to those forces destroying it.	❍	❍	❍	❍	❍
There are many radical, immoral people trying to ruin things; the society ought to stop them.	❍	❍	❍	❍	❍
Facts show that we have to be harder against crime and sexual immorality, in order to uphold law and order.	❍	❍	❍	❍	❍
If the society so wants, it is the duty of every true citizen to help eliminate the evil that poisons our country from within.	❍	❍	❍	❍	❍

Q17.3

	strongly disagree	disagree	neutral	agree	strongly agree
Our country needs free thinkers, who will have the courage to stand up against traditional ways, even if this upsets many people.	❍	❍	❍	❍	❍
Our society would be better off if we showed tolerance and understanding for untraditional values and opinions.	❍	❍	❍	❍	❍
The society needs to show openness towards people thinking differently, rather than a strong leader, the world is not particularly evil or dangerous.	❍	❍	❍	❍	❍
Many good people challenge the state, criticize the church and ignore 'the normal way of living.'	❍	❍	❍	❍	❍

People ought to put less attention to the Bible and religion, instead they ought to develop their own moral standards.	❍	❍	❍	❍	❍
It is better to accept bad literature than to censor it.	❍	❍	❍	❍	❍
The situation in the society today would be improved if troublemakers were treated with reason and humanity.	❍	❍	❍	❍	❍

Q18.1 Using the scale below, please indicate the extent to which you agree or disagree with each of the following statements by rating each statement on the following scale.

Q18.2

	strongly disagree	disagree	neutral	agree	strongly agree
You can substantially improve your intelligence.	❍	❍	❍	❍	❍
No matter who you are, you can significantly change your personality.	❍	❍	❍	❍	❍
No matter who you are, you can significantly change your morality level.	❍	❍	❍	❍	❍
You can substantially change your body weight.	❍	❍	❍	❍	❍

Q18.3

	strongly disagree	disagree	neutral	agree	strongly agree
You have a certain amount of intelligence, and you can't really do much to improve it.	❍	❍	❍	❍	❍
To be honest, you can't really change your personality.	❍	❍	❍	❍	❍
To be honest, you can't really change your morality.	❍	❍	❍	❍	❍
You have a certain amount of body weight (because of your bones and muscles), and you can't really do much to change it.	❍	❍	❍	❍	❍

Q19.1 Please think about how you feel right now and rate each item on the following scale.

Q19.2

	not at all	a little bit	somewhat	very much	extremely
I feel happy.	❍	❍	❍	❍	❍
I feel proud.	❍	❍	❍	❍	❍
I feel calm.	❍	❍	❍	❍	❍
I feel pleasant.	❍	❍	❍	❍	❍
I feel kind.	❍	❍	❍	❍	❍
I feel warm.	❍	❍	❍	❍	❍
I feel pleased.	❍	❍	❍	❍	❍
I feel affectionate.	❍	❍	❍	❍	❍
I feel encouraged.	❍	❍	❍	❍	❍

Q19.3

	not at all	a little bit	somewhat	very much	extremely
I feel discouraged.	❍	❍	❍	❍	❍
I feel embarrassed.	❍	❍	❍	❍	❍
I feel depressed.	❍	❍	❍	❍	❍
I feel ashamed.	❍	❍	❍	❍	❍
I feel unpleasant.	❍	❍	❍	❍	❍
I feel sad.	❍	❍	❍	❍	❍
I feel guilty.	❍	❍	❍	❍	❍
I feel humiliated.	❍	❍	❍	❍	❍
I feel tense.	❍	❍	❍	❍	❍

Q19.4

	not at all	a little bit	somewhat	very much	extremely
I feel that I am accomplished.	❍	❍	❍	❍	❍
I feel that I am successful.	❍	❍	❍	❍	❍
I feel fulfilled.	❍	❍	❍	❍	❍
I feel that I have self-worth.	❍	❍	❍	❍	❍
I feel confident.	❍	❍	❍	❍	❍
I feel productive.	❍	❍	❍	❍	❍

Q19.5

	not at all	a little bit	somewhat	very much	extremely
I feel that I am snobbish.	❍	❍	❍	❍	❍
I feel that I am pompous.	❍	❍	❍	❍	❍
I feel that I am stuck-up.	❍	❍	❍	❍	❍
I feel that I am conceited.	❍	❍	❍	❍	❍
I feel arrogant.	❍	❍	❍	❍	❍
I feel smug.	❍	❍	❍	❍	❍

Q20.1 Please tell us what you are thinking at this moment. There is, of course, no right answer for any statement. The best answer is what you feel is true of yourself at this moment. Be sure to answer all of the items, even if you are not certain of the best answer. Again, answer these questions as they are true for you RIGHT NOW.

Q20.2

	not at all	a little bit	somewhat	very much	extremely
I feel confident about my abilities.	❍	❍	❍	❍	❍
I feel satisfied with the way my body looks right now.	❍	❍	❍	❍	❍
I feel that others respect and admire me.	❍	❍	❍	❍	❍
I feel as smart as others.	❍	❍	❍	❍	❍
I feel good about myself.	❍	❍	❍	❍	❍
I am pleased with my appearance right now.	❍	❍	❍	❍	❍
I feel confident that I understand things.	❍	❍	❍	❍	❍

Q20.3

	not at all	a little bit	somewhat	very much	extremely
I am worried about whether I am regarded as a success or failure.	❍	❍	❍	❍	❍
I feel frustrated or rattled about my performance.	❍	❍	❍	❍	❍
I feel that I am having trouble understanding things that I read.	❍	❍	❍	❍	❍
I am dissatisfied with my weight.	❍	❍	❍	❍	❍
I feel self-conscious.	❍	❍	❍	❍	❍
I feel displeased with myself.	❍	❍	❍	❍	❍
I am worried about what other people think of me.	❍	❍	❍	❍	❍
I feel inferior to others at this moment.	❍	❍	❍	❍	❍
I feel unattractive.	❍	❍	❍	❍	❍
I feel concerned about the impression I am making.	❍	❍	❍	❍	❍
I feel that I have less scholastic ability right now than others.	❍	❍	❍	❍	❍
I feel like I'm not doing well.	❍	❍	❍	❍	❍

I am worried about looking foolish.	❍	❍	❍	❍	❍

Q21.1 Thank you for your time! Participation in this survey has brought you 3 credit points. Please click the >> button now in order to complete the survey.

Appendix 4b

(This is the version of the online survey that was used to collect the prejudice data reported in Chapter 4,

including the experience data reported in Chapters 3 and 4.

This version of the survey is a modified version of Appendix 4a)

https://uwsssap.co1.qualtrics.com/jfe/form/SV_6mLeE3nKjekxgAB

Q1.1 We are conducting an online survey about student perspectives. You will be asked to mainly rate items on given scales. Occasionally, you will be asked to answer some short-answer or multiple-choice questions. By participating, you will gain first-hand experience in psychological research.   All aspects of the survey, including results, will be confidential and only the researchers will have access to information on participants, and the identity of participants will not be disclosed. The findings from this survey will constitute part of a thesis and may be submitted for publication to a journal article, and presented at conferences. This survey should not provide any discomfort or harm to you, and participation is entirely voluntary: you are not obliged to be involved and - if you do participate - you can withdraw at any time without giving any reason and without any consequences.   This study has been approved by the University of Western Sydney Human Research Ethics Committee (H10467). If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Research Ethics Office (Tel: +61 2 4736 0229 or Email: humanethics@uws.edu.au). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.   This survey is optimized for you to do on your desktop or laptop. It will take you approximately half an hour to finish, please do not begin the survey until you have this amount of time available. Partially completed surveys will not be saved, and the survey page may automatically expire after some time of inactivity. Your consent to participate is given once you click on the option 'Yes, I agree to participate' below. NB1: The `back' button is not available for this survey, so please choose your answers carefully before clicking on the double arrow button to proceed to the next page. NB2: If, by any chance, you have already taken this survey before, please do not take it again.   Do you agree to participate in this survey?

❍ Yes, I agree to participate.
❍ No, I do not agree to participate.

If No, I do not agree to parti... Is Selected, Then Skip To End of Survey If Yes, I agree to participate. Is Selected, Then Skip To Full name

Q1.2 The following questions are for demographic purposes.

Q1.3 Full name

Q1.4 Date of birth (dd/mm/yyyy)

Q1.5 Please indicate your gender
❍ male
❍ female
❍ other

Q1.6 Ethnic origin (please select 1)
❍ European
❍ East and Southeast Asian (e.g., China, Japan, Korea, Vietnam)
❍ South Asian (e.g., India, Pakistan, Bangladesh, Sri Lanka)
❍ Middle Eastern
❍ African
❍ Latin, Central, and South American
❍ Pacific Islander
❍ Indigenous Australian (i.e., Aboriginal, Torres Strait Islander)
❍ other - please specify ____________________

Q1.7 Do/did you have a hearing impairment, or reading difficulties (e.g., difficulties learning to read), or

language development or speaking difficulties (e.g., delayed language development, stuttering, lisping, etc.)?
- ❍ yes
- ❍ no

Display This Question:
If Do/did you have a hearing impairment, or reading difficulties (e.g., difficulties learning to read), or language development or speaking difficulties (e.g., delayed language development, stuttering... Yes Is Selected

Q1.8 Please click any/all that apply (hold the Ctrl key as you click if you want to select multiple options)
- ❑ hearing
- ❑ reading
- ❑ speaking
- ❑ language

Q2.1 Where was your father born?

Q2.2 What is your father's native language?

Q2.3 Does your father speak another language apart from his native language?
- ❍ yes
- ❍ no

Display This Question:
If Does your father speak another language apart from his native language? Yes Is Selected

Q2.4 What language?

Q2.5 Where was your mother born?

Q2.6 What is your mother's native language?

Q2.7 Does your mother speak another language apart from her native language?
- ❍ yes
- ❍ no

Display This Question:
If Does your mother speak another language apart from her native language? Yes Is Selected

Q2.8 What language?

Q2.9 Where were you born? Please list the city, state, and country. If the country is Australia, specifically list the suburb. For example: Bankstown - Sydney, New South Wales, Australia

Q2.10 List cities (and countries) you've lived since you were born. If the country is Australia, specifically list the suburbs. For example: Madrid (Spain), West End - Brisbane (Australia), Bangkok (Thailand), Vancouver (Canada), Surry Hills - Sydney (Australia)

Q2.11 Spoken language: Please tell us what other languages you can speak besides English, at what age you started speaking the language and at what age you stopped (if you still speak the language, then put in the current age and the word 'now'), and how often and how well you spoke/speak it. If you don't speak any language besides English, please put 'n/a' into the cells. If you've already filled in 'n/a' into a row but the page keeps asking you to answer the question, please just click on 'Continue without answering'.

	Language	Age (start)	Age (end)	How often did/do you speak it? (1=rarely, 5=always)	How well do you speak it? (1=hardly at all, 5=highly fluent)
Example					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Q2.12 Have you ever tried the following cuisine? Please tick accordingly.

	yes	no
Chinese cuisine	❍	❍
Mexican cuisine	❍	❍
Vietnamese cuisine	❍	❍
Italian cuisine	❍	❍
Thai cuisine	❍	❍
Lebanese cuisine	❍	❍
Korean cuisine	❍	❍
French cuisine	❍	❍
Japanese cuisine	❍	❍
Indian cuisine	❍	❍

Display This Question:

If Have you ever tried the following cuisine? If yes, please tick accordingly. - yes Is Greater Than 0

Carry Forward Selected Choices from "Have you ever tried the following cuisine? Please tick accordingly."

Q2.13

	When you tried out the following cuisine, how much did you enjoy it?					In the past year, how many times did you try it? (please type in a number, for example, 5 to indicate 5 times)
	not at all	a little bit	somewhat	very much	extremely	
Chinese cuisine	❍	❍	❍	❍	❍	
Mexican cuisine	❍	❍	❍	❍	❍	
Vietnamese cuisine	❍	❍	❍	❍	❍	
Italian cuisine	❍	❍	❍	❍	❍	
Thai cuisine	❍	❍	❍	❍	❍	
Lebanese cuisine	❍	❍	❍	❍	❍	
Korean cuisine	❍	❍	❍	❍	❍	
French cuisine	❍	❍	❍	❍	❍	
Japanese cuisine	❍	❍	❍	❍	❍	
Indian cuisine	❍	❍	❍	❍	❍	

Q2.14 Were there any other people from a different city/town/region of Australia, or from a different country, who lived at your house or frequently spent time with you during your infant/childhood years (for example, grandmother watched over you while your parents were at work, a housekeeper or nanny)?

❍ yes
❍ no

Display This Question:
If Were there any other people from a different city/town/region of Australia, or from a different country, who lived at your house or frequently spent time with you during your infant/childhood years... Yes Is Selected

Q2.15 If there were any other people from a different city/town/region of Australia, or from a different country, who lived at your house or frequently spent time with you during your infant/childhood years (for example, grandmother watched over you while your parents were at work, a housekeeper or nanny),

who were they?
what country/city/town were they from?
what was their native language?
what ages were you when they were often around?

Q2.16 For each of the following foreign accents of English, please indicate whether and what type of experience you have had with speakers of the accent. For example, you may have had little to no real experience with it, or you may regularly watch a TV show in which the characters are from that country and speak that accent, or you may have a good friend or a family member or a co-worker at your job who speaks that accent, etc.For each foreign accent below, please indicate whether you have experience with it or not, and what type of experience you have had with it.

	Do you have experience with the following accent?		What type of experience have you had with the following accent?
	yes	no	
Chinese	❍	❍	
Mexican	❍	❍	
Vietnamese	❍	❍	
Italian	❍	❍	
Thai	❍	❍	
Lebanese	❍	❍	
Korean	❍	❍	
French	❍	❍	
Japanese	❍	❍	
Indian	❍	❍	

Q2.17 Have you ever travelled to these countries? Please tick accordingly.

	yes	no
China	❍	❍
Mexico	❍	❍
Vietnam	❍	❍
Italy	❍	❍
Thailand	❍	❍
Lebanon	❍	❍
Korea	❍	❍
France	❍	❍
Japan	❍	❍
India	❍	❍

Q2.18 Have you recently returned from a long period in another country or in another city in Australia, longer than 3 months?

❍ yes
❍ no

Display This Question:
If Have you recently returned from a long period in another country or in another city in Australia, longer than 3 months? Yes Is Selected

Q2.19 If so, which country/city did you visit, and when did you return to Sydney?

Q3.1 Here are a number of personality traits that may or may not apply to you. Please indicate the extent to which you agree or disagree with each statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.I see myself as:

	disagree strongly	disagree moderately	disagree a little	neither agree nor disagree	agree a little	agree moderately	agree strongly
____ extraverted, enthusiastic	❍	❍	❍	❍	❍	❍	❍
____ critical, quarrelsome	❍	❍	❍	❍	❍	❍	❍
____ dependable, self-disciplined	❍	❍	❍	❍	❍	❍	❍
____ anxious, easily upset	❍	❍	❍	❍	❍	❍	❍
____ open to new experiences, complex	❍	❍	❍	❍	❍	❍	❍
____ reserved, quiet	❍	❍	❍	❍	❍	❍	❍
____ sympathetic, warm	❍	❍	❍	❍	❍	❍	❍
____ disorganized, careless	❍	❍	❍	❍	❍	❍	❍
____ calm, emotionally stable	❍	❍	❍	❍	❍	❍	❍
____ conventional, uncreative	❍	❍	❍	❍	❍	❍	❍

Q4.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Aboriginal Australians seem to be striving to become number one.	❍	❍	❍	❍	❍	❍
Aboriginal Australians commit less time to socializing than others do.	❍	❍	❍	❍	❍	❍
In order to get ahead of others, Aboriginal Australians can be overly competitive.	❍	❍	❍	❍	❍	❍
Aboriginal Australians do not usually like to be the centre of attention at social gatherings.	❍	❍	❍	❍	❍	❍
Most Aboriginal Australians have a mentality that stresses gain of economic power.	❍	❍	❍	❍	❍	❍
Aboriginal Australians can sometimes be regarded as acting too smart.	❍	❍	❍	❍	❍	❍
Aboriginal Australians put high priority on their social lives.	❍	❍	❍	❍	❍	❍
Aboriginal Australians do not interact with others smoothly in social situations.	❍	❍	❍	❍	❍	❍
As a group, Aboriginal Australians are not constantly in pursuit of more power.	❍	❍	❍	❍	❍	❍

Q4.2 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
When it comes to education, Aboriginal Australians aim to achieve too much.	❍	❍	❍	❍	❍	❍
Aboriginal Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍	❍
A lot of Aboriginal Australians can be described as working all of the time.	❍	❍	❍	❍	❍	❍
The majority of Aboriginal Australians tend to be shy and quiet.	❍	❍	❍	❍	❍	❍
Aboriginal Australians are not very 'street smart.'	❍	❍	❍	❍	❍	❍
Aboriginal Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍	❍
Most Aboriginal Australians are not very vocal.	❍	❍	❍	❍	❍	❍
Aboriginal Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍	❍

Q4.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Aboriginal Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍	❍
Oftentimes, Aboriginal Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍	❍
Aboriginal Australians enjoy a disproportionate amount of economic success.	❍	❍	❍	❍	❍	❍
Aboriginal Australians are not as social as other groups of people.	❍	❍	❍	❍	❍	❍
Aboriginal Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍	❍
Most Aboriginal Australians function well in social situations.	❍	❍	❍	❍	❍	❍
Many Aboriginal Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍	❍
Aboriginal Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍	❍

Q5.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Anglo Australians seem to be striving to become number one.	❍	❍	❍	❍	❍	❍
Anglo Australians commit less time to socializing than others do.	❍	❍	❍	❍	❍	❍
In order to get ahead of others, Anglo Australians can be overly competitive.	❍	❍	❍	❍	❍	❍
Anglo Australians do not usually like to be the centre of attention at social gatherings.	❍	❍	❍	❍	❍	❍
Most Anglo Australians have a mentality that stresses gain of economic power.	❍	❍	❍	❍	❍	❍
Anglo Australians can sometimes be regarded as acting too smart.	❍	❍	❍	❍	❍	❍

Anglo Australians put high priority on their social lives.	❍	❍	❍	❍	❍	❍
Anglo Australians do not interact with others smoothly in social situations.	❍	❍	❍	❍	❍	❍
As a group, Anglo Australians are not constantly in pursuit of more power.	❍	❍	❍	❍	❍	❍

Q5.2 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
When it comes to education, Anglo Australians aim to achieve too much.	❍	❍	❍	❍	❍	❍
Anglo Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍	❍
A lot of Anglo Australians can be described as working all of the time.	❍	❍	❍	❍	❍	❍
The majority of Anglo Australians tend to be shy and quiet.	❍	❍	❍	❍	❍	❍
Anglo Australians are not very 'street smart.'	❍	❍	❍	❍	❍	❍
Anglo Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍	❍
Most Anglo Australians are not very vocal.	❍	❍	❍	❍	❍	❍

Anglo Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍	❍

Q5.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Anglo Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍	❍
Oftentimes, Anglo Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍	❍
Anglo Australians enjoy a disproportionate amount of economic success.	❍	❍	❍	❍	❍	❍
Anglo Australians are not as social as other groups of people.	❍	❍	❍	❍	❍	❍
Anglo Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍	❍
Most Anglo Australians function well in social situations.	❍	❍	❍	❍	❍	❍
Many Anglo Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍	❍
Anglo Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍	❍

Q6.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Arab Australians seem to be striving to become number one.	❍	❍	❍	❍	❍	❍
Arab Australians commit less time to socializing than others do.	❍	❍	❍	❍	❍	❍
In order to get ahead of others, Arab Australians can be overly competitive.	❍	❍	❍	❍	❍	❍
Arab Australians do not usually like to be the centre of attention at social gatherings.	❍	❍	❍	❍	❍	❍
Most Arab Australians have a mentality that stresses gain of economic power.	❍	❍	❍	❍	❍	❍
Arab Australians can sometimes be regarded as acting too smart.	❍	❍	❍	❍	❍	❍

Arab Australians put high priority on their social lives.	❍	❍	❍	❍	❍	❍
Arab Australians do not interact with others smoothly in social situations.	❍	❍	❍	❍	❍	❍
As a group, Arab Australians are not constantly in pursuit of more power.	❍	❍	❍	❍	❍	❍

Q6.2 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
When it comes to education, Arab Australians aim to achieve too much.	❍	❍	❍	❍	❍	❍
Arab Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍	❍
A lot of Arab Australians can be described as working all of the time.	❍	❍	❍	❍	❍	❍
The majority of Arab Australians tend to be shy and quiet.	❍	❍	❍	❍	❍	❍
Arab Australians are not very 'street smart.'	❍	❍	❍	❍	❍	❍
Arab Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍	❍
Most Arab Australians are not very vocal.	❍	❍	❍	❍	❍	❍

Arab Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍	❍

Q6.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Arab Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍	❍
Oftentimes, Arab Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍	❍
Arab Australians enjoy a disproportionate amount of economic success.	❍	❍	❍	❍	❍	❍
Arab Australians are not as social as other groups of people.	❍	❍	❍	❍	❍	❍
Arab Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍	❍
Most Arab Australians function well in social situations.	❍	❍	❍	❍	❍	❍
Many Arab Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍	❍
Arab Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍	❍

Q7.1 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Asian Australians seem to be striving to become number one.	❍	❍	❍	❍	❍	❍
Asian Australians commit less time to socializing than others do.	❍	❍	❍	❍	❍	❍
In order to get ahead of others, Asian Australians can be overly competitive.	❍	❍	❍	❍	❍	❍
Asian Australians do not usually like to be the centre of attention at social gatherings.	❍	❍	❍	❍	❍	❍
Most Asian Australians have a mentality that stresses gain of economic power.	❍	❍	❍	❍	❍	❍
Asian Australians can sometimes be regarded as acting too smart.	❍	❍	❍	❍	❍	❍

Asian Australians put high priority on their social lives.	❍	❍	❍	❍	❍	❍
Asian Australians do not interact with others smoothly in social situations.	❍	❍	❍	❍	❍	❍
As a group, Asian Australians are not constantly in pursuit of more power.	❍	❍	❍	❍	❍	❍

Q7.2 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
When it comes to education, Asian Australians aim to achieve too much.	❍	❍	❍	❍	❍	❍
Asian Australians tend to have less fun compared to other social groups.	❍	❍	❍	❍	❍	❍
A lot of Asian Australians can be described as working all of the time.	❍	❍	❍	❍	❍	❍
The majority of Asian Australians tend to be shy and quiet.	❍	❍	❍	❍	❍	❍
Asian Australians are not very 'street smart.'	❍	❍	❍	❍	❍	❍
Asian Australians know how to have fun and can be pretty relaxed.	❍	❍	❍	❍	❍	❍
Most Asian Australians are not very vocal.	❍	❍	❍	❍	❍	❍

Asian Australians are a group not obsessed with competition.	❍	❍	❍	❍	❍	❍

Q7.3 Below are a number of statements with which you will agree or disagree. There are absolutely no right

or wrong answers. Use the specified scale to indicate your response to each statement.

	strongly disagree	moderately disagree	slightly disagree	slightly agree	moderately agree	strongly agree
Asian Australians spend a lot of time at social gatherings.	❍	❍	❍	❍	❍	❍
Oftentimes, Asian Australians think they are smarter than everyone else is.	❍	❍	❍	❍	❍	❍
Asian Australians enjoy a disproportionate amount of economic success.	❍	❍	❍	❍	❍	❍
Asian Australians are not as social as other groups of people.	❍	❍	❍	❍	❍	❍
Asian Australians are motivated to obtain too much power in our society.	❍	❍	❍	❍	❍	❍
Most Asian Australians function well in social situations.	❍	❍	❍	❍	❍	❍
Many Asian Australians always seem to compare their own achievements to other people's.	❍	❍	❍	❍	❍	❍
Asian Australians rarely initiate social events or gatherings.	❍	❍	❍	❍	❍	❍

Q8.1 Does the Aboriginal Australian group make your group feel____?

	not at all	a little bit	somewhat	very much	extremely
sympathetic	❍	❍	❍	❍	❍
envious	❍	❍	❍	❍	❍
uneasy	❍	❍	❍	❍	❍
proud	❍	❍	❍	❍	❍
angry	❍	❍	❍	❍	❍
disgusted	❍	❍	❍	❍	❍
respectful	❍	❍	❍	❍	❍
pitying	❍	❍	❍	❍	❍
hateful	❍	❍	❍	❍	❍
frustrated	❍	❍	❍	❍	❍
jealous	❍	❍	❍	❍	❍
admiring	❍	❍	❍	❍	❍
resentful	❍	❍	❍	❍	❍
inspired	❍	❍	❍	❍	❍
contemptuous	❍	❍	❍	❍	❍
ashamed	❍	❍	❍	❍	❍
fond	❍	❍	❍	❍	❍

Q9.1 Does the Anglo Australian group make your group feel____?

	not at all	a little bit	somewhat	very much	extremely
sympathetic	❍	❍	❍	❍	❍
envious	❍	❍	❍	❍	❍
uneasy	❍	❍	❍	❍	❍
proud	❍	❍	❍	❍	❍
angry	❍	❍	❍	❍	❍
disgusted	❍	❍	❍	❍	❍
respectful	❍	❍	❍	❍	❍
pitying	❍	❍	❍	❍	❍
hateful	❍	❍	❍	❍	❍
frustrated	❍	❍	❍	❍	❍
jealous	❍	❍	❍	❍	❍
admiring	❍	❍	❍	❍	❍
resentful	❍	❍	❍	❍	❍
inspired	❍	❍	❍	❍	❍
contemptuous	❍	❍	❍	❍	❍
ashamed	❍	❍	❍	❍	❍
fond	❍	❍	❍	❍	❍

Q10.1 Does the Arab Australian group make your group feel____?

	not at all	a little bit	somewhat	very much	extremely
sympathetic	❍	❍	❍	❍	❍
envious	❍	❍	❍	❍	❍
uneasy	❍	❍	❍	❍	❍
proud	❍	❍	❍	❍	❍
angry	❍	❍	❍	❍	❍
disgusted	❍	❍	❍	❍	❍
respectful	❍	❍	❍	❍	❍
pitying	❍	❍	❍	❍	❍
hateful	❍	❍	❍	❍	❍
frustrated	❍	❍	❍	❍	❍
jealous	❍	❍	❍	❍	❍
admiring	❍	❍	❍	❍	❍
resentful	❍	❍	❍	❍	❍
inspired	❍	❍	❍	❍	❍
contemptuous	❍	❍	❍	❍	❍
ashamed	❍	❍	❍	❍	❍
fond	❍	❍	❍	❍	❍

Q11.1 Does the Asian Australian group make your group feel____?

	not at all	a little bit	somewhat	very much	extremely
sympathetic	❍	❍	❍	❍	❍
envious	❍	❍	❍	❍	❍
uneasy	❍	❍	❍	❍	❍
proud	❍	❍	❍	❍	❍
angry	❍	❍	❍	❍	❍
disgusted	❍	❍	❍	❍	❍
respectful	❍	❍	❍	❍	❍
pitying	❍	❍	❍	❍	❍
hateful	❍	❍	❍	❍	❍
frustrated	❍	❍	❍	❍	❍
jealous	❍	❍	❍	❍	❍
admiring	❍	❍	❍	❍	❍
resentful	❍	❍	❍	❍	❍
inspired	❍	❍	❍	❍	❍
contemptuous	❍	❍	❍	❍	❍
ashamed	❍	❍	❍	❍	❍
fond	❍	❍	❍	❍	❍

Q12.1 Please indicate the degree to which you like Aboriginal Australians:

	1	2	3	4	5	6	7	8	9	10
no liking at all:extreme liking	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

Q13.1 Please indicate the degree to which you like Anglo Australians:

	1	2	3	4	5	6	7	8	9	10
no liking at all:extreme liking	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

Q14.1 Please indicate the degree to which you like Arab Australians:

	1	2	3	4	5	6	7	8	9	10
no liking at all:extreme liking	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

Q15.1 Please indicate the degree to which you like Asian Australians:

	1	2	3	4	5	6	7	8	9	10
no liking at all:extreme liking	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

Q16.1 Using the scale below as a guide, indicate how much you agree with each statement.

	not true			somewhat true			very true
My first impressions of people usually turn out to be right.	❍	❍	❍	❍	❍	❍	❍
It would be hard for me to break any of my bad habits.	❍	❍	❍	❍	❍	❍	❍
I don't care to know what other people really think of me.	❍	❍	❍	❍	❍	❍	❍
I have not always been honest with myself.	❍	❍	❍	❍	❍	❍	❍
I always know why I like things.	❍	❍	❍	❍	❍	❍	❍
When my emotions are aroused, they bias my thinking.	❍	❍	❍	❍	❍	❍	❍
Once I've made up my mind, other people can seldom change my opinion.	❍	❍	❍	❍	❍	❍	❍
I am not a safe driver when I exceed the speed limit.	❍	❍	❍	❍	❍	❍	❍

I am fully in control of my own fate.	❍	❍	❍	❍	❍	❍	❍
It's hard for me to shut off a disturbing thought.	❍	❍	❍	❍	❍	❍	❍

Q16.2 Using the scale below as a guide, indicate how much you agree with each statement.

	not true			somewhat true			very true
I never regret my decisions.	❍	❍	❍	❍	❍	❍	❍
I sometimes lose out on things because I can't make up my mind soon enough.	❍	❍	❍	❍	❍	❍	❍
The reason I vote is because my vote can make a difference.	❍	❍	❍	❍	❍	❍	❍
My parents were not always fair when they punished me.	❍	❍	❍	❍	❍	❍	❍
I am a completely rational person.	❍	❍	❍	❍	❍	❍	❍
I rarely appreciate criticism.	❍	❍	❍	❍	❍	❍	❍
I am very confident of my judgments.	❍	❍	❍	❍	❍	❍	❍
I have sometimes doubted my ability as a lover.	❍	❍	❍	❍	❍	❍	❍

It's all right with me if some people happen to dislike me.	❍	❍	❍	❍	❍	❍	❍
I don't always know the reasons why I do the things I do.	❍	❍	❍	❍	❍	❍	❍

Q16.3 Using the scale below as a guide, indicate how much you agree with each statement.

	not true			somewhat true			very true
I sometimes tell lies if I have to.	❍	❍	❍	❍	❍	❍	❍
I never cover up my mistakes.	❍	❍	❍	❍	❍	❍	❍
There have been occasions when I have taken advantage of someone.	❍	❍	❍	❍	❍	❍	❍
I never swear.	❍	❍	❍	❍	❍	❍	❍
I sometimes try to get even rather than forgive and forget.	❍	❍	❍	❍	❍	❍	❍
I always obey laws, even if I'm unlikely to get caught.	❍	❍	❍	❍	❍	❍	❍
I have said something bad about a friend behind his or her back.	❍	❍	❍	❍	❍	❍	❍
When I hear people talking privately, I avoid listening.	❍	❍	❍	❍	❍	❍	❍

I have received too much change from a salesperson without telling him or her.	❍	❍	❍	❍	❍	❍	❍
I always declare everything at customs.	❍	❍	❍	❍	❍	❍	❍

Q16.4 Using the scale below as a guide, indicate how much you agree with each statement.

	not true			somewhat true			very true
When I was young, I sometimes stole things.	❍	❍	❍	❍	❍	❍	❍
I have never dropped litter on the street.	❍	❍	❍	❍	❍	❍	❍
I sometimes drive faster than the speed limit.	❍	❍	❍	❍	❍	❍	❍
I never read sexy books or magazines.	❍	❍	❍	❍	❍	❍	❍
I have done things that I don't tell other people about.	❍	❍	❍	❍	❍	❍	❍
I never take things that don't belong to me.	❍	❍	❍	❍	❍	❍	❍
I have taken sick-leave from work or school even though I wasn't really sick.	❍	❍	❍	❍	❍	❍	❍
I have never damaged a library book or store merchandise without reporting it.	❍	❍	❍	❍	❍	❍	❍
I have some pretty awful habits.	❍	❍	❍	❍	❍	❍	❍

I don't gossip about other people's business.	❍	❍	❍	❍	❍	❍	❍

Q17.1 Please think about how you feel right now and rate each item on the following scale.

	not at all						very much
I feel happy.	❍	❍	❍	❍	❍	❍	❍
I feel proud.	❍	❍	❍	❍	❍	❍	❍
I feel calm.	❍	❍	❍	❍	❍	❍	❍
I feel pleasant.	❍	❍	❍	❍	❍	❍	❍
I feel kind.	❍	❍	❍	❍	❍	❍	❍
I feel warm.	❍	❍	❍	❍	❍	❍	❍
I feel pleased.	❍	❍	❍	❍	❍	❍	❍
I feel affectionate.	❍	❍	❍	❍	❍	❍	❍
I feel encouraged.	❍	❍	❍	❍	❍	❍	❍

Q17.2 Please think about how you feel right now and rate each item on the following scale.

	not at all						very much
I feel discouraged.	❍	❍	❍	❍	❍	❍	❍
I feel embarrassed.	❍	❍	❍	❍	❍	❍	❍
I feel depressed.	❍	❍	❍	❍	❍	❍	❍
I feel ashamed.	❍	❍	❍	❍	❍	❍	❍
I feel unpleasant.	❍	❍	❍	❍	❍	❍	❍
I feel sad.	❍	❍	❍	❍	❍	❍	❍
I feel guilty.	❍	❍	❍	❍	❍	❍	❍
I feel humiliated.	❍	❍	❍	❍	❍	❍	❍
I feel tense.	❍	❍	❍	❍	❍	❍	❍

Q18.1 Please think about how you feel right now and rate each item on the following scale.

	not at all	a little bit	somewhat	very much	extremely
I feel that I am accomplished.	❍	❍	❍	❍	❍
I feel that I am successful.	❍	❍	❍	❍	❍
I feel fulfilled.	❍	❍	❍	❍	❍
I feel that I have self-worth.	❍	❍	❍	❍	❍
I feel confident.	❍	❍	❍	❍	❍
I feel productive.	❍	❍	❍	❍	❍

Q18.2 Please think about how you feel right now and rate each item on the following scale.

	not at all	a little bit	somewhat	very much	extremely
I feel that I am snobbish.	❍	❍	❍	❍	❍
I feel that I am pompous.	❍	❍	❍	❍	❍
I feel that I am stuck-up.	❍	❍	❍	❍	❍
I feel that I am conceited.	❍	❍	❍	❍	❍
I feel arrogant.	❍	❍	❍	❍	❍
I feel smug.	❍	❍	❍	❍	❍

Q19.1 Has anyone told you about this study before?

❍ yes

❍ no

Q20.1 Thank you for your time! Please click the >> button now in order to complete the survey.

Appendix 5:

The Scale of Anti-Asian American Stereotypes (SAAAS) as an appropriate tool to measure variability in affective attitudes towards Asians in Australia

1 Introduction

1.1 What do we know about attitude?

In Social Psychology, an attitude is defined as “an organized and consistent manner of thinking, feeling, and reacting with regard to people, groups, social issues, or, more generally, any event in one’s environment” (Lambert & Lambert, 1964, p.50). This definition of attitude reflects the components of attitudes, whether there are three of them (as in this definition: (1) thinking, (2) feeling, and (3) reacting) or less than three (as discussed in the paragraph that follows). Attitude, however, can also be defined without referring to its respective components (e.g., Eagly & Chaiken, 1998; Stephan, 1985; Zanna & Rempel, 1988; among others). In Eagly and Chaiken (1998), attitude is defined as “a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (p. 269). It can be measured on semantic differential items such as *unfavorable – favorable* (Stangor, Sullivan, & Ford, 1991) or *extremely negative – extremely positive* (Stephan & Stephan, 1993).

There are different views which conceptualize attitude in terms of its components. The first view reflects the classic tripartite view of attitude, which claims that an attitude is made up of three components: cognitive (“thoughts and beliefs”), affective (“feelings (or emotions)”), and conative (“tendencies to react”) (e.g., Katz & Stotland, 1959; Lambert & Lambert, 1964; Rosenberg & Hovland, 1960). Although this tripartite view reflects the three aspects of human experience according to Greek philosophies (McGuire, 1969) and is popular in Social Psychology textbooks (Breckler, 1984; Zanna & Rempel, 1988), it makes debatable assumptions such as that all three components must be manifested in an attitude and there must be consistency between them (Fazio & Olson, 2003). The view also assumes the existence of a correspondence between attitude and behaviour (Fazio & Olson, 2003; Zanna & Rempel, 1988). The other dominant definition of attitude reflects the unitary view. It defines attitude as having one single component: either cognitive (Fishbein & Ajzen, 1975), affective (Zajonc, 1980), or conative (Bem, 1972). By stressing on one attitude component only, this view strips away the problematic assumptions made by the tripartite view (Zanna & Rempel, 1988). However, it has its own problem: over-simplifying the attitude concept (Zanna & Rempel, 1988). As these two influential definitions are not satisfactory, several other views that attempt to solve the conflicts between the two views have been proposed. One alternative perspective is the two-component view that acknowledges the cognitive and affective components only (e.g., Zajonc & Markus, 1982). Another alternative perspective falls somewhere between the tripartite view and the unitary view (e.g., Eagly & Chaiken, 1998; Esses, Haddock, & Zanna, 1993; Petty & Wegener, 1998; Zanna & Rempel, 1988). In essence, this perspective is similar to the tripartite view in that it acknowledges the three components of attitudes, but similar to the unitary view in that it relaxes the assumptions made by the tripartite view (Fazio & Olson, 2003; Zanna & Rempel, 1988).

Overall, it seems social psychologists agree that the three attitude components exist; however, there is debate about the degree to which they acknowledge the three components as part of the attitude concept. This thesis therefore follows the view that an attitude consists of the cognitive, affective, and conative components. In discussing attitude towards another group "perceived to differ significantly from one's own" (Fiske, 1998, p. 357), or intergroup attitudes, the three components are termed: stereotypes, prejudice[29], and discrimination accordingly (e.g., Fiske, 1998).

The conceptualization of the three individual attitude components is generally agreed upon, with the exception of the affective component, which has been conceptualized in several different ways. The affective component of attitude commonly includes feelings, emotions, moods, and the like (e.g., Fiske, 1998; Lambert & Lambert, 1964; Stephan & Stephan, 1993). However, to the extent that those feelings, emotions, and moods are valenced, this valence carries an evaluative element to it. In other words, the evaluation is intertwined with the valence. Therefore, another way to conceptualize the affective component of attitude is that, apart from feelings, emotions, and moods, the affective component also includes evaluations (Fiske, 1998), perhaps as its cognitive representation (Stephan & Stephan, 1993). As mentioned in the beginning paragraph, evaluations are also considered attitude in and of themselves (e.g., Eagly & Chaiken, 1998; Stephan, 1985; Zanna & Rempel, 1988; among others). The question of whether evaluations are attitude itself or whether evaluations are only secondary to the affective component of attitude has not been resolved. Because the scale used to measure the affective component (described later in this document) was constructed by the social psychologists who considered evaluations part of the affective component, the view of evaluations that this thesis follows is that evaluations belong to the affective component, not the whole attitude. That is to say, in this thesis, the affective component of attitude includes feelings, emotions, moods, and evaluations.

[29] Note that some social psychologists use the term "prejudice" to refer to an entire attitude itself rather than simply its affective component (e.g., Allport, 1954; Stangor et al., 1991).

1.2 Explicit versus implicit measures of attitudes

Attitude components can be measured explicitly or implicitly. Explicit measure of attitude often involve self-reports, using ratings on semantic differentials, feeling thermometers, published scales, or scales constructed by the authors of the studies (see Karpinski & Hilton, 2001; McConnell & Leibold, 2001; Neto, 2009; Pantos & Perkins, 2012). These are introspective and claimed to be "direct, deliberate, controlled, and intentional self-assessments" (Nosek, Hawkins, & Frazier, 2011, p. 153). In contrast, implicit measures such as the Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998) and the Go/No-Go association task (Nosek & Banaji, 2001) often measure reaction time (but see Ball, 1983, for an implicit measure not involving reaction time). Implicit measures of attitude are indirect, spontaneous, automatic, and unintentional (but do not necessarily tap into the same construct as the explicit measures are thought to tap into; see the paragraph below). Implicit attitude is defined as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action towards social objects" (Greenwald & Banaji, 1995, p. 8).Both explicit and implicit measures have been claimed to be sensitive to the context of the measures such as the experimental environment, the standards given to the participants to make evaluations, and the background of stimulus presentation (Fazio & Olson, 2003a). Since Greenwald and Banaji's (1995) promotion of the use of implicit measures, the years after have seen a rise in the development and refinement of these types of measures (e.g., Fazio & Olson, 2003a; LeBel & Paunonen, 2011; Nosek et al., 2011; among others). Implicit measures are sometimes preferred over explicit measures for many practical and theoretical reasons, including their capacity to obtain responses relatively free from participants' social desirability concerns (Fazio & Olson, 2003a; Uhlmann et al., 2012). Nevertheless, there is still uncertainty about their validity (Nosek et al., 2011). In addition, implicit measures have been shown to have lower reliability than explicit measures (Fazio & Olson, 2003a), and findings from studies that have used implicit measures have lower replicability than studies that have used explicit measures (LeBel & Paunonen, 2011).

Another controversy in Social Psychology that involves explicit and implicit measures is whether they measure (1) the same underlying representation of attitude (e.g., Fazio, 1990, 2007; Perugini et al., 2010), or (2) two different underlying representations (i.e., explicit attitude measured by explicit measures and implicit attitude measured by implicit ones) (e.g., Greenwald & Banaji, 1995; Strack & Deutsch, 2004). The empirical evidence for these two possibilities is mixed: Evidence supporting (1) was found by Banse, Seise, and Zerbes (2001) and McConnell and Leibold (2001), and that supporting (2) was found by Karpinski and Hilton (2001), Neto (2009), and Pantos and Perkins (2012). Nevertheless, Fazio and Olson (2003a) discourage the view of two different types of attitude because, according to them, there is not enough evidence that the division between "explicit" and "implicit" exists for attitude. Attitude could be analogous to icebergs in which explicit measures assess the portion above the surface of the water and implicit measures assess the part below the surface of the water (as reviewed by Karpinski & Hilton, 2001). Recently, there has been correlational and experimental evidence that attitudes assessed via explicit and implicit measures are more strongly associated when the focus is on the affective component of attitude, as opposed to when the focus is on the cognitive component (e.g., Banse et al., 2001; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Smith & Nosek, 2011). This finding seems to suggest a third view that explicit and implicit measures assess the same underlying affective component of attitude, but they measure different underlying cognitive components. As this thesis aims to examine the relationship between the affective component of attitude, or prejudice, and speech perception, this third view has a potential implication for the selection of an appropriate tool to investigate the affective component: If the selection of either explicit or implicit measures does not matter for the investigation of the affective component, then the selection of an appropriate tool for prejudice measure in this thesis depends on other factors such as the compatibility of the attitude tool with the other experimental tools in terms of the similarity of mental processes (e.g., controlled vs. automatic processes) that they tap in participants.

Gender, age, and race are the three physical and social categories that are the most studied in stereotyping, prejudice, and discrimination research (Fiske, 1998). The focus of this thesis is between groups with different ethnic backgrounds in Australia. Therefore, of the categories of gender, age, and race, a review of racial attitude research in Australia is the most relevant. In addition, as the stimuli in the speech perception tasks reported in the content chapters involved Asian-accented speech, the following review will focus on racial attitudes towards Asians (rather than the other minority groups such as Aboriginal Australians or Arabs).

1.3 Racial attitude research on Asians in Australia

The appearance of Asians in Australia (i.e., Chinese gold miners) dates back to as early as the mid-1800s, and the end of the White Australia policy towards the end of the 1900s facilitated an influx of Asian immigrants including Vietnamese refugees. Research into White Australians' attitudes towards minority groups in Australia started at around the end of the White Australia policy (Western, 1969), but it was not until the influx of Asian immigrants that researchers began to take notice of the relationship between White Australians and Asians (e.g., Islam & Jahjah, 2001; McGrane & White, 2007; Walker, 1994; Wan, Crookes, Reynolds, Irons, & McKone, 2015). While researchers identified White Australians' prejudice towards Asians when the influx of immigrants arrived, there are still few published studies on the topic.

Many of the tools developed to measure prejudice in other communities have yet to be adapted to the Australian context. One of the few studies that did adapt existing prejudice measures to the Australian context was conducted by Islam and Jahjah (2001). The prejudice measure used in the study is very direct. It involves asking participants to rate their emotions in relation to the target minority groups (Aboriginal Australians, Asians, and Arabs). However, social norms have changed such that participants are no longer willing to endorse the kinds of blatant attitude that they were in the Katz and Braly's (1933) days, but have since moved to more subtle or "modern" forms of racism (McConahay, 1986). The direct self-report method in Islam and Jahjah (2001), therefore, poses the risk of collecting non-genuine responses from participants due to participants' concerns about the social desirability of their responses (Orne, 1959). This thesis constitutes the first attempt to modify measures about Asians to the Australian context that are designed for more indirect assessment of prejudice (e.g., instead of directly assessing prejudice, the measures would assess stereotypes that have been associated with various affective items and evaluations in past research). This would reduce the possibility of social desirability concerns in participants.

One such tool is the Scale of Anti-Asian American Stereotypes, or SAAAS, developed by Lin, Kwan, Cheung, and Fiske (2005). The scale consists of 25 items tapping stereotypes organized around two orthogonal dimensions: Competence and Sociability (see Figure A5.1). Response options are *strongly disagree* (coded as 0), *moderately disagree* (1), *slightly disagree* (2), *slightly agree* (3), *moderately agree* (4), and *strongly agree* (5). Six items are reverse-coded. SAAAS was constructed based on the Stereotype Content Model (SCM; Fiske, Cuddy, Glick, & Xu, 2002). The model suggests a link between stereotypes and prejudice by claiming that the combination of high Competence and low Warmth stereotypes attributed to a group is associated with a combination of feelings of admiration for and envy towards that group, which in turn results in a disliked evaluation of that group (i.e., prejudice). Evidence for the Stereotype Content Model comes from a correlational study where participants were asked to rate 24 items indicating their emotions (grouped into four factors: admiration, contempt, envy, and pity) towards 24 social groups (grouped onto the four quadrants of the Competence and Warmth dimensions; Fiske et al., 2002). Results showed that the six groups that were mapped onto the high Competence and low Warmth quadrant (i.e., rich people, men, Jews, Asians, professionals, educated people) indeed received the highest mean ratings for the "admiration" and "envy" feelings, as predicted by the SCM. When the SCM was translated into SAAAS, the Competence dimension was named in the same way but the Warmth dimension was changed to be called Sociability. The Sociability items of SAAAS were designed such that high Sociability scores indicate low Sociability (or Unsociability), and high total SAAAS scores (i.e., sums of scores of all SAAAS items) indicate negative prejudice. In other words, SAAAS prejudice is measured in an indirect way, which could mitigate the social desirability side effect[30] in other prejudice measures. In short, prejudice investigated in an indirect way and the theoretical underpinnings of SAAAS make the scale promising for prejudice research in the Australian context.

2 Methods

2.1 Participants

The Qualtrics survey link was opened 543 times by participants in the undergraduate Psychology pool of Western Sydney University; it was not finished 148 times. The survey was finished 59 times by people who had more than one attempt. To avoid any confounding effects that might arise from participants' responses to the survey on more than one attempt, only responses from 336 participants who finished the survey on their *first* attempt were considered for data analyses. As the survey collected Australians' attitude towards Asians, it was crucial to include in the analyses data from Australia-born participants only. Therefore, four participants were further removed as they were born in England, Brazil, Canada, and New Zealand. In addition, responses

[30] Social desirability could still affect participants' endorsement on the stereotype contents of the SAAAS items. However, since prejudice is inferred from the stereotype items (rather than directly collected), the side effect of social desirability on the prejudice measure could be minimized.

were screened for duration checks. The pilot administration of the survey revealed that the survey took approximately 30min if completed in one sitting. Consequently, 25 participants were further filtered out due to the time it took them to complete the survey (i.e., less than 30min) and the uniformity in their responses (i.e., same response choice for either 20 Blatant and Subtle Prejudice items or 25 SAAAS items). The final total sample was 307 Australia-born participants who finished the survey on their first attempt and who were deemed to have reliable responses for data analyses.

The age of these 307 participants ranged from 18 to 57 years (M = 22, SD = 6). Participants' ethnic backgrounds were not collected, but questions on participants' parents' native languages were included in the survey. Of 307 participants, 272 had both parents who spoke English as their first language, and 35 had either of the parents or both parents who spoke another language natively.

2.2 Scale administration

Figure A5.1 shows the 6-point-Likert SAAAS as it appears in Lin et al. (2005). The scale was adapted to the Australian context by changing the term "Asian Americans" to "Asian Australians." It was then administered as an online survey with Qualtrics Survey Software on the platform of Western Sydney University. To further minimize the potential social desirability effects in the endorsement of SAAAS items, a number of filler questions and scales were added to distract participants from the true purpose of the investigation. These fillers were questions on personal details and language backgrounds, and scales on handedness (Oldfield, 1971), religious ideologies (Putney & Middleton, 1961), disgust (Haidt, McCauley, & Rozin, 1994), and self-esteem (Rosenberg, 1965). They all appeared before SAAAS, in the same order as above so that all participants would have the same experience before responding to SAAAS. The survey also included scales such as right-wing authoritarianism (Zackrisson, 2005), implicit theories (6 of the original items and 12 modified filler items: Dweck, Chiu, & Hong, 1995), emotional responses (Pinkus, Lockwood, Schimmack, & Fournier, 2008), emotional responses – pride scale (Tracy & Robins, 2007), and state self-esteem (Heatherton & Polivy, 1991), which appeared after SAAAS, in a randomized order. The Asian focus of the SAAAS scale itself was also disguised by the inclusion of two-thirds of SAAAS items for several filler groups such as females, young people, Aboriginal Australians, African Australians, and Arab Australians. SAAAS items were randomized among themselves to avoid order effects.

APPENDIX
The Scale of Anti–Asian American Stereotypes (SAAAS)

Below are a number of statements with which you will agree or disagree. There are absolutely no right or wrong answers. Use the specified scale to indicate the number that best matches your response to each statement.

0	*1*	*2*	*3*	*4*	*5*
strongly disagree	*moderately disagree*	*slightly disagree*	*slightly agree*	*moderately agree*	*strongly agree*

(C)	1.	Asian Americans seem to be striving to become number one.
(S)	2.	Asian Americans commit less time to socializing than others do.
(C)	3.	In order to get ahead of others, Asian Americans can be overly competitive.
(S)	4.	Asian Americans do not usually like to be the center of attention at social gatherings.
(C)	5.	Most Asian Americans have a mentality that stresses gain of economic power.
(C)	6.	Asian Americans can sometimes be regarded as acting too smart.
(S)[a]	7.	Asian Americans put high priority on their social lives.
(S)	8.	Asian Americans do not interact with others smoothly in social situations.
(C)[a]	9.	As a group, Asian Americans are *not* constantly in pursuit of more power.
(C)	10.	When it comes to education, Asian Americans aim to achieve too much.
(S)	11.	Asian Americans tend to have less fun compared to other social groups.
(C)	12.	A lot of Asian Americans can be described as working all of the time.
(S)	13.	The majority of Asian Americans tend to be shy and quiet.
(S)	14.	Asian Americans are not very “street smart.”
(S)[a]	15.	Asian Americans know how to have fun and can be pretty relaxed.
(S)	16.	Most Asian Americans are not very vocal.
(C)[a]	17.	Asian Americans are a group *not* obsessed with competition.
(S)[a]	18.	Asian Americans spend a lot of time at social gatherings.
(C)	19.	Oftentimes, Asian Americans think they are smarter than everyone else is.
(C)	20.	Asian Americans enjoy a disproportionate amount of economic success.
(S)	21.	Asian Americans are not as social as other groups of people.
(C)	22.	Asian Americans are motivated to obtain too much power in our society.
(S)[a]	23.	Most Asian Americans function well in social situations.
(C)	24.	Many Asian Americans always seem to compare their own achievements to other people’s.
(S)	25.	Asian Americans rarely initiate social events or gatherings.

NOTE: S = sociability item, C = competence item. Scoring instructions are as follows: Sociability and competence scores on the Scale of Anti–Asian American Stereotypes can be calculated separately by adding up the score for all items on the relevant subscale after reverse-scoring the items listed below. The sociability and competence subscales also can be combined to form a total anti–Asian American prejudice score. Reverse-scored items (0 = 5, 1 = 4, 2 = 3, 3 = 2, 4 = 1, 5 = 0): 7, 9, 15, 17, 18, 23. Sociability score = total of all the sociability items: 2, 4, 7, 8, 11, 13, 14, 15, 16, 18, 21, 23, 25. Competence score = total of all the competence items: 1, 3, 5, 6, 9, 10, 12, 17, 19, 20, 22, 24.
a. Indicates a reverse-scored item.

Figure A5.1: *SAAAS (Lin et al., 2005)*

In one American sample, SAAAS was reported to correlate closely with the Subtle Prejudice[31] Scale (SPS) (r = .57, p < .001; Lin et al., 2005), which was proposed alongside the Blatant Prejudice[32] Scale (BPS), by Pettigrew and Meertens (1995). The SPS and BPS were administered with Blacks as the attitude target in Lin et al. (2005) to examine the commonalities between prejudice against Asians and prejudice against Blacks in America as well as the forms of anti-Asian prejudice (i.e., whether they can be in subtle forms). To examine

[31] Note that the use of the term “prejudice” used in the name of this scale (as well as BPS) is not the same as how the term “prejudice” is used in this thesis: As mentioned earlier, in this thesis “prejudice” means the affective component of attitudes when applied to the context of relationships between groups (or intergroup relations). In contrast, the term “prejudice” in the SPS and BSP scales means the overall attitudes (involving all the three cognitive, affective, and conative components) when applied to the context of intergroup relations.
[32] Same as Footnote 31.

this correlation in an Australian sample, both the SPS and BPS were also included in the Qualtrics survey, right after the self-esteem filler scale and before SAAAS. As the focus of the Qualtrics survey is only on Asians, a correlation between SAAAS and SPS would not provide information about prejudices between groups, but it could be informative about the forms of anti-Asian prejudice in Australia. As the SPS and BPS had originally been proposed to investigate attitude towards West Indians (Pettigrew & Meertens, 1995) and then adapted to examine Australians' attitude towards Asians in Australia by McGrane (2003) and McGrane and White (2007), the scales administered in the Qualtrics survey were based on the Australia-adapted version by McGrane (2003). The SPS and BPS consist of 20 items divided into five subscales. BPS consists of two subscales investigating blatant attitude manifesting in the perception of threat and rejection towards another group ("Threat and rejection" subscale) as well as in strong disagreement towards intimate contact with members of another group ("Intimacy" subscale). SPS deals with more subtle attitude in three subscales: defending traditional values ("Traditional values" subscale), exaggerating cultural differences ("Cultural differences" subscale), and denying positive emotions ("Positive emotions" subscale). Response options range from 1 (*strongly disagree*) to 5 (*strongly agree*), with some variations such as 4-point rating scales for the other types of responses (e.g., *very similar* to *very different*; *not at all bothered* to *very bothered*; *never* to *very often*). Five items are reverse-coded. In order to reduce the potential for confusion with different responses options across the two scales of interest in the survey (i.e., SPS and SAAAS), the original 6-point Likert response options in SAAAS were changed to 5-point Likert response options: *strongly disagree* (0), *disagree* (1), *neutral* (2), *agree* (3), and *strongly agree* (4). Similarly, the 4-point rating scales for some responses in the Blatant and Subtle Prejudice Scales were also adjusted to 5-point rating scales.

Compared with Pettigrew and Meertens' (1995) original version, McGrane's (2003) Australia-adapted version produced correlations in wider range, especially at the low end (Blatant: McGrane's r = .53 to .73 vs. Pettigrew & Meertens's r = .84 to .89; Subtle: McGrane's r = .54 to .76 vs. Pettigrew & Meertens's r = .70 to .81[33]). The correlations at the low end in McGrane's version might have been due to some small wording changes between McGrane's version and the original version (e.g., "Asian persons" vs. "West Indians"; "persons" vs. "people"). Consequently, in this Qualtrics survey, the wording was reversed back to Pettigrew

[33] Values reported in Footnote 2 of Pettigrew and Meertens (1995).

and Meertens's wording (e.g., "Asians," "people") in an effort to improve the correlations. In addition, the first item in the Intimacy subscale in McGrane's (2003) Australia-adapted version ("I would be willing to have sexual relationships with Asians") was changed into "I would not mind having a serious romantic relationship with an Asian" in response to feedback from the pilot administration of this Qualtrics survey. This change was probably a major departure from both Pettigrew and Meertens' and McGrane's versions. However, as the focus of this Qualtrics survey is on SPS (for SAAAS validity checking) while that intimacy item belongs to BPS, this change presumably did not affect SAAAS validity checking. As above, the Asian focus of the BPS and SPS was also disguised by the inclusion of two-thirds of these items for filler groups such as Aboriginals, Africans, and Arabs. These items were also randomized among themselves to avoid order effects. Table A5.1 shows the wording of BPS and SPS items as appearing in McGrane (2003).

Table A5.1: *Modified Blatant and Subtle Prejudice items (McGrane, 2003)* [Red indicates where changes were made to the items in the Qualtrics survey.]

Threat & Rejection Factor Items – Anglo version (Blatant prejudice scale)

1. Asians have jobs that the Anglos should have.
2. Most Asians living here who receive support from welfare could get along without it if they tried.
3. Anglo people and Asians can never really be comfortable with each other, even if they are close friends.
4. Most politicians in Australia care too much about Asians and not enough about the average Anglo person.
5. Asian persons come from less able races and this explains why they are not as well off as most Anglo persons.
6. How different or similar do you think Asians living here are to other Anglos in how honest they are?[a]

Intimacy Factor Items – Anglo version (Blatant prejudice scale)

1. I would be willing to have sexual relationships with Asians. (reversed scoring)
2. I would not mind if a suitably qualified Asian person was appointed as my boss. (reversed scoring)
3. I would not mind if an Asian person who had a similar economic background as mine joined my close

family by marriage. (reversed scoring)

4. How bothered would you be if a child of yours had children with a person of very different colour and physical characteristics than your own, and your grandchildren did not physically resemble the people on your side of the family?[a]

Traditional Values Factor Items – Anglo version (Subtle Prejudice Scale)

1. Asians living here should not push themselves where they are not wanted.
2. Many other groups have come to Australia and overcome prejudice and worked their way up. Asians should do the same without special favour.
3. It is just a matter of some people not trying hard enough. If Asian Australians would only try harder they could be as well off as Anglo people.
4. Asian people living here teach their children values and skills different from those required to be successful in Australia.

Cultural Differences Factor Items – Anglo version (Subtle Prejudice Scale)

1. How different or similar are Asians living here to other Anglo people in the values that they teach their children?[a]
2. How different or similar are Asians to other Anglo people in their religious beliefs and practices?[a]
3. How different or similar are Asians living here to other Anglo people in their sexual values or sexual practices?[a]
4. How different or similar are Asians living here to other Anglo people in the language that they speak?[a]

Positive Emotions Factor Items – Anglo[34] version (Subtle Prejudice Scale)

1. How often have you felt sympathy for Asians living here? [a] (reversed scoring)
2. How often have you felt admiration for Asians living here?[a] (reversed scoring)

Note: Items that have the superscript “a” were rated on a 4-point scale; all other items were rated on 5-point scales.

[34] In McGrane (2003), this was inadvertently labeled as an Asian version of the scale.

2.3 Procedure

Participants signed up to do the survey on the Qualtrics survey link through the Research Participation System run by the School of Psychology, Western Sydney University, for course credit. After they signed up, a URL link became visible to them on the sign-up page that led them straight to the Qualtrics survey, which was described as a survey exploring student perspectives in order to disguise the true purpose of the attitude investigation to mitigate social desirability effects. Participants could complete the survey anywhere and at any time at their convenience, and received the credit immediately after they finished the survey.

2.4 Preliminary analyses

Assumption checks (e.g., normality, skewness, and outliers) were conducted on the final sample. The Shapiro-Wilk test confirmed that, of all the scales and subscales, only SPS and Competence were normally distributed (p = .31 and p = .09, respectively), although, according to histograms, Traditional values and Positive emotions also looked approximately normally distributed. The distributions of BPS, Threat and rejection, and Intimacy were positively skewed. In contrast, the distributions of Cultural differences, total SAAAS, and Sociability were slightly negatively skewed. Outliers were first identified by box plots, then confirmed by standardized scores larger than |3.29| (Tabachnick & Fidell, 2007). These outliers were treated following Tabachnick and Fidell's (2007) recommendations: (1) Outliers were checked to see if they belonged to the desired population, and (2) If the outliers did belong to the desired population, their values were changed to one unit larger (or smaller) than the next most extreme values. The purpose of this adjustment is to reduce the impact of the outliers while still keeping intact the nature of the sample. As all the outlier cases in the survey sample belonged to the desired Australian population and could not be deleted, they were treated according to (2). However, when the values of the outliers could not be reduced further (e.g., an outlier had a score of 20 while the next most extreme score was 19), they remained the same in the reliability and correlation analyses. Five out of 10 outliers in total (i.e., one from BPS, one from Threat and rejection subscale, two from Intimacy subscale, and one from SAAAS) could not be treated.

3 Results

3.1 Reliability

Cronbach's alpha for the modified SAAAS was very high ($\alpha = .92$), comparable to the alpha for the original 6-point Likert SAAAS in Study 1 of Lin and colleagues (2005) ($\alpha = .94$). The alpha coefficients for the two SAAAS subscales were also comparable to the published results (Competence: $\alpha_{\text{present study}} = .90$ vs. $\alpha_{\text{Lin et al. (2005)}} = .92$; Sociability: $\alpha_{\text{present study}} = .86$ vs. $\alpha_{\text{Lin et al. (2005)}} = .91$). The alpha obtained for BPS ($\alpha = .86$) was higher than the high end of the alpha range reported in McGrane and White (2007) ($\alpha = .53$ - $.73$), and within the alpha range reported in Pettigrew and Meertens (1995) ($\alpha = .84$ - $.89$). The alpha obtained for SPS ($\alpha = .75$), however, fell within both the alpha range in McGrane and White (2007) ($\alpha = .54$ - $.76$) and the alpha range in Pettigrew and Meertens (1995) ($\alpha = .70$ - $.81$). The alphas obtained for Intimacy ($\alpha = .79$) and Traditional values ($\alpha = .60$) also fell within the alpha ranges reported for the corresponding subscales in Pettigrew and Meertens (1995) ($\alpha = .70$ - $.93$ & $\alpha = .53$ - $.67$, respectively). The alphas for Threat and rejection ($\alpha = .83$), Cultural differences ($\alpha = .74$), and Positive emotions ($\alpha = .75$) were higher than the published alphas in Pettigrew and Meertens (1995) ($\alpha = .73$ - $.81$, $\alpha = .57$ - $.72$, & $\alpha = .61$ - $.73$, respectively).

3.2 Correlation

The Pearson correlation coefficient[35] between SAAAS and SPS in this Qualtrics survey ($r = .47$, $p < .01$) was lower than the coefficient obtained in Study 3 of Lin et al. (2005) ($r = .57$, $p < .001$). According to Fisher's two-tailed r-to-z transformation, this difference is not significant ($z = -1.61$, $p = .107$). Similarly, the coefficient between Competence and Sociability in this survey ($r = .58$, $p < .01$) was lower than the coefficient obtained in Study 1 of Lin et al. (2005) ($r = .71$, $p < .001$). However, Fisher's two-tailed r-to-z transformation revealed that this difference is significant ($z = -2.74$, $p < .01$). Table A5.2 summarizes how the Cronbach's alphas and Pearson correlation coefficients obtained for BPS and SPS in this Qualtrics survey compare to those reported in Pettigrew and Meertens (1995) and McGrane and White (2007). The correlation coefficients among BPS, SPS, and their subscales were mostly within the tested range in Pettigrew and Meertens (1995),

[35] All the correlations reported in this section are significant at the .01 level (two-tailed).

except for the correlations between the Intimacy subscale and BPS, Intimacy and the Threat and rejection subscale, and Intimacy and the Positive emotions subscale. These correlations were lower than the corresponding tested ranges.

Table A5.2: *Comparison between the Cronbach's alphas and Pearson correlation coefficients obtained for BPS and SPS in this Qualtrics survey and those obtained in previous studies*

	Scale		Subscales				
	BPS	SPS	Threat & rejection	Intimacy	Traditional values	Cultural differences	Positive emotions
BPS	*.86* *.84 - .89*[36] *(.87 - .90)*[37] *.53 - .73*						
SPS	.63 .48 - .70	*.75* *.70 - .81* *(.73 - .82)* *.54 - .76*					
Threat & rejection	.91 .90 - .94	.62 .41 - .68	*.83* *.73 - .81*				
Intimacy	.81 .85 - .91	.46 .37 - .60	.51 .59 - .71	*.79* *.70 - .93*			
Traditional values	.67 .36 - .68	.82 .80 - .88	.68 .40 - .70	.45 .26 - .54	*.60* *.53 - .67*		
Cultural differences	.33 .08 - .48	.74 .63 - .77	.33 .06 - .45	.23 .09 - .40	.33 .21 - .43	*.74* *.57 - .72*	
Positive	.36	.62	.31	.33	.41	.20	*.75*

[36] Values reported in Footnote 2 of Pettigrew and Meertens (1995).
[37] Values reported in Table 2 of Pettigrew and Meertens (1995).

emotions	.35 - .56	.55 - .68	.29 - .50	.35 - .52	.30 - .44	.12 - .31	*.61 - .73*

Note: Black font indicates the values obtained in the Qualtrics survey, blue font indicates reported ranges in Pettigrew and Meertens (1995), and green font indicates reported ranges in McGrane and White (2007). Yellow-highlighted black indicates that the values obtained in the Qualtrics survey are outside of the previously reported range/s. Italic values on the diagonal cells indicate Cronbach's alphas.

3.3 Variability in prejudice towards Asians

Recall that higher SAAAS scores reflect stronger prejudice towards Asians. The results of total SAAAS scores in this Qualtrics survey show varied levels of prejudice, ranging from a low of 9 to a high of 85 (theoretical range: 0 to 100) (M = 53.96, SD = 13.38). As mentioned earlier, the original 6-point Likert response options in Lin et al. (2005) were changed to 5-point Likert response options in the Qualtrics survey: *strongly disagree* (0), *disagree* (1), *neutral* (2), *agree* (3), and *strongly agree* (4). However, this difference is not enough to explain the difference in *SD*. As a result, a rough comparison reveals that the obtained SAAAS mean was comparable with those reported for the two American samples in Study 5 of Lin et al. (2005) ($M_{\text{sample 1}}$ = 51.39, $M_{\text{sample 2}}$ = 53.32), but the standard deviation in this survey was much lower ($SD_{\text{sample 1}}$ = 21.29, $SD_{\text{sample 2}}$ = 22.60). Figure A5.2 is a histogram of the SAAAS score distribution. There are two modes in the distribution: one is at 50 and the other one is at 60. According to the Shapiro-Wilk test, there is evidence that the distribution is not normal. It can be seen from the distribution that the majority of participants had medium prejudice towards Asians, a few participants had very low prejudice towards Asians, and the number of participants who had high prejudice towards Asians is bigger than the number of those on the low end.

Turning to the SAAAS subscales, compared to the American sample in Study 1 (Competence: M = 26.84, SD = 12.16; Sociability: M = 24.26, SD = 11.23), the Australian sample assigned higher scores for both Asians' competence (M = 28.54, SD = 8.13) and sociability (M = 25.41, SD = 6.92), reinforcing the stereotypes. However, the results are not directly comparable as SAAAS response options in this Qualtrics survey were changed from 6-point Likert to 5-point Likert, and thus some psychometric properties of SAAAS might have been changed (i.e., the introduction of a neutral point in this study).

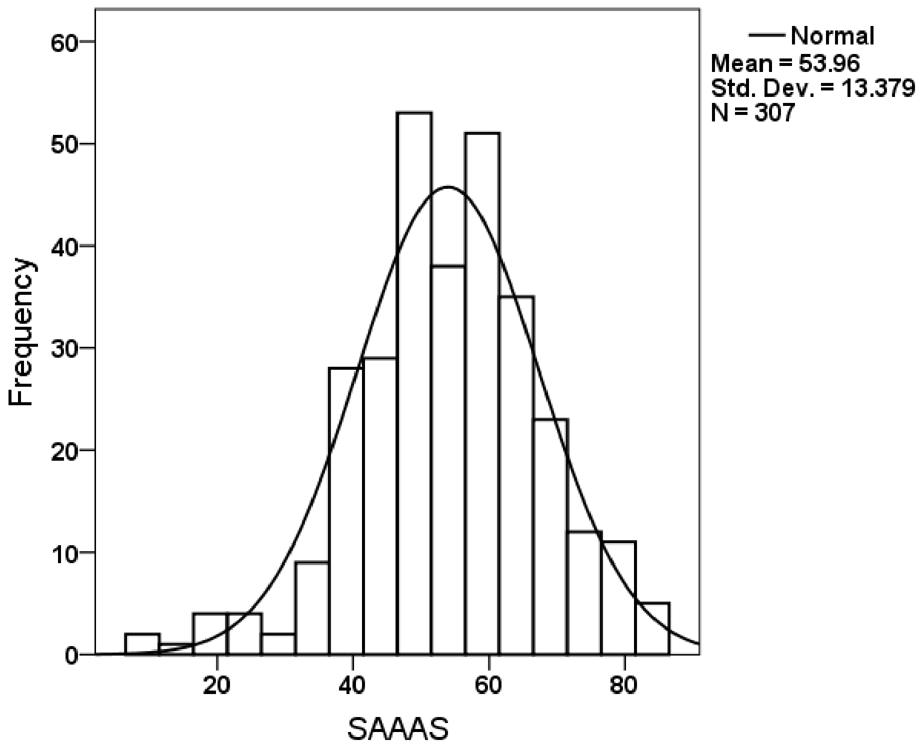


Figure A5.2: *SAAAS distribution*

4 Discussion and conclusion

Despite the introduction of a neutral point as well as less fine-grained response options (i.e., 6-point Likert reduced to 5-point one), the results of the SAAAS version administered in this Qualtrics survey make the scale look promising as a robust and valid tool for the indirect measure of anti-Asian prejudice. These results still reflected those of the original version in Lin et al. (2005). Examples include high alpha values for the whole scale and the two subscales, and correlation between the two subscales. This shows that the scale is robust enough to handle small changes in the scale construct. In addition, the finding that there is no significant difference in the correlation between SAAAS and SPS in this Qualtrics survey and Study 3 of Lin et al. (2005) could serve as evidence for more subtle or "modern" forms of racism in Australia, the types described in McConahay (1986). Subtle racism could entail a high level of social desirability concerns, and therefore, indirect approaches to the investigation (e.g., SAAAS) are better suited than direct ones (e.g., the prejudice measure used in Islam & Jahjah, 2001). Moreover, the variability in the range of SAAAS scores suggests that there is a variety of prejudice levels in the Western Sydney area, which creates the right environment to explore the relationship between prejudice and speech perception, as reported in Chapter 4 of this thesis.

A few points are worth noting when employing SAAAS for prejudice investigation:

(1) Social desirability concerns: SAAAS, although being indirect, is still a self-report measure. It can mitigate participants' social desirability concerns regarding their prejudice because participants are not aware that their prejudice is being examined. However, with the scale comprising stereotype items, participants may still have social desirability concerns over their stereotypes when responding to the scale, which in turn may indirectly affect the interpretation of prejudice. As a result, it is important to help reduce social desirability concerns in participants such as stressing on the anonymity of the data collection process, mixing SAAAS items with other scale items, and having the scale administered by a non-Asian researcher. These ideas were in fact applied for the experiment in Chapter 4, and should also be considered for future research.

(2) Neutral point: The introduction of the neutral point in the SAAAS scale, although not affecting the validity of the scale in general, might affect some participants' responses. In fact, this response option might have facilitated some participants' avoidance of giving specific answers, and thus reduced the sensitivity of the scale. As a result, future studies are recommended to use the original version of the scale, as reported in Lin et al. (2005). This original version was actually used to collect prejudice data in Chapter 4 of this thesis and, together with the experimental data, produced meaningful results.

(3) SCM: The significantly lower Pearson correlation coefficient between Competence and Sociability in this Qualtrics survey, compared to the coefficient obtained in Lin et al. (2005), may result from the applicability of the SCM in the Australian context. SAAAS is a proxy measure of prejudice which was built on SCM principles. Based on the data from American participants, the SCM maps the stereotypes about Asians in the high Competence - low Warmth quadrant onto the feelings of admiration and envy, which are then mapped onto the evaluation of dislike towards Asians. The existence of a considerable correlation between Competence and Sociability for the Australian sample may indicate that the SCM also works for the Australian context. However, the fact that this correlation is a little weaker than the correlation obtained for the American sample may suggest that the SCM may not apply to the Australian context to the same degree that it applied to the American context. As a result, follow-up research should investigate how the SCM's tenets apply in Australia and modify the model, as well as SAAAS, to fit the Australian context if necessary. This would

strengthen the predictions from the model as well as potentially contribute to the advancement of attitude research in Australia.

References

Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Ball, P. (1983). Stereotypes of Anglo-Saxon and Non-Anglo-Saxon accents: Some exploratory Australian studies with the matched guise technique. *Language Sciences*, *5*(2), 163–183.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, *48*(2), 145–160.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology (Vol. 6)* (pp. 1–62). New York: Academic Press.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*(6), 1191–1205.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*(4), 267–285.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., Vol. 1, pp. 269–322). New York: McGraw-Hill.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, Cognition, and Stereotyping Interactive Processes in Group Perception* (pp. 137–166). Academic Press.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

Fazio, R. H., & Olson, M. A. (2003). Attitudes: Foundations, functions, and consequences. In M. A. Hogg, & J. Cooper (Eds.), *The Sage Handbook of Social Psychology* (pp. 139–160). London: Sage.

Fazio, R. H., & Olson, M. A. (2003a). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: an introduction to theory and research*. Reading, MA: Addison-Wesley.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., Vol. 1, pp. 357–414). New York: McGraw-Hill.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu. J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, *16*(5), 701–713.

Hamilton, D. L., & Mackie, D. M. (1993). Cognitive and affective processes in intergroup perception: The developing interface. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, Cognition, and Stereotyping Interactive Processes in Group Perception* (pp. 1–11). Academic Press.

Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*(6), 895–910.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.

Islam, M. R., & Jahhah, M. (2001). Predictors of young Australians' attitudes toward Aboriginals, Asians and Arabs. *Social behaviour and Personality*, *29*(6), 569–580.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*(5), 774–788.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, *28*(3), 280–290.

Katz, I., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A Study of a Science (Vol. 3)*. New York: McGraw-Hill.

Lambert, W. W., & Lambert, W. E. (1964). The social significance of attitudes. In R. S. Lazarus (Ed.), *Foundations of Modern Psychology Series Social Psychology* (pp. 49–69). New Jersey: Prentice-Hall.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*(4), 570–583.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of Anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31*, 34–47.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 91–125). San Diego, CA: Academic Press.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442.

McGrane, J. A. (2003). *The significance of attitudinal ambivalence to attitude theory and measurement*. Unpublished Ph.D. thesis, University of Sydney.

McGrane, J. A., & White, F. A. (2007). Differences in Anglo and Asian Australians' explicit and implicit prejudice and the attenuation of their implicit in-group bias. *Asian Journal of Social Psychology*, *10*, 204–210.

McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., Vol. 3, pp. 136–314). Reading, MA: Addison-Wesley.

Neto, F. (2009). Implicit and explicit attitudes among students. *Actitudes implícitas y explícitas en estudiantes*, *8*(3), 849–857.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*(6), 161–176.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: from measures to mechanisms. *Trends in Cognitive Sciences*, *15*(4), 152–159.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Orne, M. T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal Psychology*, *58*(3), 277–299.

Pantos, A. J., & Perkins, A. W. (2012). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, *32*(1), 3–20.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition Measurement, Theory, and Applications* (pp. 255–277). The Guildford Press.

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*, 57–75.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., Vol. 1, pp. 323–390). New York: McGraw-Hill.

Pinkus, R. T., Lockwood, P., Schimmack, U., & Fournier, M. A. (2008). For better and for worse: Everyday social comparisons between romantic partners. *Journal of Personality and Social Psychology*, *95*(5), 1180–1201.

Putney, S. & Middleton, R. (1961). Dimensions and correlates of religious ideologies. *Social Forces*, *39*(4), 285–290.

Rosenberg, M. J. (1960). Cognitive reorganization in response to hypnotic reversal of attitudinal affect. *Journal of Personality*, *28*(1), 39–63.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude Organization and Change: An Analysis of Consistency Among Attitude Components* (pp. 1–14). New Haven, CT: Yale University Press.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*(4), 300–313.

Stangor, C., Sullivan, L. A., & Ford, T. E. (1991). Affective and cognitive determinants of prejudice. *Social Cognition*, *9*(4), 59–80.

Stephan, W. G. (1985). Intergroup relations. In G. Lindzey, & E. Aronson (Eds.), *The Handbook of Social Psychology* (pp. 599—658). Hillsdale, NJ: Erlbaum.

Stephan, W. G., & Stephan, C. W. (1993). Cognition and affect in stereotyping: Parallel interactive networks. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping interactive processes in group perception* (pp. 111–136). San Diego, CA: Academic Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behaviors. *Personality and Social Psychology Review*, *8*(3), 220–247.

Tabachnick, B. G., & Fidell, L. S. (2007). Cleaning up your act: Screening data prior to analysis. In S. Hartman (Ed.), *Using multivariate statistics* (pp. 60–116). Boston: Allyn & Bacon.

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. *Journal of Personality and Social Psychology, 92*(3), 506–525.

Trawalter, S., & Shapiro, J. R. (2010). Racial bias and stereotyping. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition Measurement, Theory, and Applications* (pp. 375–391). New York, London: The Guilford Press.

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods*, *5*(4), 553–600.

Walker, I. (1994). Attitudes to minorities: Survey evidence of Western Australians' attitudes to Aborigines, Asians, and Women. *Australian Journal of Psychology*, *46*(3), 137–143.

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition*, *144,* 91–115.

Western, J. S. (1969). What white Australians think. *Race*, 10, 411–434.

Zackrisson, I. (2005). Construction of a short version of the right-wing authoritarianism (RWA) scale.

Personality and Individual Differences, 39(5), 863–872.

Zajonc, R. B. (1980). Feeling and thinking Preferences need no inferences. *American Psychologist, 35*(2), 151–175.

Zajonc, R. B., & Markus, H. (1982). Affective and cognitive factors in preferences. *Journal of Consumer Research, 9*(2), 123–131.

Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal, & A. Kruglanski (Eds.), *The Social Psychology of Knowledge* (pp. 315–334). New York: Cambridge University Press.

Appendix 6:

Affective attitudes towards Asians influence perception of Asian-accented vowels

This appendix has been published as:

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (2015). Affective attitudes towards Asians influence perception of Asian-accented vowels. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0561. Retrieved https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0561.pdf

AFFECTIVE ATTITUDES TOWARDS ASIANS INFLUENCE PERCEPTION OF ASIAN-ACCENTED VOWELS

Nhung Nguyen[a], Jason A. Shaw[a,b], Michael D. Tyler[a,c], Rebecca T. Pinkus[a,c], Catherine T. Best[a,b]

[a] The MARCS Institute, University of Western Sydney, Sydney, Australia;
[b] School of Humanities and Communication Arts, University of Western Sydney, Sydney, Australia;
[c] School of Social Sciences and Psychology, University of Western Sydney, Sydney, Australia
nhung.nguyen@uws.edu.au, j.shaw@uws.edu.au, m.tyler@uws.edu.au, r.pinkus@uws.edu.au, c.best@uws.edu.au

ABSTRACT

Previous studies have shown that speech perception can shift because of cognitive attitudes about the speaker, but little is known about affective attitudes in speech perception. In our study, we investigated how affective attitudes towards Asians relate to the perception of Asian-accented vowels by native Australian English listeners. Affective attitudes were assessed with an established scale from the social psychology literature adapted to our specific purpose. Vowel perception was assessed using a vowel categorization task. Results show that the degree of dislike towards Asians negatively correlates with listeners' accuracy in vowel categorization. The results also provide evidence that cognitive attitudes elicit affective attitudes, and suggest the appropriateness of using social psychological tools to explicitly evaluate the role of affective attitudes in speech perception.

Keywords: sociophonetics, speech perception, social cognition, affective attitudes, modelling

1. INTRODUCTION

According to Allport, 'most of the business of life can go on with less effort if we stick together with our own kind. Foreigners are a strain...' [1]. Attitudes towards outgroups are known to influence a wide range of behaviour. In discussing how intergroup attitudes influence behaviour, social psychologists distinguish among three components: a cognitive component called stereotypes ('exaggerated [beliefs] associated with a category' [1]), an affective component called prejudice ('[a] feeling, favorable or unfavorable, toward a person or thing, prior to, or not based on, actual experience' [1]), and a behavioural component called discrimination ('[prejudice]-driven behavior' [13]). There is a growing body of evidence that the cognitive component (e.g., activation of beliefs about groups via priming) influences low level speech perception [7, 8, 10]. For example, although listening to the same Detroit speaker, Detroit listeners in [10] perceived 'standard' vowels when they were told to expect a Detroit speaker but perceived raised vowels (as the vowels actually were) when they were told to expect a Canadian speaker. Similarly, New Zealand listeners exposed to the concept of Australia by the word 'Australia' written on the test sheets [8], or stuffed kangaroos and koalas surreptitiously presented in the test rooms [7], reported more Australian-like vowels on a vowel matching task, even when they knew the voice was from a New Zealander. Therefore, extra-linguistic factors should also be considered for inclusion in speech perception models [7].

The studies above suggest that vowel perception can be affected by the cognitive component of intergroup attitudes. The other two components - the affective component and the behavioural component - have not been extensively investigated in speech perception. Cognitive attitudes have been shown to elicit affective attitudes which then shape other types of behavior such as recall tasks in [4]. After going through lists of traits describing English Canadians and French Canadians, English Canadians with negative affective attitudes towards French Canadians recalled more negative traits for the French Canadian outgroup and more positive traits for their own group. The categories 'English Canadian' and 'French Canadian' were activated in English Canadians, which in turn elicited their negative affective attitudes towards French Canadians to influence their performance on the recall task. Given this finding, we ask here whether cognitive attitudes might also elicit affective attitudes which then relate to performance in speech perception tasks.

There are few demonstrations of this association in speech perception. [10] is perhaps the closest although affective information was not directly measured. In this study, we measured affective attitudes towards the speaker group and assessed the relation between these affective attitudes and speech perception behaviour. Affective attitudes towards Asians were measured by the Scale of Anti-Asian American Stereotypes [9] modified for the Australian context (SAAAS). These attitudes were then elicited by revealing the Vietnamese identity of the speaker. Listeners' perception of this Asian-accented English speech was tested via a standard vowel categorization task [14].

2. METHOD

2.1. Participants

Thirty-one participants from the Greater Sydney community took part in the experiment for either course credit or cash payment. Data from seven participants were excluded because they did not pass the training phase of the vowel categorization task. The analyses were conducted on the remaining 24 participants who were all native listeners of English and born in Australia.

2.2. SAAAS survey

SAAAS was constructed based on the Stereotype Content Model, which contains two dimensions - Competence and Sociability [5]. The combination of the stereotypes along these two dimensions invokes different ingroups' mixed feelings towards outgroups. For example: outgroups that are perceived as high on Competence but low on Sociability are respected but disliked, which is the case for Asian Americans [6, 9]. Participants were asked to rate 25 SAAAS items [9] (12 Competence items and 13 Sociability items) from 'strongly disagree' (coded as 0) to 'strongly agree' (coded as 4). The survey was hosted online by Qualtrics Survey Software on the platform of the University of Western Sydney.

Several filler items were included to mask the purposes of the survey. Firstly, to mask the target Asian group, 16 SAAAS items were used for five distractor groups - females, young people, Aborigines, Africans, and Arabs. Secondly, to mask the investigation on affective attitudes, eight other scales were included: handedness, religious ideologies, disgust, self-esteem, blatant and subtle prejudice, right-wing authoritarianism, implicit theories, and state self-esteem. The whole survey took approximately 30 min to complete.

For data analyses, the SAAAS items were recoded and summed, following [12], to yield a composite SAAAS score for each participant. Higher SAAAS scores indicated stronger dislike towards the Asian outgroup. The SAAAS scores were normally distributed and ranged from -24 to 24 (M = -0.38, SD = 11.97).

2.3. Vowel categorization task

2.3.1. Speakers

There were two accents in the categorization task: Australian-accented English for the training phase and Vietnamese-accented English for the test phase. The Australian-accented English speaker was a female in her 20s. She was born and raised in Western Sydney, and spoke Australian English only. The Vietnamese-accented English speaker was a female in her 30s. She learned English in Vietnam with Vietnamese teachers, had lived in Australia for 19 years at the time of the recording, and rated herself as having intermediate fluency in English.

2.3.2. Nonce word auditory stimuli

Auditory stimuli were recorded in a sound-attenuated booth using a Shure SM10A-CN headset microphone and a MOTU 896 mk3 sound card. The sampling rate was 44.1 kHz. Materials were 13 Australian English monophthongs (as classified in [3]) embedded in the /hVdə/ context: heeda /hiːdə/, hidda /hɪdə/, hedda /hedə/, hadda /hædə/, harda /hɐːdə/, hudda /hɐdə/, hodda /hɔdə/, horda /hoːdə/, hoodda /hʊdə/, who'da /hʉːdə/, hurda /hɜːdə/, heerda /hɪədə/, and hairda /heːdə/. In each accent, four tokens out of 10 repetitions per monophthong were selected based on subjective judgements of similarity in speaking rate and loudness. The only exception was Australian-accented 'hudda' where only two tokens were chosen (and therefore repeated twice in the training phase) since pilot results confirmed that the rest of Australian-accented 'hudda' tokens were confused with Australian-accented 'hadda' tokens.

2.3.3. Choice word visual display

Participants were presented with a grid of 13 choice words: 'bad,' 'bard,' 'bead,', 'beard,' 'bed,' 'bid,' 'bird,' 'book,' 'bored,' 'bud,' 'paired,' 'pod,' and 'rude.' The letter(s) corresponding to the monophthong in each word were highlighted to maximize clarity. The position of the words on the screen was randomized across participants.

2.4. Procedure

At the lab, participants were greeted by either of two associate researchers who were Australians and not Asians. They were asked to do the online survey, followed by the vowel categorization task, in which they listened to the /hVdə/ nonce words and selected a choice word that contained the same vowel as the nonce word. After making a selection, their chosen word reappeared with the question ‘How good is the match? (1 = very poor, 7 = excellent)’ and the seven numbers from 1 to 7 beneath it. Participants then selected a number to indicate the goodness of fit.

There were 52 trials per phase. In the training phase, participants heard the Australian-accented English stimuli, and they were given feedback on incorrect responses. Participants continued cycling through 50 vowel tokens (four tokens per vowel, with the exception of ‘hudda’ as mentioned in 2.3.2) until they answered at least 10 out of 13 vowels correctly. Those who failed to achieve that criterion after four cycles of training were excluded from subsequent analysis. In the test phase, participants heard Vietnamese-accented English (four tokens per vowel) and no feedback was provided. They could take as long as they wished before responding, but the auditory stimuli were played only once for the choice word selection step. For the rating step, the stimuli were played only once also, but participants could click anywhere on the screen (except the rating numbers) to hear the stimuli again.

After finishing the tasks, participants were debriefed on the purposes of the speech task only. The entire lab session lasted from 50 min to 1.5 hrs, depending on how quickly participants completed the training phase.

3. RESULTS

In discussing the results, we refer to vowels according to the lexical sets of Wells [15]. Figure 1 shows mean accuracy by vowel for the Vietnamese-accent test phase. Performance was near ceiling for two vowels (START and KIT) and near floor for two others (NORTH and TRAP). The remaining nine vowels show more variation across participants. Of these nine, the GOOSE, STRUT, and NEAR vowels had a large number of inaccurate responses (37.5% of participants had at least two inaccurate responses for GOOSE tokens, 41.7% for STRUT, and 41.7% for NEAR tokens) in the Australian-accent training phase. Therefore, these were excluded from further analyses. Categorization accuracy for the remaining six vowels (FLEECE, DRESS, NURSE, FOOT, SQUARE, and LOT) was very high in the training phase (with no more than one inaccurate response for the majority of participants) but exhibited a large amount of cross-participant variation in the test phase, indicated by the error bars in Figure 1. We focus on these vowels to evaluate how affective attitudes relate to vowel perception accuracy.

Figure 1: Mean categorization accuracy by vowel. Error bars indicate 95% confidence intervals.

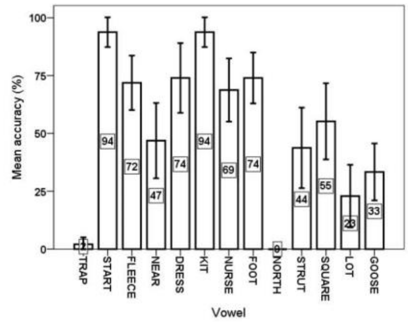
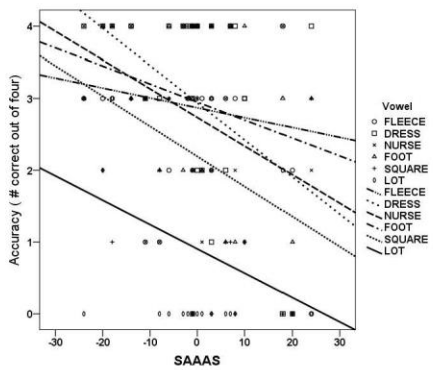


Figure 2 shows a negative correlation between SAAAS scores and speech perception accuracy - the higher the SAAAS score, the poorer the speech perception performance. Regression lines were fitted to each vowel separately. The downward slope of these lines indicates that people who score high on the SAAAS generally have lower vowel categorization accuracy than people who score low on the SAAAS. It can also be seen that this correlation is fairly uniform across the six vowels.

Figure 2: Categorization accuracy in the test phase as a function of SAAAS.



Even though we excluded participants who failed to reach the accuracy criterion in training and we eliminated vowels that showed a high number of errors in training (GOOSE, STRUT, and NEAR), there was still some variation across participants in vowel categorization accuracy in training. Some participants passed only 10 out of 13 vowels (n = 11) in training, while others passed 11 (n = 4), 12 (n = 7), or 13 out of 13 (n = 2). Because variation in the number of vowels passed in training (V_Passed) is likely to correlate with categorization accuracy in test, we also included this as a factor in the model.

We fitted a binomial mixed effects model to the accuracy data in R (version 3.1.2) [11] using lme4 package [2] with optimizer NM2 (the nloptr package version of Nelder-Mead):

(1) Accuracy $\sim SAAAS + Vowel + V\_Passed + (1|Participant) + (1 + SAAAS|Token)$

SAAAS, Vowel, and V_Passed were the three fixed factors in the model. The random factors were Participant (random intercepts) and Token (both random slopes and intercepts). Table 1 lists the significant predictors in the best-fitting model. The β values indicate the magnitude and direction of the associations. The Z values and Pr(>|z|) indicate statistical significance. SAAAS has a statistically significant association with accuracy. The negative β value (-0.04) for SAAAS indicates that, for every unit increase in dislike towards Asians, the odds of accurately categorizing Vietnamese-accented English monophthongs decrease by a multiplicative factor of 0.96 ($e^{-0.04}$). The table also shows that accuracy on LOT was worse than the other five vowels. Finally, and as expected, the positive β coefficient for V_Passed indicates that participants who passed more vowels in training had higher accuracy in test. The interaction between SAAAS and individual vowels was also tested but did not lead to significant improvement over the model.

Table 1: Summary of significant effects in the best-fitting model.

predictor	β	S.E.	Z value	Pr(>|z|)
(intercept)	-6.06	2.12	-2.85	< 0.01
SAAAS	-0.04	0.02	-2.21	0.03
LOT	-2.68	0.49	-5.53	< 0.001
V_Passed	0.64	0.19	3.35	< 0.001

4. DISCUSSION

Cognitive attitudes are known to elicit affective attitudes which then shape intergroup behaviour. We have presented evidence that suggests that the same is true for speech perception. In our study, the category 'Asians' was presumably activated in listeners when they were told they would be hearing Vietnamese-accented speech. This priming aspect of the experimental set-up was quite similar to [10]. This set-up also elicited our listeners' affective attitudes towards Asians, which were directly measured by the SAAAS survey beforehand. The SAAAS scores allowed us to quantify affective attitudes towards Asians so that we could directly evaluate whether they correlated with vowel perception. As has been observed in other types of intergroup behaviour, attitudes towards Asians had a significant negative correlation with the accurate perception of Asian-accented speech. Although we are careful about interpreting our results in terms of direct causation, the findings are consistent with the conclusion drawn in other areas of social psychology [4], that activation of beliefs about a group also involves activation of favourability towards the group and that both of these components can influence speech perception.

To our knowledge, this is the first direct demonstration of the role of affective attitudes towards speaker groups in a speech perception task. We are optimistic about the prospects of using tools such as SAAAS to evaluate the role of affective attitudes in speech perception. To facilitate future research in this area, we would like to close the discussion with a methods point that may be helpful. According to the Stereotype Content Model, the combination of high Competence and low Sociability traits elicits dislike towards outgroup members [9]. However, these two stereotypic components do not have equal power in predicting the levels of liking and attention that outgroup members receive. An Asian who is perceived to be less sociable is more disliked and paid less attention to [9]. Accordingly, we explored the separate Competence and Sociability subscales of the SAAAS (instead of the composite score) and found that the Sociability score was a significant predictor of vowel accuracy whereas Competence was not.

5. CONCLUSIONS

We have established that affective attitudes, just like cognitive attitudes, influence speech perception. This result supports the conclusion in [7] that speech perception models should take both linguistic and extra-linguistic factors into account. Moreover, it goes further by demonstrating that, in addition to beliefs about a group, listeners' favourability towards a group can also be informative for speech perception. The result also suggests the appropriateness of using SAAAS and vowel categorization tasks to assess the role of affective attitudes in speech perception.

6. REFERENCES

[1] Allport, G. W. 1954. *The Nature of Prejudice*. Cambridge: MA: Addison-Wesley.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. _lme4: Linear mixed-effects models using Eigen and S4_. http://CRAN.R-project.org/package=lme4.

[3] Cox, F. 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Aust. J. of Ling.* 26, 147–179.

[4] Dutta, S., Kanungo, R. N., Freibergs, V. 1972. Retention of affective material: Effects of intensity of affect on retrieval. *Journal of Personality and Social Psychology* 23, 64–80.

[5] Fiske, S. T., Cuddy, A. J. C., Glick, P., Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 878–902.

[6] Glick, P., Fiske, S. T. 2001. Ambivalent sexism. In: Zanna, M. P., (ed), *Advances in Experimental Social Psychology Vol. 33*. Thousand Oaks, CA: Academic Press 115–188.

[7] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865–892.

[8] Hay, J., Nolan, A., Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review* 23, 351–379.

[9] Lin, M. H., Kwan, V. S. Y., Cheung, A., Fiske, S. T. 2005. Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin* 31, 34–47.

[10] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18, 62–85.

[11] R Core Team. 2014. R: A language and environment for statistical computing. http://www.R-project.org/.

[12] Schweitzer, A., Lewandowski, N. 2013. Convergence of articulation rate in spontaneous speech. *Proc. 14th Interspeech* Lyon. 525–529.

[13] Smith, E. R. 1993. Social identity and social emotions: Toward new conceptualizations of prejudice. In: Mackie, D. M., Hamilton, D. L., (eds), *Affect, Cognition, and Stereotyping Interactive Processes in Group Perception*. Academic Press 297–315.

[14] Tyler, M. D., Best, C. T., Faber, A., Levitt, A. G. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21.

[15] Wells, J. C. 1982. *Accents of English I: An Introduction*. Cambridge: Cambridge University Press.