Article:
Foulkes, Paul orcid.org/0000-0001-9481-1004 (2023) Revealing perceptual structure through input variation: cross-accent categorization of vowels in five accents of English. Laboratory Phonology. ISSN: 1868-6354

https://doi.org/10.16995/labphon.6436

Laboratory Phonology
Journal of the Association for Laboratory Phonology

Shaw, J. A., Foulkes, P., Hay, J., Evans, B. G., Docherty, G., Mulak, K. E., & Best, C. T. (2023). Revealing perceptual structure through input variation: cross-accent categorization of vowels in five accents of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 14(1), pp. 1–38. DOI: https://doi.org/10.16995/labphon.6436

Open Library of Humanities

Revealing perceptual structure through input variation: cross-accent categorization of vowels in five accents of English

Jason A. Shaw*, Dept. of Linguistics, Yale University, New Haven, CT, USA, jason.shaw@yale.edu

Paul Foulkes, Dept. of Language & Linguistic Science, University of York, Heslington, York, UK, paul.foulkes@york.ac.uk

Jennifer Hay, New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Christchurch, New Zealand, jen.hay@canterbury.ac.nz

Bronwen G. Evans, Dept. of Speech, Hearing and Phonetic Sciences, University College London, London, UK, bronwen.evans@ucl.ac.uk

Gerard Docherty, Arts, Education and Law, Griffith University, Brisbane, Australia, gerry.docherty@griffith.edu.au

Karen E. Mulak, The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, Australia, K.Mulak@westernsydney.edu.au

Catherine T. Best, The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, Australia, C.Best@westernsydney.edu.au

*****Corresponding author.

This paper characterizes the perceptual structure of vowel systems in five regional accents of English, from Australia (A), New Zealand (Z), London (L), Yorkshire (Y), and Newcastle upon Tyne (N), on the basis of "whole system" vowel categorization experiments. We established patterns of within-accent vowel confusions, and then explored cross-accent perception, assessing how listeners from one accent background categorize vowels from another. Our experimental task required mapping continuous phonetic dimensions to perceptual categories in the absence of phonotactic and lexical cues to vowel identity and socio-indexical information about the talker. Our results show that, without these sources of information, there is uncertainty in vowel categorization, even for native accent vowels, and that this degree of uncertainty increases for unfamiliar accents. The patterns of cross-accent perception largely reflect the accent-specific perceptual structure of the listener, as opposed to adaptations to the stimulus accents. This finding contrasts with the type of active talker adaptation found with tasks offering lexical information about vowel identity and indexical information about the talker.

1. Introduction

There is broad consensus that models of speech perception must encapsulate the benefits listeners gain both from phonetically detailed exemplars (i.e., episodic memory for speech via encoding of continuous phonetic dimensions), and also abstract phonological representations, reflected in so-called "hybrid" models of spoken word recognition (Cutler, Eisner, McQueen, & Norris, 2010; Goldinger, 2007; Pierrehumbert, 2016). However, identifying the perceptually relevant phonetic dimensions within a particular speech community remains a challenge. Researchers have tended to infer perceptual relevance from production patterns as opposed to perception patterns. But listeners may pay more attention to some dimensions of speech than others, or weight exemplars differently depending on the social context in which they are heard (Foulkes & Docherty, 2006; Foulkes & Hay, 2015). These factors complicate models of speech categorization that are based solely on acoustic/articulatory similarities, as we need to discover empirically the dimensions of relevance to listeners in different speech communities (e.g., Montgomery & Moore, 2018) and the factors that influence listener attention to phonetic detail.

In vowel perception, it is clear that certain acoustic phonetic properties (e.g., the first three formants, vowel duration, f0) play an important role. How formants vary as a function of supra-laryngeal vocal tract shapes is well understood (e.g., Carré & Chennoukh, 1995; Chiba & Kajiyama, 1941; Fant, 1960; Stevens, 1998; Story, 2005; Whalen, Chen, Tiede, & Nam, 2018). In vowel perception, the information about vocal tract configuration that is conveyed through formants can interact with vowel duration (Ainsworth, 1972; Gottfried & Beddor, 1988) and pitch (Hirahara & Kato, 1992). However, acoustic/auditory properties offer only a partial account of vowel perception. Social expectations also enter into how vowels and other phonological categories are perceived (Drager, 2010; Hay & Drager, 2010; Hay, Nolan, & Drager, 2006; Hay, Warren, & Drager, 2006; Hurring, Hay, Drager, Podlubny, Manhire, & Ellis, 2022; Nguyen, Shaw, Pinkus, & Best, 2016; Nguyen, Shaw, Tyler, Pinkus, & Best, 2015; Niedzielski, 1999; Strand, 1999; Walker, Szakay, & Cox, 2019). Crucially, the social interpretation of phonetic variation can be decidedly local, depending on the meaning assigned *in situ* by a particular speech community or listener (Clopper & Pisoni, 2005; Clopper & Pisoni, 2006; Eckert, 2008). Recovering the socio-indexical properties of the talker from variation may make complementary use of the same phonetic dimensions used to recognize phonological categories (Best, 2015; Docherty, Foulkes, Tillotson, & Watt, 2006; Kleinschmidt, 2019; Kleinschmidt, Weatherholtz, & Jaeger, 2018). Associations between social information and phonetic patterns, once learned, may facilitate interpretation of phonetic variation in terms of phonological categories. Accordingly, the social use of variation to perceive and convey aspects of talker identity, social identity, or speech style, may interact with mappings between phonetic properties of the speech signal and phonological categories.

In one of the first large-scale studies involving perception and acoustic measurements of vowels, Peterson and Barney (1954) showed that listener categorization of vowels, in /hVd/ frames, was near ceiling (96%) despite several vowels showing overlap in F1 and F2 space. Hillenbrand, Getty, Clark, & Wheeler (1995) replicated the study with similar results. In these studies, both of which examined vowels in rhotic American English accents, native listeners showed a high level of categorization performance, suggesting that there is more to vowel perception than similarity based on F1 and F2. Listeners may have relied on some latent representation of the talkers, and/or of the accent, effectively normalizing the vowels. Indeed, apparent effects of F1 and F2 normalization on vowel perception have been modelled using richer acoustic/auditory representations incorporating additional dimensions that are posited to contain talker-specific information, such as f0 and higher formants, e.g., F4 (Johnson, 1997). Johnson's (1997) exemplar model mimics F1 and F2 normalization by disproportionately activating exemplars of contextually relevant talkers. Selective priming of exemplars can also be achieved through situational or social priming. If listeners are primed to expect a talker of a different sex/gender (Strand, 1999), regional accent (Hay & Drager, 2010; Hurring et al., 2022; Niedzielski, 1999) or foreign accent (Nguyen et al., 2016), they have been shown to shift expectations in ways that influence how acoustic/auditory information is interpreted in terms of phonological categories. That listeners can perceive vowel categories veridically in the presence of apparently overlapping or ambiguous F1/F2 values may reflect how listeners make use of expectations about the talker to guide interpretation of the acoustic signal in terms of phonological categories. That is, acoustic-phonetic variation is not random; it is socially structured in ways that listeners are aware of and can make use of in speech perception. Nor is the relevant acoustic information limited to F1/F2 values alone.

The present paper builds on vowel perception results reported on in our past work, in which *no* information was provided about the talkers beyond the acoustic details in the tokens themselves (Shaw, Best, Docherty, Evans, Foulkes, Hay, & Mulak, 2018). Australian listeners in that study showed a rather low level of categorization accuracy on their native vowels as well as on vowels produced by speakers of less familiar accents of English: Regional accents of London, New Zealand, Yorkshire, and Newcastle upon Tyne. Notably, that study involved categorization of vowels presented in disyllabic nonsense words, thus providing no lexical information, as produced by multiple talkers (not blocked by talker), whose regional origins were not revealed nor their talker identities tagged in any way. Under these circumstances, in a 19-alternative forced choice task (all English monophthongs and diphthongs), accuracy was well below ceiling, even on Australian (native accent) vowels. Introducing natural variation from multiple talkers without providing a mechanism to factor talker characteristics into the perception process, and excluding information about vowel identity coming from word identity, made for a highly challenging task. Interestingly, though, accuracy patterns by vowel showed surprising consistency across accents.

To a greater degree than expected based on sociophonetic descriptions of the accents, Australian listener accuracy in categorizing vowels across accents resembled the pattern of accuracy on their own vowels. That is, vowel categories that achieved a relatively high categorization accuracy in the native accent (Australian) also showed high accuracy for other accents, despite acoustic-phonetic differences. Moreover, there was a stability to the accuracy pattern – which vowels were categorized more-or-less accurately – in that it persisted even after multi-talker exposure to the other accents via a pre-test story passage. Notably, the pre-test exposure to the accent came from different talkers than the test items and there was no explicit indication that the pre-test and test talkers came from the same region. In the absence of a social dimension to link these talkers, no accent-level adaptation took place (cf., Maye, Aslin, & Tanenhaus, 2008). Although the perceptual responses were not modelled in terms of expected acoustic values, it was clear from the measurements reported in the paper that vowel formants and durations offered only a partial explanation for listener behavior.

Rather, listener ability to adapt to accent variation following exposure may have been hindered by talker variation and lack of any clear social information about the talkers. In the absence of social information, vowels of other accents, although phonetically different, may assimilate in perception to the categories of the native listener in ways that are similar to perceptual assimilation in cross-language (e.g., Best, 1995) and L2 speech perception (e.g., Best & Tyler, 2007). Perception across the vowel system may be flexible to natural vowel variation, within or between regional accents, unless there is a clear cue to adjust, displaying perceptual assimilation except when clear social information is available for attribution of the variation. Given this type of perceptual flexibility, the differences across accents may not have been large enough to impact vowel perception patterns, at least as reflected in categorization accuracy alone. The relative consistency in how Australian listeners categorized vowels from other accents may reflect the perceptual structure of Australian listeners, which we see clearly when we take away the social information that allows a listener to tailor perception to the context.

However, Shaw et al. (2018) provided no baseline on how vowels in the non-Australian accents are perceived by listeners of those accents (i.e., how London/New Zealand/Yorkshire/Newcastle listeners perceive the vowels of their own accents and the Australian accent). The lack of this baseline limits how those previous results can be interpreted in terms of perceptual assimilation. In addition, Shaw et al. (2018) focused narrowly on "accuracy", defined as whether the listener correctly identified the phonemic target intended by the talker. The pattern of accuracy across vowels was similar regardless of whether the Australians were listening to their own accent or one of the four others. However, accuracy of vowel categorization is only one part of the broader pattern of perceptual confusion represented in a complete confusion matrix. The *pattern of errors* may provide additional information about

listener behavior, above and beyond accuracy, a consideration that motivates reanalysis with new methods. In particular, the apparent consistency of vowel categorization accuracy across accents might be an artifact of looking narrowly at accuracy. That is, the patterns of confusions could be different across accents even if overall accuracy remains consistent. The present study thus investigates perceptual similarity in the five accents of English that Shaw and colleagues examined. This includes analysis of data from some conditions that Shaw et al. (2018) did not report on, along with reanalysis of some of the conditions they did examine, using new methods.

Regional variation in English vowels provides an example of higher order constancy, or phonological systematicity, in the face of phonetic variation (see, e.g., Best, 2015; Best, Tyler, Gooding, Orlando, & Quann, 2009), which we aimed to tap into with our new analyses. In a survey of regional accent variation in English, Wells (1982) introduced the Lexical Set framework to express the phonological coherence of the lexicon across accents. Large parts of the English lexicon preserve the "same" phonological vowel category across groups and talkers, despite widespread variation in its phonetic realizations between regions, social groups, and individual talkers. In the Lexical Set framework, KIT, for example, refers to the stressed vowel in all words sharing the same vowel as *kit* across all accents. We adopted this framework to compare perceptual patterns across accents, referring to Wells' keywords for Lexical Sets in small caps (e.g., KIT), and orthographic representations of words in angled brackets (e.g., <kit>). There are a small number of vowel mergers and splits across accents, resulting in differences at a phonological level of description. For example, the STRUT and FOOT distinction, arising from a split, did not take place in our Northern UK varieties, Yorkshire and Newcastle; and New Zealand has merged NEAR and SQUARE, which remain distinct in the other varieties in our study. There are also some lexically specified differences, most notably the patterning of the BATH lexical set with the vowel pronunciations in either TRAP or START/PALM. However, in large part, the network of *phonological* contrasts supported by the English lexicon is overwhelmingly preserved across regional accents despite substantial *phonetic* variation across accents in vowel quality realization for a given lexical set. Although there are rich data on vowel F1/F2 acoustic patterns across accents, which take the whole vowel system into consideration, comparable data on cross-accent vowel *perception* are in short supply. The extent to which system-wide variation affects accent-specific perception is largely unknown.

Our experiment consisted of analyses of eight conditions collected but not analyzed in Shaw et al. (2018), together with new analyses of five conditions examined in that study, for a total of 13 conditions. In the within-accent conditions for the present study, listeners from five non-rhotic regional accents of English categorized vowels from 20 distinct lexical sets. These five accents have similar numbers of phonological vowel categories, but the phonetic realization of those vowels differs substantially. Since they are spoken in different regions, we also expect different

local socio-indexical evaluations of phonetic variation, which may factor into categorization behavior. The within-accent conditions allow us to evaluate the degree to which patterns of vowel confusion remain stable across accents. Additionally, they provide a baseline against which we assess two cases of cross-accent perception.

We compared our new within-accent baseline to the cross-accent conditions of Shaw et al. (2018), in which different groups of Australian listeners heard vowels produced by speakers of the same five regional accents (Australian, New Zealand, and the three UK varieties). We predicted that the Australian listeners would show different patterns of confusion from the 'native' listeners of those accents, although this has yet to be confirmed. We also ran conditions in which each listener group categorized Australian vowels. This allows us to look at how different listener groups respond to the same stimuli. We predicted group differences (i.e., that the same stimuli would be categorized differently by different listener groups, owing to accent-specific differences in the perceptual systems across groups).

Both of these cross-accent scenarios allow us to probe the degree to which variation in a listener's accent influences their perceptual behavior in a task involving ecological system-wise variation in vowels. Notably, listeners were not told which accent they were hearing nor given any reason to suspect an unfamiliar accent. For this reason, we view the cross-accent stimuli as probes to the listener perceptual system, revealing the bounds of perceptual categories. In the absence of relevant socio-indexical information about the talkers or lexical information about the vowel category, we anticipated that each stimulus would function as a probe to listener perceptual structure, which we conceptualize as categories in a multi-dimensional phonetic space (e.g., Pierrehumbert, 2003). Assuming that listeners are unable to adapt to vowels without lexical or socio-indexical information, responses in the cross-accent conditions should largely reflect the structure of listener perceptual categories. That is, if two categories are relatively difficult to distinguish in the native accent in the absence of talker information, they should be relatively difficult in other unfamiliar accents as well. Therefore, we first established such perceptual patterns in each accent independently and then compared cross-accent conditions against that baseline.

The analytical methods that we used to compare conditions take advantage of our experimental design, which elicited responses to the entire vowel space in terms of a large, and possibly exhaustive, number of response categories. Perceptual similarity is typically assessed using pairwise confusions (i.e., how often two categories are confused with each other). Often, there are asymmetries in pairwise confusability whereby one category, p_1, is confused with another, p_2, more often than p_2 is confused with p_1 (e.g., Polka & Bohn, 2003; Schwartz, Abry, Boë, Ménard, & Vallée, 2005). The rich set of response options in our paradigm, together with anticipated task difficulty, motivate a more global characterization of perceptual similarity across accents that

makes use of the complete distribution of responses. To this end, we illustrate perceptual patterns through Hierarchical Cluster Analysis (HCA) based on vector similarity across all responses to a vowel, cf., pairwise similarity. We then compare conditions by calculating a normalized Euclidean Distance between entire confusion matrices.

The remainder of the paper is organized as follows. Section 2 describes the methods for the experiment (2.1–2.3) and data analysis (2.4) in detail. Section 3 reports the results, starting with a comparison of accents (3.1), and then comparison of cross-accent conditions (3.2). Section 4 provides a general discussion and Section 5 provides a brief conclusion.

2. Method

We report a vowel categorization experiment with five different regional accent listener groups: Australia [A], New Zealand [Z], London [L], Yorkshire [Y], and Newcastle [N]. The complete set of experimental conditions for the current report is summarized in **Table 1**. As noted above, this includes five conditions from Shaw et al. (2018) reanalyzed using new analytical methods, and eight new conditions that were not evaluated in that study. In the "within accent" conditions, each listener group categorized vowels from their own accent. This provides an important baseline for reanalysis of the conditions from Shaw et al. (2018), in which Australians categorized vowels from the other four accents as well as from Australian. Lastly, in four new cross-accent conditions, Z, L, Y, and N listeners categorized Australian vowels. Thus, 13 conditions are presented.

All conditions involved listeners first hearing a meaningful exposure passage in their own accent and then completing a vowel categorization and ratings task with nonsense words (i.e., phonotactically licit but unattested/meaningless words) spoken by different speakers of either (1) their own regional accent or (2) some other regional accent. As a shorthand label for each of the conditions we use two letter combinations.[1] The first letter indicates the listener group (i.e., A (Australia), Z (New Zealand), L (London), Y (Yorkshire), N (Newcastle)), and the second letter indicates the accent used in the categorization test stimuli. Thus, L-L refers to London listeners hearing London stimuli while L-A refers to London listeners hearing Australian stimuli. The A-A condition, at the top of the table, serves as a baseline for several of our cross-accent comparisons. The other accents are listed in the table in their order of similarity to Australian – L, Z, Y, and N – as based on sociophonetic descriptions (summarized in Shaw et al., 2018).

[1] Shaw et al. (2018), which only tested Australian listeners, also used two letter combinations to represent conditions but in a different way. In that paper, the first letter represented the accent of the exposure passage and the second represented the accent of the nonsense word stimuli that were classified.

Experimental Conditions				
	Listener accent	**Stimulus accent**	**Condition code**	**# of participants**
Shaw et al. (2018) same listener group	Australian	Australian	A-A	16
	Australian	London	A-L	16
	Australian	New Zealand	A-Z	16
	Australian	Yorkshire	A-Y	16
	Australian	Newcastle	A-N	16
New data within-accent	London	London	L-L	12
	New Zealand	New Zealand	Z-Z	16
	Yorkshire	Yorkshire	Y-Y	16
	Newcastle	Newcastle	N-N	12
New data same stimulus accent	London	Australian	L-A	14
	New Zealand	Australian	Z-A	17
	Yorkshire	Australian	Y-A	16
	Newcastle	Australian	N-A	12

Table 1: Accents of listener groups and stimulus items for each condition

2.1 Participants

A total of 195 listeners participated in the study: 12–17 per condition across 13 test conditions. The number of participants in each condition is given in the rightmost column of **Table 1**. Listeners were recruited from local university communities in western Sydney (A), Christchurch (Z), London (L), Yorkshire (Y), and Newcastle (N), the same communities from which talkers (see below) were recruited. We note that while the talkers were selected because they demonstrate features characteristic of their regional accent, the criteria for listeners were less stringent. We required only that they were long-time residents, born and raised in the accent region, on the assumption that this would ensure majority exposure to the target regional accents. No participants reported speech/hearing/language problems, and all reported having minimal long-term regular exposure to other languages or other regional English accents.

2.2 Stimuli

2.2.1 Test stimuli: Nonsense words

The phonotactically permissible frame /z**V**bə/ was used to elicit nonsense words for each of the English vowels in 20 lexical sets: FLEECE, KIT, DRESS, TRAP, START, STRUT, GOOSE, FOOT, NORTH, LOT, NEAR, SQUARE, NURSE, CURE, FACE, GOAT, PRICE, CHOICE, MOUTH, and BATH. None of the resulting nonsense items in the /z**V**bə/ frame forms a real word. The very low phonotactic probability of /zVbə/ also minimized lexical biases in perception of the target items.

The targets were produced six times each by two female and two male talkers from each region. Two tokens per nonsense word per talker per accent were selected on the basis that the target vowel was judged to be representative of the accent by a phonetically trained researcher experienced with that accent. Tokens were extracted with a 100 ms buffer of inter-stimulus silence at the beginning and end of the nonsense word. An amplitude ramp and damp were imposed on the initial and final 20 ms of each file, respectively, and tokens were normalized to 65 dB.

2.2.2 Pre-test accent exposure passage

Immediately preceding the experimental vowel categorization tasks, participants were asked to listen to a meaningful passage (~10 minutes) read by four talkers of their own regional accent.[2] None of the talkers were those used for the nonsense word stimuli in the same accent. The passages were created from recordings of two male and two female talkers of each accent by selecting three non-adjacent subsections of the passage for each talker and concatenating the subsections in sequence to form the complete story. These were the same passages used for exposure in Shaw et al. (2018). They serve in the current study only to standardize the experience of listeners across conditions, cf. Shaw et al. (2018), which also manipulated whether the exposure accent matched that of the listener or that of the other-accent test stimulus nonsense words.

2.2.3 Talkers

The stimuli for the experiment, including the pre-exposure passage and nonsense words, were produced by 12 talkers (six female), recruited from each region represented in the experiment. From this larger set of recordings, eight talkers (four female) were selected for inclusion in the

[2] The notion of the "same" regional accent is never entirely unproblematic, but the case of London requires special mention. For London, both the passage talkers and the listeners were recruited from London. All talkers had a Popular London/Cockney accent characteristic of working-class London communities (cf. Wells, 1982). However, our criteria for listeners were not as stringent. All of our London listeners were highly familiar with the Popular London/Cockney accent – they'd grown up in London – but most were talkers of a more standard accent, intermediate between Popular London and Received Pronunciation, and similar to what Wells (1982) describes as "London Regional Standard" (cf. Tollfree, 1999).

experiment — four talkers (two female) were used for the exposure passage and four different talkers (two female) were used for the nonsense words. All talkers were recorded in their native accent region: London talkers in London, Newcastle and Yorkshire talkers from their respective districts in Northern England, New Zealand talkers in Christchurch, and Australian talkers in Sydney. All talkers selected for inclusion in the stimulus materials were judged to be representative of their regional accent by one or more of the authors and a research assistant familiar with the regional accent of the group. The talkers representing a regional accent were consistent for all conditions involving that accent (e.g., the Australian materials were identical in A-A, L-A, Z-A, Y-A, and N-A conditions (**Table 1**)). The age ranges of these talkers (both nonsense words and exposure passage talkers) were: Western Sydney (17.0–26.4 years, $M = 21.7$, $SD = 3.9$), Christchurch, New Zealand (18.5–20.6 years, $M = 19.6$, $SD = 1.0$), London (20.2–50.6 years; $M = 38.0$, SD = 14.3), Yorkshire (19.5–31.7 years; $M = 24$, SD = 5.4), and Newcastle (21.5–45.9 years, $M = 31.6$, $SD = 11.7$).

2.3 Procedure

After listening to the pre-test exposure story, listeners heard and judged nonsense words in either their own regional accent or a less familiar regional accent. For L, Z, Y, and N listeners, the less familiar accent was Australian. Conditions in which Australians heard one of the other four less familiar regional accents reported in Shaw et al. (2018) were reanalyzed using the techniques in this paper. Participants were not told which accent(s) they would be listening to, nor did we inform them that accents were the focus of the task. They first completed the pre-test exposure phase in which they listened to our short story passage in their local regional accent. Next, they completed the vowel categorization and goodness rating task in the accent of their condition (see **Table 1**).

On each trial of the categorization task, participants heard a nonsense token. They then saw a grid on a computer monitor containing the vowel keywords for this 19-alternative forced choice (19AFC) task (see **Figure 1**). The keywords were real words selected to serve as a printed choice for listeners to use in categorizing the nonsense word vowels in the assimilation task. There were keywords for 19 of the 20 target vowels. We did not have a separate keyword for the BATH vowel, as it systematically groups with one of the other vowels for all accents: In A/L/Z, BATH and START group together; in N/Y, BATH and TRAP group together. Keywords were presented together on a grid in the form of real /b**V**d/ words, with exceptions made if the context did not result in an easily recognizable word (e.g., we used <c**o**de> instead of <b**o**de>, <r**u**de> instead of <b**oo**ed>). The keywords were: <bead, bid, bed, bad, bard, bud, rude, hood, bored, pod, beard, paired, bird, toured, paid, code, hide, boyd, proud>, respectively, for the lexical sets FLEECE, KIT, DRESS, TRAP, START (same vowel as PALM in all five accents), STRUT, GOOSE, FOOT, NORTH (same vowel as THOUGHT in these accents), LOT, NEAR, SQUARE, NURSE, CURE, FACE, GOAT, PRICE, CHOICE, and MOUTH.

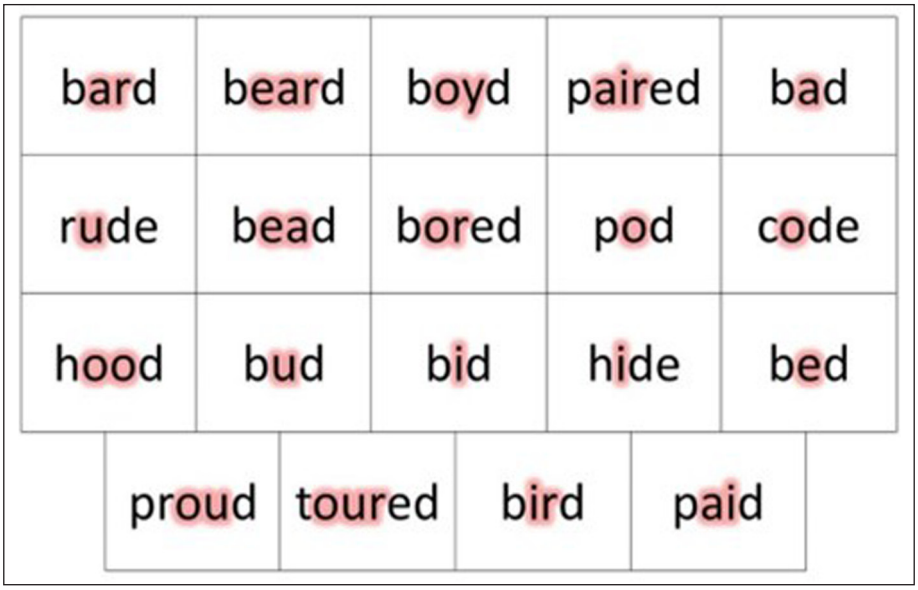


Figure 1: Example of a randomized keyword grid presented to a participant for choosing the vowel they heard in each nonsense word during the categorization and goodness rating task.

Participants clicked on the keyword whose highlighted vowel they considered the best match for the target vowel in the nonsense token they had heard. The layout of keywords on the grid was randomized across participants, but the order for a given participant remained constant throughout the task. After selecting a keyword, they were asked to rate how well the vowel they heard matched the vowel in the keyword they had chosen (following, e.g., Tyler, Best, Faber, & Levitt, 2014). Participants rated the goodness of fit on a 7-point Likert scale (7 = excellent, 1 = poor). These ratings did not enter into the current analysis (see Shaw et al., 2018 for a discussion of goodness ratings in the context of cross-accent perception). To familiarize participants with the task and their randomized grid, prior to the categorization task they completed training trials (without feedback) with nonsense tokens produced by the talkers of the exposure story they had heard, arranged so that they received one token per grid item. After training, participants completed the categorization test (160 trials = 20 nonsense words x 2 tokens x 4 talkers), presented in random order via e-Prime (v. 2.0.8.22). There were four breaks built into the categorization task. The total task took between 40–60 minutes, depending on the pace of the individual participant and the amount of time that they chose to rest during breaks. Participants were compensated with a small cash payment at the end of the experiment.

2.4 Analysis

To visualize and compare the perceptual similarity of vowels across conditions, we used hierarchical cluster analysis (HCA). All analyses were conducted in R version 3.6.2

(R Foundation for Statistical Computing) and are included as supplementary materials. Our methods for HCA follow the general description in Gries (2009, pp. 306–318) and were implemented using the *hclust* function in R. To compare across conditions, we calculated a normalized Euclidean distance between confusion matrices, analytical steps that are explained in detail in the remainder of this section.

2.4.1 Vector-based vowel similarity metric

An important consideration for HCA methods is how to compute perceptual similarity. A commonly used method in speech perception studies is to calculate perceptual similarity pairwise, from the confusability of one speech sound with the other and vice versa (e.g., Johnson, 2011; Shepard, Romney, & Nerlove, 1972). Given two speech sounds, p_1 and p_2, perceptual similarity, on this approach, is a function of how often p_1 is confused with p_2 and how often p_2 is confused with p_1, as confusion may be asymmetrical. We take a different approach, one that makes use of the complete range of participant responses to each vowel stimulus.

Instead of comparing just the confusions of p_1 and p_2 with each other, we treat the entire set of responses to each vowel as contributing to its perceptual characterization. Accordingly, we represent each vowel as an *N*-item vector of responses. Each item of the vector corresponds to a response choice, that is, the number of times that the stimuli for a particular vowel were classified as one of the choice words. Since there were 19 response options in our experiment, there are 19 items in each vowel vector. Thus, the vector representing a vowel is the total set of behavioral responses to that vowel distributed across the response space (i.e., one row of the confusion matrix).

To illustrate a vector representation of the perceptual categorization of a vowel, we include three rows from the A-A confusion matrix in (1). These show results pooled across the 16 participants in this condition. The item *zahba,* included to represent the START/PALM lexical set[3] and produced with a vowel close to [a:] by Australian talkers, was categorized as <bad> (i.e., TRAP) a total of 35 times and as <bard> (i.e. START/PALM) a total of 81 times. Many of the other response options were never selected for *zahba* and some were selected only rarely. For example, <paired> was selected seven times out of a total of 128 responses. In our approach, the entire row of responses, including <bad>, <bard>, <paired>, and also items that were never selected, are incorporated into the analysis. By doing so, we are not forced into making a (perhaps arbitrary) cutoff between responses that are simply noise (e.g., possibly the single

[3] We devised unique English orthographic representations of each of the nonsense stimuli (e.g., *zahba*) for the purpose of eliciting recordings from naïve (non-phonetician) participants. These orthographic representations are somewhat awkward but they played no role in the actual perception experiment, where the only orthographic representations were of the choice words.

<bud>, <rude>, and <toured> responses to *zahba*), and responses that are simply less frequent confusions. This is particularly important for vowels with more dispersed response profiles. For example, responses to *zubba*, included to represent the STRUT vowel, include equal numbers of <bad> and <bud>, and also included non-trivial numbers of <bard> and <rude> responses. On the other hand, *zeeba,* included to represent FLEECE, had a clear majority of responses falling in the <bead> category, but there were still many (n = 27) selections of <bed>. Rather than set aside the dominant responses, we incorporate all the experimental trials into the analysis by representing each vowel as a vector of responses. The complete set of confusion matrices is included in the supplementary materials, along with the source data and R code to generate them.

(1) Example of vector representations of three vowels from the A-A condition

	bad	bard	bead	beard	bed	bid	bird	bored	boyd	bud	code	hide	hood	paid	paired	pod	proud	rude	toured
zahba	35	81	0	0	0	0	0	0	0	1	0	0	0	2	7	0	0	1	1
zubba	43	23	0	1	1	0	0	0	0	43	0	1	0	0	1	2	1	12	0
zeeba	2	1	72	8	27	14	0	0	0	0	0	0	0	2	1	1	0	0	0

To calculate the perceptual similarity between two vowels, we therefore calculated the similarity between response vectors. There are a number of ways to do this, including the inner product (sum of the products of corresponding cells) of the vectors and normalized variations of the inner product (e.g., cosine similarity, Pearson's correlation). Here, we chose the Euclidean Distance of the vectors as our similarity metric, for reasons we lay out below. The numerical expression is given in (2), where $\overrightarrow{V1}$ and $\overrightarrow{V2}$ are the vectors for two vowel categories, $r1_i$ is an item of $\overrightarrow{V1}$ and $r2_i$ is an item of $\overrightarrow{V2}$. Each item, $r2_i$, of the vowel vector, $\overrightarrow{V2}$, is subtracted from the corresponding item, $r1_i$, in the comparison vowel vector, $\overrightarrow{V1}$. The square root of the sum of square differences between all items in the vector is the Euclidean distance between vectors and serves as our similarity metric for subsequent analysis.

(2) Vowel similarity based on vector distance:

$$Vdist(\overrightarrow{V1}, \overrightarrow{V2}) = \sqrt{\sum_{i=1}^{N} (r1_i - r2_i)^2}$$

The distance metric in (2) is appropriate for our case for a number of reasons. First, the vectors being compared are always the same length (N = 19), so it is not necessary to normalize for vector length. More importantly, because the differences are squared, this particular method of calculating similarity as distance has the property of enhancing large differences, which also

distinguishes it from methods based on the inner product of vectors (e.g., cosine similarity, Pearson's correlation). This is an advantage for us because we do wish to amplify large differences between vectors. This practice is commonplace in the study of speech perception; at the extreme, only the most frequent response is incorporated for analysis while less frequent responses are discarded as noise. In many cases, analysis of perceptual confusion focuses only on the most frequent response options (i.e., the dominant responses, or limits analysis to pairwise confusions, as described above). These methods are extreme versions of focusing on large differences across vowels. This is most appropriate for speech sounds that are perceived in a categorical or near categorical manner, which tends not to be the case for vowels (e.g., Faris, Best, & Tyler, 2016). By adopting the Euclidean distance metric of similarity, we bias total vector similarity towards response differences of a large magnitude (i.e., a single large response difference has a greater effect than several small response differences). Thus, our calculation is naturally biased towards large differences, which is desirable, while still factoring all responses into the analysis.

To exemplify vowel similarity based on vector distance, consider again the vowels in (1). Based on our metric the distance between *zahba* and *zeeba* (i.e., START/PALM versus FLEECE) is 116; the distance between *zahba* and *zubba* (START/PALM versus STRUT) is 72; and the distance between *zeeba* and *zubba* (FLEECE versus STRUT) is 101. The distances computed in this way are driven by large differences without ignoring dispersion across response options.

More broadly, a general advantage of the vector-based calculation of vowel similarity over pairwise calculation is that it better controls for certain task-based factors that may leave their imprint on the data. The task we report on here is a 19AFC task. Some task-based factors that could influence behavior include the specific choice words used to represent vowels, the orthographic representation of the vowel category, the lexical statistics of the choice words, and the location of the choice words on the screen (which we randomized across subjects). Any one of these factors could introduce a response bias toward some particular choice word, which is independent of the vowel stimulus. Such a bias would skew calculation of similarity based on a pairwise method. Shaw et al. (2018) attempted to correct for this by factoring *a priori* biases for selecting a choice word into their dependent variable, *accuracy'* ("accuracy prime").[4] In a vector-based approach, any such bias remains omnipresent across all vowels, and, since similarity calculation is based on distance, as in (1), will be cancelled out as one vector is subtracted from the other. By filtering task-based effects out of the similarity calculation in this way, our analytical method offers heightened potential for comparison across studies with

[4] For reference, the *accuracy'* results from that study are included in the supplementary materials.

different methods (e.g., similarity based on forced choice tasks with different choice words or different numbers of alternatives).

2.4.2 Visualizing perceptual structure

The similarity/distance matrix, calculated by applying (1) to all combinations of vowels in a condition, served as the input to HCA, which was used to visualize the data.

Our application of HCA progressively fused the distance matrix into binary clusters according to an objective function: Minimize variance of each cluster, a common technique for clustering (Ward, 1963). This method is iterative and ultimately imposes the number of splits needed to differentiate each vowel. The perceptual structure in the data is reflected in the shape of the resulting dendrogram.

2.4.3 Comparing conditions

To compare across conditions, we build on the vector-based Euclidean Distance method in (2). The spirit of the approach is to add up the Euclidean Distances between the vowels in different conditions to get a measure of the total distance across conditions. However, since not all conditions had exactly the same number of participants (N = 12–17), the total number of responses to a vowel also varies across conditions. To normalize the response vectors, we divided each cell by the total number of responses. Each cell in the vector is thus represented as the proportion of total responses to that vowel. An example of normalized rows of the confusion matrix is shown in (3), cf., the unnormalized counterpart of the same data in (1).

(3) Example of a normalized vector representations of three vowels from the A-A condition:

	bad	bard	bead	beard	bed	bid	bird	bored	boyd	bud	code	hide	hood	paid	paired	pod	proud	rude	toured
zahba	0.27	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.05	0.00	0.00	0.01	0.01
zubba	0.34	0.18	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.34	0.00	0.01	0.00	0.00	0.01	0.02	0.01	0.09	0.00
zeeba	0.02	0.01	0.56	0.06	0.21	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00

The formula for calculating normalized vector distance is provided in (4). The only difference from (3) is that the responses are divided by the total number of responses, T.

(4) Similarity based on normalized vector distance:

$$Vdistn(\overrightarrow{V1}, \overrightarrow{V2}) = \sqrt{\sum_{i=1}^{N} (\frac{r1_i - r2_i}{T})^2}$$

To compare conditions, the normalized distance between vowel vectors was added together. The minimum difference for a vowel across conditions is 0 — this is the case when both vowels

have the same proportion of responses to each of the choice words. The maximum difference is $\sqrt{2}$ — this is the case when all responses are different (i.e., no similarity). Since there are a total of 20 stimulus vowels per condition (including the BATH vowel as a stimulus but not as a choice word), the total difference between conditions can range between 0 and 28.28427 ($20*\sqrt{2}$). To make this similarity scale more intuitive, we normalized it from 0 to 1, so that 0 indicates no difference and 1 indicates maximum difference. We did this by dividing the sum of normalized vector distances by the theoretical maximum, according to the equation in (5).

(5) Normalized difference between conditions:

$$Cdistn(Cond1, Cond2) = \frac{\sum_i^N Vdistn_i}{N\sqrt{2}}$$

The methods described above allow us to compare conditions in our experiments holistically, based on the complete set of participant responses to each vowel, to arrive at a set of structured relationships between vowel categories. Such methods make optimal use of our experimental task, which provides perceptual data on the entire vowel system, including all monophthongs and diphthongs in each accent. The results provide a perceptual companion to studies describing production patterns in each of these accents (Blackwood Ximenes, Shaw, & Carignan, 2017; Cole & Evans, 2020; Cox & Fletcher, 2017; Docherty & Foulkes, 1999; Elvin, Williams, & Escudero, 2016; Foulkes & Docherty, 1999; Haddican, Foulkes, Hughes, & Richards, 2013; Hay, Maclagan, & Gordon, 2008; Tollfree, 1999; Watt & Milroy, 1999).

3. Results

We begin the results by presenting perceptual patterns in each accent and pairwise comparisons across accents. These results document the perceptual structure of each accent. At one extreme, vowels may maintain perceptual distance across accents. If this was the case, then we should see the same pattern of responses for each accent, despite acoustic-phonetic differences in their production. At the other extreme, acoustic-phonetic differences in vowels across accents may condition different patterns of confusion, such that vowels that are relatively distinct in one accent may be confusable in another. Thus, this first set of results reports listener categorization of their own accent vowels and will establish what perceptual differences exist across accents. We report these patterns in section 3.1.

Notably, any differences in how listeners perceive the vowels of their own accent could be driven by the distinctiveness of the vowels in the materials. Our first cross-accent condition provides a check on this. In section 3.2, we report how Australian vowels are categorized by listeners from each of the other accents. If behavior in this task is driven by the distinctiveness of the stimulus materials, then we should expect to see the same perceptual patterns on the

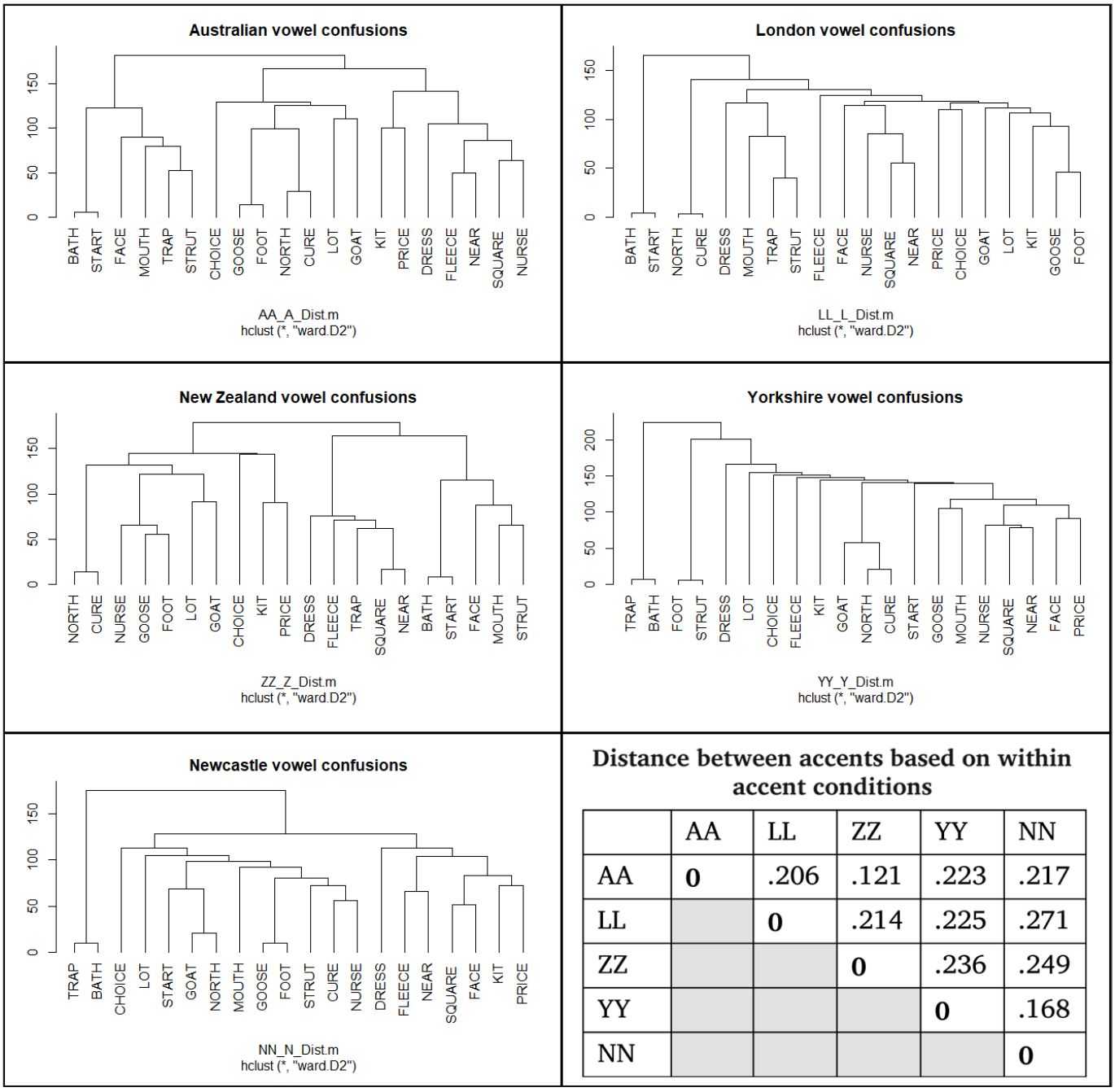
same materials across listener groups. Differences across listener groups on the same materials can only be attributed to variation in listener perceptual structure. To assess this, we compare the cross-accent condition – Australian stimuli and other accent listeners – to two baselines: How other listeners perceive their own vowels and how Australian listeners perceive Australian vowels. If perceptual behavior in this cross-accent condition drifts towards Australian perceptual patterns and away from own-accent behavior, it indicates a clear stimulus effect. Finally, in section 3.3, we return to the cross-accent conditions of Shaw et al (2018), in which Australian listeners categorized the vowels of the other accents, in light of the baselines.

3.1 Within-accent perception and comparisons

Figure 2 provides a dendrogram for each accent, based on listeners hearing their own accent, and a table of pairwise accent comparisons, based on normalized Euclidean distance between confusion matrices. Smaller numbers indicate greater similarity (less distance) between accents: 0 indicates no distance and 1 indicates the maximum possible distance. The pairwise similarity across accents in this sample ranges from .121 to .271, on a scale from 0 (maximally similar) to 1 (maximally dissimilar). The most similar accents, in terms of perceptual responses, are Australian and New Zealand (.121) followed by Newcastle and Yorkshire (.168). The least similar pair is London and Newcastle (.271).

Many of the perceptual differences across accents serve to substantiate accent differences previously described on the basis of impressionistic listening and acoustic measures of similarity. Before continuing, we highlight some of these patterns for each accent.

As expected, perception of the BATH vowel patterns differently across accents. It is fused with START in the dendrogram for Australian, London, and New Zealand, but with TRAP for Yorkshire and Newcastle, a pattern expected from descriptions of BATH in these varieties (e.g., Cole & Evans, 2020; Tollfree, 1999; Wells, 1982). The FACE vowel also patterns quite differently across accents. In Australian and New Zealand accents, FACE is fused to a cluster containing MOUTH and STRUT presumably owing to the relatively low front onset of the diphthong in these varieties (Cox & Palethorpe, 2007; Sóskuthy, Hay, Maclagan, Drager, & Foulkes, 2017). In the UK varieties, FACE is perceived as more similar to mid vowels (e.g., NURSE, NEAR, SQUARE). Another point of difference is whether NEAR is more confused with FLEECE, as in Australian and Newcastle, or with SQUARE as in the other three accents. Predictably, based on past descriptions (see references at end of 2.4.3), GOAT was similar to LOT in Australian, New Zealand, and London but not in the northern UK varieties, where GOAT was perceptually similar to NORTH. Note that GOAT tends to be a mid-back monophthong in Yorkshire and Newcastle, as NORTH is in all five varieties. New Zealand stands alone in having FLEECE directly fused with a branch containing TRAP, which is a rather more raised front vowel in New Zealand than in most other varieties



Distance between accents based on within accent conditions

	AA	LL	ZZ	YY	NN
AA	**0**	.206	.121	.223	.217
LL		**0**	.214	.225	.271
ZZ			**0**	.236	.249
YY				**0**	.168
NN					**0**

Figure 2: Dendrogram of each accent, based on listeners hearing their own accent, and a table showing the normalized distance between each accent (within-accent conditions). Greater numbers indicate a larger distance between responses. Distances are normalized to fall between 0 and 1, where 0 indicates identical responses and 1 indicates maximally different responses.

(Hay, Pierrehumbert, Walker, & LaShell, 2015). Similarly, London stands alone in having KIT fused directly to a cluster containing GOOSE. This presumably reflects a greater extent of GOOSE and FOOT fronting in southern UK accents (Alderton, 2020) than in the other accents in the sample. In particular, FOOT in Australian maintains a relatively back position (Blackwood

Ximenes et al., 2017). Thus, several trends in the perceptual differences found across accents also correspond straightforwardly to trends in vowel production documented for these varieties.

3.2 Cross-accent perception

We now turn to cross-accent perception. We have already seen that listeners from the different accents in our study show different patterns of vowel confusions for their own vowels, and that these patterns of confusion reflect closely the phonetic differences in the vowel systems of the accents. That is, accent-specific vowel perception is closely attuned to accent-specific acoustic patterns of vowel production. This is unsurprising. Of interest to understanding the boundaries of the vowel perceptual categories are the cross-accent conditions, which allow us to probe how perceptual attunement to one accent — the listener's native accent — influences perception of another less familiar accent. That is, the cross-accent conditions allow us to decouple the accent-specific production-perception cycle. We ask how the perceptual structure that has developed with primary input from one accent influences perception of vowels produced by speakers of another less familiar accent. We pursue two separate comparisons: The first (3.2.1) examines how listener groups from the five accent backgrounds categorize the Australian vowels; the second (3.4.2) examines how Australian listeners categorize vowels from each of the five accents.

3.2.1 Perception of Australian vowels by other accent groups

Our first cross-accent comparison allows us to investigate the degree to which the differences in perceptual behavior observed above across accents can be attributed strictly to stimulus properties of the vowels. Two vowels that are perceptually similar in one accent, because of similar acoustic-phonetic properties, might be expected to be perceptually similar for listeners of another accent. At the extreme, listeners could show perceptual behavior with other accents that matches that of the talkers of those accents. This would indicate that the perceptual differences observed above are not entrenched aspects of perceptual structure in the listener but rather are more local (short-term) responses to the acoustic properties of accent-particular vowels.

To evaluate this scenario, we made use of conditions in which listeners from the non-Australian locations categorized Australian vowels. In particular, we are interested in whether listeners from the other accents responded to Australian vowels in the same way as Australian English speakers, or instead differed by listener accent. We therefore compared A-A to four other conditions: L-A, Z-A, Y-A, N-A; these hold the stimulus accent (A) constant while varying the listener group. The results are shown in **Table 2** (second column). To interpret the patterns, we included two points of reference. First, we repeated the distances for the accent comparisons (from **Figure 2**) in the first column. Second, we provided the distance between the cross-accent condition and the Within-Other conditions (in which listeners of the non-Australian accents categorized their own accent vowels). We are interested in whether the numbers in the second column are lower than these baselines (first and third columns). This would indicate that listeners categorize an unfamiliar accent more like speakers of that accent than how they categorize their own accent vowels.

The smallest distances (lowest numbers) in each row in **Table 2** are in bold. Comparing the cross-accent conditions (second column) to the accent differences (first column) and own accent baseline (third column), we see that two of the four accents — Yorkshire and Newcastle — show a decreased distance in the cross-accent condition. For these listener groups, the distance between conditions is smaller in the second column (cross-accent perception) of **Table 2** than in the first column (within accent perception) or third column. Patterns of confusion for Australian vowels by these two listener groups are more similar to Australian listener patterns than to confusion patterns for their own vowels. These results indicate that some portion of the accent differences observed may follow simply from the patterns of similarity that are present in the stimulus items. It is notable, however, that this effect was found only for Yorkshire and Newcastle, the accents with the largest baseline difference from Australian. The effect was not found for the other two listener groups, from London and New Zealand.

Figure 3 breaks down the results by vowel. It shows which vowels contribute most to the differences across conditions (column two comparison), for each listener accent. These results were obtained by subtracting the cross-accent normalized Euclidean distance (e.g., Cdistn(AA,ZA)), from the accent distance, (e.g., Cdistn(AA,ZA)). Negative numbers indicate that perceptual distance is bigger for that vowel in the cross-accent scenario, relative to within-accent listening. Positive numbers indicate that perceptual distance is smaller for that vowel in the cross-accent scenario. That is, the stimulus is classified more like the Australian listeners would classify it, rather than the within-accent group.

We briefly comment on the individual vowels that contributed to the overall pattern in **Table 2**. New Zealand has several vowels that are more dissimilar in the cross-accent condition, with KIT having the largest effect. This is because New Zealanders and Australians are both relatively successful at classifying their own KIT, but New Zealanders have a tendency to categorize the Australian vowel differently (most often as DRESS). This means the distance between New Zealanders listening to themselves and New Zealanders listening to Australians is much bigger for this vowel than the distance between the two within-accent conditions. London's vowels, in contrast, tend towards the weak positive end of the scale, led by START and CURE. London responses to the Australian version of these vowels drifted towards the Australian responses (i.e., a possible stimulus effect). For Yorkshire, the overall pattern is driven largely by positive effects for FOOT and BATH. Responses to the Australian pronunciations of these vowels shifted dramatically towards the Australian response pattern. This makes sense as it reflects the difference in acoustic properties of vowel production across the two varieties. Newcastle also shows a strong positive cross-accent effect for BATH. However, for Newcastle, compared to Yorkshire, there are also several vowels with small negative effects.

	Accent differences	Cross-accent perception of Australian vowels by other accent listeners	
	A-A vs. Within-Other	**A-A vs. Other-A**	**Other-A vs. Within-Other**
London	0.206	0.166	**0.158**
New Zealand	**0.121**	0.197	0.212
Yorkshire	0.223	**0.173**	0.246
Newcastle	0.217	**0.208**	0.273

Table 2: Comparison across conditions based on normalized Euclidean Distance. The smallest numbers, indicating least perceptual distance, in each row are in bold.

To summarize, this first cross-accent comparison allowed us to investigate whether the non-Australian listeners got closer to the Australian listener confusion pattern when they were listening to Australian vowels. We found this for two of four listener groups, but the strength of the effect varied across accents and across vowels. The strongest accent level effect of this type was found for Yorkshire. Confusion patterns moved from 0.223 (A-A vs. Y-Y) to 0.173 (A-A vs. Y-A), a change of 0.05. This was driven largely by responses to FOOT and BATH. The other group that showed a shift towards Australian perceptual patterns when listening to Australian vowels was Newcastle but the shift was tiny, just a 0.009 change, and it was also driven primarily by response to a single lexical set, BATH. New Zealand listeners did not shift towards Australians. London listeners did show a shift, a .04 decrease from column 1 to column 2 (**Table 2**), but there was also not a decrease relative to column 3. Thus, the change in the direction of A-A when listening to Australian vowels also brought London listeners closer to L-L. We cannot clearly interpret the L-A confusion pattern as moving towards A-A when it also moves toward L-L.[5]

Overall, then, we have mixed results, with only the Yorkshire listeners, and to a small degree Newcastle listeners, showing more Australian-like response patterns on Australian vowels. Moreover, this effect was driven largely by just a few vowels that show large accent differences in acoustic-phonetic properties. All accents had some vowels with positive shifts and some with negative shifts. There are Australian vowels that are categorized by other accent listener groups

[5] We're not sure why only London shows this pattern. One possibly related note is that this is the condition that likely has the greatest accent difference between the talkers who produced the stimuli and the listeners who participated in the task, as the talkers were selected on the basis of having particular characteristics representative of working-class London accent while the listeners were indigenous members of the London community but not subject to the same production-based screening.

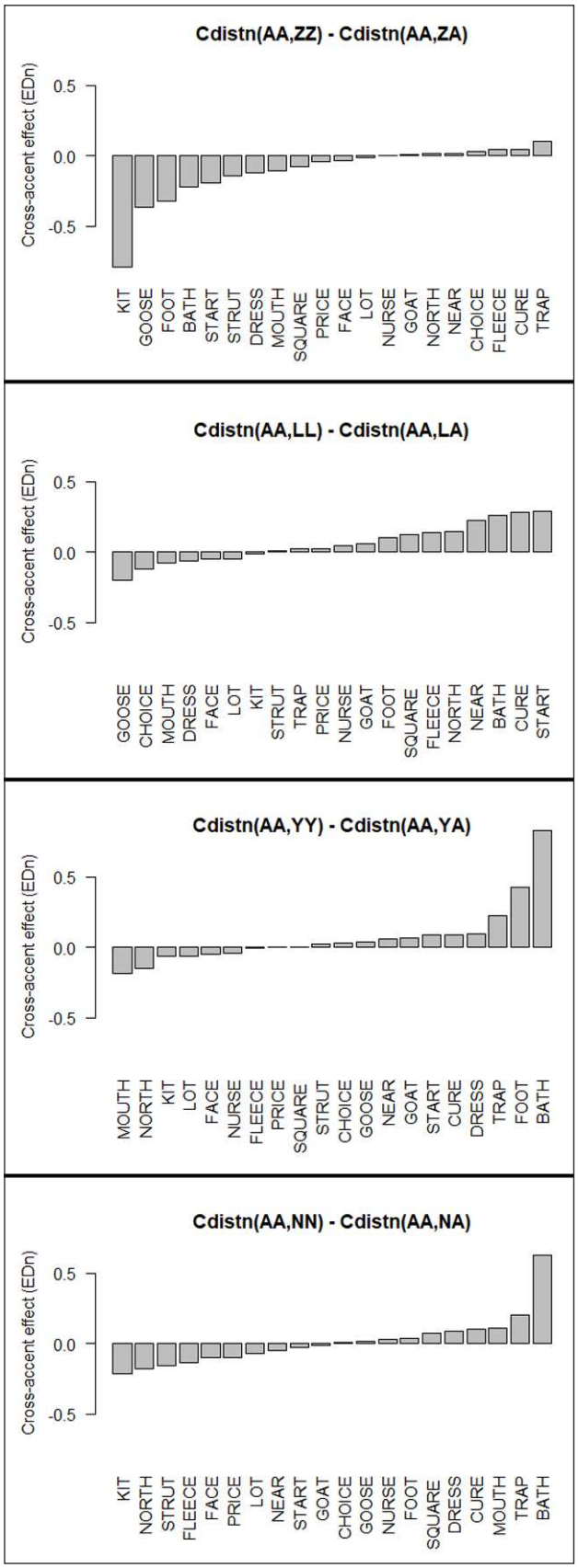


Figure 3: The effect of cross-accent perception by vowel, obtained by subtracting the cross-accent distance (e.g., Cdistn(AA,AZ)) from the baseline accent distance (e.g., Cdistn(AA,ZZ)). Negative numbers indicate that perceptual distance in the cross-accent condition (e.g., Cdistn(AA,AZ)) is larger than the within-accent distance (e.g., Cdistn(AA,ZZ)).

just as Australians listeners categorize them, even though this is a different pattern from the listeners' within-accent perception of that lexical set. These vowels, the ones at the positive end of the scales in **Figure 3**, require no adaptation to be perceived as the Australian speakers intended them. On the other hand, there are Australian vowels that are categorized by other accent listener groups in ways that differ both from the Australian listeners and from their within-accent categorization. These are vowels that, to be perceived as intended by the speakers, would require some perceptual learning on the part of the listener. On the whole, Australian patterns of confusion on Australian vowels did not carryover to other-accent listeners, except in a small number of isolated cases.

3.2.2 Perception of other accent vowels by Australian listeners

Our next set of comparisons holds the listener group constant, while varying the accent of the vowel stimuli. We report the A-L, A-Z, A-Y, and A-N conditions. In these conditions, the listener group is always Australian. This allows us to probe the stability of perceptual categories across variation in vowel stimuli. We again note that accuracy across these conditions was reported in Shaw et al. (2018), who found that accuracy by vowel was surprisingly consistent across accents. The new metrics developed in this paper allow for a more comprehensive look at this behavior.

We calculated the distance between the cross-accent conditions and A-A. This enables us to quantify differences from the target (Australian) accent. We also calculated the distance between each cross-accent condition and its own "other-other" (L-L, Z-Z, Y-Y, N-N) condition. This enables us to quantify how responses to the target accent differ from those for native accent vowels. A summary of the results is provided in **Table 3**. We present the results alongside the accent differences (A-A vs. within-other) in column one. At the extreme, if Australian listeners confuse other accent vowels in the same way as they confuse their own vowels, then the second column (A-other vs. A-A) will be zero. If there is a numerical trend in this direction, then the second column will be smaller than the first (and third) column baselines. This is the case for London and Yorkshire accents (in each row of **Table 3**, the smallest number is in bold). For New Zealand and Newcastle, it is the first column (accent difference) that is smallest. Listening to these accents results in a unique categorization (i.e., one that is neither like the listener accent nor the speaker accent). Notably, the third column was never the smallest. That is, it was never the case that cross-accent perception (A-other) was closer to the stimulus accent (within-other). This reinforces a conclusion from the last section (i.e., a relative dearth of stimulus effects in these data).

Australian listeners categorized vowels in all accents in ways that were closer to how they categorized their own accent than to how those vowels were categorized by other accent listeners. In other words, the type of uncertainty that Australians have about categorizing their own vowels is reflected in their categorization of the other accent vowels.

Figure 4 indicates which vowels contributed to the overall trends. This was calculated in the same way as for our first cross-accent comparison (**Figure 3**). The normalized distance between within-accent and cross-accent conditions (e.g., Cdistn(AA, AZ)) was subtracted from the baseline, the normalized difference between within-accent conditions (e.g., Cdistn(AA, ZZ)). Negative numbers indicate that the cross-accent perceptual distance, Cdistn(AA, AZ), was larger than the within-accent perceptual distance, Cdistn(AA, ZZ). For New Zealand, perception of DRESS and KIT contributed the most to the cross-accent perceptual difference. For Yorkshire and London, most vowels trended positive, meaning that responses to the Yorkshire and London vowels tended towards Australian responses. For Yorkshire, the largest positive effects were for BATH and FOOT. For London, SQUARE and FLEECE had the strongest positive effects. Newcastle had the most polarized responses across vowels. BATH had a strong positive effect, but several vowels had negative effects, with KIT, START and FACE showing the strongest negative effects.

	Accent differences	**Cross-accent perception by Australian listeners**	
	A-A vs. Within-Other	**A-other vs. A-A**	**A-other vs. Within-Other**
London	0.206	**0.131**	0.201
New Zealand	**0.121**	0.174	0.143
Yorkshire	0.223	**0.172**	0.246
Newcastle	**0.217**	**0.217**	0.234

Table 3: Comparison across conditions based on normalized Euclidean Distance. The smallest numbers, indicating least perceptual distance, in each row are in bold.

To summarize this set of comparisons, we found that cross-accent vowel perception by Australian listeners was closer to Australian vowel perception than it was for two of the accents providing comparator vowel stimuli, London and Yorkshire. For the others, New Zealand and Newcastle, perceptual distance increased in the cross-accent scenario. It was never the case, in these conditions (Australian listeners), that perception shifted in the direction of the stimulus accent.

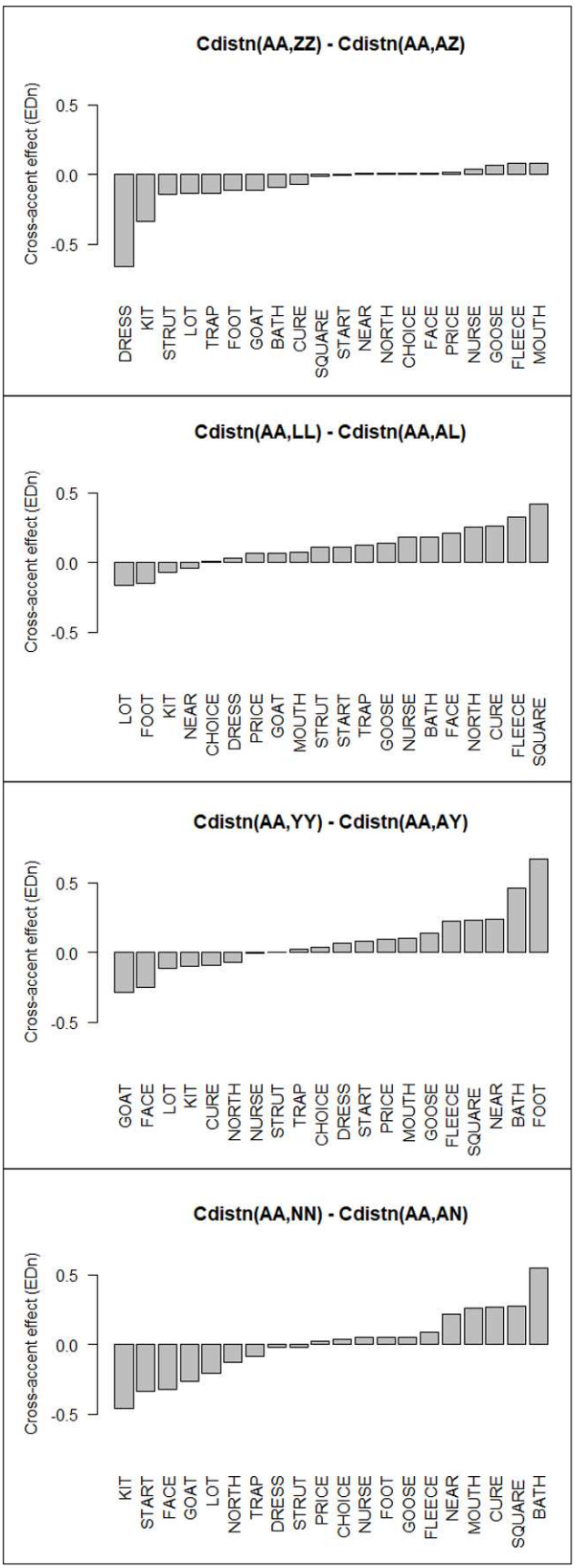


Figure 4: The effect of cross-accent perception by vowel, obtained by subtracting the cross-accent distance (e.g., Cdistn(AA,ZA)) from the baseline accent distance (e.g., Cdistn(AA,ZZ)). Negative numbers indicate that perceptual distance in the cross-accent condition (e.g., Cdistn(AA,ZA) is larger than the accent distance (e.g., Cdistn(AA,ZZ)).

4 General Discussion

4.1 Perceptual differences across accents

One goal of this study was to offer a perceptual description of different English speech varieties. We presented the perceptual responses for each accent and computed pairwise comparisons of similarity. Our perceptual characterization of the accents is based on patterns of confusion across all vowels. Notably, two accents can have the same pattern of perceptual confusion even if the vowels are phonetically different. To illustrate, consider responses to the DRESS vowel. The patterns of confusion for DRESS in within-accent conditions (A-A, Z-Z, Y-Y, L-L, N-N) are provided in **Table 4** (for reasons of space, choice words with zero responses, <rude>, <toured>, <code> are excluded). Unsurprisingly, <bed> is the most common response for this vowel for all accents. Now, consider the second most common response. For four out of five accents, this was <bead>. Notably, DRESS is a relatively high vowel in New Zealand that thus encroaches upon high front FLEECE. In Australian, Yorkshire and London, DRESS is mid-front and thus more distant from high front FLEECE. Despite the phonetic differences in the realization of the vowels, the confusion pattern remains similar across the four accents. Newcastle is different in this respect – there were few non-DRESS responses overall with <paid> (FACE) being the most common. It is noteworthy that FACE tends to be realized as a close front monophthong in this accent, thus similar in quality (if not duration) to Australian DRESS. In Yorkshire FACE is also monophthongal, but usually more open. The broader point is that vowels can maintain perceptual distance to some degree across accents, even as they shift in quality within the acoustic vowel space, if we consider their relatively maintained topological relationships within the whole system. Similarities in perceptual structure across accents can derive from phonetic similarities in the vowels as well as from maintenance of relative perceptual distances between vowels.

	Vowel Choices															
DRESS (zebba)	**bad**	**bard**	**bead**	**beard**	**bed**	**bid**	**bird**	**bored**	**boyd**	**bud**	**hide**	**hood**	**paid**	**paired**	**pod**	**proud**
A-A (N = 16)	4	3	**10**	0	102	0	0	0	1	0	1	3	0	2	1	1
Z-Z (N = 16)	2	0	**12**	9	97	0	0	1	1	0	0	0	2	4	0	0
Y-Y (N = 16)	2	0	**3**	0	121	1	0	0	0	1	0	0	0	0	0	0
L-L (N = 12)	2	0	**4**	1	87	0	1	0	0	0	0	1	0	0	0	0
N-N (N = 12)	1	0	0	0	90	1	0	0	0	0	0	0	**2**	1	0	1

Table 4: Perceptual categorization responses to 'zebba' (DRESS VOWEL) across accents. The second most frequent response in each row is in bold.

We evaluated the pairwise distances between all five accents. Our metric, normalized Euclidean Distance, varies from 0 (identical) to 1 (maximally different). On this scale, pairwise accent comparisons ranged from .121 (the most similar pair: Australia and New Zealand) to .271 (the least similar pair: Newcastle and London). These relatively low numbers indicate that there is a high degree of perceptual similarity across accents likely driven by listeners often choosing the talker's intended vowel. Interestingly, the cross-accent conditions (as compared to listener and stimulus accents) generally stayed within this range as well. The one exception was Newcastle listeners to Australian vowels (N-A), which differed from Newcastle listeners to Newcastle vowels (N-N) by .273, just slightly higher than the largest cross-accent difference (.271). Despite the relative similarity of perception across conditions, there were still some differences (i.e., accent-specific perceptual patterns).

Some of the accent-specific perception patterns in this study correspond quite clearly to accent-specific acoustic-phonetic patterns, as documented in the sociophonetics literature. For example, in Australian English, the FACE and GOAT vowels are realized as diphthongs that begin with relatively low (open) vowels (Cox & Palethorpe, 2007). These vowels have a similar realization in New Zealand (Hay et al., 2008, pp. 25–26). In perception, FACE and GOAT were grouped with relatively low vowels by Australian/New Zealand listeners, a pattern that mirrors their acoustic-phonetic details. In all accents, NORTH is higher (closer) than LOT (for a compact comparison of acoustic vowel spaces, see Shaw et al., 2018, **Figure 1**). GOAT was grouped with LOT by Australian and New Zealand listeners but with the higher vowels NORTH/CURE in Newcastle and Yorkshire. FACE was grouped with TRAP/STRUT/MOUTH by Australian listeners and STRUT/MOUTH by New Zealand listeners. Presumably, FACE was confused more with TRAP by Australian listeners than by New Zealand listeners because TRAP is higher in New Zealand and therefore more acoustically distinct from FACE. In the UK accents, FACE was categorized as higher vowels. FACE was confused with SQUARE in Newcastle, with NURSE/SQUARE/NEAR in London, and with PRICE in Yorkshire. In Yorkshire and Newcastle, FACE and GOAT are generally produced as relatively high monophthongs. In perception, Yorkshire and Newcastle listeners group GOAT with NORTH. The difference in how FACE is perceived across Yorkshire and Newcastle listeners – grouped with PRICE in Yorkshire, and with SQUARE in Newcastle – might be related to the relative height of this vowel in production. The Newcastle FACE vowel is somewhat higher than the Yorkshire FACE vowel. The existence in Newcastle of a monophthongal realization of FACE that makes it similar to realizations of SQUARE (Watt & Milroy, 1999) may also contribute to the pattern. These aspects of the accent-specific perceptual patterns mirror accent-specific acoustic-phonetic patterns.

To summarize the previous two points: (1) Some aspects of the vowel perception data, such as the DRESS~FLEECE confusion pattern noted above, are relatively consistent across accents, even in the presence of acoustic-phonetic variation, and (2) some accent-specific perceptual patterns appear to follow from accent-specific acoustic-phonetic differences. These findings provide an important baseline for considering cross-accent perception patterns.

4.2 Cross-accent perception

A second goal of the study was to expand the scope of our work on cross-accent perception. Past work showed that Australian listeners have similar patterns of vowel categorization accuracy across regional accent vowel systems (i.e., similar accuracy patterns when categorizing their own vowels or those of some other regional accent), suggesting some degree of perceptual assimilation to native accent categories (Shaw et al., 2018). Moreover, accuracy patterns remained stable even after short-term multi-talker exposure to each accent. For ease of reference, the accuracy results are provided in the supplementary materials. The present study builds on Shaw et al. (2018) by adding the new baseline conditions, with each accent group categorizing their own vowels (described above), and new cross-accent conditions, with each accent group categorizing Australian vowels, and by developing new methods for comparing perceptual confusions across accents. These additions extend previous results and allow us to sharpen our interpretation of the perception data.

As mentioned above, there were some accent-specific perceptual patterns (in the within-other conditions). Which vowels are confusable with each other varies to some degree across accents. This was important to confirm. It rules out one possible explanation for the systematicity with which Australians categorized other accent vowels in Shaw et al. (2018) – this behavior is not because each accent maintains a consistent perceptual distance between categories.

Additionally, the new cross-accent conditions provided evidence against stimulus-driven effects. Listeners from each accent categorized Australian vowels differently. Non-Australian listeners did not generally show Australian-like perceptual patterns. The one exception was Yorkshire listeners, whose perceptual patterns were more Australian-like when listening to Australian than when listening to their own vowels. However, this shift towards the Australian perceptual pattern was driven primarily by just two lexical sets, BATH and FOOT, which are vowels that are stereotypically different in production in the two accents. Newcastle listeners showed a tiny shift towards Australian-like perceptual patterns, again driven by BATH. These results indicate that the same set of stimuli (Australian) can be perceived in systematically different ways by listeners of other accents. They also suggest that we can rule out the possibility that differences in perceptual behavior across accents are due to the particular stimuli selected for each accent. That is, we find different perceptual behavior across listener groups even when the stimuli are held constant. Thus, the only possible explanation for accent-specific perceptual patterns is accent-specific perceptual structures. Listeners from different regional accent backgrounds interpret the same acoustic information differently.

The new baseline conditions sharpen our understanding of cross-accent conditions with Australian listeners. Comparison of how Australian listeners perceive their own vowels, in the A-A condition, with how they perceived other accents in the A-L, A-Z, A-Y, and A-N conditions, revealed some imposition of native accent perceptual structure on the stimuli produced

by speakers of the unfamiliar accents. Categorization of vowels in two accents, London and Yorkshire, shifted towards the Australian accent pattern. This indicates that the uncertainty that Australians have about their own vowel categories influenced how they categorized London and Yorkshire vowels. Aspects of Australian-specific perception patterns surfaced when Australians listened to London and Yorkshire accents. This was indicated by a smaller normalized distance between cross-accent (e.g., A-L) and own-accent (e.g., A-A) conditions than between within-accent conditions (e.g., L-L and A-A). In other words, the errors that Australian listeners made in categorizing vowels in London and Yorkshire accents were similar to the errors that they made with their own accent, but different from the errors that native listeners of those other accents made on their own accent.

To illustrate this pattern, we focus on the vowel with the biggest perceptual difference between Australian and London accents (A-A vs. L-L). This vowel was SQUARE, which had a normalized Euclidean distance of 0.536 between accents. In the Australian dendrogram (**Figure 2**), SQUARE is first fused with NURSE, then with FLEECE/NEAR and then with DRESS. London is different, particularly in the separation of SQUARE from DRESS. In London SQUARE is fused with NEAR, then NURSE, then FACE (with SQUARE still three branches away). The response vectors for this vowel for A-A, L-L, and A-L conditions are offered for comparison in **Table 5** (for reasons of space, choice words with zero responses, <bid>, <bud>, <hood>, <rude>, are excluded). Consistent with the dendrogram, London listeners mostly selected <paired> (SQUARE) and sometimes selected <beard> (NEAR) as responses to 'zairba' (SQUARE). Australian listeners very frequently selected <bed> (DRESS), followed by <paired> (SQUARE) but they also selected <paid> (FACE), <bead> (FLEECE), and <beard> (NEAR) at least 10 times each. This variability in perception may be related to observed formant variability in the production of this vowel by Australian speakers (Nguyen & Shaw, 2014). The perceptual distance in the cross-accent condition A-L, Australians hearing London, dropped substantially, from 0.536 (A-A vs. L-L) to 0.118 (A-A vs. A-L). Australians responded to London SQUARE in a similar manner to their own SQUARE, distributing responses across <bed>, <paired>, <beard>. In this case, it appears that the degree of uncertainty that Australians have about their own SQUARE is thus somewhat reflected in their responses to the London vowel.

	Vowel Choices														
SQUARE (zairba)	**bad**	**bard**	**bead**	**beard**	**bed**	**bird**	**bored**	**boyd**	**code**	**hide**	**paid**	**paired**	**pod**	**proud**	**toured**
A-A (N = 16)	5	2	10	10	45	2	1	1	0	1	11	36	1	2	1
L-L (N = 12)	3	0	1	12	6	3	0	0	0	0	2	69	0	0	0
A-L (N = 16)	1	3	6	19	43	4	1	0	1	0	5	44	0	0	1

Table 5: Responses to 'zairba' SQUARE across accents.

The vowel most responsible for driving the cross-accent perception of Yorkshire vowels by Australian listeners closer to Australian patterns was FOOT. The normalized distance for FOOT between accents is 0.820 (AA-YY). In the cross-accent condition (A-Y), Australians responded more like they did to their own-accent FOOT vowel, reducing the distance to 0.151. This pattern follows from the uncertainty that Yorkshire listeners had in categorizing FOOT. Yorkshire listeners categorized FOOT as <bud> (STRUT), reflecting the lack of contrast between these two vowels in northern accents of England. When Australian listeners heard Yorkshire stimuli, they did not select <bud> as often as Yorkshire listeners, which brought the Australian response pattern to Yorkshire FOOT closer to the Australian pattern.

The cross-accent conditions involving Yorkshire and London present a new type of example of how native-accent perceptual structure can influence perception of another accent. Notably, this type of pattern was not seen for every accent combination, at least not in the aggregate measure. For Newcastle vowels, the cross-accent conditions with Australian listeners (A-N) did not inch towards the Australian (A-A) pattern. For New Zealand vowels too, perceptual distance for the cross-accent condition did not decrease — in fact, it increased — relative to the baseline accent difference. This illustrates a different way in which listener perceptual structure influences perception. The vowel that contributed most to the increase was DRESS. As shown in **Table 4**, Australian (A-A) and New Zealand (Z-Z) responses to DRESS were similar: The normalized Euclidean distance was 0.090. When Australians heard New Zealand DRESS, they categorized it primarily as <bid> (KIT), a sharp departure from both A-A and Z-Z conditions, increasing distance from 0.090 to 0.754 for this vowel. This is a case in which the phonetic properties of a vowel as produced in one accent locate it in a different perceptual category of another accent (i.e., a Category-Shifting difference) (Best, Shaw, & Clancy, 2013; Best et al., 2015a; Best et al., 2015b; Best, Tyler, Gooding, Orlando, & Quann, 2009; Faris et al., 2016; Tyler et al., 2014; Ying, Shaw, & Best, 2013). While there are only a handful of true Category-Shifting differences across our accents, they had a significant impact on the results, a point we elaborate on in the next sub-section.

4.3 Revealing perceptual structure through input variation

In more ecological speech perception scenarios, listeners are typically aided by lexical, indexical, and other contextual factors that facilitate mapping between the speech signal and a phonological category. These sources of information can support novel mappings between the speech signal and phonological categories, as in the case of adaptation to unfamiliar talkers or accents (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Maye et al., 2008; Norris, McQueen, & Cutler, 2003; Sumner, 2011; Sumner & Samuel, 2009).

In the absence of these sources of information, the condition that our experiment was designed to simulate, listeners appear to interpret the signal rather superficially. This is similar

to what can be observed in cross-language speech perception (e.g., Best, 1995; Polka, 1991; Sebastián-Gallés, 2005; Werker & Tees, 1984), whereby listeners tend to categorize tokens based upon phonetic similarity to their own phonological categories. Our listeners were not told that they would be listening to other accents, and indeed there was nothing about the experiment that would indicate this. The design of the experiment was thus particularly well-suited to exposing the role of phonological boundaries in the native accent, even though this was exposed through naturally produced stimuli from other accents. Unsurprisingly, the category boundaries between vowels tended not to be sharp. Even for the native accent, listener responses typically fell across multiple categories. The largest contributors to cross-accent differences in perception tended to be stimulus-category mismatches of three main types, on which we now elaborate.

The first type relates to large "realizational" differences, to use Wells' (1982) term. When vowel locations shift together — as in the locations of the short front vowels of New Zealand relative to, e.g., Australian — they largely maintain perceptual distance within accents. Thus, responses to New Zealand vowels in the Z-Z condition and to Australian vowels in the A-A condition were quite similar, the most similar of any two accent pairs. This is despite the realizational differences in the vowels. However, in the cross-accent conditions, Z-A and A-Z, many vowels were mis-categorized because of how the respective vowel spaces align. The largest differences were for DRESS and KIT, where the realization in one accent maps to a different category in the other. New Zealand DRESS was classified as KIT most often by Australian listeners; Australian KIT was classified as DRESS by New Zealand listeners. To perceive these vowels as intended by the talker (i.e., veridically across accents) identification of these category-shifting differences would require perceptual adaptation.

Besides accent-specific realizational shifts, differences in the number of vowels – as in the Yorkshire FOOT-STRUT overlap, discussed above – and differences in the patterning of the BATH lexical set account for the largest deviations in the cross-accent conditions. The primary reason why the distance between Y-A and A-A conditions decreased relative to Y-Y and A-A conditions is the lack of FOOT-STRUT contrast in Yorkshire (which reflects a historical split of STRUT from FOOT in the main ancestors of the Australian accent, a split that did not occur in ancestral Yorkshire). The BATH vowel was a major contributor to cross-accent perceptual patterns involving Newcastle and Yorkshire. BATH has the same quality as TRAP in Yorkshire and Newcastle but the same quality as START/PALM in the other three accents. When listeners from Yorkshire and Newcastle accents heard Australian BATH, they categorized it as START, which accounted for the biggest difference from how they categorized their own accent. Had the Australian listeners known that they were listening to Yorkshire or Newcastle speakers or had they had lexical information, they might easily have adapted their expectations. In the absence of explicit indexical or lexical information, the data primarily reveal the phonological category boundaries of the listeners. Our current perceptual data can therefore be taken as a baseline against which cross-accent perception

in less informationally impoverished situations can be understood. Specifically, the confusions we report isolate the specific contribution of phonological categorization in the absence of lexical or indexical factors.

On the whole, we interpret our results as evidence for accent-specific perceptual structure. Each accent imposes a unique pattern of perception, on even the same stimuli. Moreover, listener uncertainty about perceptual categories – reflected in confusions made on their own accent vowels – often persist when listening to other accents. Finally, we observed a small number of category-shifting differences, following from differences in either the number of phonological vowel categories across accents or large realizational differences. Such differences provide opportunities to assess perceptual adaptation, which may occur in more ecological listening tasks.

5. Conclusion

We reported vowel categorization experiments with five listener groups covering the entire vowel system of each accent. We compared patterns of perception across the five accents, Australian, New Zealand, London, Yorkshire, Newcastle, as well as two sets of cross-accent perceptual conditions: (1) Australians categorizing vowels from the other four accents, and (2) the other four accents categorizing Australian vowels. The cross-accent conditions served to demonstrate boundaries between perceptual categories. In large part, listeners were tolerant of small realizational differences across accents, showing patterns of responses on unfamiliar accents that were similar to how they respond to their native accent vowels. Large deviations across accents come primarily from three sources: (1) Realizational differences whereby the phonetics of a vowel in one accent maps to a different category in another accent, (2) category-level differences, such as the split of FOOT-STRUT in Australian and lack of split in Yorkshire and Newcastle, and (3) lexically-determined differences (i.e., the BATH lexical set). Aside from these cases, responses to vowels in the cross-accent conditions largely served to delineate listeners' own accent category boundaries. Each accent has a unique perceptual structure, which we exposed through "whole-system" analyses, comparing complete confusion matrices and vowel response vectors across conditions. The data reveal perceptual behavior in the absence of higher-level influences from context, including absence of lexical information and socio-indexical information about the talker. Our results indicate which vowels might be subject to adaptation in more ecological listening settings and, more broadly, that vowel perception in the absence of lexical and socio-indexical information comes with a high degree of uncertainty, even in a listener's native accent. Future work should address how closely the phonetic properties of the vowels alone can derive the perceptual categorization patterns and how integration of lexical, phonotactic, and socio-indexical information can shift behavior relative to the baseline we have reported here.

Additional file

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

- **Supplementary Materials.** Supplementary A to Supplementary C. DOI: https://doi.org/10.16995/labphon.6436.s1

Acknowledgements

We would like to thank participants in this study and several research assistants who contributed to the collection of the data reported here — in New Zealand: Ksenia Gnevsheva, Mike Peek, Alia Hope-Wilson, Nathan Taylor, and Vicky Watson; in Northern England: Jalal Al-Tamimi, Sophie Wood, Vincent Hughes, Amanda Cardoso, Hannah How, Justin Lo, Ella Jeffries and James Tompkinson; in London: Anita Wagner, Katharine Mair, Gisela Tomé Lourido; in Sydney, Sarah Wright, Sara Fenwick, and Mark Lathouwers. The exposure passage was written by Meg Mundell. We thank audiences at the *19th International Congress of Phonetic Sciences* in Melbourne and the workshop of the Berkeley Linguistics Society *Phonology Representations: at the crossroads between gradience and categoricity*, where aspects of this research were presented. The project was supported by Australian Research Council grant DP120104596.

Competing Interests

The authors have no competing interests to declare.

References

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *The Journal of the Acoustical Society of America, 51*(2B), 648–651. DOI: https://doi.org/10.1121/1.1912889

Alderton, R. (2020). Speaker gender and salience in sociolinguistic speech perception: GOOSE-fronting in Standard Southern British English. *Journal of English Linguistics, 48*(1), 72–96. DOI: https://doi.org/10.1177/0075424219896400

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174–EL180. DOI: https://doi.org/10.1121/1.4789864

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: Johns Benjamins. DOI: https://doi.org/10.1075/lllt.17.07bes

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistics experience: Issues in cross-language research* (pp. 171–204). York, Timonium, MD.

Best, C. T. (2015). Devil or angel in the details? In J. Romero & M. Riera (Eds.), *The phonetics–phonology interface: Representations and methodologies* (pp. 3–32). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.335.01bes

Best, C. T., Shaw, J. A., & Clancy, E. (2013). Recognizing words across regional accents: the role of perceptual assimilation in lexical competition. *Proceedings of Interspeech 2013,* 2128–2132. DOI: https://doi.org/10.21437/Interspeech.2013-504

Best, C. T., Shaw, J. A., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., Mulak, K.E. & Wood, S. (2015a). From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels. *Proceedings of Interspeech 2015,* 1932–1936. DOI: https://doi.org/10.21437/Interspeech.2015-426

Best, C. T., Shaw, J. A., Mulak, K. E., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K. & Wood, S. (2015b). Perceiving and adapting to regional accent differences among vowel subsystems. *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow, UK, University of Glasgow.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native-and Jamaican-accented words. *Psychological Science, 20*(5), 539–542. DOI: https://doi.org/10.1111/j.1467-9280.2009.02327.x

Blackwood Ximenes, A., Shaw, J., & Carignan, C. (2017). A comparison of acoustic and articulatory methods for analyzing vowel differences across American and Australian dialects of English. *The Journal of Acoustical Society of America, 142*(2), 363–377. DOI: https://doi.org/10.1121/1.4991346

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729. DOI: https://doi.org/10.1016/j.cognition.2007.04.005

Carré, R., & Chennoukh, S. (1995). Vowel-consonant-vowel modeling bysuperposition of consonant closure on vowel-to-vowel gestures. *Journal of Phonetics, 23*(1), 231–241. DOI: https://doi.org/10.1016/S0095-4470(95)80045-X

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its nature and structure*. Tokyo: Tokyo-Kaiseikan Publishing Co.

Clopper, C. G., & Pisoni, D. (2005). Perception of dialect variation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 312–337). Oxford: Blackwell. DOI: https://doi.org/10.1002/9780470757024.ch13

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18*(2), 193. DOI: https://doi.org/10.1017/S0954394506060091

Cole, A., & Evans, B. G. (2020). Phonetic variation and change in the Cockney Diaspora: The role of place, gender, and identity. *Language in Society*, 1–25. DOI: https://doi.org/10.1017/S0047404520000640

Cox, F., & Fletcher, J. (2017). *Australian English pronunciation and transcription*: Cambridge University Press. DOI: https://doi.org/10.1017/9781316995631

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

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory Phonology, 10*, 91–111. DOI: https://doi.org/10.1515/9783110224917.1.91

Docherty, G. J., & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 47–71). London: Arnold.

Docherty, G. J., Foulkes, P., Tillotson, J., & Watt, D. J. L. (2006). On the scope of phonological learning: issues arising from socially structured variation. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology 8: Varieties of phonological competence* (pp. 393–421). Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197211.2.393

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*(7), 473–480. DOI: https://doi.org/10.1111/j.1749-818X.2010.00210.x

Eckert, P. (2008). Variation and the indexical field 1. *Journal of Sociolinguistics, 12*(4), 453–476. DOI: https://doi.org/10.1111/j.1467-9841.2008.00374.x

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America, 140*(1), 576–581. DOI: https://doi.org/10.1121/1.4952387

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America, 139*(1), EL1–EL5. DOI: https://doi.org/10.1121/1.4939608

Foulkes, P., & Docherty, G. J. (Eds.) (1999). *Urban voices: Accent studies in the British Isles*. London: Arnold.

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., & Hay, J. B. (2015). The emergence of sociophonetic structure. In B. WacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (Vol. 87, pp. 292–313). Oxford: Blackwell. DOI: https://doi.org/10.1002/9781118346136.ch13

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences,* 49–54, Saarbrücken, Germany.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*(1), 57–75. DOI: https://doi.org/10.1177/002383098803100103

Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110216042

Haddican, B., Foulkes, P., Hughes, V., & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change, 25*(3), 371–403. DOI: https://doi.org/10.1017/S0954394513000197

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., Maclagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press. DOI: https://doi.org/10.1515/9780748630882

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

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition, 139*, 83–91. DOI: https://doi.org/10.1016/j.cognition.2015.02.012

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

Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp 89–112). Tokyo: Ohmsha; Amserdam: lOS Press.

Hurring, G., Hay, J., Drager, K., Podlubny, R., Manhire, L., & Ellis, A. (2022). Social priming in speech perception: Revisiting kangaroo/kiwi priming in New Zealand English. *Brain Sciences, 12*(6), 684. DOI: https://doi.org/10.3390/brainsci12060684

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

Johnson, K. (2011). *Acoustic and auditory phonetics*. Malden, MA: John Wiley & Sons.

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience, 34*(1), 43–68. DOI: https://doi.org/10.1080/23273798.2018.1500698

Kleinschmidt, D. F., Weatherholtz, K., & Jaeger, T. F. (2018). Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science, 10*(4), 818–834. DOI: https://doi.org/10.1111/tops.12331

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32*(3), 543–562. DOI: https://doi.org/10.1080/03640210802035357

Montgomery, C., & Moore, E. (2018). Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language, 94*(3), 629–661. DOI: https://doi.org/10.1353/lan.2018.0038

Nguyen, N., & Shaw, J. A. (2014). Why the SQUARE vowel is the most variable in Sydney. *15th Australasian International Conference on Speech Science and Technology (SST2014)*, 36–39.

Nguyen, N., Shaw, J. A., Pinkus, R. T., & Best, C. T. (2016). Intergroup dynamics in speech perception: Interaction among experience, attitudes and expectations. *University of Pennsylvania Working Papers in Linguistics, 22*(2), 1–16.

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R., & Best, C. T. (2015). Affective attitudes towards Asians influence perception of Asian-accented vowels. *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow, UK, University of Glasgow.

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

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238. DOI: https://doi.org/10.1016/S0010-0285(03)00006-9

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115–154. DOI: https://doi.org/10.1177/00238309030460020501

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

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America, 89*(6), 2961–2977. DOI: https://doi.org/10.1121/1.400734

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*(1), 221–231. DOI: https://doi.org/10.1016/S0167-6393(02)00105-X

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication, 45*(4), 425–434. DOI: https://doi.org/10.1016/j.specom.2004.12.001

Sebastián-Gallés, N. (2005). Cross-Language Speech Perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Malden, MA: Blackwell. DOI: https://doi.org/10.1111/b.9780631229278.2004.00025.x

Shaw, J. A., Best, C., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., & Mulak, K. E. (2018). Resilience of English vowel perception across regional accent variation. *Laboratory Phonology, 9*(1), 1–36. DOI: https://doi.org/10.5334/labphon.87

Shepard, R. N., Romney, A. K., & Nerlove, S. B. (1972). *Multidimensional scaling: Theory and applications in the behavioral sciences: I. Theory*. New York: Seminar Press.

Sóskuthy, M., Hay, J., Maclagan, M., Drager, K., & Foulkes, P. (2017). Early New Zealand English: The closing diphthongs. In R. Hickey (Ed.), *Listening to the past: Audio records of accents of English* (pp. 529–561). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781107279865.023

Stevens, K. (1998). *Acoustic phonetics*. Cambridge: MIT Press.

Story, B. H. (2005). A parametric model of the vocal tract area function for vowel and consonant simulation. *The Journal of the Acoustical Society of America, 117*(5), 3231–3254. DOI: https://doi.org/10.1121/1.1869752

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

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition, 119*(1), 131–136. DOI: https://doi.org/10.1016/j.cognition.2010.10.018

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

Tollfree, L. (1999). South East London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 163–184). London: Arnold.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21. DOI: https://doi.org/10.1159/000356237

Walker, M., Szakay, A., & Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners? *Laboratory Phonology, 10*(1). DOI: https://doi.org/10.5334/labphon.90

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association, 58*(301), 236–244. DOI: https://doi.org/10.1080/01621459.1963.10500845

Watt, D., & Milroy, L. (1999). Patterns of variation and change in three Newcastle vowels: is this dialect levelling. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 25–46). London: Arnold.

Wells, J. C. (1982). *Accents of English, vol.2: The British Isles*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511611759

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866–1878. DOI: https://doi.org/10.1121/1.390988

Whalen, D. H., Chen, W.-R., Tiede, M. K., & Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics, 68*, 1–14. DOI: https://doi.org/10.1016/j.wocn.2018.01.003

Ying, J., Shaw, J. A., & Best, C. T. (2013). L2 English learners' recognition of words spoken in familiar versus unfamiliar English accents. *Proceedings of Interspeech 2013, 2108–2112*. DOI: https://doi.org/10.21437/Interspeech.2013-500