TRACKING VOWEL CATEGORISATION BEHAVIOUR LONGITUDINALLY: A STUDY ACROSS THREE X THREE YEAR INCREMENTS (2012, 2015, 2018)

Debbie Loakes[1], Paola Escudero[2], Josh Clothier[1] and John Hajek[1]

[1, 3] The University of Melbourne; [2, 3] Western Sydney University,
[3] ARC Centre of Excellence for the Dynamics of Language
dloakes@unimelb.edu.au; paola.escudero@westernsydney.edu.au; johnth; joshuajc@unimelb.edu.au

ABSTRACT

Longitudinal data provide a unique opportunity to address questions around language change, and speaker/listener behaviour. Processing behaviour is considered subject to change over time, but it remains an open question as to over what time period incremental changes might occur. This study compares responses to a forced-choice listening test over three x three-year increments (2012, 2015, 2018), from a set of the same ten mainstream Australian English listeners. The listeners are from a small town (Warrnambool, Australia), where crucially, a distinction between /el/-/æl/ is lost for many. Here we focus on the contrasts between /ɪ e æ/ in /hVt/ and /CVL/ environments. Despite our predictions, overall results show that the increments, which span six years in total, are too small for any changes to arise. This study contributes to our understanding of longitudinal processing behaviour, showing overall consistency across 2012-2018, even in the context of a merger in-progress.

Keywords: perception, longitudinal, vowels, vowel categorisation, Australian English.

1. INTRODUCTION

It is well attested that human speech behaviour within an individual is relatively consistent. Overall consistency is shown, for example, across non-contemporaneous recordings in forensic research, see e.g. the review in [15] which shows relative stability over non-contemporaneous recordings. While this is the case, it is also true that gradual shifts occur in production, which has been shown most famously in the analysis of the Queen's annual Christmas broadcasts, described in [5;6]. In sound change research, it has been noted that "[i]ncremental sound change due to mutual imitation is … predicted by exemplar models of speech … in which phonological categories are modelled as statistical abstractions across remembered exemplars" [5:416]. Such a theory also predicts changes in processing because remembered exemplars should be gradually evolving and updating over time, which can only occur via perception, as listeners interact with different speakers on a daily basis. While processing behaviour is considered subject to incremental change in this way, it remains an open question over what time scales such changes might occur.

2. BACKGROUND

2.1 Processing variability

Processing studies which compare how listeners respond on non-contemporaneous occasions are rare. Where the same listeners are tested more than once, the research tends not to be on "typical" populations, but rather focused on developmental, maturation, and learning-based changes which are likely to occur due to the special conditions of the listener populations [e.g. 13]. One study in which longitudinal listener data were analysed amongst a typical (highly mobile) population was carried out in the UK [4]. This study tested whether people who had moved dialect areas for reasons of study would show production or perception changes over time. Participants were tested at various increments across their university experience (spanning just over two years). Crucially, incremental accent changes were observed in production, but in this particular time period perceptual changes were not observed.

For Australian English, there has been relatively little longitudinal research into processing behaviour. One important study, however, has analysed vowel categorisation behaviour across a 16-year time span (1998 - 2004) [10]. That study focused on *different* listener populations and so was not longitudinal in the same sense as the current study (a panel study, which tracks exactly the same participants over time). However, [10] still showed a shift in perceptual boundaries across this period, linked to diachronic production shifts in the community (see 2.2).

The amount of variation in a listener's environment is known to affect their responses to processing tasks. In noting how "mobile" listeners

react to speech processing tasks, Clopper [2:71] states "individuals with exposure to relatively little sociolinguistic variation … have relatively less variable distributions of phonological … exemplars, whereas individuals with exposure to more sociolinguistic variation … have more variable distributions". In the current study, "non-mobile" listeners from a small town (see details in 3.1), should have relatively invariable distributions with variation largely be limited to what they hear in their immediate community.

2.2 Variability in Australian English

Sound changes have been described as "nature's speech perception experiment" [12]. While Australian English has been described as relatively uniform, there are some important changes occurring in phonological categories and boundaries, meaning that listeners receive variable input; lax front vowels have been raised "to a peak" and have subsequently lowered and retracted in a relatively short time span [3]. Likely connected to this is a regionally defined merger where /el/ -> /æl/ for some [1, 8, 9]. This occurs in the state of Victoria, in the south-east of Australia, and is found in the southern-most communities in that state including the regional town of Warrnambool where the present study is focused [8,9].

Conditions in which where some sounds are in flux and some are merging afford a perfect opportunity to investigate whether incremental changes in perception can be tracked in individuals and their community. That is, listeners encounter wide variation when processing speech from various talkers in the community, hearing some (typically older speakers) who use more compressed vowel spaces and some who merge /el/->/æl/. As will be described in 3.1, the participants are relatively stable monolingual Australian English listeners, so by keeping mobility at a constant, if incremental changes were to be observed across the chosen time points (see 2.3), then we may expect to find them under these conditions.

2.3 Aims and Research Questions

The aim of this research is to analyse longitudinal listening behaviour in a small town in Australia, via a panel study. Reactions to a forced-choice vowel categorisation task over 3 x 3-year increments (2012, 2015, 2018) are analysed. We know that two years is likely insufficient time to show longitudinal changes in processing behaviour [i.e. 4], while 16 years is ample time [e.g. 10, which used different listener populations]. Three- and six-year time spans as studied here should shed further light on when

changes in linguistic processing might occur over time. Research questions are:

1. Are listeners consistent in their responses when performing a listening task more than once?
2. Is there gradual (rather than abrupt) category change in processing behaviour for listeners across the following increments: 2012-2015, 2015-2018 and across the entire time span of 2012-2018?

Predictions are that incremental changes may be evident in the merger context, at least between the first and third exposure to the task which spans six years. In general, the study will contribute to our understanding of longitudinal behaviour in processing, of which very little is known, in a community where sound change is expected.

3. METHOD

3.1. Listeners

Participants are ten monolingual Australian English listeners who at the time of the first data collection (2012) had lived all of their lives in Warrnambool (approximately 3 hours from the state capital, Melbourne, by road). Warrnambool is a small community of approximately 35,000 people in the south-west of the state of Victoria. In 2012, 15 listeners took part in the study, but only 10 took part in 2015. This same set of ten participants also took part in the third data collection in 2018. Information about the participants is shown in Table 1 below which gives the participant's code, shows whether they were male or female, and also shows their age at the first participation in 2012, as well as their age category ("Y" for younger and "O" for older).

Table 1: participant information (ID, Sex, Age).

Participant ID	Sex	Age category and age in 2012
WN01_2012	F	Y (29)
WN04_2012	M	Y (23)
WN08_2012	F	Y (28)
WN09_2012	M	Y (31)
WN11_2012	F	O (65)
WN12_2012	M	O (67)
WN13_2012	F	O (56)
WN14_2012	M	Y (36)
WN15_2012	M	O (73)
WN16_2012	F	O (75)

A natural split is observable in age, with the young listeners aged 23-36 in 2012, and older listeners

aged between 56-75. There is also an even split between male and female listeners. Of these ten listeners, only one (the youngest, WN04) later moved away from Warrnambool, but remained in the study. By 2015 he had moved to a town an hour and 20 minutes away by road for work (Portland), and in 2018 he lived in a Melbourne suburb (Point Cook). He was not excluded from the study because merger occurs in all these regions, but he has been exposed to more variability than other participants.

3.2. Task

Listeners took part in a categorisation task on three occasions separated by three years. The vowel categorisation task is a forced-choice identification task, presented on an iPad using a custom app designed specifically by the first author [e.g. 8]. The items were created using real word tokens produced by a 40-year-old female speaker from north-east Australia for whom no merger is present. The contrast pairs consisted of real words, and equidistant continua were produced using Akustyk in Praat [8]. The items analysed in this study are designed to focus on the highly variable and changing /ɪ/-/e/ and /e/-/æ/ contrasts [e.g. 3,10], in various phonetic contexts. Here we analyse *hill-hell*, *hit-het, mill-Mel*, and *hell-Hal, het-hat* and *Mel-Mal*.

Using *hill-hell* as an example, we took a naturally produced *hill* as Step 1, a naturally produced *hell* as Step 7, and five synthetic tokens were created in between (equidistant in F1, F2, F3), resulting in a 7-step continuum. At random, listeners heard each item four times, and the orthographic representations were presented twice on the left, and twice on the right. Here, we present on 168 trials per listener, per year. All listeners completed all of the trials, giving 504 responses for each individual. The experiment was conducted under fieldwork conditions; across all trials, listeners were visited at home by the first author. Shure SRH840 Reference Studio headphones were used, and experimental volume was kept constant. Hearing difficulties were not reported, but this was not independently tested.

4. RESULTS

4.1. Community (group) results

Response curves for each continuum, fit using logistic functions in the *quickpsy* [7] package for R [13], are shown for /ɪ/-/e/ in Figure 1, and /e/-/æ/ in Figure 2. Years 2012, 2015, 2018 are all overlaid on the same image for ease of comparison. In the figures below, the stimuli are arranged from left-to-right for three phonological conditions, i.e. in Figure 1 for *hill-hell*, *hit-het*, and *mill-Mel*. The group

proportion of responses is on the y-axis, and the steps for each of the three curves are shown on the x-axis. Curves indicate responses to the left-most item; Step 1 is always the left-most item in the contrast pair, while Step 7 is right-most (Steps 2-6 are the synthesised tokens). For example, for the *hill-hell* curve, 100% of listeners responded that they heard hill at Step 1, and this gradually declines until the crossover (around Step 5), where a majority of response are now for *hell*. By Step 7, almost, but not all, listeners heard *hell*. Comparing curves on the same plots, the crossbars represent 95% CIs and are used to assess differences between groups (in this case, years). Where the crossbars touch, there are no statistically significant differences between them. If they do not touch, statistically significant differences exist. While the images are relatively small, Figure 1 shows that the main finding is in fact similarity across the years, with no significant differences in any crossovers for /ɪ/-/e/.

Figure 1: Response curves for the /ɪ/-/e/ contrast across 2012 (red), 2015 (green), 2018 (blue).
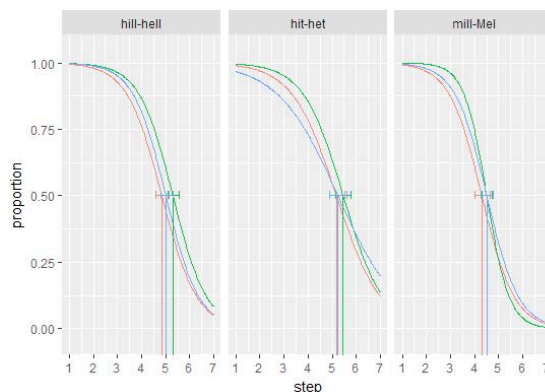

Comparing across the three time points, the ten listeners respond in essentially the same way in 2012, in 2015, and in 2018. For each contrast pair, the majority of listeners in the majority of cases chose the intended response (most Step 1 responses have 100% agreement). However, there is some disagreement for *hell* and *het* at Step 7. *Mel* on the other hand, is almost always heard as *Mel*.

Figure 2 shows a similar pattern of behaviour, with no significant differences in the *hell-Hal* or *Mel-Mal* contexts. However, somewhat surprisingly, the *het-hat* contrast has a significant difference for the middle year (2015), with the curve shifted significantly to the right (overall, more *het* responses) whereas 2012 and 2018 pattern together and are not significantly different. This warrants further analysis and will be addressed in Section 4.2. These particular items were designed to test listener responses to the merger, and merger behaviour is clearly evident when we focus on the *hell-Hal* curve, with listeners not reaching 100% agreement for Step

1, and only around 75% at Step 7 (i.e. a quarter of responses were *Hal* when *hell* was presented). There is also variability in the crossover points for *hell-Hal*, but this does not reach significance. Results for *Mel-Mal* show 25% "incorrect" responses at either end, indicating this contrast presents greater confusability. Remarkably though, listeners are extremely consistent across all years for this pair.

Figure 2: Response curves for the /e/-/æ/ contrast across 2012 (red), 2015 (green), 2018 (blue).
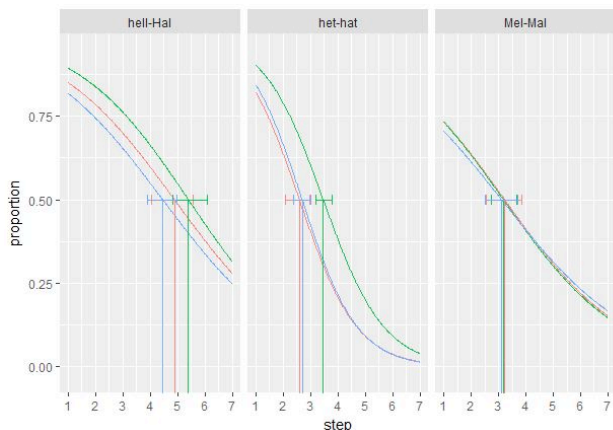
4.2. Individual differences

It is useful to look to individual listener behaviour to explain the 2015 *het-hat* result in Figure 2. The 30 bar graphs below (Fig. 3) show results for all ten individuals on the x-axis, with proportions of responses to *het-hat* in individual bars on the y-axis (where the bar is one colour, the listener always chose the same item). For each year and listener there are three bar graph images (2012, 2015, 2018).

Figure 3: Bar graphs for individual listeners: het (aqua) hat (red) in 2012, 2015, 2018.
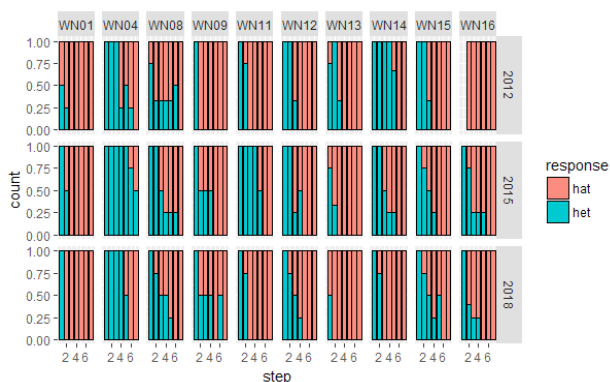
It is clear from this figure that WN04 and WN11 act differently in 2015, whereas the other listeners are effectively consistent. In 2012 and 2018, WN04 has a majority of *hat* responses by Step 7, whereas in 2015 he had a majority of *het* responses. WN04 was the youngest listener, and the only one that could

have been considered in any way "mobile", which this may well explain this result. WN11, who is an older listener, had more *het* responses in 2015, but still clearly switches to *hat* by Step 6. This particular individual reported working as a volunteer in a tourist organisation, and so is exposed to multiple talkers from various backgrounds, which may explain the change from 2012 to 2015 for this listener [2]. There are some other minor variations for individuals (WN09, WN14, WN16) but they are limited, and in differing directions, not affecting overall group behaviour, as seen in Figure 2.

5. DISCUSSION AND CONCLUSION

The study showed that listeners are, for the most part, consistent when performing a listening task more than once (RQ1). With respect to whether there are observable category shifts over time in processing behaviour (RQ2), no differences were observed. Contrary to the prediction of at least gradual change over time for the merger context, the present study showed that no cross-group changes in response patterns across increments of three or six years; these primarily "non-mobile" listeners in fact behave remarkably similarly when repeating the same task. As shown in [10] a 16-year time span can be enough to show category shifts. In this current study, small individual differences were observed and discussed (4.2) but these are not consistent. It is, of course, possible that clear changes in behaviour might become evident in another three or six years.

Crucially, the study shows that listener behaviour is more consistent than inconsistent; when listeners are exposed to a task once, they tend to act similarly when exposed again (and again), even with a merger-in-progress and variable vowel productions in their community. While three years is a relatively long time between the repeated exposures to the listening experiment, it is also possible that individuals learnt to perform the task in one way in 2012, and then used this same strategy at later times.

While these findings are useful in and of themselves for understanding processing behaviour, further research will also compare results with those from a new sample of listeners from the same region who had never been exposed to the task previously (collected in 2018). This will be an important comparison, to understand whether there is indeed consistency in processing behaviour in the longitudinal data. Future research will also analyse response times. We might predict faster responses in 2015 and 2018, although perhaps three-year increments might mean older participants slow down in their responses, especially given the oldest participant was 81 in 2018.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Bradley, D. 2004. Regional Characteristics of Australian English: Phonology. In Burridge, K. Kortmann, B. Eds. A *Handbook of Varieties of English. Vol. 1: Phonology*. Berlin: Mouton de Gruyter, 645–655.

[2] Clopper, C. G. 2014. Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*. 5, 69-90.

[3] Cox, F., Palethorpe, S. 2008. Reversal of Short Front Vowel Raising in Australian English. In Janet Fletcher, Deborah Loakes, Roland Göcke, Denis Burnham and Michael Wagner (eds.) *Proceedings of Interspeech 2008 incorporating ASSTA 2008*. Bonn: ISCA, 342- 345.

[4] Evans, B.G., Iverson, P. 2007. Plasticity in vowel perception and production: a study of accent change in young adults. *J. Acoust Soc Am*. 121, 3814-3826.

[5] Harrington, J., Schiel, F. 2017. /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language* 93.2, 414-445.

[6] Harrington, J., Palethorpe, S., Watson, C. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas Broadcasts. *JIPA* 30. 63-78

[7] Linares, D., López i Moliner, J., 2016. quickpsy: An R package to fit psychometric functions for multiple groups, *The R Journal*, 8, p. 122-131.

[8] Loakes, D., Clothier, J. Hajek, J., Fletcher, J. 2014. An Investigation of the /el/-/æl/ Merger in Australian English: A Pilot Study on Production and Perception in South-West Victoria. *AJL* 34, 436–452.

[9] Loakes, D., Fletcher, J. Hajek. J. 2017. Can you t[æ]ll I'm from M[æ]lbourne? Merger of the DRESS and TRAP vowels before /l/ as a regional accent marker in Australian English. *EWW*. 38 29-49.

[10] Mannell, R. 2004. Perceptual vowel space for Australian English lax vowels: 1988 and 2004. *Proc. SST2004*. Sydney, 221-226.

[11] Microsoft, R Core Team, 2018. *Microsoft R Open*. 3.4.4 ed.

[12] Ohala, J. J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13, 155-161.

[13] Peterson, N.R., Pisoni, D.B., Miyamoto, R.T. 2010. Cochlear implants and spoken language processing abilities: review and assessment of the literature. *Restor Neurol Neurosci*. 28, 237-250.

[14] RStudio Team, 2016. *RStudio: Integrated Development Environment for R*. 1.1.447 ed. Boston, MA: RStudio, Inc.

[15] Rose. P. 2002. *Forensic Speaker Identification*. London, New York: Taylor & Francis.