JOURNAL ARTICLE

Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Michael Walker, Anita Szakay and Felicity Cox
Centre for Language Sciences, Department of Linguistics, Macquarie University, Sydney, AU
Corresponding author: Michael Walker (michael.walker@mq.edu.au)

The presence of culturally significant objects has been shown to induce biases in speech perception consistent with features of the dialect relevant to the object. Questions remain about the transferability of this effect to different dialect contexts, and the efficacy of the task in inducing the effect. This paper details an Australian-context experiment modelled on Hay and Drager's (2010) New Zealand-context stuffed toy study. Seventy-five listeners heard spoken Australian English (AusE) phrases with phrase-final monosyllabic words containing either KIT, DRESS, or TRAP vowels. Each phrase was followed by audio presentation of a six-step synthesized vowel continuum, from New Zealand English (NZE)-like to exaggerated AusE-like tokens. Listeners attempted to match one of the synthesized variants to the speaker's realization of the target vowel. Listeners were exposed to one of two priming conditions, established by stuffed toy kiwis (New Zealand) and stuffed toy koalas (Australia), or a control condition (no toy). Contrary to Hay and Drager (2010), token selections did not differ significantly between the New Zealand and Australian priming conditions. However, reversing the order of continuum presentation did significantly affect token selection for KIT vowels, raising questions about the task design itself. Results suggest that the influence of regional primes on speech perception may be more limited than previously thought.

Keywords: speech perception; vowel perception; sociophonetics; Australian English; New Zealand English

1. Introduction

It has been widely shown that listeners make use of linguistic cues to inform knowledge of a speaker's social or regional background, suggesting a close relationship between linguistic and indexical information (see e.g., Foulkes, Scobbie, & Watt, 2010; Thomas, 2002 for reviews). There is also evidence that this relationship is bidirectional, with indexical information also able to influence speech perception. Listeners' sensitivity to the phonetic consequences of a speaker's social characteristics has been demonstrated in lexical access and phoneme identification tasks where the implied age (Drager, 2011; Hay, Warren, & Drager, 2006; Koops, Gentry, & Pantos, 2008), gender (Johnson, Strand, & D'Imperio, 1999; May, 1976; Munson, 2011; Strand & Johnson, 1996), or ethnicity (Babel & Russell, 2015; Rubin, 1992) of a speaker is manipulated. These tasks rely on a listener's awareness of sociophonetic variation (or stereotyped variation) related to the social characteristics to generate predictable response variance.

1.1. Speech perception and regional priming

Niedzielski (1997, 1999) observed a regional priming effect in a phoneme matching task where listeners' perception of vowels was shown to shift when information about the geographic origin of a speaker was manipulated. Niedzielski (1997) showed that the dialect of English spoken by Detroiters included features similar to Canadian English, such as a raised /aʊ/. However, Detroiters believed their dialect to be equivalent to Standard American English (SAE), which crucially does not feature raised /aʊ/. Niedzielski (1999) found that this belief biased Detroiters' perception. Participants, from Detroit, were assigned one of two conditions (Canadian or Michigan) and were told that they would be listening to a speaker from that region. Participants then completed a perceptual matching task that required a vowel token from a continuum of six synthesized vowel variants to be matched to the vowel in a target word contained within a recorded phrase. The speech data were from a single Detroit speaker who produced raised /aʊ/. For target words containing /aʊ/, participants in the Canadian condition selected a more raised variant of the synthetic stimuli than those in the Michigan condition. Niedzielski (1999) argues that participants in the Michigan condition anticipated a SAE vowel space which biased perception towards variants that were congruent with their mental representations of SAE vowels.

Two replications of Niedzielski's (1999) procedure were conducted in a New Zealand context: Hay, Nolan and Drager (2006) and Hay and Drager (2010). Both observed a similar priming effect when NZE-speakers were primed towards AusE. In Hay et al. (2006), response sheets labelled with either *Australia* or *New Zealand* were used to establish the two conditions. Participants heard recorded sentences spoken by a single NZE speaker, each containing a monosyllabic target word featuring a KIT, DRESS, or TRAP vowel. These three short front vowels differ considerably between AusE and NZE (Gordon et al., 2004). Following each sentence, participants heard a synthetic six-step continuum, representing a sequence of vowels from AusE-like to NZE-like. In Hay and Drager (2010), stuffed toy animals (kiwis for New Zealand and kangaroos and koalas for Australia) were used instead of explicit labels to establish each condition.

In both Hay et al. (2006) and Hay and Drager (2010), female participants who were exposed to the Australian condition selected significantly more AusE-like KIT tokens than those exposed to the New Zealand condition. This effect was also observed in token selections for TRAP vowels in Hay et al. (2006); however, in Hay and Drager (2010) participants in the Australian condition selected more NZE-like TRAP vowels. Neither study reported significant response variance for DRESS vowels—not surprising given the greater similarity between NZE and AusE DRESS compared to KIT and TRAP (see **Figure 1.1**). In both studies linear regression was used without taking listener into account as a random factor; therefore, it is possible that the significance of condition as a predicting factor was overstated by the model, particularly given the sample size in Hay and Drager (2010) (n = 26). Results from both studies found that male participants selected more AusE-like KIT vowels in the New Zealand condition. As no gender difference had been observed in Niedzielski's (1999) study, Hay et al. (2006) and Hay and Drager (2010) attributed this difference to a competitive relationship between Australians and New Zealanders (see also: Drager, Hay, & Walker, 2010). In a similar experiment, Walker, Hay, Drager, and Sanchez (2018) found that exposing New Zealanders to negative facts about Australia resulted in a perceptual shift towards more AusE-like KIT vowels.

A key divergence from Niedzielski (1999) in Hay et al.'s (2006) experiment lies in the establishment of the priming conditions. While participants in Niedzielski (1999) were directly told the speaker's origin, and may very well have believed this information, Hay et al. (2006) used only the written label on participants' response sheets and observed, in results from a post-task questionnaire, that the priming condition did not influence
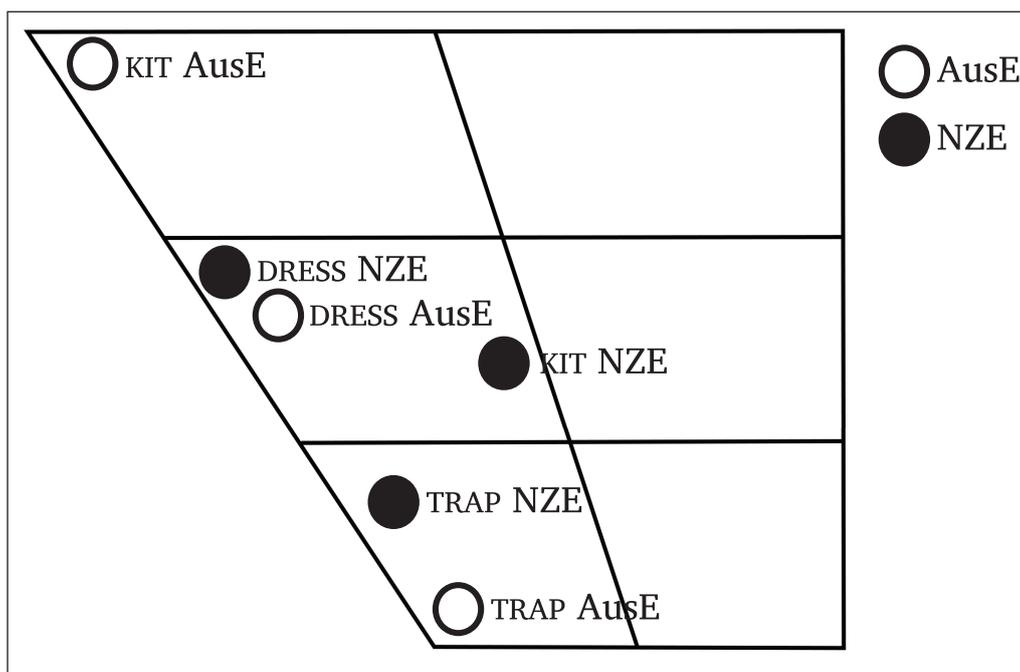
Figure 1.1: Approximate locations in the vowel space for the target vowels in AusE (Cox & Palethorpe, 2007) and NZE (Bauer et al., 2007).

participants' belief about the speaker's origin. Hay et al. (2006) proposed that exposure to the concept of Australia was enough to initiate the perceptual shift. This hypothesis found further support when the stuffed toys used in Hay and Drager (2010) also resulted in the priming effect. It would seem unlikely that the presence of stuffed toy kangaroos and koalas would cause participants to think the speaker was Australian; however, the toys may induce the concept of Australia, influencing the responses.

Further evidence of a regional priming effect was observed by Jannedy, Weirich, and Brunner (2011). In the German multi-ethnolect spoken in large urban areas of Germany, /ç/ is realized as palatalized [ʃ]. Participants identified items on a continuum from *Fichte* /fɪçtə/ to *fischte* /fɪʃtə/. Written prompts were used to imply a speaker was from either Kreuzberg where the [ʃ] variant is common or Zehlendorf, where the variant is not. Those in the Kreuzberg condition perceived more items as *fischte* than those in the Zehlendorf condition. This priming-induced shift in category boundary reflects earlier work by Strand and Johnson (1996) who were able to show that a speaker's perceived gender could influence classification of /s/ and /ʃ/, even when it conflicted with acoustic information. Participants were able to accurately determine whether a speaker was male or female in an audio only context as the perceptual boundary between /s/ and /ʃ/ is higher in female speakers due to differences in vocal tract size (May, 1976). However, when a voice was presented with an image of a female, participants identified more tokens as /ʃ/ than when the same voice was presented with an image of a male. In a similar experiment, Johnson et al. (1999) found perceived speaker gender influenced the location of the perceptual boundary between /ʊ/ and /ʌ/ with the effect found to be greater when the voice was more stereotypically male or female.

More broadly, the notion that social information attributed to a speaker can influence the categorization of speech has been supported by experiments demonstrating that the perceived age of a speaker can also induce perceptual biases in listeners. Drager (2011) observed that the perceptual boundary between DRESS and TRAP vowels for NZE-speakers shifted according to the implied age of a speaker with older listeners perceiving more

tokens as TRAP when the speech was accompanied by a photo of a younger person. This result may suggest that participants are responding to a trend towards raised variants of both DRESS and TRAP in younger NZE speakers. Although Drager (2011) concedes the effect is subtle, this result is supported by Hay et al. (2006), and Koops et al. (2008) who found that perceived age influences anticipated vowel qualities in speakers.

The apparent ethnicity of a speaker has also been shown to influence how speech is received by a listener. In Rubin (1992), listeners responded to a recorded lecture spoken by a single Standard American English (SAE) speaker paired with an image of either a female Caucasian or Asian face. The lecture was then rated on various intelligibility, accent, and social scores with responses from those participants exposed to the image of an ethnically Asian woman in a direction indicating nonstandard or 'accented' speech. Similar results have been described in Babel and Russell (2015) as well as McGowan (2015), who argues that a reduction in transcription accuracy was more likely to be a consequence of misleading social cues, rather than the result of an inherent bias against non-native English speakers.

Importantly, there is evidence to suggest that the manipulation of social information may not consistently result in the aforementioned response variation. Lawrence (2015) found no evidence that regional labels influenced speakers of Standard Southern British English in a replication of Niedzielski (1999) and Hay et al. (2006). Participants heard sentences produced by a speaker from Sheffield, Northern England featuring either a BATH or STRUT vowel in sentence final position. In Southern British English BATH is realized as [ɑ:] and in Northern British English as [a]. The Southern STRUT is realized as [ʌ] and Northern, either [ʊ] or [ə]. Continua design was largely consistent with Hay et al. (2006), and the priming conditions were established via an on-screen label (*Sheffield, Northern England* or *London, Southern England*). Participants were also explicitly told that the speaker was from either Sheffield or London. Although participants did exhibit some variability consistent with the priming condition, results from chi-squared tests revealed no significant difference between token selections in the two conditions. This raises questions about the generalizability and strength of a regional priming effect, particularly given that the BATH and STRUT vowels are "widely acknowledged as highly salient markers of regional identity in British English" (Lawrence, 2015, p. 1). This result echoes Squires' (2013) argument in favor of a more limited expectation of bidirectionality between linguistic and indexical influences on perception. In a morphosyntactic paradigm, Squires (2013) found that linguistic information influenced impressions about a speaker's social status but socioeconomic cues did not influence perception of non-standard speech.

1.2. Exemplar theory

Results from the sociophonetic research discussed above present strong evidence that listeners make use of indexical information to anticipate how a speaker will, or should, sound. Listeners display an understanding of how speech varies according to age, gender, ethnicity, and other social features, particularly when this variation is socially salient. As seen in Drager (2011), Rubin (1992), and Strand and Johnson (1996) for example, manipulation of available social information may result in perceptual biases in phoneme and word recognition.

Exemplar theory proposes that phonological knowledge is represented in memory as a continually updated aggregation of phonetically rich perceptual memories (Bybee, 2006; Goldinger, 1996; Johnson, 1997; Lacerda, 1997; Pierrehumbert, 2001; Wedel, 2006). Importantly, exemplar theory also accounts for the indexical context in which speech input was encountered. An exemplar representation therefore encompasses simultaneous indexing of the propositional, allophonic, and indexical properties of speech. In this way,

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 5 of 29

the auditory properties that distinguish speakers, as well as the source of these properties such as age, dialectal background, or gender are retained (Johnson, 1997).

Within an exemplar system, categories emerge from clusters of similar input formed within a cloud of remembered exemplars (Pierrehumbert, 2001; Wedel, 2006). These categories represent a class of equivalent perceptual experiences shaped by frequency information and density distributions (Pierrehumbert, 2001, 2006). Accordingly, more frequently encountered categories become more substantially and richly represented and a lack of invariance in input builds a more explicit representation of variation (Pierrehumbert, 2006). Social categories are similarly represented and are acquired "through cluster analysis over perceived properties of people and social interactions" (Pierrehumbert, 2006, p. 527). Over time, associations between linguistic information and relevant indexical information begin to develop (Foulkes & Docherty, 2006). These associations are strengthened when the link between the phonological variant and social category is more transparent (Foulkes & Docherty, 2006). Although the precise mechanisms for category assignment vary between exemplar models, it is generally agreed that all novel input automatically updates the entire category system in some way (Wedel, 2006).

More recent applications of exemplar-based perception argue for multi-level representation, fully incorporating elements from traditional abstractionist theory (McLennan, 2007; Pierrehumbert, 2006, 2016). Hybrid models such as this import the central claims of exemplar theory but borrow, from generative models, the concept of multiple levels of representation. Pierrehumbert (2016) argues that an abstract level of representation is necessary because phonological representation is inherently abstract. Further, abstract representation is required for the processing of novel word forms. At this level, phonetic detail and indexical information is likely to be disregarded. However, there is enough evidence to suggest a second level of representation, where fine phonetic detail and indexical features are retained.

Hay et al. (2006) and Hay and Drager (2010) contend that the observed regional priming effect can be explained by an exemplar model of speech perception. A prime, such as the stuffed toy koalas, would activate exemplars associated with 'Australia' prior to the listener receiving any speech input. This would then raise their resting activation level, in anticipation of AusE input. When speech input is received, provided there is sufficient acoustic similarity, exemplars associated with Australia (i.e., those labelled as AusE) would then be the quickest to reach full activation, leading to a system bias (Drager, 2011; Hay et al., 2006; Hay & Drager, 2010). In the context of the matching task, this bias results in identification of the phoneme as slightly more AusE-like, potentially due to the weight of other remembered tokens in the assigned category influencing memory of the vowel.

This explanation does, of course, rely on a listener having sufficient exposure to develop the requisite associations between social and phonological categories. It would seem apparent that the participants in Niedzielski (1999), Hay et al. (2006), and Hay and Drager (2010) did have sufficient representation of the social and phonological categories relevant to the experiment. However, would Australians show the same sensitivity to a New Zealand prime and NZE as New Zealanders did with the Australian prime and AusE? As seen in the results reported in Lawrence (2015), experimental manipulation of speaker information may not produce significant perceptual shifts even when listeners should have sufficient dialect exposure.

1.3. Australian and New Zealand English

Consistent with Hay et al. (2006) and Hay and Drager (2010), the target vowels in this study are KIT, DRESS, and TRAP. **Figure 1.1** gives a schematic of the approximate relative positioning in the vowel space of KIT, DRESS, and TRAP for AusE and NZE (according to Cox

& Palethorpe, 2007 and Bauer, Warren, Bardsley, Kennedy, & Major, 2007). For consistency and clarity, Wells' (1982) lexical set labels will be used when referring to vowels.

The KIT vowel is relatively centralized in NZE (Bauer et al., 2007). Easton and Bauer (2000) suggest that the ongoing centralization of KIT in NZE has mostly manifested as diachronic lowering. However, Watson, Maclagan, and Harrington (2000) argue that the distinctiveness of the NZE KIT is due to ongoing retraction. In AusE, KIT is a high front vowel. The AusE KIT is believed to have raised throughout the 20th century with some recent signs of a reversal of the raising trend (Cox & Palethorpe, 2008). Stressed KIT represents the most identifiable and socially salient difference between the two dialects (Maclagan, Gordon, & Lewis, 1999; Watson, Harrington, & Evans, 1998). Indeed, the difference between the respective KIT vowels is the source of much humor and mockery between Australians and New Zealanders, perhaps best exemplified by the exaggerated imitation of the others' pronunciation of 'fish and chips': Australians are said to pronounce "feesh and cheeps," while New Zealanders, "fush and chups" (Bauer & Warren, 2004).

The DRESS vowel is considered to be raised in both AusE and NZE relative to other dialects of English (Watson et al., 1998); however, the NZE DRESS is typically more raised than in AusE and there is ongoing raising in the NZE DRESS, particularly in younger speakers (Easton & Bauer, 2000; Maclagan & Hay, 2007; Watson et al., 2000). In contrast, the AusE DRESS is lowering (Cox & Palethorpe, 2008).

According to Watson et al. (1998), both AusE and NZE TRAP have been traditionally considered raised, although the AusE TRAP has undergone extensive lowering and retraction over a period of at least 25 years (Cox, 1999; Cox & Palethorpe, 2008, 2014) so that it is now the most open vowel in AusE. These recent sound changes in AusE and NZE represent chain shifts in opposite directions.

1.3.1. Recognition and attitudes between AusE and NZE speakers

Speakers of both AusE and NZE are said to consider the other accent to be not only easily identified, but undesirable (Weatherall, Gallois, & Pittam, 1998). According to Bayard, Weatherall, Gallois, and Pittam (2001) and Weatherall et al. (1998), Australians and New Zealanders are generally accurate at identifying both dialects. Ludwig (2007) found that speakers of both dialects could accurately differentiate between AusE and NZE productions of isolated words. Australians were found to be more accurate than New Zealanders when identifying DRESS and TRAP vowels as either AusE or NZE. Australians also identified the NZE KIT with a high level of accuracy, finding KIT to be one of the most salient identifiers of NZE (Ludwig, 2007).

Sensitivity to the differences between AusE and NZE vowels has also been observed in production tasks. Babel (2010) showed that speakers of NZE shifted to more AusE-like vowel productions in the presence of an Australian. This was particularly true for speakers with a positive attitude towards Australia. Drager et al. (2010) showed that the level of convergence displayed by New Zealanders primed towards Australia varied according to their level of sports fandom, arguing that Australia/New Zealand sporting rivalries have resulted in an inherently more competitive and negative view of Australia. Sanchez, Hay, and Nilson (2015) showed that NZE speakers shifted their production of KIT and TRAP vowels towards more AusE-like qualities through temporal proximity to Australia-related topics and lexical items. Production of DRESS vowels was only found to shift towards more AusE-like qualities when preceded by an Australia-related word in speakers with a higher level of exposure to AusE (Sanchez et al., 2015). This suggests that familiarity to the primed dialect may determine the extent to which an individual accommodates.

1.4. Predictions

Given the results reported in Hay and Drager (2010), it is predicted that culturally significant stuffed toys will influence participants' performance in a vowel matching task. To test this prediction, we therefore designed an experiment based on that described in Hay and Drager (2010), using the same target vowels—KIT, DRESS, and TRAP, but tested in an Australian context. Previous research has indicated that speakers of AusE are able to identify NZE (Bayard et al., 2001; Ludwig, 2007; Weatherall et al., 1998), particularly the short front vowels, with most salience attributed to KIT (Maclagan et al., 1999; Watson et al., 1998). Thus, we expect that AusE-speaking participants will be influenced by a New Zealand prime and will therefore select more NZE-like continuum tokens than those exposed to an Australian prime. Consistent with Hay et al. (2006) and Hay and Drager (2010), this shift should be present for target words containing a KIT vowel but may also be observed for target words containing a DRESS or TRAP vowel.

However, as indicated in Sanchez et al. (2015), there is reason to believe a regional priming effect may be dependent on the level of familiarity shown towards the primed dialect. For this reason, we predict that the priming effect would be affected by participants' level of familiarity to New Zealand and NZE. Thus, our hypotheses are as follows:

> H1. Stuffed toy kiwis will influence Australian listeners' selections in the vowel matching task.
> H2. Familiarity with New Zealand and NZE will affect listeners' selections in the vowel matching task.

2. Method
2.1. Participants

Seventy-five female speakers of AusE from 18–27 years with a mean age of 20.2 years ($SD = 2.03$) participated in the perception task. All participants were native speakers of AusE who were born and educated in Australia. Participants received either course credit or a $20 gift voucher for their time. The task took approximately 45 minutes. Three male participants were excluded as they were too small a sample for any meaningful comparison to be made between genders. Five female participants from the original pool of 80 were also excluded from analysis, two who were considered to be outliers by age (42 and 46 years of age), one who declared familiarity with the purpose of the study, and two who recognized the speaker's voice.

2.2. Stimuli and materials
2.2.1. Target vowels and target words

Consistent with Hay et al. (2006) and Hay and Drager (2010), the present study focuses on the KIT, DRESS, and TRAP vowels. Each target vowel was represented in 10 unique/CVt/ words; however, due to lexical restrictions, five target words contained a complex onset (*grit, skit, slat, Brett, threat*). The use of a coda /t/ for all target words ensured target vowels were presented in a consistent phonetic environment. This set of target words (presented in **Table 2.1**) represents a more controlled set of stimuli than those in Hay et al. (2006) and Hay and Drager (2010) who used monosyllabic target words with a range of coda consonants. The target words used by Niedzielski (1999) also varied in coda identity as well as syllable number.

This list of target words does not contain *fish*, a particularly well-known distinguisher of AusE and NZE speakers. Hay et al. (2006) found that when *fish* appeared as a target word, the priming effect was stronger. However, this effect was not observed by Hay and Drager (2010). As *fish* violates the coda /t/ constraint in the present study, it was excluded.

Table 2.1: Target words in KIT, DRESS, and TRAP sets.

KIT	DRESS	TRAP
bit	bet	bat
fit	Brett	cat
grit	debt	chat
hit	jet	fat
kit	net	gnat
knit	pet	hat
mitt	set	Matt
pit	threat	rat
skit	vet	slat
wit	wet	vat

2.2.2. Sentences

Each target word was embedded in phrase-final position in a unique carrier sentence. This facilitated target identification and ensured participants would not be exposed to any additional vowels between the target and continuum. Unlike the stimuli used in Hay et al. (2006) and Hay and Drager (2010), the sentences used here contained no other instances of the target vowels in stressed position. This minimized any additional priming effect by reducing overt identifiers of AusE (when contrasted with NZE). Example sentences, with identified target words, are shown below (for a complete list, see Appendix A):

1. The new movie was a huge summer **hit**
2. She's studying to become a **vet**
3. She called her mum for a short **chat**

A 19-year-old male monolingual speaker of Standard AusE from Sydney was recorded reading the carrier sentences. The speaker and his parents were Australian-born and the speaker had completed the entirety of his primary and high school education in Australia. At the time of recording, the speaker was an undergraduate student at Macquarie University. Sentences were recorded in a soundproof room with an AKG C535 condenser microphone and a PreSonus StudioLive 16.4.2 digital mixer using Pro Tools 11.3.1 at a 48 kHz sampling rate. The room contained no potential regional primes that might have influenced the speaker's vowel production beyond typical variation. The speaker's vowels were compared acoustically with mean AusE vowel formant values provided by Cox, Palethorpe, Miles, and Davies (2014) and were determined to be a representative of Standard AusE by a highly experienced AusE phonetician.

2.2.3. Continua design

Each continuum comprised six synthesized variants of the vowel from each sentence-final target word. The six tokens in each continuum were numbered, with token 1 representing the most NZE-like values and tokens 5 and 6 representing exaggerated AusE. Token 4 matched the F1 and F2 values of the speaker's actual vowel in each sentence. While this continuum structure was generally consistent with that in Hay et al. (2006) and Hay and Drager (2010), some important departures from their design were made. In the previous studies, continuum tokens for all target words were based on a single KIT, DRESS, and TRAP vowel produced by the speaker in an isolated /hVd/ frame. These three vowels were synthesized to create token 4 of the respective continuum, then first and second formant values (F1 and F2) were manipulated in equal Hertz steps to create the additional five

tokens. This meant that continuum tokens were consistent across all target words but, as acknowledged in Hay et al. (2006, p. 9), "While the tokens are generally positioned near token 4 in the continua, some individual tokens are closer to tokens 3 or 5." Therefore, token 4 would not align exactly with the speaker's vowel in each target word. We felt that it was important for token 4 to be consistently the most similar to the speaker's actual production and therefore elected instead to create a unique continuum for each target word whereby the F1 and F2 values of token 4 matched those of the vowel in the target word produced by the speaker in each separate sentence. We acknowledge that this continuum manipulation protocol represents a divergence from Niedzielski (1999), Hay et al. (2006), and Hay and Drager's (2010) experimental design. Rather than matching a target vowel to a single set of tokens for each vowel type, participants in the present study were required to match the target vowel to continuum tokens that varied according to the speaker's actual production of that vowel. The intention of the unique continuum was to minimize variance in token selection that could be attributed to factors other than a priming effect, such as the influence of coarticulation.

2.2.4. Vowel synthesis

The speaker's production of the target vowel in each sentence was analyzed in Praat (Boersma & Weenink, 2014). Using criteria from Cox (2006), the first and second formant values (F1 and F2) were extracted at the point where the vowel was considered to be least influenced by its surrounding phonetic context. This was typically approximately midway through the nucleus for the short front vowels examined here. The extracted F1 and F2 values were used to define the parameters for token 4 in each continuum and the baseline from which each of the additional five continuum tokens was derived. Step intervals between continuum tokens were calculated using the bark scale (Zwicker, 1961) in which equal bark distance is an approximation to perceptually equal distance. Bark values were then converted to Hertz to give formant values for each token. For KIT vowels, F1 intervals were 0.35 bark steps and F2 intervals 0.6 bark steps. For DRESS, the intervals were 0.8 (F1) and 0.3 (F2) and TRAP, 0.5 (F1) and 0.5 (F2). Step intervals for each vowel were determined in order to ensure that the F1 and F2 values for token 1 (most NZE-like) would remain within 2 standard deviations of mean formant values produced by male speakers aged 15–19 years (Easton & Bauer, 2000). Note that step intervals vary between F1 and F2 as well as across KIT, DRESS, and TRAP due to variance in the dialectal differences between these vowels. **Table 2.2** shows the mean and standard deviations for F1 and F2 of the most NZE-like token (Token 1) for KIT, DRESS, and TRAP compared to the mean values given by Easton and Bauer (2000) for these vowels. The procedure for using equal bark steps also represents a departure from Hay et al. (2006) and Hay and Drager (2010) who used equidistant Hertz steps to manipulate the continuum tokens and therefore adjacent tokens would not have been perceptually equidistant.

Table 2.2: Mean and Standard Deviation (in brackets) values in Hertz for the most NZE-like tokens in the continuum (Token 1) compared to values provided by Easton and Bauer (2000) for young NZE-speaking males.

Token	Token 1 – NZE-like		Easton & Bauer (2000)	
	F1 mean (Hz)	F2 mean (Hz)	F1 mean (Hz)	F2 mean (Hz)
KIT	516 (37)	1577 (77)	486 (36)	1619 (109)
DRESS	374 (26)	2014 (73)	417 (50)	2195 (220)
TRAP	645 (27)	1947 (93)	530 (72)	1890 (312)

As the step intervals were kept uniform for all continua related to a specific vowel, some F1 and F2 values violated the two standard deviation constraint for the most NZ-like tokens. For example, the coarticulatory effects on the post-rhotic KIT vowel in *grit* predictably resulted in a lower F2 than would be typical for the reference set in Easton and Bauer (2000). Similar violations also occurred with F1 values for *bit, grit, knit,* and *debt*. F1 values for the tokens of KIT in these words were less than 5 Hz beyond 2 standard deviations from the mean and for the word *debt* the F1 value was 14 Hz below 2 standard deviations of the mean for NZE DRESS. These violations were deemed acceptable on the basis that the synthesized vowels represented coarticulatory appropriate values, and the vowel steps maintained a consistent perceptual distance between tokens. **Figures 2.1–2.3** show the mean formant values of each of the six steps for the three vowel continua in relation to the vowel space of our male speaker of Standard AusE. For KIT, DRESS, and TRAP the mean values from the speaker's sentence data are plotted, while the rest of the vowels are taken from a hVd word list not used as stimulus in our experiment. The NURSE vowel has been removed for clarity in the centre of the vowel space. These figures also illustrate that the most NZE-like tokens on our synthesized vowel continua (Step 1) fit within 2 *SD* of the mean formant values reported for NZE KIT, DRESS, and TRAP by Easton and Bauer (2000).

To investigate potential differences between the extent of each vowel set continua, a linear regression analysis was fitted with Euclidean distance between the two end points of each continuum as the dependent variable and vowel set as the predictor, with DRESS as the default factor level. This analysis revealed that the KIT continua were significantly longer than the DRESS continua (intercept estimate = 1796.4, KIT estimate 596.6, $p < .0001$), which in turn were significantly longer than the TRAP continua (TRAP estimate = –455.1, $p < .0001$). This point will be revisited in Section 3.1.
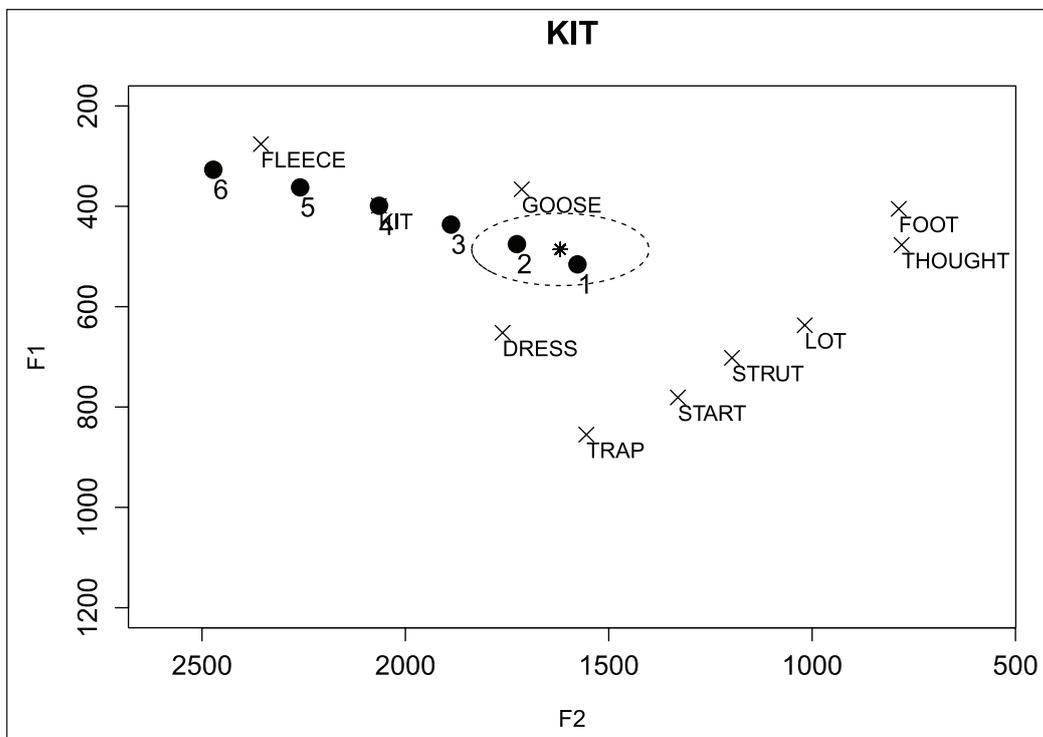

Figure 2.1: Mean formant values for each step (1 to 6) on the synthesized KIT vowel continua in relation to the male Standard AusE speaker's vowel space. The star represents the mean formant values reported for NZE KIT by Easton and Bauer (2000), with the ellipse showing 2 *SD* away from the mean.
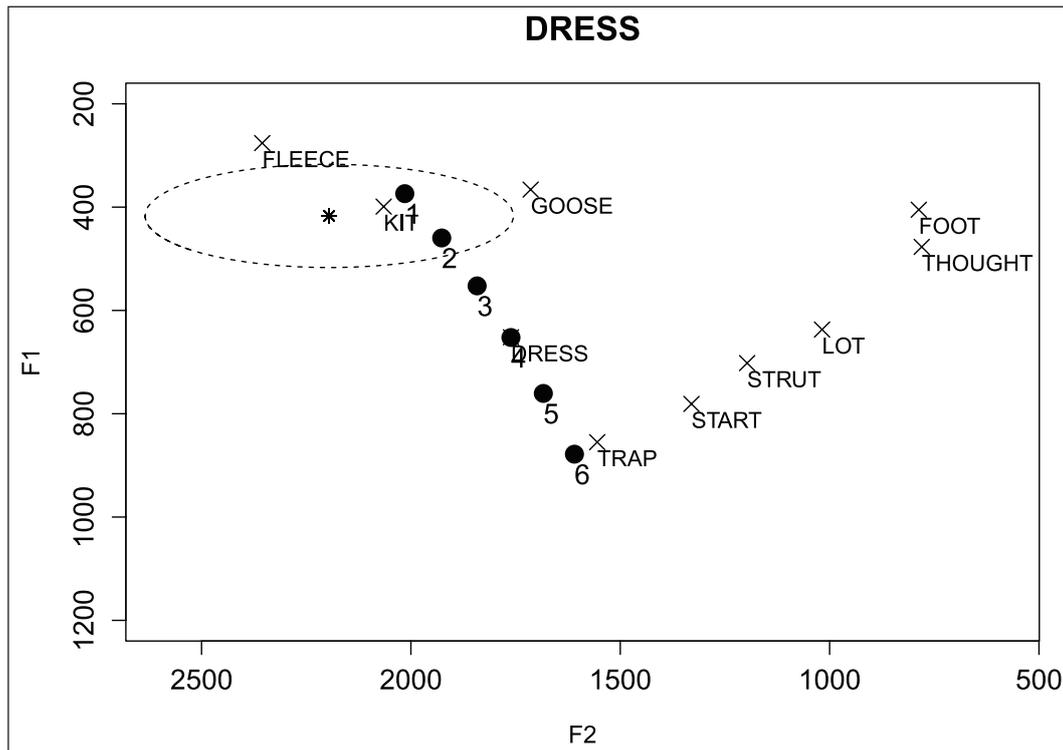
Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian
English speaking listeners?

Art. 7, page 11 of 29

Figure 2.2: Mean formant values for each step (1 to 6) on the synthesized DRESS vowel continua
in relation to the male Standard AusE speaker's vowel space. The star represents the mean
formant values reported for NZE KIT by Easton and Bauer (2000), with the ellipse showing 2 *SD*
away from the mean.
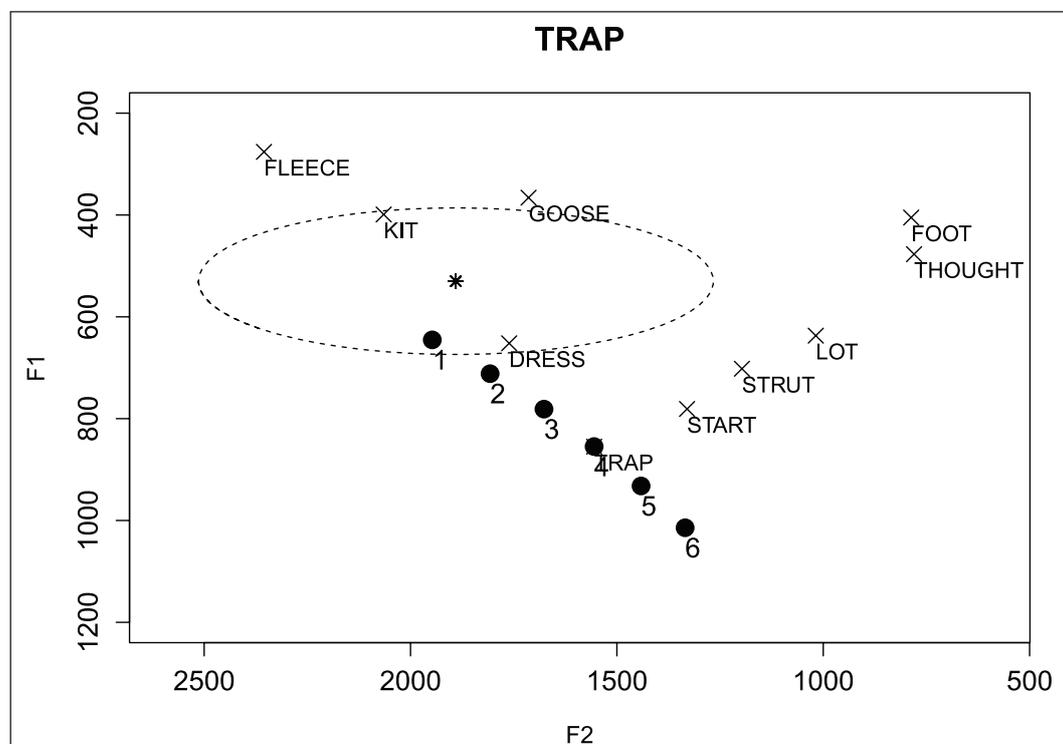
Figure 2.3: Mean formant values for each step (1 to 6) on the synthesized TRAP vowel continua
in relation to the male Standard AusE speaker's vowel space. The star represents the mean
formant values reported for NZE KIT by Easton and Bauer (2000), with the ellipse showing 2 *SD*
away from the mean.

Vowels were synthesized using the vowel editor function in Praat (Boersma & Weenink, 2014) which generates a vowel from inputted formant values, F0, and duration parameters. The task was piloted with synthesized token durations consistent with that of the speaker's original vowel, as well as with durations of 180 ms and 250 ms. Pilot participants indicated that 180 ms was the most suitable length for the task. The vowels synthesized with the same duration as the vowel in the target word were judged to be too short (between 68 ms and 149 ms), making the task extremely difficult. The 250 ms tokens were judged to be unnecessarily long as this duration would be consistent with intrinsically long rather than short vowels. All synthesized tokens were therefore created to be 180 ms long. F0 at the vowel onset was 160 Hz which represented the mean onset F0 for the speaker's 30 tokens. An F0 slope of –1 octaves per second was also applied to all synthesized vowels. Consistent with Niedzielski (1999), Hay et al. (2006), Hay and Drager (2010), and Lawrence (2015), F3 was not manipulated in any way. **Figures 2.4–2.6** show spectrograms of the continuum tokens for the same vowels. A full set of formant values for tokens is given in Appendices B and C.
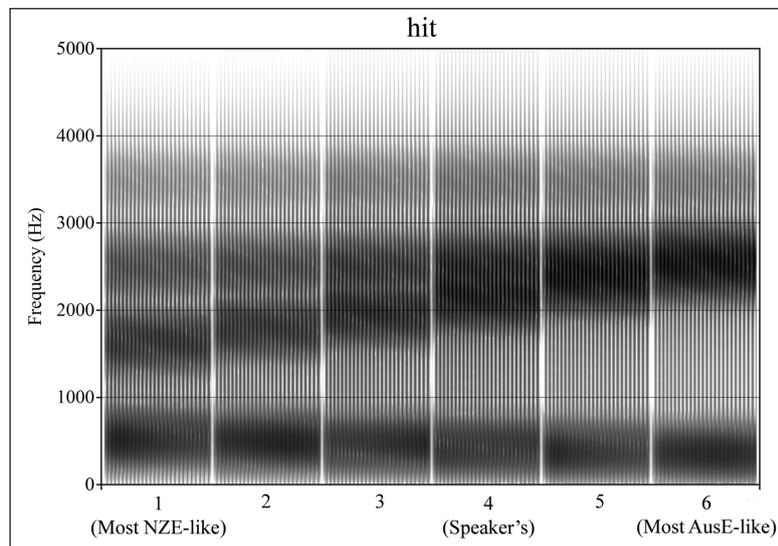

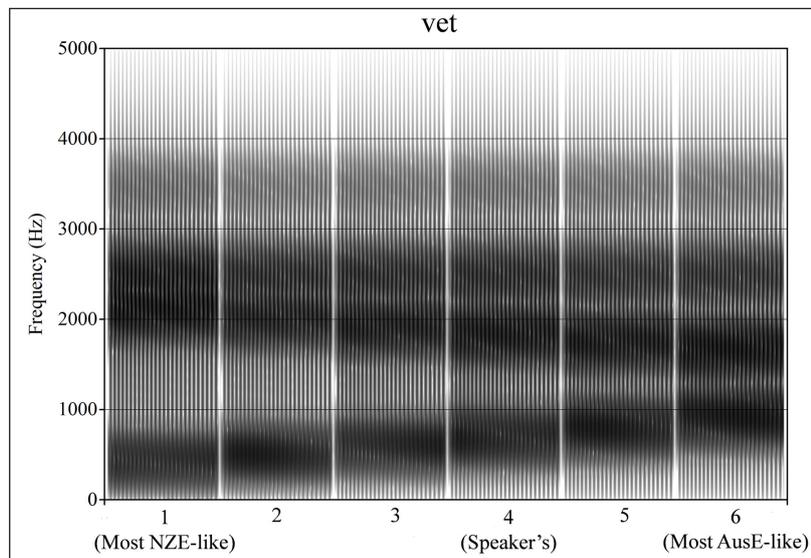**Figure 2.4:** Spectrograms of continuum tokens – hit.

Figure 2.5: Spectrograms of continuum tokens – vet.

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?
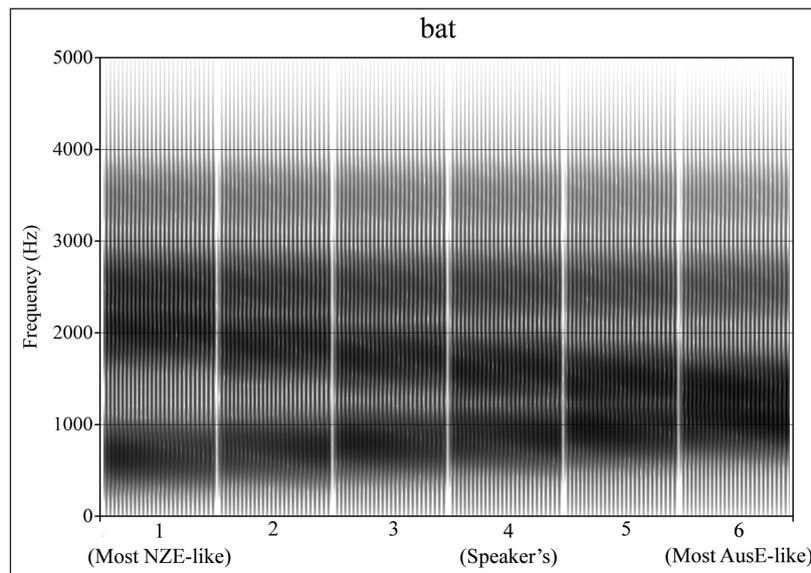
Art. 7, page 13 of 29

Figure 2.6: Spectrograms of continuum tokens – bat.

2.3. Priming

Participants completed the perception task in one of three conditions: Australian (n = 25), New Zealand (n = 25), and Control (n = 25). Priming for the conditions was cued by the presence of stuffed toy koalas (Australian) or kiwis (New Zealand) with the Control group exposed to no toys. Primes consisted of two koalas and two kiwis of approximately equivalent combined size (See **Figure 2.7**). Each participant completed the task individually and in the same sound-attenuated room. The room layout was identical for all participants, containing no other potential regional primes. All participants interacted with the same experimenter—a 29-year-old male, born in New Zealand who had immigrated to Australia at age 10—therefore any potential priming effect from the experimenter was consistent for all participants. The experimenter was the first author and aware of the hypotheses being tested. Walker et al. (2018) attributed response variance in a similar experiment to have been influenced by a possible shift in production from an experimenter who was aware of the purpose of the task. While this could be a concern for the present experiment, an experimenter who is blind to the hypotheses could still conceivably be influenced by the priming condition. Without acoustic analysis of the experimenter's speech during the experiment, the extent to which there is any shift is unclear.

To introduce the prime, the experimenter 'found' the headphones required for the task in a drawer, under the toys. Participants were told that the toys were being used in another experiment and placed them on the table, in the participant's line of sight, where they remained for the duration of the experiment. The toys were placed at approximately arm's length from the participant, alongside the laptop being used for the task. The intention was to draw attention to the prime, ensuring it had been seen without making it obvious that the toy was related in any way to the experiment. This was similar to the procedure used in Hay and Drager (2010).

As none of the previous iterations of the experiment had used a control, it was deemed important for a group to complete the perception task in an un-primed context. This would allow us to determine if the New Zealand and Australian primes were both producing an effect and, if so, which effect was the strongest. For example, there is no way of knowing whether the New Zealanders participating in New Zealand condition of the Hay et al. (2006) experiment were influenced in any way by the presence of the label New

Figure 2.7: Stuffed toy koalas and kiwis used in the experiment.

Zealand, or if the divergence shown in responses to KIT and TRAP vowels was a result of participants in the Australian condition alone. Further, a control group would also allow us to determine the accuracy with which participants could complete the task without having to account for a priming effect.

2.4. Perception task

Consistent with Hay et al. (2006) and Hay and Drager (2010), participants were told the purpose of the experiment was to determine the accuracy of synthesized vowel sounds and that they would hear a sentence of recorded human speech containing an identified target word followed by synthesized vowels. Participants were instructed by the experimenter to select the synthesized vowel that was the closest match to the vowel in the recorded target word. Prior to the experimental phase, a familiarization task containing three practice questions was presented. To reduce any potential priming effects the familiarization task used a different voice, a different set of target vowels, and contained no instances of KIT, DRESS, or TRAP vowels. Participants received no feedback in either the familiarization or experimental phase and no information was provided about the speaker.

In the experimental phase, participants were informed orthographically (without audio) that they would hear each sentence once, immediately followed by the six isolated synthesized vowel sounds and then a selection screen. Participants heard each sentence while it was simultaneously presented orthographically on-screen (Hay et al., 2006, and Hay and Drager, 2010, used written response sheets which included the sentences). The phrase-final word was in bold and underlined, identifying it as the target. The sentence remained visible for the duration of the recorded sentence plus an additional 2000 ms. Niedzielski (1997) reports that a 3000 ms break was used between sentence and continuum but no detail of presentation timing is given in Hay et al. (2006) or Hay and Drager (2010). Immediately following the 2000 ms, each 180 ms continuum token was played with its corresponding number visible on the screen. The number remained on screen for an additional 820 ms before the presentation of the next token (a total of 1000 ms). Thus, the entire continuum

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 15 of 29

was 6000 ms (6 tokens × 1000 ms). The experimental procedure is illustrated in **Figure 2.8**.

When the selection screen appeared, participants made their selection by pressing the corresponding number key on a keyboard. A selection could not be made until after all six continuum tokens had been played, and any key press by the participant prior to the appearance of the selection screen was not recorded. Presentation of the synthesized tokens with on-screen label numbers differs from Hay et al. (2006) and Hay and Drager (2010) who labelled tokens with spoken numbers only. Visual labelling was preferred here as it reduced any additional priming effect from the token labels. This was particularly important for label six, which contains a KIT vowel and pronunciation of six is a highly salient identifier of NZE when compared with AusE. In Lawrence (2015), continuum tokens were represented by dots, rather than numbers. However, this could potentially complicate token identification.

Sentences were presented once with the continuum tokens played in order from NZE-like to exaggerated AusE and once with the token order reversed (i.e., exaggerated AusE to NZE-like). However, in both orders, tokens were presented to the participant labelled from 1–6. In other words, items labelled 1 in the original order were the most NZE-like and those labelled 1 in the reverse order were the most exaggerated AusE tokens. The different presentation orders were used to discourage any potential selection patterning. Sentences were presented in two blocks with each block containing all 30 sentences. Block one contained half of the sentences presented with their continua in the original order and half with their continua in reversed order with block two containing each continuum in the opposite order. The two blocks were identical for all participants but sentences were presented at random. **Figure 2.9** illustrates the continuum for *hit* in the original order and reversed presentation order.
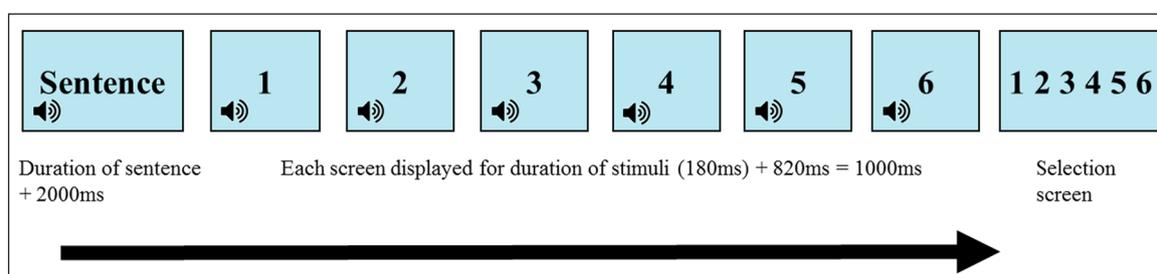

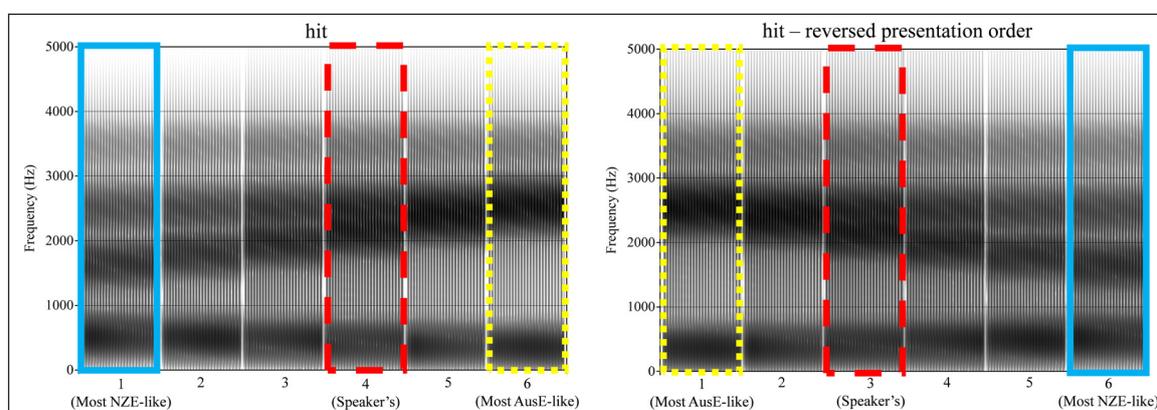**Figure 2.8:** Illustration of experimental procedure.

Figure 2.9: Spectrograms for continuum tokens for 'hit' in original order on the left and reversed order on the right. Blue solid box indicates the most NZE-like vowel. Red dashed box indicates the synthesized vowel based on the speaker's production. Yellow dotted box represents the exaggerated AusE vowel.

The perception task was presented on a Sony Vaio laptop using the E-Prime 2.0 software (Psychology Software Tools, 2012). All participants used the same pair of Sennheiser HD 461i closed over-ear headphones and were able to adjust their volume to a comfortable level. Token selections for all participants, including reaction time measured from the appearance of the selection screen, were extracted from E-prime. Individual selections were represented by a number from 1–6, corresponding to the actual token selected by the participant. For those trials presented in reverse order, selections were re-coded (i.e., $1 \rightarrow 6$, $2 \rightarrow 5$, etc.). This meant that coding of selections was consistent according to the acoustic features of the selected token (i.e., F1 and F2), rather than its numeric label. Following the perception task, participants completed a questionnaire modelled on those used in Hay et al. (2006) and Sanchez et al. (2015). The questionnaire concerned the participant's impressions of the speaker's age, occupation and education level, a free-choice question requiring participants to state where they believed the speaker was from, as well as questions designed to assess the participant's level of exposure to New Zealand and NZE.

2.5. Data analysis

A total of 4500 token selections were recorded (60 selections × 75 participants). Although participants were not given any explicit instruction regarding response timing, we elected to exclude any selection that was outside three standard deviations of the mean response time within each vowel set. In total, 41 selections were excluded (21 KIT, 15 DRESS, and 5 TRAP vowels). Reaction times for all remaining selections were 7914 ms or less (mean = 926 ms; SD = 913 ms). The final data set included 4459 selections.

2.5.1. New Zealand Exposure Score

Questionnaire responses were used to create a New Zealand Exposure Score (NZES) intended to weight a participant's likely exposure to, or familiarity with, NZE. This score was custom designed for the present study. Participants received two points if they had been to New Zealand with an additional two points if they had spent more than one month in New Zealand and two points if they had been to New Zealand in the eighteen months prior to the experiment. Participants received another two points if they were personally acquainted with any New Zealanders, with an additional two if they were in weekly contact with one-five New Zealanders, three points for six-ten New Zealanders, and four points for more than ten. Finally, an additional point was awarded if the participant could name any New Zealand media. A maximum score of 13 points was possible.

Overall, 26.67% of participants had been to New Zealand, all but one in the last ten years. 38.67% of participants answered that they were personally acquainted with New Zealanders (not all of whom had been to New Zealand). Almost half of the participants responded that they spoke with New Zealanders on a weekly basis, 45.33% spoke with one-five, and 2.67% spoke with ten or more New Zealanders. However, 52% reported speaking with no New Zealanders on a weekly basis. Finally, 29.33% of respondents could name any New Zealand media they had recently seen or heard. Using the metric described above, each participant was given a NZES between 0 and 13. Participants scored from 0 to 12 points with a mean score of 3.00 (SD = 2.98).

2.5.2. Statistical analysis

To test the influence of the priming condition on token selection we fitted cumulative link mixed models for ordinal data using the ordinal package (Christensen, 2015) in R (R Core Team, 2016), with no restrictions of equidistance or symmetry imposed on the

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 17 of 29

thresholds.[1] The dependent variable was token selection (1, 2, 3, 4, 5, 6). The model was kept maximal with respect to both its random and fixed effects structure. Random intercepts were included for participant and word, and random slopes were included for continuum presentation order (original and reversed) by participant, and for continuum presentation order in an interaction with experimental condition (Australia, New Zealand, and Control) by word. Fixed effects included experimental condition, participant NZES, and continuum presentation order and were entered into the model as a three-way interaction. The reference level for condition was the Australian toy condition. The syntax for this model was as follows: clmm(response ~ (1 + order|participant) + (1 + order*condition|word) + cond*order*NZES). Although participant Socioeconomic Index (SEI) was found to be a significant effect in Hay et al. (2006) and Hay and Drager (2010), it was suggested that SEI was a predictor of New Zealanders' experience with AusE. As we created a metric to score our participants' exposure to NZE (the NZES) we elected not to include participant SEI in our model. In line with Hay et al. (2006) and Hay and Drager (2010) each vowel set was analyzed separately.

2.5.3 Power analysis

A power analysis was carried out in R (R Core Team, 2016) using the simR package (Green & MacLeod, 2016). We ran a power simulation based on our own data using the effect size from Hay and Drager (2010). After running 1000 simulations the results indicated that our study needs 25 participants per condition to reach 80% power.

3. Results

Table 3.1 outlines the mean token selections for all vowel contexts in each of the three conditions. A value of 1 represents the most NZE-like vowel, 4 represents the vowel synthesized from the speaker's actual vowel production, and 6 represents the most exaggerated AusE vowel. Although tokens 5 and 6 represent exaggerated AusE vowels, for simplicity, higher token numbers will be referred to as more AusE-like and lower token numbers as more NZE-like for the remainder of this analysis.

On average, for each of the three target vowels, participants selected tokens representing more phonetically raised (i.e., lower F1) and fronted (i.e., higher F2) vowels than the synthesized speaker's vowels (token 4). This explains why the mean token selection for KIT (5.19) was so different from the mean selections for DRESS (3.10) and TRAP (3.6). For KIT vowels, higher token numbers represent raised and fronted variants (more AusE-like) however, for the DRESS and TRAP vowels, lower token numbers represent raised and fronted variants (more NZE-like). **Figure 3.1** shows the overall distribution of selections for KIT, DRESS, and TRAP vowels.

Table 3.1: Mean token selection and Standard Deviation (in brackets) – all conditions.

Condition	KIT	DRESS	TRAP
Australia	5.28 (0.78)	3.12 (0.85)	3.40 (1.08)
New Zealand	5.25 (0.78)	3.01 (0.83)	3.49 (1.02)
Control	5.03 (0.97)	3.16 (0.86)	3.49 (1.07)
Total	5.19 (0.86)	3.10 (0.85)	3.46 (1.05)

---

[1] An alternative, linear mixed effects analysis was also run on the data where the dependent variable was the formant value corresponding to the selected token. Details of these models are given in Appendix D. The results from the two different statistical approaches show the same predictors to be significant.
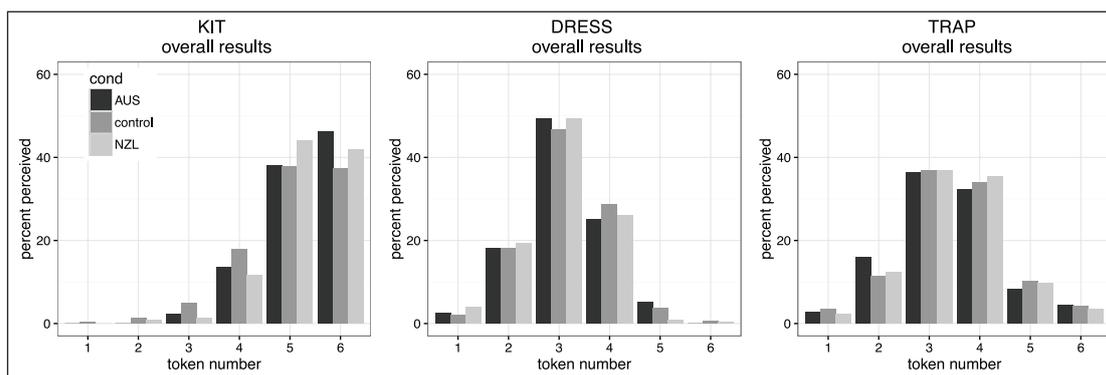
Figure 3.1: Token selections – all conditions. Lower numbers represent more NZE-like tokens.

Table 3.2: Coefficients table for KIT vowel responses. Cumulative link mixed model for ordinal data with Australia as default condition. Model syntax: clmm(response ~ (1+order|participant) + (1+order*condition|word) + cond*order*NZES).

	Estimate	*SE*	*z* value	Pr(>|z|)	
condition=control	−0.545	0.600	−0.908	0.364	
condition=NZ	0.208	0.605	0.344	0.731	
order=reversed	1.376	0.451	3.050	0.0023	**
NZES	0.107	0.121	0.883	0.377	
cond=control:order=reversed	−0.444	0.622	−0.713	0.476	
cond=NZ:order=reversed	0.190	0.634	0.300	0.764	
cond=control:NZES	−0.099	0.151	−0.656	0.512	
cond=NZ:NZES	−0.169	0.150	−1.124	0.261	
order=reversed:NZES	−0.136	0.125	−1.092	0.275	
cond=control:order=reversed:NZES	0.262	0.155	1.688	0.091	
cond=NZ:order=reversed:NZES	0.078	0.153	0.510	0.610	

3.1. Statistical analysis

Coefficients from the mixed effects ordinal logistic regression models described in Section 2.5.2 are presented in **Tables 3.2–3.4**. Regarding the first hypothesis—that stuffed toy kiwis will influence Australian listeners' selections in the vowel matching task—no main effect of condition was found. There was no significant difference between selections in the Australian and New Zealand conditions for any of the three target vowels (KIT, DRESS, and TRAP). Further, no significant difference was found between token selections in the Control condition and either primed condition for any of the three target vowels. In addition, we also found no support for our second hypothesis—that familiarity with New Zealand and NZE would affect listeners' token selections, as there was no effect of NZES. There were insufficient numbers of participants to conduct an analysis of only those listeners with elevated NZES.

As shown in the above tables, continuum presentation order had a significant effect in the KIT model ($p = .0023$), and a nearly significant effect in the DRESS model ($p = .0518$). Participants selected higher numbered tokens when the continuum order was reversed (KIT: original 4.99, reversed 5.38; DRESS: original 2.94, reversed 3.26; TRAP original 3.44, reversed 3.47). That is, more AusE-like tokens were selected when the continuum presented the most AusE-like token first and the most NZE-like token last. These data

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 19 of 29

Table 3.3: Coefficients table for DRESS vowel responses. Cumulative link mixed model for ordinal data with Australia as default condition. Model syntax: clmm(response ~ (1+order|participant) + (1+order*condition|word) + cond*order*NZES).

	Estimate	SE	z value	Pr(>|z|)
condition=control	−0.454	0.407	−1.116	0.265
condition=NZ	−0.512	0.424	−1.208	0.227
order=reversed	0.920	0.473	1.945	0.0518
NZES	−0.035	0.083	−0.414	0.679
cond=control:order=reversed	0.168	0.605	0.277	0.782
cond=NZ:order=reversed	0.111	0.615	0.180	0.857
cond=control:NZES	0.109	0.103	1.061	0.289
cond=NZ:NZES	0.021	0.103	0.198	0.843
order=reversed:NZES	−0.105	0.123	−0.857	0.392
cond=control:order=reversed:NZES	0.095	0.152	0.624	0.533
cond=NZ:order=reversed:NZES	0.097	0.152	0.637	0.524

Table 3.4: Coefficients table for TRAP vowel responses. Cumulative link mixed model for ordinal data with Australia as default condition. Model syntax: clmm(response ~ (1+order|participant) + (1+order*condition|word) + cond*order*NZES).

	Estimate	SE	z value	Pr(>|z|)
condition=control	0.012	0.477	0.026	0.979
condition=NZ	−0.119	0.476	−0.251	0.802
order=reversed	−0.529	0.456	−1.160	0.246
NZES	0.019	0.097	0.193	0.847
cond=control:order=reversed	0.872	0.616	1.416	0.157
cond=NZ:order=reversed	0.710	0.626	1.134	0.257
cond=control:NZES	−0.013	0.120	−0.109	0.913
cond=NZ:NZES	0.064	0.120	0.538	0.591
order=reversed:NZES	0.119	0.127	0.938	0.348
cond=control:order=reversed:NZES	−0.156	0.156	−0.998	0.318
cond=NZ:order=reversed:NZES	−0.190	0.157	−1.212	0.225

suggest that if a participant selected, for example, token 5 as the best match to the target vowel in the sentence 'The new movie was a huge summer *hit*,' when the continuum was presented in the original order, they would not necessarily select the equivalent token 2 when the continuum was presented in reverse order. Instead, participants were more likely to select the more AusE-like token 1 (i.e., token 6 in the original order). The fact that this continuum presentation order effect is most pronounced for the KIT vowel set might be due to the relative extent of differences between end points for each of the three synthesized vowel set continua. As shown in Section 2.2.4, the Euclidean distance between token 1 and token 6 was significantly larger for the KIT continuum than for the DRESS continuum, while the TRAP continuum was significantly shorter than both the KIT and DRESS continua.

As mentioned above, Hay and Drager (2010) used a linear fixed effects regression model without taking random effects, such as listener and item, into account. To compare our

Table 3.5: Coefficients table for the linear fixed effects only model. Coefficients table for KIT vowel responses as predicted by a linear fixed effects model with Australia as default condition. Model syntax: lm(response ~ cond + NZES).

	Estimate	SE	t value	Pr(>|z|)	
(Intercept)	5.143	0.05996	85.775	<0.001	***
cond=NZ	−0.091	0.07102	−1.277	0.202	
NZES	−0.003	0.01250	−0.278	0.781	

results more directly with Hay and Drager (2010), we also ran a separate linear regression analysis on our KIT vowel data set with fixed effects only (see **Table 3.5**). This model included condition (Australia and New Zealand) and NZES as predictors. To keep the data set as similar as possible to the previous study, we excluded responses from the Control condition and to continua presented in the reversed presentation order. Neither condition nor NZES had a significant effect on listeners' vowel perception.

3.2. Speaker origin

When participants were asked to identify the speaker's origin, 86.67% correctly responded Australia with only 4% responding New Zealand (one participant did answer Australia or New Zealand). Consistent with the findings reported in Hay et al. (2006) the presence of the stuffed toy(s) did not influence participant's belief about the speaker's origin. Responses across the three conditions were as follows. In the Control condition: Australia (21), New Zealand (2), Australia or New Zealand (1), India (1). In the Australian condition: Australia (23) and European (2). In the New Zealand condition: Australia (21), New Zealand (1), European (1), Middle East (1), United Kingdom (1).

4. Discussion

Hay and Drager (2010, p. 883) suggest that "subtle differences in experimental environment can influence subjects' responses." Our aim was to establish whether the previously observed regional priming effect could be reproduced in an Australian context with an AusE-speaking listener sample. Using priming conditions similar to those in Hay and Drager (2010), we hypothesized that participants exposed to a New Zealand prime (stuffed toy kiwis) would select continuum tokens that represent more NZE-like vowels than those exposed to an Australian prime (stuffed toy koalas).

No support was found for this hypothesis. Token selections in our matching task did not differ significantly between the Australian and New Zealand conditions. Although some variance was observed between selections in the Control condition and both priming conditions (see **Figure 3.1**), this variance was not found to be significant using a mixed effects regression model. Our second hypothesis, that exposure to NZE would be a significant predictor for response variability between the conditions, was also not supported by the analysis. NZES did not significantly influence token selection.

Lawrence's (2015) replication of Niedzielski (1999) using BATH and STRUT vowels with Southern British English listeners also failed to find support for the idea that priming listeners towards another dialect could influence their performance in a matching task. Although results in Lawrence (2015) appeared to indicate some shift consistent with the priming condition, this shift was not found to be significant. Lawrence (2015) argues that the influence of a regional prime is either more limited than was previously suggested or the influence exists but is highly contextually specific. While the results reported in the present study might be seen to add support to this conclusion, there are a number of other possible explanations for the observed result.

One possible explanation is an ineffective or unpredictable priming condition. This explanation has been offered in previous experiments. For example, Squires (2013) acknowledged that the lack of an observable priming effect may have been due to the primes not working as expected. If the effect reported in Hay and Drager (2010) relies on a strong association between the toy, its elicited dialect, and relevant dialectal variation, then it is possible that the toy kiwis were not culturally significant enough to activate 'New Zealand' for our participants. In other words, there is no way of knowing whether those participants in the New Zealand condition associated the toy kiwis with New Zealand. It may be that the kangaroos and koalas used in Hay and Drager's (2010) experiment are more salient signifiers of Australia for New Zealanders than our kiwis are signifiers of New Zealand for Australians. One way to overcome this problem in the future would be to test the identification or recognition of the kiwis and other toys after their use in the experiment.

4.1. Exposure and sensitivity to NZE

Another possible explanation for our lack of demonstrable regional priming is that the participants may not have the level of NZE exposure required to complete the matching task as predicted. It has been shown that dialect recognition and feature identification is facilitated by a listener's familiarity with that dialect (Clopper & Pisoni, 2007; Sumner & Samuel, 2009). Further, Labov (2010) found a 'significant local advantage' whereby listeners from three dialect areas of North America were better at identifying vowels produced by a speaker from their own area than from one of the other two areas. Questionnaire responses in the present study indicated that 20 of the 75 participants (eight in the Control condition, five in the Australian condition, and seven in the New Zealand condition) had never been to New Zealand, didn't speak with, or know, any New Zealanders, and couldn't name any New Zealand media. Although these individuals might be generally aware of NZE and how it differs from AusE, it is possible that they are not.

Expecting that exposure to NZE would improve a listener's representation of NZE vowels, as well as increase the likelihood of prime recognition, we predicted that familiarity with New Zealand and NZE would affect listeners' selections in the vowel matching task. Yet, NZES did not emerge as a significant factor in our model. This suggests that our result cannot be attributed to a lack of exposure to NZE alone. There is evidence to support the idea that Australians are generally sensitive to the differences between NZE and AusE and are aware of the most salient differences between the dialects (Bayard et al., 2001; Ludwig, 2007; Weatherall et al., 1998). This suggests that if New Zealanders showed the priming effect, it is reasonable to expect that Australians might do so too.

In addition, naïve listeners have been shown to make use of reliable acoustic-phonetic properties to identify dialects of American English (Clopper & Pisoni, 2004, 2007; Preston, 1993) and Welsh English (Williams, Garrett, & Coupland, 1999), as well as variation in Dutch (Van Bezooijen & Gooskens, 1999). The salience of the difference between AusE and NZE KIT (Maclagan et al., 1999; Watson et al., 1998) suggests a reliable acoustic-phonetic difference between the two which should be detectable even for those participants who may otherwise lack frequent exposure to NZE. It could be that, despite the level of exposure assumed to our participants, they did not have the requisite fine phonetic knowledge that would result in meaningful influence from the priming condition. Future studies would benefit from assessing participants' ability to identify AusE and NZE in a post-experiment task. If participants could reliably identify a NZE speaker on the basis of KIT, DRESS, and TRAP vowels and still fail to show any influence of the prime, this would strengthen the null result.

4.2. Implications for exemplar theory

The lack of a significant priming effect in the present study does not in itself contradict Hay et al. (2006) and Hay and Drager's (2010) support for an exemplar-based model of speech perception. According to exemplar models, speech input is categorized by comparing the relative activation levels for each candidate category (Hay et al., 2006; Pierrehumbert, 2001). There is evidence to suggest that socially salient cues may bias categorization in favor of acoustic variants associated with that indexical information (Foulkes & Docherty, 2006). The system may even be directed towards a particular categorization when the acoustic information does not match (Niedzielski, 1999). Hay and Drager (2010) argued that, for New Zealanders, stuffed toy koalas and kangaroos raise the activation level of exemplars indexed as 'Australian.' Phonetic input is then more likely to be classified along with those raised exemplars, because the activated portion of the category distribution is centered around Australian exemplars (Hay et al., 2006, p. 24). Hence, the bias towards more AusE-like vowels. It may be that for our experiment, any existing New Zealand or NZE exemplars were not sufficiently activated by the kiwi to compete with the resting activation level of AusE exemplars. In order to produce a shift in token selection towards NZE-like variants and further test the predictions of exemplar theory, a more overt prime might have been required to activate the relevant indexical and linguistic categories.

4.3. Working memory and paradigm issues

A potential issue with the design of this task lies in the assumption that participants are able to give equal consideration to all six tokens when selecting the best match. Research into the limitations of working memory suggests that the requirements of this task may be beyond the capabilities of untrained listeners. This may have led to accuracy issues which, given the subtleties of the effect overall, are troubling. Miller (1956) proposed that working memory was limited to lists of seven items, plus or minus two, a conclusion supported by Kinsbourne and Cohen (1971). Additional studies investigating working memory capacity support an even more modest limit of four items (Cowan, 2001; Luck & Vogel, 1997; Sperling, 1960). Li, Cowan, and Saults (2012) also found that listeners struggled to retain more than four tones in memory. Further, Baddeley (1992, 2010) suggested that a phonological similarity effect impairs recall of words that are similar in sound, although Medin and Bettger (1994) found that there may be a processing benefit when stimuli are presented in a way that maximizes similarity between successive items. It is worth mentioning that none of the pilot participants used for the present study indicated that the task was difficult. Although, one pilot participant did mention that she had trouble earlier on in the task until she started repeating the target word to herself between continuum tokens. This strategy itself may be problematic for the task, as the mental representation of the speaker's target word could be influenced by the listener's repetitions.

  It could be that the continuum design used in the present study changed the focus of the matching task. Hay and Drager (2010) proposed that variation in responses between the priming conditions was evidence of regional information biasing the categorization of phonemes. By presenting a single set of continuum tokens within each vowel context, with no acoustic match, the participant is required to match the target vowel to a token that is either more AusE-like or more NZE-like than the target vowel. In contrast, our continua were unique to each target word and did include an acoustic match. As discussed in Section 2.2.3, our intention was to present continuum tokens that were acoustically modelled on the speaker's actual realization of each target word to minimize variance in token selection that could be attributed to factors other than a priming effect, such as the influence of coarticulation. Although the results presented in **Figure 3.1** demonstrate that our participants did not simply select the acoustic match and the overall distribution

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 23 of 29

of responses is comparable to results from previous studies, it is possible that our design reduced the size of any effect that may exist. A second consideration is that participants in the previous studies may have recognized consistency in the continua; they then may have been able to anticipate their response upon hearing the target word. Any observable priming effect in the present study might have been masked by the potential for the novel continua to increase the demand on a participant's working memory. However, it is worth restating that no trends towards a priming effect were found in our analysis that would indicate our design reduced the size of a regional priming effect. Either way, our efforts to reduce the possibility for ambiguity or uncertainty in responses by creating a more signal-driven matching task may provide another explanation as to why our results differ from Hay and Drager (2010), Hay et al. (2006), and Niedzielski (1999).

4.4. Ordering effect

Our analysis did reveal that KIT vowel selections were significantly more AusE-like when the continuum was played in the reversed order (from AusE-like to NZE-like). This may be further evidence that the task is undermined by the limitations of participants' working memory. As discussed above, it is possible that participants did not have the working memory capacity to hold six acoustically similar synthetic tokens equally in memory in order make an accurate comparison to a target vowel. Our results indicate that participants favored the two exaggerated AusE tokens for KIT vowels, which represent more peripheral selections than those made for either DRESS or TRAP vowels. When tokens were presented in the reversed presentation order, these two exaggerated AusE tokens would be heard first. Participants would then hear the four additional continuum tokens before being able to make their selection. Unsure of which of the first two tokens they preferred, participants may have simply selected the first token as a more certain option.

4.5. Limitations

Unfortunately, we did not attract enough male participants to test the gender effect observed in Hay et al. (2006) and Hay and Drager (2010). In both of these previous studies, males showed the opposite effect to females, selecting more NZE-like tokens in the Australian condition. Hay and Drager (2010) argued that many New Zealand males have an inherently negative association towards Australia due to a sporting rivalry between the two countries. This resulted in males displaying a divergence response in the Australian condition. It would not have been surprising if this attitude was reciprocated by Australian males towards New Zealand.

As we elected to restrict the target words used in the experiment to the forms /CVt/ or /CCVt/, we were unable to control for lexical frequency. Although the constraints on our target words resulted in a more phonetically controlled set of stimuli than those used in previous experiments, in order to include an adequate sample of items we did include some less frequent lexical items (such as *grit*). It is possible that lexical frequency could influence the experiment and further research should take this into account, either in experimental design or analysis.

Throughout this discussion, we have offered potential ideas for continuation of research in this paradigm. Despite the findings of the present study supporting the null result reported in Lawrence (2015), the extent to which regional priming influences speech perception warrants ongoing experimentation. Given the methodological and procedural issues identified in this analysis, there are two main areas that would require modification: presentation of stimuli and introduction of the priming condition. The task might be best suited to four token continua with participants required to make a comparison to a vowel in an isolated target word. Alternatively, each token could be immediately preceded by

the target word with participants required to rate the similarity of the token to that target word. In such a task, the ordering or labelling of tokens would be unnecessary. Incorporating an additional post-task questionnaire requiring participants to identify the prime, state whether it was noticed, and identify features of the primed dialect would assist evaluation of the success of the priming condition and its effect.

5. Conclusions

The stuffed toy priming effect observed in Hay and Drager (2010) was not replicated in this study. For each of the three target vowels: KIT, DRESS, and TRAP, token selections did not vary significantly with the priming condition. This may be a result of cultural asymmetry in recognition and familiarity between New Zealanders and Australians; however, even those participants who indicated frequent contact with NZE did not show sensitivity to the New Zealand prime. It may simply be that the effect is limited to highly contextually specific situations, such as those identified in Hay and Drager (2010) and Niedzielski (1999). In addressing the lack of statistical significance, we considered potential complications associated with the priming condition and sensitivity or exposure to NZE. These issues highlight the need for carefully considered experimental design, particularly when investigating variation at a fine phonetic level.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Full sentence list with identified target words. DOI: https://doi.org/10.5334/labphon.90.s1
- **Appendix B.** Formant 1 Values in Hertz. DOI: https://doi.org/10.5334/labphon.90.s2
- **Appendix C.** Formant 2 Values in Hertz. DOI: https://doi.org/10.5334/labphon.90.s3
- **Appendix D.** Formant analysis. DOI: https://doi.org/10.5334/labphon.90.s4

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437–456. DOI: https://doi.org/10.1017/S0047404510000400

Babel, M., & Russell, J. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823–2833. DOI: https://doi.org/10.1121/1.4919317

Baddeley, A. 1992. Working memory. *Science, 255*(5044), 556–559. DOI: https://doi.org/10.1126/science.1736359

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 25 of 29

Baddeley, A. 2010. Working memory. *Current Biology, 20*(4), R136–R140. DOI: https://doi.org/10.1016/j.cub.2009.12.014

Bauer, L., & Warren, P. 2004. New Zealand English: Phonology. In: Kortmann, B., Schneider, W., Burridge, K., Mesthrie, R., & Upton, C. (eds.), *A Handbook of Varieties of English, 1*, 580–602. Berlin: Mouton de Gruyter.

Bauer, L., Warren, P., Bardsley, D., Kennedy, M., & Major, G. 2007. New Zealand English. *Journal of the International Phonetic Association, 37*(1), 97–102. DOI: https://doi.org/10.1017/S0025100306002830

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. 2001. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics, 5*(1), 22–49. DOI: https://doi.org/10.1111/1467-9481.00136

Boersma, P., & Weenink, D. 2014. *Praat: Doing phonetics by computer* (Version 5.4). Retrieved 16 October 2014 from: http://www.praat.org/.

Bybee, J. 2006. From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733. DOI: https://doi.org/10.1353/lan.2006.0186

Christensen, R. 2015. *2015 ordinal: Regression Models for Ordinal Data.* (R Package Version. 6-28).

Clopper, C. G., & Pisoni, D. B. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*(1), 111–140. DOI: https://doi.org/10.1016/S0095-4470(03)00009-3

Clopper, C. G., & Pisoni, D. B. 2007. Free classification of regional dialects of American English. *Journal of Phonetics, 35*(3), 421–438. DOI: https://doi.org/10.1016/j.wocn.2006.06.001

Cowan, N. 2001. The magical number four in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24*, 87–114. DOI: https://doi.org/10.1017/S0140525X01003922

Cox, F. 1999. Vowel change in Australian English. *Phonetics, 56*, 1–27. DOI: https://doi.org/10.1159/000028438

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

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

Cox, F., & Palethorpe, S. 2008. Reversal of short front vowel raising in Australian English. *Proceedings of Interspeech*, 342–345. Available at: http://hdl.handle.net/1959.14/156260.

Cox, F., & Palethorpe, S. 2014. Phonologisation of vowel duration and nasalised /æ/ in Australian English. In: Hay, J., & Parnell, E. (eds.), *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, 33–36. Christchurch, New Zealand.

Cox, F., Palethorpe, S., Miles, K., & Davies, B. 2014. Is there evidence for region specific vowel variation in /hVd/ word list data from AusTalk? *Paper presented at the Australian Linguistic Society Conference*. Newcastle.

Drager, K. 2011. Speaker age and vowel perception. *Language & Speech, 54*(1), 99–121. DOI: https://doi.org/10.1177/0023830910388017

Drager, K., Hay, J., & Walker, A. 2010. Pronounced rivalries: Attitudes and speech production. *Te Reo, 53*, 27–53.

Easton, A., & Bauer, L. 2000. An acoustic study of the vowels of New Zealand English. *Australian Journal of Linguistics, 20*(2), 93–117. DOI: https://doi.org/10.1080/07268600020006021

Foulkes, P., & Docherty, G. 2006. The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409–438. DOI: https://doi.org/10.1016/j.wocn.2005.08.002

Foulkes, P., Scobbie, J. M., & Watt, D. 2010. Sociophonetics. *The Handbook of Phonetic Sciences, Second Edition,* 703–754. DOI: https://doi.org/10.1002/9781444317251.ch19

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. DOI: https://doi.org/10.1037/0278-7393.22.5.1166

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. 2004. *New Zealand English: Its Origins and Evolution.* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486678

Green, P., & MacLeod, C. J. 2016. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7,* 493–498. DOI: https://doi.org/10.1111/2041-210X.12504

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

Hay, J., Nolan, A., & Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. *The linguistic review, 23*(3), 351–379. DOI: https://doi.org/10.1515/TLR.2006.014

Hay, J., Warren, P., & Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484. DOI: https://doi.org/10.1016/j.wocn.2005.10.001

Jannedy, S., Weirich, M., & Brunner, J. 2011. The effect of inferences on the perceptual categorization of Berlin German fricatives. *Proceedings of the International Congress of Phonetic Sciences, ICPHS 2011*, 962–965.

Johnson, K. 1997. Speech perception without speaker normalization. In: Johnson, K., & Mullennix, J. W. (eds.), *Talker Variability in Speech Processing,* 145–166. New York: Academic Press.

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: https://doi.org/10.1006/jpho.1999.0100

Kinsbourne, M., & Cohen, V. 1971. English and Hebrew consonant memory span related to the structure of the written language. *Acta Psychologica, 35*(5), 347–351. DOI: https://doi.org/10.1016/0001-6918(71)90009-6

Koops, C., Gentry, E., & Pantos, A. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics, 14*(2), 12. Available at: http://repository.upenn.edu/pwpl/vol14/iss2/12.

Labov, W. 2010. A controlled experiment on vowel identification. In: *Principles of Linguistic Change,* 48–58. Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444327496.ch3

Lacerda, F. 1997. Distributed memory representations generate the perceptual-magnet effect. Manuscript, Institute of Linguistics, Stockholm University.

Lawrence, D. 2015. Limited evidence for social priming in the perception of the bath and strut vowels. In: *The Scottish Consortium for ICPhS 2015, Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: The University of Glasgow.

Li, D., Cowan, N., & Saults, J. S. 2012. Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, & Psychophysics, 75*(1), 145–160. DOI: https://doi.org/10.3758/s13414-012-0383-z

Luck, S. J., & Vogel, E. K. 1997. The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. DOI: https://doi.org/10.1038/36846

Ludwig, I. 2007. Identification of New Zealand English and Australian English based on stereotypical accent markers. Unpublished master's thesis, University of Canterbury.

Maclagan, M. A., Gordon, E., & Lewis, G. 1999. Women and sound change: Conservative and innovative behavior by the same speakers. *Language Variation and Change*, *11*(1), 19–41. DOI: https://doi.org/10.1017/S0954394599111025

Maclagan, M. A., & Hay, J. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English front vowel shift. *Language Variation and Change*, *19*(1), 1–25. DOI: https://doi.org/10.1017/S0954394507070020

May, J. 1976. Vocal tract normalization for /s/ and /ʃ/. *The Journal of the Acoustical Society of America*, *59*(S1), S25–S25. DOI: https://doi.org/10.1121/1.2002554

McGowan, K. B. 2015. Social expectation improves speech perception in noise. *Language and Speech*, *58*(4), 502–521. DOI: https://doi.org/10.1177/0023830914565191

McLennan, C. T. 2007. Challenges facing a complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, 67–70. Saarbrücken: Saarland University.

Medin, D. L., & Bettger, J. G. 1994. Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*(2), 250–254. DOI: https://doi.org/10.3758/BF03200776

Miller, G. A. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. DOI: https://doi.org/10.1037/h0043158

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

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: https://doi.org/10.1177/0261927X99018001005

Niedzielski, N. A. 1997. The effect of social information on the phonetic perception of sociolinguistic variables. Unpublished doctoral dissertation, University of California, Santa Barbara. Available from ProQuest Dissertations & Theses database.

Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J., & Hopper, P. J. (eds.), *Frequency effects and emergent grammar*, 137–158. Amsterdam & Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/tsl.45.08pie

Pierrehumbert, J. B. 2006. The next toolkit. *Journal of Phonetics*, *34*(4), 516–530. DOI: https://doi.org/10.1016/j.wocn.2006.06.003

Pierrehumbert, J. B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*(1), 33–52. DOI: https://doi.org/10.1146/annurev-linguistics-030514-125050

Preston, D. 1993. Folk dialectology. In: Preston, D. (ed.), *American dialect research*, 333–378. Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/z.68.17pre

Psychology Software Tools, Inc. 2012. E-Prime 2.0. Retrieved from: http://www.pstnet.com.

R Core Team. 2016. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531. DOI: https://doi.org/10.1007/BF00973770

Sanchez, K., Hay, J., & Nilson, E. 2015. Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics, 48*, 76–95. DOI: https://doi.org/10.1016/j.wocn.2014.10.004

Sperling, G. 1960. The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1–29. DOI: https://doi.org/10.1037/h0093759

Squires, L. 2013. It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics, 17*(2), 200–237. DOI: https://doi.org/10.1111/josl.12025

Strand, E. A., & Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D. (ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS conference, Bielefeld, October 1996*, 14–26. Berlin: Mouton. DOI: https://doi.org/10.1515/9783110821895-003

Sumner, M., & Samuel, A. G. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Thomas, E. R. 2002. Sociophonetic applications of speech perception experiments. *American Speech, 77*, 115–47. DOI: https://doi.org/10.1215/00031283-77-2-115

Van Bezooijen, R., & Gooskens, C. 1999. Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology, 18*, 31–48. DOI: https://doi.org/10.1177/0261927X99018001003

Walker, A., Hay, J., Drager, K., & Sanchez, K. 2018. Divergence in speech perception. *Linguistics 56*(1), 257–278. DOI: https://doi.org/10.1515/ling-2017-0036

Watson, C. I., Harrington, J., & Evans, Z. 1998. An acoustic comparison between New Zealand and Australian English vowels. *Australian Journal of Linguistics, 18*(2), 185–207. DOI: https://doi.org/10.1080/07268609808599567

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. DOI: https://doi.org/10.1017/S0954394500121039

Weatherall, A., Gallois, C., & Pitam, J. 1998. Australasians identifying Australasian accents. *Te Reo, 41*, 153–162.

Wedel, A. B. 2006. Exemplar models, evolution and language change. *Linguistic Review, 23*(3), 247–274. DOI: https://doi.org/10.1515/TLR.2006.010

Wells, J. C. 1982. *Accents of English*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511611759

Williams, A., Garrett, P., & Coupland, N. 1999. Dialect recognition. In: Preston, D. R. (ed.), *Handbook of perceptual dialectology*, 345–358. Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/z.hpd1.29wil

Zwicker, E. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America, 33*(2), 248–248. DOI: https://doi.org/10.1121/1.1908630

Walker et al: Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?

Art. 7, page 29 of 29