Regional priming in Australian English KIT, DRESS and TRAP vowels

Michael Walker, Anita Szakay, Felicity Cox

Department of Linguistics, Macquarie University, Australia

michael.walker@mq.edu.au, anita.szakay@mq.edu.au, felicity.cox@mq.edu.au

Abstract

40 speakers of AusE participated in a pseudo-replication of Hay and Drager's stuffed toy study [1]. Participants heard spoken AusE sentences with a phrase-final target word containing either a KIT, DRESS or TRAP vowel [2] followed by a six-step continuum of isolated synthesised vowels, from NZE-like to exaggerated AusE-like. Participants selected the token from the continuum they believed best matched the realisation of the vowel in the target word. Participants were exposed to stuffed toy kiwis, koalas, or neither (control). Contrary to results reported in [1], participants' responses were not influenced by the presence of the toys.

Index Terms: speech perception, vowel perception, sociophonetics, Australian English, New Zealand English

1. Introduction

Reference to a particular country or dialect has been shown to influence listeners' performance in perceptual tasks [1], [3], [4]. Listeners are thought to be highly sensitive to the sociophonetic consequences of a speaker's social characteristics, such as nationality, and may anticipate certain features in speech, resulting in predictable perceptual biases [4]. These findings support the idea that indexical, as well as linguistic information plays a role in speech perception.

Niedzielski [3] showed that the expected dialect of a speaker could influence the categorisation of socially salient dialectal variation. Canadians and Detroiters both typically produce a raised MOUTH vowel, consistent with vowels influenced by Canadian Raising (CR), yet Detroiters do not recognise the raising in their own speech. Participants from Detroit heard recorded speech from a single Detroit speaker but were told the speaker was from either Canada or Detroit, with response sheets also labelled *Canadian* or *Michigan*. Phrases with words containing a target MOUTH vowel were played, followed by a continuum of six synthesised tokens ranging from raised to unraised variants of the target vowel. Participants selected the token they believed was most like the vowel in the target word. Participants in the Canadian condition selected on average a more raised variant than those in the Michigan condition. [3] argues that because Detroiters only associate CR with Canadian English they would not expect to hear raised variants in a speaker of their own dialect.

Hay, Nolan and Drager [4] replicated the procedure outlined in [3] by testing New Zealanders' perception of KIT, DRESS and TRAP vowels produced by a speaker of New Zealand English (NZE). Participant answer sheets were labelled with either *Australia* or *New Zealand*. Continua consisted of six tokens from Australian English (AusE)-like to NZE-like. Perception of vowels was shown to shift in the direction of well-recognised differences between AusE and NZE, consistent with the priming condition. However, this effect was limited to female participants. KIT vowels, being the most salient difference between AusE and NZE, showed the highest variability congruence with the prime. A lesser effect was observed in TRAP vowels, and DRESS vowels showed no influence of the prime. In a post-task questionnaire, all participants, except one, identified the speaker as a NZE speaker. It was suggested that activation of the concept of Australia, rather than participants believing the speaker was Australian, was enough to shift perception of vowels towards values associated with AusE [4].

Hay and Drager [1] proposed that any culturally significant token might produce a similar effect. In [1], participants completed the same task as [4], using identical auditory stimuli. Stuffed toy koalas and kangaroos were used to activate 'Australia' resulting in NZE-speaking participants selecting more AusE-like tokens when compared with a group exposed to stuffed toy kiwis. However, the effect was again limited to female participants and KIT vowels.

In another replication of [3], Lawrence [5] found no evidence for regional labels ('Sheffield, Northern England' or 'London, Southern England') influencing responses in BATH and STRUT vowels. Despite BATH and STRUT being highly salient regional markers, no priming effect was found. This suggests the priming effect observed in [1], [3], [4] may be highly contextually specific and not generalisable.

The present study is a pseudo-replication of the experiment outlined in [1], [3], [4]. It seeks to determine whether a shift in perception as a consequence of the stuffed toy prime, as reported in [1], would occur in an AusE-speaking sample. This study also aims to address a number of design and procedural concerns in [1], [3], [4] relating to control of stimuli, continua construction and the lack of a control condition. Consistent with [1], [4] the experiment focuses on KIT, DRESS and TRAP vowels in AusE and NZE.

The NZE KIT is relatively centralised and typically realised as [ɘ] [6] while in AusE it is realised as [ɪ] [7]. The difference in KIT between the two dialects is said to be the most salient and well known to laypeople [4] as exemplified in the stereotyped "fush and chups" (NZE) vs "feesh and cheeps" (AusE). The NZE DRESS is typically a more raised variant of /e/ than the AusE counterpart [8]. TRAP is typically realised as [ɛ] in NZE and [æ] in AusE [8**]**.

Hay et al. [1], [4] draw on an exemplar model of speech perception to account for the priming effect shown in [1], [3], [4]. The model proposes that linguistic information is represented in memory as phonetically detailed exemplars. Exemplars form categories which represent a class of equivalent perceptual experiences [10]. Social information about the speaker is also retained. Categorisation of new input involves comparison to existing exemplars [10] and is said to be biased towards frequently activated acoustic and social categories [4]. Accordingly, by activating 'New Zealand' with stuffed toy kiwis, Australian listeners would be biased towards categorising speech input as NZE, provided the exemplars associated with NZE are robust.

Following the results presented in [1], [3], [4] and assuming an exemplar model of speech perception, we expect that, for those AusE-speaking participants exposed to the kiwi, the perception of target vowels would shift towards more NZE-like vowel qualities. Consistent with [1], [4] this shift should be present in words containing KIT vowels but may not be with DRESS and TRAP vowels. However, according to an exemplar model, associations between social and linguistic categories emerge only after frequent activation. For this reason, we predict that the priming effect will be limited to those participants who indicate some level of exposure to New Zealand and NZE.

2. Methods

2.1. Participants

40 female speakers of AusE participated in the perception task. 3 males also participated but their data will not be discussed here. All were born and educated in Australia and were undergraduate students at Macquarie University in Sydney receiving course credit for their time. Ages ranged from 18-27 with a mean age of 19.5.

2.2. Stimuli and materials

2.2.1. Target vowels and target words

KIT, DRESS and TRAP target vowels were represented in /CVt/ or /CCVt/ monosyllabic words. 10 different words were used for each of the three target vowels. Due to lexical restrictions, some complex onsets were included (*grit*, *skit*, *slat*, *Brett*, *threat*). The set of target words represent more controlled stimuli than those used in [1], [4], which used multiple coda consonants.

2.2.2. Sentences

Target words were each embedded in 30 unique sentences. Each sentence contained the target word in phrase-final position thus participants would not be exposed to any additional vowels between the target and continuum tokens. In addition, consistency in target word position allowed for straightforward target identification by participants. Unlike the stimuli used in [1], [4], there were no other examples of the target vowels in stressed position in any sentence. This minimised any additional priming effect by reducing overt identifiers of AusE (when contrasted with NZE). Example sentences are shown:

1. The new movie was a huge summer **hit**
2. She is studying to become a **vet**
3. The bug looked like it was a **gnat**

A 19-year-old male monolingual speaker of Standard AusE from Sydney read the sentences. Stimuli 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 48kHz sampling rate.

2.2.3. Continua

Continua consisted of six synthesised vowel tokens representing equal steps from NZE-like (token 1) to exaggerated AusE-like (token 6). Characteristics of the speaker's spoken target vowel from each sentence (i.e. F_1 and F_2) were used to synthesise token 4 in Praat [11]. F_1 and F_2 values of the synthesised vowels were then manipulated to create the additional five tokens (see Table 1 for examples).

Table 1. *Formant values for 'hit', 'vet' and 'gnat' continuum tokens.*

Token	hit		vet		gnat	
	F1	F2	F1	F2	F1	F2
1	484	1627	380	2078	655	1995
2	445	1781	466	1987	722	1852
3	407	1948	560	1900	792	1718
4	370	2130	660	1817	866	1594
5	334	2330	769	1737	944	1480
6	300	2551	887	1661	1027	1371

Step intervals between continua tokens were calculated in bark steps [12]. For KIT vowels, intervals were 0.35 bark steps for F_1 and 0.6 bark steps for F_2. For DRESS the intervals were 0.8 (F_1) and 0.3 (F_2). For TRAP, 0.5 (F_1) and 0.5 (F_2). Step intervals were calculated such that F_1 and F_2 values for token 1 were within 2 standard deviations of mean NZE values published in [13]. Mean AusE values provided by [14] were used as a reference for tokens 4-6. This differs from the continua used in [1], [4], which were based on a single vowel produced by the speaker in a /hVd/ context. This meant that in [1], [4] continua tokens were consistent but token 4 would not align exactly with the speaker's vowel in each target word. We elected to create a unique continuum for each target word where the F_1 and F_2 values of token 4 matched those of the vowel produced by the speaker in that sentence. Synthesised tokens were 180ms in duration. Initial F_0 was 160Hz (the mean initial F_0 for all tokens) with an F_0 slope of -1 octaves per second applied to each of the synthesised vowels. Figure 1 shows spectrograms of the synthesised continuum tokens for *hit*.

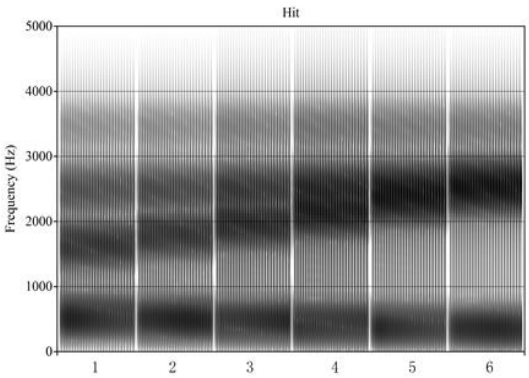

Figure 1. *Spectrogram of continuum tokens for 'hit' from most NZE-like to exaggerated AusE.*

2.3. Procedure

2.3.1. Priming

Participants were exposed to one of three conditions: Australian (n = 13), New Zealand (n = 14) and a control (n = 13). Similar to [1], priming for the conditions was cued by the presence of stuffed toy koalas (Australian) or kiwis (New Zealand) with the control exposed to no toys. To introduce the prime, the experimenter 'found' the headphones required for the task in a drawer under the toys. The experimenter mentioned the toys were being used in another experiment and

placed them on the table beside the computer being used for the task. The intent was to draw attention to the prime without it being obvious that it was related to the experiment. The toys remained in the participant's line of sight for the duration of the experiment. This was similar to the procedure used in [1]. A control was not used in [1], [3-5], however, it was deemed important to test the perception task in an un-primed context.

Each participant completed the task individually in a soundproofed room, which contained no other regional primes. The only difference between the conditions was the presence of the prime itself. All participants interacted with the same experimenter – a 29-year-old male, born in New Zealand who had moved to Australia at age 10 – so any potential priming effect from the experimenter is consistent for all participants.

2.3.2. Perception task

Participants heard each sentence while it was simultaneously presented on-screen with the target word bold and underlined. Sentences remained visible for the duration of the recorded stimuli plus an additional 2000ms. Following the sentence, continuum tokens were played with the corresponding number for each token visible for 1000ms. After hearing all six continuum tokens, participants were prompted to make their selection by pressing the corresponding number key. The task did not continue until a selection was made. Each sentence was presented once with the continuum tokens in order from token 1 to token 6, and once with the continuum tokens in reverse order. Presentation of the synthesised tokens with a visual numeric label differs slightly from [1], [4] who used spoken numbers preceding each token to label the steps. Visual labelling was preferred to reduce any additional priming effect, particularly from the word *six*, which contains a KIT vowel. Prior to the main task, participants completed three familiarisation questions using non-target vowels with stimuli read by a different male speaker.

The task was divided into two blocks. Each block contained all 30 sentences once, half with their continua presented in the original order and half with their continua presented in reversed order. This meant that the synthesised version of the speaker's vowel, as well as highly centralised or peripheral variants, would not always be in the same position, discouraging any response patterning. Sentence order was randomised within each block with each sentence heard once per block and twice overall. The perception task was presented on a Sony Viao laptop using E-Prime 2.0 [15]. All participants used Sennheiser HD 461i closed over-ear headphones and could adjust volume to a comfortable level.

Following the perception task, participants completed two questionnaires modelled on those used in [4]. The first concerned the participant's impressions of the speaker's age, nationality, occupation and education level. The second was designed to assess participant's level of exposure to New Zealand and NZE.

3. Results

Table 2 shows mean selection scores for KIT, DRESS and TRAP vowels in the three conditions, as well as the mean score across all conditions. On average, participants selected tokens that represent more phonetically raised variants of KIT, DRESS and TRAP than the vowel produced by the speaker. This explains why the mean token selection for KIT is so different to the mean selections for both DRESS and TRAP. 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 2 shows the distribution of responses for KIT vowels in the three conditions.

Table 2. *Mean token selection.*

	Aus	NZ	Control	All
KIT	5.21	5.23	5.02	5.16
DRESS	3.13	3.10	3.06	3.10
TRAP	3.40	3.55	3.38	3.45

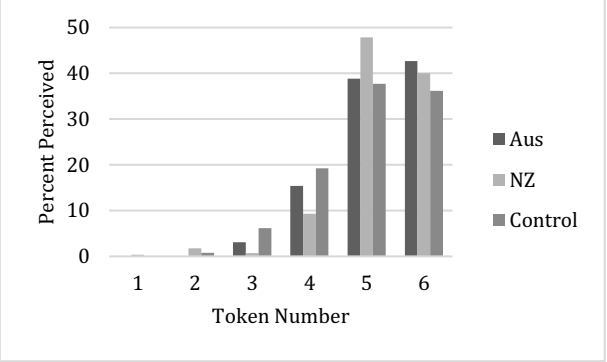

Figure 2. *Token selections for KIT vowels. 1 is NZE-like and 6 is exaggerated AusE.*

We fit a linear mixed effects model [16] to the data using the lme4 library in R [17], [18] with random intercepts for speaker and word. Fixed effects included experimental condition (Aus, NZ and control) and vowel class (KIT, DRESS and TRAP). Exposure to NZE (no frequent exposure, frequent exposure) was originally included in the model, however, it was not found to be a predictor of participants' responses, therefore it was removed. Significance was calculated using Satterthwaite's [19] approximations for the degrees of freedom using the lmerTest library [20]. As shown in Table 3, no significant difference was found between responses in the control condition and either primed condition. The model was also run with Australia as the default level of condition. This showed that the Australia and New Zealand conditions were also not significantly different from each other ($p=0.6$). Contrary to findings in [1], the presence of the stuffed toy koalas or kiwis was not shown to have a significant influence on responses in the perception task. Table 3 also shows that mean selections for KIT vowels differed significantly from TRAP. Selections also varied significantly between KIT and DRESS ($p<0.001$).

Table 3. *Fixed effects with control as default condition and TRAP as default vowel*

	Estimate	Std. error	df	t-Value	p-Value
(Intercept)	3.19	0.10	55.6	31.52	< 0.001
Condition = NZ	0.14	0.09	37	1.60	0.117
Condition = Aus	0.10	0.09	37	1.09	0.281
Vowel = DRESS	-0.35	0.11	27	-3.10	0.004
Vowel = KIT	1.71	0.11	27	15.23	< 0.001

Although the model did not show significant differences in token selection between conditions, participants did respond differently in a free-choice task regarding the speaker's nationality. Participants were less likely to identify the speaker as 'Australian' in the New Zealand condition (71%) than in the Australian or control conditions (92%).

4. Discussion

The priming effect shown in [1] was not replicated in an Australian context. Participants exposed to stuffed toy kiwis did not perform differently in the matching task when compared with those exposed to toy koalas or those in the control condition. If the effect relies on a strong association between the toy, its elicited dialect, and relevant sociophonetic variants, this result may simply suggest that the toy kiwis were not culturally significant enough to activate 'New Zealand' for the Australian participants. In other words, there is no way of knowing whether those participants in the New Zealand condition recognised the toys as kiwis and, in turn, associated them with New Zealand.

Indeed, questionnaire responses indicate that 11 participants had never been to New Zealand, didn't speak with, or know, any New Zealanders and couldn't name any New Zealand media. While these individuals may be generally aware of NZE and how it differs from AusE, NZE is a foreign dialect in Australia so phonetic sensitivity should not be assumed. If this is the case, then the lack of a result supporting [1] is not surprising. It would have been useful to have participants in the New Zealand condition tested on their ability to identify AusE and NZE in a post-experiment task. If participants completing the task could be shown to identify a NZE speaker with a level of accuracy similar to that reported in [9] yet still not show a priming effect, this would strengthen the null result.

We predicted that, in the New Zealand condition, participants with some level of exposure to NZE and New Zealanders would select more NZE-like vowels, showing influence of the prime. More than half (n = 26) of participants reported speaking with New Zealanders on a regular basis or having travelled to New Zealand. Six participants had a parent, partner or close friends from New Zealand. However, responses in the matching task from these participants did not show the predicted effect suggesting that our null result cannot be attributed to a lack of exposure alone. Either the influence of social information is more limited than suggested in [1], [3], [4] or the influence exists but is highly contextually specific. Further, it may be that even with frequent exposure to NZE, an individual may not have sufficient phonetic sensitivity to complete the matching task as predicted.

The lack of a priming effect in the present study does not necessarily contradict Hay et al.'s [1], [4] argument for exemplar-based speech categorisation. It may be that, even for those Australians with frequent exposure to NZE, NZE exemplars or categories are not activated enough, if at all, by the kiwi to compete with the resting activation level of AusE exemplars. A more overt priming condition might be required to produce a shift in Australians' responses towards more NZE-like variants. A possible priming effect from the kiwis was observed as it reduced the proportion of participants who recognised the speaker as Australian. However, responses did not correlate with the prime, as only one participant answered 'New Zealand' when identifying the speaker's nationality.

5. Conclusion

The stuffed toy priming effect observed in [1] was not replicated in an Australian context, even for those participants who had frequent contact with NZE. Further experimentation in this paradigm may be able to establish whether the results reported in [1], [3], [4] are limited to highly contextually specific situations or whether the participants tested in the present experiment and [5] simply did not have sufficient exposure to the primed dialect.

6. References

[1] J. Hay and K. Drager, "Stuffed toys and speech perception," *Linguistics*, vol. 48, no. 4, pp. 865–892, Jul. 2010.

[2] J. C. Wells, *Accents of English*. Vol. 1. Cambridge University Press, 1982.

[3] N. Niedzielski, "The Effect of Social Information on the Perception of Sociolinguistic Variables," *Journal of Language and Social Psychology*, vol. 18, no. 1, pp. 62–85, Mar. 1999.

[4] J. Hay, A. Nolan, and K. Drager, "From fush to feesh: Exemplar priming in speech perception," *Linguistic Review*, vol. 23, no. 3, pp. 351–379, Sep. 2006.

[5] D. Lawrence, "Limited evidence for social priming in the perception of the bath and strut vowels," in *International Congress of Phonetic Sciences*, Glasgow, 2015.

[6] L. Bauer, P. Warren, D. Bardsley, M. Kennedy, and G. Major, "New Zealand English," *Journal of the International Phonetic Association*, vol. 37, no. 1, pp. 97–102, Apr. 2007.

[7] J. Harrington, F. Cox, and Z. Evans, "An acoustic phonetic study of broad, general, and cultivated Australian English vowels," *Australian Journal of Linguistics*, vol. 17, no. 2, pp. 155–184, Sep. 1997.

[8] C. I. Watson, J. Harrington, and Z. Evans, "An acoustic comparison between New Zealand and Australian English vowels," *Australian Journal of Linguistics*, vol. 18, no. 2, pp. 185–207, Oct. 1998.

[9] I. Ludwig, "Identification of New Zealand English and Australian English based on stereotypical accent markers," M.A. thesis, Univ. of Canterbury, New Zealand, 2007.

[10] J. Pierrehumbert, "Exemplar dynamics: Word frequency, lenition and contrast. Frequency and the emergence of linguistic structure," in *Frequency effects and the emergence of linguistic structure*, J. Bybee and P. Hopper, Eds. Amsterdam: John Benjamins, 2001, pp. 137-57.

[11] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer* (Version 5.4) [Computer program], 2009.

[12] E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America*, vol. 33, no. 2, pp. 248–248, Feb. 1961.

[13] A. Easton and L. Bauer, "An Acoustic Study of the Vowels of New Zealand English," *Australian Journal of Linguistics*, vol. 20, no. 2, pp. 93–117, Oct. 2000.

[14] F. Cox, "Formant Frequencies and Durations for /hVd/ vowels from AusTalk," unpublished.

[15] Psychology Software Tools, Inc, *E-Prime 2.0* (Version 2.0) [Computer program], 2012.

[16] H. Baayen, *Analyzing Linguistic Data; A Practical Introduction to Statistics using R*. Cambridge University Press, 2008.

[17] D. Bates, M. Maechler, and B. Bolker, *Linear mixed-effects models using S4 classes* (R-Version 0.999375–41), 2011.

[18] R Core Team, *R: A language and environment for statistical computing* (Version 3.3.1) [Computer program], 2016.

[19] F. E. Satterthwaite, "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, vol. 2, no. 6, pp. 110–114, 1946.

[20] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, *lmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)* (R-Version: 1.1-0), 2013.