

Perceiving and adapting to regional accent differences among vowel subsystems

Catherine T. Best¹, Jason A. Shaw¹, Karen E. Mulak¹, Gerard Docherty², Bronwen G. Evans³, Paul Foulkes⁴, Jennifer Hay⁵, Jalal Al-Tamimi⁶, Katharine Mair³, Sophie Wood⁴

¹University of Western Sydney, ²Griffith University, ³University College London, ⁴University of York, ⁵University of Canterbury, ⁶Newcastle University

c.best@uws.edu.au, j.shaw@uws.edu.au, k.mulak@uws.edu.au, gerry.docherty@griffith.edu.au, bronwen.evans@ucl.ac.uk, paul.foulkes@york.ac.uk, jen.hay@canterbury.ac.nz, jalal.al-tamimi@newcastle.ac.uk, k.mair@ucl.ac.uk, slwood1191@gmail.com

ABSTRACT

Listeners perceptually assimilate nonnative phones to native phonemes, but do they do so with unfamiliar accents of their own language? We assessed Australian (Aus) listeners' assimilation of vowels in two unfamiliar English accents, and whether prior short-term exposure to the other accent would yield any adaptation to its vowels. Participants categorized the vowels of nonce words spoken in Aus or one of the unfamiliar accents (London [Lon]; Yorkshire [Yks]), after first hearing a story told round-robin by multiple speakers of Aus or of the other accent. Here we address six vowels with differing assimilation predictions for Lon vs. Yks. Results indicate that perceptual assimilation does contribute to perception of vowel variation, and that brief exposure to an unfamiliar accent can yield some degree of adaptation to its vowels.

Keywords: vowel perception, English accents, perceptual assimilation, accent adaptation

1. INTRODUCTION

A core theoretical issue regarding perception of spoken words is how *abstract* knowledge about the phonological composition of words [1] and *episodic* memory for specific tokens of words combine to support flexible word recognition [2-3]. Listeners familiarized to words spoken with "odd" variants of one phoneme generalize the deviant pronunciation to untrained words by the same speaker containing that phoneme [4-6]. This demonstrates that abstract phonemes can be a locus of perceptual adaptation, but not how episodic learning may contribute. Moreover, such localized variations cannot reveal how episodic and abstract processes combine to support adaptation to more systemic variations, as found in regional accents.

Accent variation initially decreases speech comprehension in noise [7-8], though listeners have been shown to rapidly adapt to such variation [9-10]. However, these studies typically measure

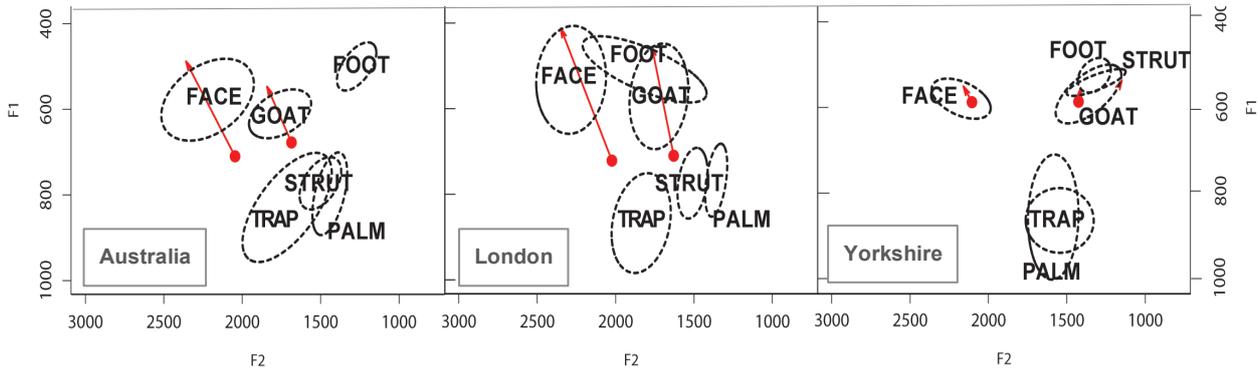
only intelligibility, not phoneme perception. Pinpointing how abstract and episodic processes contribute to spoken word recognition will rest, in part, on how listeners categorize systematic phonetic variations in native phonemes, such as familiar native-accented vowel variations vs. the less familiar variations in other accents [11-12], and on how perception is affected by short-term pre-exposure to unfamiliar accent variation.

We exploited natural vowel variation across English regional accents to test Australian (Aus) listeners' perceptual assimilation of vowels from two UK accents that are fairly unfamiliar to them: Yorkshire (Yks) and working class London (Lon) [see also 13-14]. We focus on six vowels for which Yks and Lon show different variations from the same Aus lexical sets [15]. Yks FACE and GOAT tend to be monophthongal [e: o:] [16-17], but are diphthongal in Aus [æɪ, əʊ] [18-19] and Lon ([aɪ~æɪ, əʊ]) [15-20]. Yks lacks the FOOT-STRUT split found in Aus/Lon: Yks STRUT uses the FOOT vowel [ʊ], while Aus/Lon STRUT [ɐ] is lower in vowel space and contrasts with FOOT [15-20]. Lon FOOT, however, has shifted frontward ([ɥ]) in younger speakers [20], but Aus FOOT ([ʊ]) has not. TRAP is reported to have lowered to [æ] in young Aus and Lon speakers [18-20], but to be [a] (lower, less front) in Yks. PALM is [ɛ:] in Aus/Lon, [a:] in Yks (i.e., same location as Yks TRAP) [13-17].

We estimated the properties of these vowels in our target stimuli in each accent, to compare to the published descriptions, with Lobanov-normalized [19] F1/F2 plots of our speakers' productions of 3-5 tokens of keywords with these vowels (2.2.3).

Figure 1 confirms the descriptions of the six target vowels in the literature; here we note some finer additional details. TRAP and STRUT have similar locations in normalized vowel space for Aus and Lon, but tighter ellipses in Lon where TRAP-STRUT are separated, while they overlap nearly completely in Aus. As expected, Yks STRUT is higher than in Aus/Lon and is subsumed within

Figure 1. *F1/F2 ellipses for the 50% point in the six keyword vowels produced by the target speakers of each accent. Red arrows for FACE and GOAT show the F1/F2 trajectories from 20→80% of the vowel.*



Yks FOOT, both very similar to Aus FOOT. Conversely, Lon FOOT is more fronted than Aus/Yks, consistent with [16]. Yks TRAP is farther back than Aus/Lon, and is largely encapsulated by PALM in Yks only. For FACE, while the 50% F1/F2 position is similar across the accents, and it is a similarly rising diphthong in Aus/Lon, it is nearly monophthongal in Yks as is well-known. GOAT is also diphthongal in Aus/Lon, but Lon GOAT differs from Aus in its larger 20-80% formant trajectory and less fronted offglide. The Yks monophthongal GOAT is much more retracted than in Aus/Lon, such that it overlaps Yks/Aus FOOT. PALM appears the most similar of these vowels across the accents, though not identical: it is slightly more fronted in Aus than Lon, and it is less compact, lower and slightly more fronted in Yks than Aus/Lon.

To assess perceptual assimilation we adapted a task from cross-language speech perception studies [20-21]. Listeners categorized audio nonce words spoken in one of the accents with reference to a set of printed keywords. Nonce targets restrict perceptual effects to the phoneme level (sub-lexical). To probe accent exposure effects on vowel assimilations, listeners first heard a passage in one of the accents.

Given the above, we predict lower categorization accuracy (assimilation) for Lon FOOT, GOAT and for Yks TRAP, STRUT, FACE, GOAT, relative to the Aus vowels. Higher accuracy on those vowels after hearing the same accent passage than after the Aus passage would indicate accent adaptation.

2. METHOD

2.1. Listeners

Monolingual Aus university students from greater western Sydney participated ($n = 80$; range = 17.8-42.8 yrs, $M = 22$; $SD = 4.8$). None had more than

minimal exposure to Lon/Yks or any other accents.

2.2. Stimuli

2.2.1. Nonce words. For the target nonce words, 20 English vowels (all monophthongs, diphthongs, and vowels before orthographic <R>, e.g., NORTH), were inserted into the frame /zVbə/, which yields no real words. Note that all three accents are non-rhotic, i.e., lack post-vocalic /r/. Twelve speakers each (6 male) from western Sydney (17.0-26.4 yr, $M = 22$, $SD = 3.9$), southeast/east/north London (20.2-50.6 yr; $M = 38$, $SD = 14.3$), and Sheffield/Leeds, Yorkshire (19.5-31.7 yr; $M = 24$, $SD = 5.4$), produced each nonce word six times. Two females and two males of each accent were chosen, and two tokens per nonce word per speaker were selected, judged as representative of that accent by a phonetically trained researcher familiar with the accent. Tokens were extracted with 100 ms onset and offset buffers; a ramp and damp were imposed on the initial and final 20 ms. Tokens were normalized to 65 dB.

2.2.2. Exposure passage. We modified the children's story "Chicken Little" for adult listeners, using at least ten occurrences of each vowel in stressed syllables. We recorded all speakers producing the passage. We chose four speakers (2 male) per accent, different to the nonce speakers, to create round-robin versions of the story. Three non-adjacent subsections of the story were selected per speaker, and a complete 12-section story was concatenated from them for each accent. A 1.5 s fade out/in was added between subsections (akin to the speakers' natural pauses between subsections of the story). Each passage was scaled to 65dB.

2.2.3. Keyword choices. For each target vowel, a monosyllabic word was chosen to serve as one of

the keywords presented onscreen in an array of response options for the assimilation task. Most had the form /bVd/ or /pVd/, unless that gave obscure, ambiguous, or no words (e.g., English has no such FOOT WORDS, so we used <hood>). Speakers recorded 3-5 tokens of each keyword (for Fig. 1).

2.3. Procedure

Participants first heard the Chicken Little story in one accent (*Passage*: A [Aus], L [Lon] or Y [Yks]) and answered questions about its content to ensure they had paid attention. They then categorized the nonce word vowels in one accent (*Test*: A, L or Y). Each of five Passage + Test conditions (A-A, A-L, A-Y, L-L, Y-Y) had 16 participants.

On each test trial, participants heard a single nonce token and saw the vowel keyword grid displayed on a computer monitor. They clicked on the keyword with the vowel that best matched the stressed vowel in the nonce token, then rated how well the nonce token represented the chosen vowel (1 [poor] to 7 [excellent]). Keyword order on the grid was randomized across participants, but was kept constant for a given participant. To familiarize them to the task and grid, they completed 20 randomized training trials with nonce tokens from the Aus story speakers, one per nonce word. They then completed the categorization test of 160 trials (20 nonce words x 2 tokens x 4 speakers), presented in random order via E-Prime (v. 2.0.8.22).

3. RESULTS

3.1. Cross-accent vowel categorization accuracy

We evaluated perceptual assimilation of Lon and Yks vowels by comparing correct categorizations (accuracy) of the vowels in Lon/Yks versus Aus nonce words after the Aus story (A-A, A-L, A-Y). The baseline for interpreting Lon/Yks assimilations is accuracy on the Aus vowels (A-A), which was above chance (1/20 or 5%) but well below ceiling for each vowel ($M\%$: PALM 63.3; TRAP 59.4; FACE 56.3; GOAT 44.4; FOOT 42.2; STRUT 33.6).

Generalized linear mixed modeling (GLMM) in R (version 3.0.3), *glmer* function (binomial family), was used to fit a series of binomial mixed models to the accuracy data, for separate analyses comparing A-A/A-L and A-A/A-Y. In each, the fixed factors were nonce word accent (Aus and either Lon or Yks), vowel (PALM, TRAP, FOOT, STRUT, FACE, GOAT), and the interaction. Random intercepts were included for subjects and items.

The intercept was set to the listeners' accent (Aus) and the PALM vowel, as it appears to be the vowel that is most similar across the accents (see Fig. 1). Results are shown in Table 1: β coefficients reflect performance relative to Aus for Lon/Yks, and relative to PALM for vowel (s.e. = standard error). Interactions reflect accent effects on accuracy for each vowel. Negative z indicates lower accuracy on Lon/Yks relative to Aus PALM. Pr is the probability that β deviates from zero. **Bold** = significant effect. **Bold italics** = marginal effect.

Table 1. Accuracy: Vowels in Lon and Yks re: Aus

Fixed factors	Lon: A-A vs A-L				Yks: A-A vs A-Y			
	β	s.e.	z	Pr \leq	β	s.e.	z	Pr \leq
<i>intercept</i>	0.67	0.36	1.91	.06	0.67	0.34	1.98	.05
Nonce acc	0.91	0.51	1.77	.08	0.52	0.49	1.07	.29
TRAP	-0.21	0.33	-0.63	.53	-0.21	0.35	-0.59	.56
FACE	-0.38	0.33	-1.14	.25	-0.38	0.35	-1.07	.28
GOAT	0.18	0.34	0.52	.60	0.17	0.36	0.49	.62
FOOT	-1.10	0.33	-3.31	.001	-1.10	0.35	-3.11	.002
STRUT	-1.56	0.34	-4.61	.001	-1.55	0.36	-4.34	.001
TRAP*acc	-0.74	0.48	-1.54	.12	-1.42	0.50	-2.85	.004
FACE*acc	-0.45	0.48	-0.94	.35	-1.77	0.50	-3.52	.001
GOAT*acc	-1.52	0.48	-3.15	.002	-3.65	0.55	-6.62	.001
FOOT*acc	-3.49	0.58	-6.04	.001	-0.64	0.50	-1.28	.20
STRUT*acc	-0.48	0.59	-0.99	0.32	-1.88	0.55	-3.45	.001

The FOOT and STRUT vowels were categorized less accurately than PALM (intercept) across the three accents. Indeed, they showed the lowest accuracies of the Aus items, and were expected to have even lower accuracy in Lon and/or Yks.

The vowel*accent interactions reveal how the Lon/Yks vowels were categorized in relation to the Aus vowels. As expected, accuracy was significantly lower for Lon than Aus FOOT, but not so for Yks FOOT. Conversely, also as predicted, Yks STRUT, TRAP and FACE, but not their Lon counterparts, were less accurately categorized than in Aus. The one commonality between Yks and Lon is that GOAT was less accurate than Aus, also as expected.

3.2. Cross-accent perceptual adaptation

Accent adaptation was modeled by comparing accuracy on the Lon and Yks vowels following the Lon or Yks passages, as compared to the Aus passage, in two analyses. Fixed effects were passage accent and vowel, similar to 3.1, but for conditions A-L/L-L, and A-Y/Y-Y. We included random intercepts and slopes (subjects and items) for passage accent. Results for both models are in Table 2.

Main effects with negative β coefficients for all vowels indicate each was categorized less accurately than the reference vowel PALM, across

passage accents. Of greater interest are several vowel*accent interactions with positive β values, which indicate clear cases of perceptual adaptation to Lon FOOT and to Yks STRUT, TRAP and FACE, following short-term exposure to the accent via the passage. In all these cases, the assimilation results indicated significantly lower performance on these Lon or Yks vowels than on the corresponding Aus vowels, after the Aus passage. Interestingly, the one exception to that pattern was for GOAT, which was less accurately categorized in both Lon/Yks than in Aus after the Aus passage, but failed to show adaptation after the Lon or Yks passage.

Table 2. *Adaptation: Lon/Yks versus Aus passages*

Fixed factors	Lon: A-L vs L-L				Yks: A-Y vs Y-Y			
	β	s.e.	z	Pr <	β	s.e.	z	Pr <
<i>intercept</i>	1.61	0.38	4.23	.001	1.18	0.35	3.44	.001
Passage acc	-0.81	0.46	-1.76	.08	-0.66	0.35	-1.91	.06
TRAP	-0.97	0.40	-2.43	.015	-1.62	0.42	-3.77	.001
FACE	-0.85	0.40	-2.14	.032	-2.19	0.44	-4.97	.001
GOAT	-1.37	0.40	-3.45	.001	-3.63	0.50	-7.27	.001
FOOT	-4.68	0.52	-8.93	.001	-1.72	0.43	-4.01	.001
STRUT	-2.08	0.40	-5.16	.001	-3.41	0.48	-7.12	.001
TRAP*acc	0.61	0.43	1.40	.16	1.27	0.39	3.23	.001
FACE*acc	-0.51	0.43	-1.18	.24	1.25	0.41	3.07	.002
GOAT*acc	-0.34	0.44	-0.79	.43	0.42	0.52	0.81	.42
FOOT*acc	1.80	0.58	3.10	.002	0.09	0.40	0.22	.82
STRUT*acc	-0.32	0.46	-0.71	0.48	1.49	0.46	3.24	.001

3.3. Perceptual assimilation patterns

To understand perceptual assimilation of vowels in other accents, we must consider not only accurate categorizations but also incorrect choices and goodness ratings. These richer details offer further insight into the accuracy patterns we found. They may also help explain why we found no adaptation for Lon/Yks GOAT but did find it for all of the other Lon/Yks vowels that were categorized less accurately than the Aus vowels.

Extending perceptual assimilation principles [13-14], we predicted that a vowel of an unfamiliar accent that lines up/overlaps with a different lexical set in the listener's accent will be heard as the latter vowel, rather than the one the speaker intended. This yields poor accuracy and systematic incorrect choice(s), which we call Category-Shifting (CS) assimilation. CS assimilation causes word misperceptions, as in a joke about a British general who heard an Australian soldier's "I came today" as "I came to die." In contrast, other-accent vowels that overlap, but imperfectly, with the same lexical set in Aus will be categorized correctly though with lower accuracy and/or ratings than in Aus, a

Category Goodness (CG) difference in assimilation. If accent adaptation occurs, one question is whether it is more likely for CS or CG assimilations.

The array of vowel categorization choices in the Aus passage conditions are in line with published descriptions of our accents, and with our speakers' keyword properties (Fig. 1). For Aus, Lon and Yks PALM vowels the top choice was correct, consistent with our view that it is the most similar vowel across the accents; across accents, TRAP was a less-chosen incorrect answer. The top choice for TRAP was also correct, with incorrect choices PALM and STRUT, across accents. However, TRAP accuracy was much lower and incorrect choices higher for Yks than Aus/Lon, i.e., CG assimilation. For Yks STRUT, the top choice was FOOT as expected, i.e., CS assimilation (also chosen: LOT, GOOSE, STRUT). Lon FOOT showed CS assimilation, split among STRUT, DRESS and KIT. Aus/Lon FACE was correctly identified, but Yks FACE was correct only half as often and split with SQUARE, i.e., CG assimilation. The top choice for both Aus and Lon GOAT was correct, but was chosen less often for Lon, suggesting CG assimilation (also chosen: LOT). In contrast, the top choice for Yks GOAT was THOUGHT, i.e., CS assimilation. As for adaptation, it was found for both CG (Yks TRAP, FACE) and CS assimilations (Yks STRUT, Lon FOOT), yet also failed for both types in the case of GOAT (Yks: CS, Lon: CG).

4. DISCUSSION and CONCLUSIONS

For Lon and Yks vowels in the Aus passage conditions, categorization accuracy and keyword selections including incorrect choices were largely compatible with the characteristics of vowels in these accents, and consistent with our predictions, specifically for CS and CG assimilations. Results thus support the idea that perceptual assimilation operates in the native language, guiding listeners' perception of less-experienced vowel variation, such as found in unfamiliar accents. Exposure to a story in an unfamiliar accent, moreover, can yield perceptual adaptation to many, though not all, of its vowels that are categorized less accurately than the same native vowel. But adaptation and lack of adaptation do not map neatly to CS vs. CG cases.

Further study will be needed to discover how assimilation and adaptation are involved in perceiving spoken words. We argue that such work will be important to learning how both episodic and abstract processes contribute to word recognition.

5. ACKNOWLEDGEMENTS

This research was funded by Australian Research Council (ARC) grant DP120104596.

6. REFERENCES

- [1] Cutler, A. 2010 Abstraction-based efficiency in the lexicon. *Lab. Phon.* **1**, 301-318.
- [2] Goldinger, S. 1998 Echoes of echoes? An episodic theory of lexical access. *Psy. Rev.* **105**, 251-279.
- [3] Goldinger, S. 2007 A complementary-systems approach to abstract and episodic speech perception. *Proc 16th ICPHS Saarbrücken* **16**, 49-54.
- [4] Maye, J., Aslin, R. N., Tanenhaus, M. K. 2008 The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cog Sci* **32**, 543-562.
- [5] Dahan, D., Drucker, S. J., Scarborough, R. A. 2008 Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* **108**, 710-718.
- [6] Samuel, A.G., Kraljic, T. 2009 Perceptual learning for speech. *Atten, Perc & Psychophys* **71**, 1207-1218.
- [7] Clopper, C. G., & Bradlow, A. 2006 Effects of dialect variation on speech intelligibility in noise. *J Acous Soc Am* **119**, 3424.
- [8] Floccia, C., Goslin, J., Girard, F., Konopczynski, G. 2006 Does a regional accent perturb speech processing? *J Exp Psy: Human Perc & Perf* **32**, 1276-1293.
- [9] Bradlow, A. R., Bent, T. 2008 Perceptual adaptation to non-native speech. *Cognition* **106**, 707-729.
- [10] Clarke, C. M., Garrett, M. 2004 Rapid adaptation to foreign accented speech. *J Acous Soc Am* **116**, 3647-3658.
- [11] Evans, B. G., Iverson, P. 2004 Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *J Acous Soc Am* **115**, 352-361.
- [12] Cutler, A., Smits, R., Cooper, N. 2005 Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Commun* **47**, 32-42.
- [13] Wells, J. C. 1982 *Accents of English (Vols 1-3)*. Cambridge: Cambridge University Press.
- [14] Stoddart, J., Upton, C., Widdowson, J. 1999 Sheffield dialect in the 1990s: Revisiting the concept of NORMs. In: P. Foulkes & G. Docherty (Eds) *Urban Voices: Accent Studies in the British Isles*. London: Arnold.
- [15] Khattab, G. 1999 A sociophonetic study of English-Arabic bilingual children. *Leeds Working Papers in Linguistics and Phonetics* **7**, 79-94.
- [16] Cox, F., Palethorpe, S. 2007 Australian English. *Int J Phon Assoc* **37**, 341-350.
- [17] Cox, F. 2008 The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Aus J Ling* **26**, 147-179.
- [18] Tollfree, L. 1999 Southeast London English. In: P. Foulkes & G. Docherty (Eds) *Urban Voices: Accent Studies in the British Isles*. London: Arnold.
- [19] Lobanov, B. M. 1971 Classification of Russian vowels by different speakers. *J Acous Soc Am* **49**, 606-607.
- [20] Best, C. T. 1995 A direct realist perspective on cross-language speech perception. In: W. Strange (ed), *Speech Perception and Linguistic Experience*. Timonium MD: York Press 167-200.
- [21] Best, C. T., Tyler, M. D. 2007 Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, M., Bohn, O.-S. (eds) *Second Language Speech Learning* Amsterdam: Benjamins, 13-34.
- [22] Wright, S., Shaw, J. A., Best, C. T., Docherty, G., Evans, B., Foulkes, P., Hay, J., Mulak, K. 2014 Improvements to vowel categorization in non-native regional accents resulting from single- versus multiple-talker training: A computational approach. *Australasian Speech Science and Technology (SST)*, Christchurch NZ, 3-6 December.
- [23] Shaw, J., Best, C. T., Mulak, K., Docherty, G., Evans, B., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., Peek, M., Wood, S. 2014 Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants. *Proc Australasian Speech Sci & Tech (SST)*, Christchurch, 3-6 December.