15th Australasian International

SPEECH SCIENCE & TECHNOLOGY

CONFERENCE // SST2014 CHRISTCHURCH NEW ZEALAND

Proceedings of the 15th Australasian International Conference on Speech Science and Technology

2-5 December 2014 // Rydges Latimer Hotel

Christchurch, NEW ZEALAND

Editors: Jennifer Hay and Emma Parnell

ASSTA

The Australasian Speech Science & Technology Association Incorporated

ISSN 1039-0227

www.nzilbb.canterbury.ac.nz/SST.shtml

Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants

Jason A. Shaw[1], Catherine T. Best[1], Karen E. Mulak[1], Gerard Docherty[2], Bronwen G. Evans[3], Paul Foulkes[4], Jennifer Hay[5], Jalal Al-Tamimi[6], Katharine Mair[3], Mike Peek[5], Sophie Wood[4]

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

J.Shaw@uws.edu.au, C.Best@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, mikepeek28@gmail.com, slwood1191@gmail.com

Abstract

We evaluated how Australian listeners perceive consonants spoken in two unfamiliar accents of English (Cockney, Yorkshire) and how consonant perception is influenced by short-term exposure to those accents. Results indicate that Australians misperceive some consonants from these accents and that short-term pre-exposure to them actually leads to further degraded performance in consonant categorization for these unfamiliar accents (relative to native Australian). These results rule out an account of perceptual adaptation in terms of the perceptual remapping of one consonant to another.

Index Terms: regional accents, consonant categorization, perceptual assimilation, multi-talker passage adaptation

1. Introduction

A central theoretical debate in psycholinguistics has been how *abstract* knowledge about the phonological composition of words [1] and *episodic* memories of specific utterances of words (exemplars) interact to support flexible spoken word recognition [2-3]. When phonetic variation is localized to a specific consonant of familiarization-phase words, listeners immediately generalize the "odd" variant to untrained words containing that consonant in subsequent word recognition tasks [4]-[6]. These studies provide evidence that sub-lexical units (phonemes) can be a locus for perceptual adaptation.

Real world analogs of such localized phonetic variation can be found in consonant variation across regional accents of English. For example, "Cockney" English has certain consonantal features which may cause perceptual confusions for listeners of another accent who are unfamiliar with Cockney [*cf* 7, 8]. These include *th-fronting* (/θ,ð/→[f,v], e.g.,<thing, wreathe> realized as [fɪŋ, riv]), *h-dropping* (/h/→[∅], e.g., <hand> realized as [ænd]), *t-glottalization* (t/→[ʔ], e.g., <litter> as [lɪʔə]), and *r-labialization* (/r/→ [ʋ], e.g., <rich> as [ʋɪtʃ]). Due to *th-fronting*, an Australian English (AusE) listener may assimilate the initial consonant in words such as <thick> produced in Cockney to their own /f/ category, giving rise to the non-word percept */fɪk/ instead of /θɪk/. We asked, firstly, whether listeners' experience such perceptual assimilations when listening to unfamiliar English accents that show these types of phonetic variations from their native accent. Secondly, we tested whether, as would be expected, consonant assimilations vary across unfamiliar accents in which the consonant pronunciations differ from each other and from the listeners' native accent (here, AusE). Thirdly, we asked whether prior exposure to a meaningful story told in the unfamiliar accent would lead to perceptual adaptation at the phonemic level.

In addition to Cockney-accented English, we assessed the same listeners' categorization of consonants spoken in a Yorkshire accent. Though some consonantal processes reported for Cockney can also surface variably in this accent, its consonant realizations are more often similar to AusE. Both accents are fairly unfamiliar to AusE listeners.

Several reports indicate listeners can adapt perceptually to unfamiliar accents and can do so quickly, improving their speed and accuracy of comprehension [9, 10]. But these reports did not address the locus of the adaptation, particularly the contribution of abstract consonantal units. We isolate effects at this level by using a task that requires abstract phonological judgments. Well-established in the L2 perceptual assimilation literature [11, 12], the task requires listeners to categorize nonce words by selecting a reference word starting with the same consonant. For example, listeners hear a nonce word /foɪb/ and select from a grid of 18 printed choice words, e.g., <**f**ive, **h**igh, **th**igh, **t**ie, **v**ibe …>, the corresponding consonant word <**f**ive>. Owing to *th-fronting*, however, Australian listeners might also choose <**f**ive> for nonce word /θoɪb/ spoken in a Cockney accent. We examined whether pre-exposure to a story presented in the unfamiliar accent would lead to improved performance on this abstract task.

2. Method

2.1. Participants

Sixty-four undergraduates at the University of Western Sydney, aged 17.8-42.8 years ($M = 21.8$; $SD = 5.3$), participated for course credit. All were native monolingual AusE speakers without hearing/language problems and no regular exposure to other languages or accents. All were raised in monolingual AusE homes in Greater Western Sydney.

2.2. Stimuli

2.2.1. Nonce words

Seventeen consonants were recorded in English nonce word contexts. Target consonants occurred in initial (/**C**ɔɪb/), medial (/ɔɪ**C**ə/), and/or final (/zɔɪ**C**/) position.

The nonce targets were produced multiple times by two female and two male speakers each from Greater Western Sydney (17.0-26.4 years, $M = 21.7$, $SD = 3.9$), southeast, east and north London (20.2-50.6; $M = 37.7$, $SD = 14.3$), and Sheffield and Leeds, Yorkshire (19.5-31.7; $M = 24.3$, *SD* = 5.4). Speakers recorded six tokens of each nonce target. Two tokens of each nonce word per speaker were selected for the perceptual study, on the basis that the target consonant or vowel was produced satisfactorily as judged by a phonetically trained researcher experienced in the respective accent. Tokens were extracted with a 100 ms buffer at the beginning and end. A ramp and a damp were imposed on the initial and final 20 ms of each file, and tokens were normalized to 65 dB.

2.2.2. Exposure passage

A version of the children's story "Chicken Little" was developed that contained at least ten occurrences, in stressed syllables, of each of the nonce consonants. This experiment used the passage as produced by two female and two male speakers each of AusE (18.8-43.9 years, *M* = 31.8, *SD* = 10.3), London (20.1-41.2, *M* = 30.7, *SD* = 9.3), and Yorkshire (23.6-45.1, *M* = 30.0, *SD* = 9.3). None of the speakers were the same as those used for the nonce tokens.

The exposure passage was made by combining sections produced by each of the four speakers of an accent. For each accent, three non-adjacent subsections of the passage were chosen from each speaker, and concatenated in sequence to form a complete story. A 1.5 s fade out and fade in was added between subsections (corresponding to the natural pauses speakers left between subsections). The final passage was scaled to 65dB.

2.2.3. Choice words

For each target consonant, a real word was selected to serve as a printed category choice for the listeners in the assimilation task. For target consonants, choice words generally were in the form of /**C**aɪ/, though exceptions were made if the context did not result in a real word (e.g., <**k**ite> was used instead of *<**k**ie>).

2.3. Procedure

Participants completed one of the five conditions shown in Table 1. Each participant first completed an exposure phase; they listened to the Chicken Little passage in the accent of their condition (AusE, London or Yorkshire) and then answered five multiple choice questions about the story to ensure they had paid attention. Next, they completed the phoneme categorization task in the designated nonce token accent for their condition (AusE, London or Yorkshire).

Table 1. *Exposure passage x nonce tokens conditions.*

Condition	Exposure passage	Nonce tokens
A-A	Australian	Australian
A-L	Australian	London
L-L	London	London
A-Y	Australian	Yorkshire
Y-Y	Yorkshire	Yorkshire

On each trial, participants heard a nonce token. They then saw a grid on a computer monitor containing the consonant or vowel choice words. Participants clicked on the choice word whose highlighted consonant best matched the target consonant in the nonce token they had heard. The layout of words on the grid was randomized across participants, but the order for a given participant remained constant throughout the task. To familiarize participants with the task and their randomized choice grid, prior to the categorization task, they completed training trials with nonce tokens produced by the speakers of the AusE passage. The training comprised 17 trials and was arranged so that they received one token per grid item. After training, participants completed the test, which was blocked by consonant position: initial, medial, and final. There were two trials for each nonce token (two tokens x four speakers) for a total of 328 test trials in the consonant categorization task. Stimuli were presented using E-Prime (version 2.0.8.22), which also recorded their choice word responses.

3. Results

We first inspected the Australian listeners' assimilations of the consonants in the AusE, London and Yorkshire nonce words. Figure 1 illustrates responses to the medial targets that were associated with accent differences in assimilations. The figure shows which consonant categories were selected by Australian listeners in response to nonce words in the five conditions. Differences between A-A and A-L and between A-A and A-Y indicate that AusE listeners select different consonantal choice words depending on the accent of the nonce words. This happened for /θ/: The nonce word [oi**θ**a] was categorized as <**th**igh> in the A-A condition but as <**f**ive> in the A-L condition. To evaluate the effect of condition on consonant categorization, we fit a series of linear mixed models to the *accuracy* data. A response was coded as correct when the selected category matched the category speakers were asked to produce in the nonce words, e.g., choosing <**th**igh> for target /oi**θ**a/ and <**f**ive>for /oi**f**a/. Categorization accuracy was first assessed for consonants of each accent in the AusE passage conditions (A-A, A-L, A-Y) to provide the baseline for evaluating whether exposure to the London or Yorkshire passage (Y-Y and L-L conditions) affected subsequent consonant judgments in those accents.

3.1. Effect of accent on categorization accuracy

Generalized linear mixed modeling (GLMM) in R (version 3.0.2) was used to fit a series of binomial mixed models to the accuracy data using the *glmer* function (binomial family). The first model tested for an over-arching effect of accent on consonant categorization accuracy. The model predicted responses across consonants and positional contexts (initial, medial, final). The fixed factor was the accent of the nonce words participants categorized, and thus had three levels. Random slopes and intercepts were included for participants and items. Results are summarized in Table 2. AusE nonce words are the reference category (intercept). Negative β coefficients for London and Yorkshire indicate lower categorization accuracy than for AusE. The decrease is greater for London than York; however, the accent effect was not significant in this overall analysis, which combined across all consonants and in all environments.

Table 2. *Effects of accent across all consonants.*

Accent	β	S.E.	Z value	Pr (>|z|)
(intercept)	2.02	0.20	10.3	< 0.001
London	- 0.37	0.33	- 1.12	0.26
Yorkshire	- 0.02	0.28	- 0.08	0.94

We next added positional allophone (consonants in initial, medial, final position) to the model as a fixed factor. Allophone [$F(2,40)=29.0$] and the accent x allophone interaction [$F(2,80)=6.01$] showed large effects. To follow up the interaction we fit separate models to each positional allophone. Those that contributed strongly to the interaction are associated with the four noted consonantal variations of London and Yorkshire accents from other English accents AusE: *th-fronting*, *h-dropping*, *t-glottalization*, and *r-labialization*. In the subsequent sections, we group the allophone-specific models according to these four accent-differentiating processes. The intercept is always set to the AusE condition so the β coefficients can be interpreted as improved or degraded from the AusE baseline.

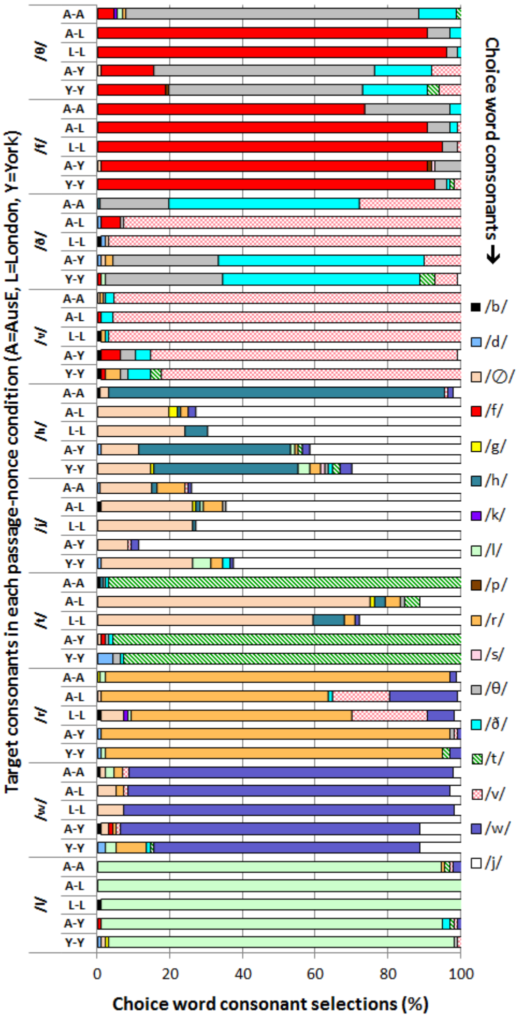

Figure 1. *Assimilation of medial consonants of interest*

3.1.1. *th-fronting*

This set included the positional allophones associated with *th-fronting*: /θ, ð/ and /f, v/ (see Table 3; in this and subsequent tables, **p* < .05, ***p* < .01, ****p* < .001, †*p* < .10). Responses to /v/ are not shown, as it was very accurately categorized across accents, was tested only in medial and final position, and did not contribute to the accent x allophone interaction. Negative beta coefficients indicate that accuracy on London and Yorkshire /θ/ was degraded in all allophone positions. The effect size, however, was much larger for London, where it was significant at the .01 level for all positions. The accent effect for /ð/ was not as reliable. Positive coefficients for Yorkshire imply *improved* accuracy over the AusE baseline, but the effect was small and did not reach significance. For London, the effect was negative but only significant for word-final position. The smaller effect for /ð/ is due in part to a lower AusE baseline. AusE listeners performed poorly on /ð/ in their native accent, frequently categorizing it as /θ/, a different error compared with Yorkshire and London tokens, where /ð/ was heard as /v/ (Figure 1). Effects of accent on /f/ were also mixed. Categorization accuracy for /f/ was degraded in initial position, particularly for Yorkshire. In medial and final positions, however, the non-significant trend was toward *improved* accuracy relative to the AusE baseline.

Table 3. *Accent effects associated with th-fronting.*

allophone	London		Yorkshire	
	β	Z value	β	Z value
initial /θ/	-6.40	-5.64***	-1.10	-1.50
medial /θ/	-4.75	-7.32***	-1.22	-1.34
final /θ/	-9.21	-2.77**	-0.71	-1.31
initial /ð/	-31.1	0.00	0.47	0.45
medial /ð/	-29.9	0.00	0.39	0.55
final /ð/	-4.54	-3.66***	0.20	0.26
initial /f/	-0.91	-0.86	-2.78	-2.66**
medial /f/	2.12	1.87	0.87	1.02
final /f/	0.26	0.26	1.03	0.84

3.1.2. *h-dropping and t-glottalization*

Accuracy on /h/ was also degraded for both London and Yorkshire nonce words, consistent with an *h-dropping* pattern (Table 4). However, the decline was significant only in medial position, where London and Yorkshire /h/ targets were often misheard as /j/ (Figure 1), which itself also showed marginally decreased accuracy (p = .06) for London. In initial position, London and Yorkshire /h/ were most often misheard as /Ø/ (no consonant onset). For both London and Yorkshire nonce items, categorization accuracy on medial /t/ was degraded (marginally so in both cases: p = .06; see Table 6). Final /t/ accuracy was significantly decreased for only the London accent; it was slightly *improved* for the Yorkshire tokens.

Table 4. *Accent effects for h-dropping and t-glottalization*

allophone	London		Yorkshire	
	β	Z value	β	Z value
initial /h/	-32.9	0.00	-0.80	-1.11
initial /Ø/	- 3.49	-1.12	-1.71	-0.44
medial /h/	- 8.17	-6.48***	-6.54	-2.91**
medial /j/	- 2.28	-1.84†	0.19	0.13
medial /t/	-19.6	-1.82†	-5.36	-1.85†
final /t/	- 7.48	-5.37***	2.47	1.41

3.1.3. *r-labialization*

Accuracy was significantly degraded for initial /r/ in the London nonce words, but not at all for the Yorkshire items. London initial /r/ was most often misheard either as /w/, which also displayed a significant decrease in accuracy, or as /l/; these patterns are consistent with *r-labialization* (Table 5).

Table 5. *Accent effects associated with r-labialization.*

allophone	London		Yorkshire	
	β	Z value	β	Z value
initial /r/	-6.18	-2.62**	19.4	0.80
medial /r/	5.54	1.07	0.46	0.53
initial /w/	16.2	2.33*	0.03	0.03
initial /l/	74.0	0.47	125.7	0.577

3.2. Effect of pre-exposure to an unfamiliar accent

To assess accent exposure effects on consonant categorization, we compared conditions in which listeners heard the exposure passage in the AusE accent and categorized consonants in an unfamiliar accent (London or Yorkshire) against those in which they heard the exposure passage in the same unfamiliar accent as the tokens they categorized. Separate logistic regression models were fit to the London and Yorkshire accents to evaluate perceptual adaptation to each accent (Table 6). Random slopes and intercepts were included for subjects and items in both models.

Table 6. *Accent exposure effects on categorization*

Accent	β	S.E.	Z value	Pr (>|z|)
London	- 0.28	0.13	- 2.19*	< 0.05
(intercept)	1.23	0.24	5.15	< 0.001
York	- 0.12	0.31	- 0.38	0.70
(intercept)	1.92	0.20	9.62	< 0.001

Negative β coefficients for both accents indicate that exposure to the target accent actually decreased overall accuracy on consonant categorization. This effect is small for Yorkshire, but is larger and significant for the London accent.

We also fit a model that included positional allophones and the interaction between allophone and exposure accent. For both accents the interaction effects were modest (London [$F(1,40)=1.42$]; Yorkshire [$F(1,40)=2.09$]) indicating that the negative effect of exposure is largely uniform across consonants. Finally, we examined the effect of unfamiliar accent exposure on each allophone separately. None of the Yorkshire consonants showed significant effects. For London, only word-final /f/ showed a significant adaptation effect, and it was *positive* (London: β=2.63, S.E.=1.33, Z=1.97, Pr (>|z|)=< 0.05; intercept: β=1.96, S.E.=0.68, Z=2.89, Pr (>|z|)=< 0.0).

4. Discussion

Our results showed that consonant categorization is indeed degraded in unfamiliar accents. Degraded performance is not distributed evenly across consonants but instead is localized in specific allophones and positional environments, e.g., medial /θ/, initial /r/, final /t/ in London, and medial /h/ in both London and Yorkshire accents. Overall, and as expected, AusE listeners performed worse on London than Yorkshire consonants. Second, our results showed that short-term exposure to the target accent, ~10 minutes of a story told by multiple talkers, did not improve consonant categorization. On the contrary, it led largely to even worse accuracy.

Past work has shown that listeners can adapt quickly to different talkers, even to accented talkers, and that exposure to an accented talker can lead to benefits in understanding that generalize to other speakers of the accent [e.g., 13]. In these studies, though, it was not possible to identify the level or relation between levels at which adaptation occurs. The abstract categorization task that we have deployed allows us to rule out certain possibilities. For example, even after exposure from multiple talkers producing *thick* as [fɪk], our listeners do not appear to construct rules at the level of consonants, e.g., adjusting percepts so that Cockney [f] for /θ/ is *heard* as /θ/. Adaptation may require a close link to the lexicon, which was not directly tapped by our nonce phoneme categorization task.

One possible explanation for the *negative* influence of short-term exposure is that listeners adjust their expectations about consonant frequency while listening to the passage and then rely on these updated frequencies in categorization. Due to the Cockney /f/~/θ/ merger, AusE listeners presumably hear a larger number of [f]s in the London passage than in the AusE passage. Degraded performance after exposure to London could be due to adjustments of prior expectation and reliance on those priors when uncertain about a consonant. Put simply, listeners exposed to Cockney might answer /f/ when uncertain about a fricative because [f] is more frequent in that accent. As a side-effect, however, /f/ categorization should actually *improve* after exposure to London. Indeed, though improvements after unfamiliar-accent exposure were rare, /f/ in final position did show exactly this type of adaptation.

5. Conclusions

Differences in regional accents can significantly hinder consonant perception, but this effect is restricted to unfamiliar positional allophones that are confusable with other native consonants. Short-term exposure to a multi-talker story in an unfamiliar accent may further hinder, rather than improve, categorization of its consonants. These results rule out a perceptional adaptation account in terms of rules that re-map one consonant to another in the presence of positive evidence. Instead, it seems that a role for abstract categories in accent adaptation may be to track distributional statistics in the input.

6. Acknowledgements

This research was funded by ARC grant DP120104596.

7. References

[1] Cutler, A. “Abstraction-based efficiency in the lexicon”, Lab Phon, 1: 301-318, 2010.

[2] Goldinger, S. “Echoes of ehoes? An episodic theory of lexical access”, Psych Review, 105:251-279, 1998.

[3] Goldinger, S., “A complementary-systems approach to abstract and episodic speech perception”, Procs ICPhS, 16: 49-54, 2007

[4] Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. “How abstract phonemic categories are necessary for coping with speaker-related variation. LabPhon, 10:91-111, 2010.

[5] Samuel, A. G., and Kraljic, T. “Perceptual learning for speech”, Atten, Percep & Psychophysics, 71:1207-1218, 2009.

[6] Norris, D., McQueen, J. M., and Cutler, A. “Perceptual learning in speech. Cognitive Psychology, 47:204-238, 2003.

[7] Foulkes P, and Docherty, G. [Eds] “Urban Voices: Accent Studies in the British Isles”, Arnold, 1999.

[8] Docherty, G. and Foulkes, P. “Glottal variants of /t/ in the Tyneside variety of English”, In Hardcastle W, Mackenzie Beck J [Eds], A Figure of Speech, 173-199, Erlbaum, 2005.

[9] Evans, B. G. and Iverson, P. “Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences”, JASA 115: 352-361, 2004.

[10] Cutler, A., Smits, R. and Cooper, N. “Vowel perception: Effects of non-native language vs. non-native dialect”, Speech Commun, 47:32–42, 2005.

[11] Best, C. T. “A direct realist perspective on cross-language speech perception”, In W. Strange [Ed] Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research, 167-200, York Press, 1995.

[12] Best, C. T. and Tyler. M. D. “Nonnative and second-language speech perception: Commonalities and complementarities”, In M. Munro & O.-S. Bohn [Eds] Second Language Speech Learning, 13-34, Benjamins, 2007.

[13] Bradlow, A. R. and Bent, T. “Perceptual adaptation to non-native speech,” Cognition, 106:707-729, 2008.