Phonologisation of vowel duration and nasalised /æ/ in Australian English

Felicity Cox[1,3,4] *and Sallyanne Palethorpe*[2,3,4]

[1] Department of Linguistics, [2]Department of Cognitive Science, [3]Centre for Language Sciences, Macquarie University, Australia
[4]ARC Centre of Excellence in Cognition and its Disorders

felicity.cox@mq.edu.au, Sallyanne.palethorpe@mq.edu.au

Abstract

An allophonic split in height between oral and nasalised /æ/ is an ongoing sound change in Australian English. Speakers participating in this change produce phonetically raised [æ̃] that overlaps the F1/F2 /e/ space, achieving [æ̃]/[ẽ] contrast through duration. We tested listeners' sensitivity to this production change using forced-choice identification. Listeners responded to long and short synthetic /bVn/ and /bVd/ tokens constructed to simulate variation from /æ/ to /e/. Oral vowels were primarily identified according to F1 whereas listeners relied on length for nasalised vowels. This finding confirms the primacy of duration in cueing [æ̃]/[ẽ] contrast and indicates phonologisation of length.

Index Terms: Australian English, vowels, duration, speech perception, sound change

1. Introduction

Previous work on Australian English (AusE) has revealed substantial synchronic variability associated with the vowel /æ/ [1]. This variability has two sources related to sound change; phonetic lowering of the oral allophone [1] and, for some speakers, phonetic raising of the nasal allophone [2] creating an extensive allophonic split in the production of nasalised and oral /æ/ [3]. Hyman [4] refers to the development of language specific allophony as phonologisation. Phonologisation occurs when a low level 'automatic' phonetic effect becomes exaggerated to the extent that it takes on a contrastive role (which may have been previously reserved for another phonetic feature). That is, phonologisation occurs when a phonetic process becomes phonological [4]. When the integrity of a contrast is threatened, this can disrupt the existing weighting of phonetic cues and act as the catalyst for phonologisation. In AusE, acoustic studies have confirmed that the phonetic height of /æ/ is extremely variable, particularly in nasal contexts [3]. Cox and Palethorpe propose, in line with the Probabilistic Enhancement Hypothesis [5] that, in the absence of the more usual F1/F2 difference, those speakers who produce raised nasalised /æ/ enhance vowel duration as a mechanism for maintaining contrast with /e/. They propose that duration has been phonologised to become the primary contrastive cue to distinguish [æ̃] and [ẽ]. However, any proposal for considering such enhancement to be a phonologising change must be based not only on production but on perception as well.

In this paper we examine listeners' responses to a set of stimuli designed to replicate the variability associated with changes to /æ/. The aim of this perceptual experiment is to help us understand how spectral and durational cues are variously weighted according to coda nasality in the differentiation of the short front vowels in AusE.

Establishing listener responses to a set of carefully controlled stimuli will provide a baseline for discussing the relationship between perception and production in this sound change. It may also help to inform the debate about functional (speaker-based) vs. non-functional (listener-based) approaches to understanding change.

1.1. Australian English intrinsic vowel duration

The AusE vowel system comprises 18 stressed vowels including twelve monophthongs (six short and six long) [2]. The six short vowels are approximately 60% the length of long vowels in citation or phrase final context ([6], [7]) with /æ/ being the longest of the inherently short vowels ([6], [8], [9]). Bernard [8] proposes that the increased length of /æ/ is licensed because its spectral separation from other long vowels reduces the chance that it might compromise existing contrasts. One interesting feature of AusE /æ/ is its participation in the bad/lad split which is a lexical effect found in some varieties of English where certain adjectives (e.g. *sad, mad, glad, bad*) are produced with long /æ/ ([10], [11] and references therein). Long /æ/ also precedes tautosyllabic /m/ *ham* and /n/ *pan* (but not /ŋ/) and /g/ *bag*.

AusE vowels are primarily differentiated spectrally [2] although there are at least two vowel pairs for which duration is the primary contrastive cue. Cox [6] and Bernard [8] have shown that the central open vowels /ɐː/ and /ɐ/ (as in *heart* and *hut*) and the half close front vowels /eː/ and /e/ (as in *shared* and *shed*) have minimal spectral differences in the F1/F2 plane. Evidence for the primacy of duration in these contrasts comes from various sources including Watson and Harrington [12] who in Gaussian classification experiments established that certain vowels were best classified when durational data were included. Early work by Bernard [8] in a study where vowel duration was manipulated, also confirmed the importance of this phonetic feature in listener differentiation of /ɐː/ and /ɐ/. AusE speakers are therefore attuned to duration as a cue to vowel identification [7]. Recent work on acquisition has found that AusE-speaking children as young as 18 months [13] and 3 years [7] have the ability to successfully use duration in their speech production and perception.

1.2. Australian English /æ/ change

Many dialects of English, particularly in North America, exhibit allophonic raising and fronting of /æ/ in nasal contexts ([14], [15]). In AusE, the nasalisation of /æ/ and the resulting raised quality of the allophone has been the subject of comment for over 100 years. McBurney [16] describes the vowel in *dance* and *hand* as frequently nasalised and raised and Mitchell and Delbridge [17] make the same observation. In historical data from speakers born in the late part of the 19th century, the raising of nasalised /æ/ was found to be present but not excessive [18].

Recent change in AusE has seen substantial lowering of the oral allophone of /æ/ so that it now occupies an extreme open position in the vowel space in non-lateral contexts [2]. This change has resulted in a reconfiguration of the AusE vowel space (see Figure 1 from [3]). The newly initiated extreme allophonic split that results in the acoustic overlap between nasalised /æ/ and /e/ is interesting in that contrast in production is maintained by duration rather than F1/F2 for certain speakers [3]. It has also been found that speakers who have the greatest acoustic spectral overlap between /e/ and /æ/ appear to enhance duration as a contrast maintenance strategy [3]. These findings support Kirby's Probabilistic Enhancement hypothesis [5] which predicts that redundant cues (in this case, duration) can be redeployed to preserve a highly functional contrast when primary contrastive cues (e.g. F1 & F2) have been reduced through sound change. The level of enhancement is a function of the degree to which contrast precision is compromised. Production data provide compelling evidence that length has been phonologised. However, listener sensitivity to this effect has yet to be established.

In this paper we will examine listeners' use of spectral and/or durational data to identify /æ/ in oral and nasal contexts. Based on previous work on production [3], we hypothesise for listener perception that spectral cues will take primacy in oral contexts but that temporal cues will take primacy in nasal contexts. Previous work on perception [19] found that AusE listeners did not use duration to distinguish American English oral /æ/ and /ɛ/. We might therefore also expect perception of oral AusE vowels to rely less on duration than spectral information.

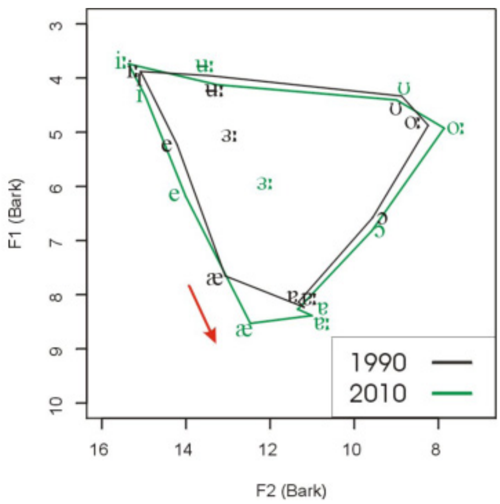

Figure 1: *Comparison between AusE-speaking young females' oral context monophthongs - 1990 to 2010.*

2. Method

2.1. Stimuli

Stimuli were created using the HLSyn (High Level Parameter Speech Synthesis System [20]). Eight nine-step synthetic continua were created, each of which contained vowels constructed with a stable F0 of 105 Hz and F1 varying from 400 to 800 Hz in equal 0.40 Bark-size steps (400 444 489 536 584 635 688 743 800). Four of the continua contained /bVd/ with oral vowels and four contained /bVn/ with nasal vowels. For each of the nasal and oral contexts two different F2 values (1900Hz or 2100Hz) and two different vowel durations from onset to closure (short – 180 msec or long – 300 msec) were used. The eight conditions are summarised as follows:

- 180msec, 1900Hz F2, oral context and nasal context
- 300msec, 1900Hz F2, oral context and nasal context
- 180msec, 2100Hz F2, oral context and nasal context
- 300msec, 2100Hz F2, oral context and nasal context

The full set of stimuli from each condition were randomised and presented in 5 blocks divided for presentation purposes into two halves. There were 72 items in each block: two nasality conditions (nasal/oral) X two length conditions (180msec/300msec) X two F2 conditions (1800Hz/2100Hz) X nine F1 steps resulting in a total of 360 items. Listeners were engaged in a forced-choice closed set identification task and were instructed to select one of the following words from a list: BAD, BED, BAN, BEN upon auditory presentation of a synthetic token. Tokens were presented in a sound treated room over high quality headphones at 80 dB SPL. Each item was preceded by a recorded identifying number produced by a natural-voice followed by a 1500msec interval then the test token presentation followed by a 2000msec interval.

We selected *bad/ban* vs *bed/ben* as our test words specifically to control for the duration of oral /æ/ which in *bad* is long as discussed in Section 1.1. Had we selected *pad/pan,* for example, a confound with length would occur in that *pad* contains a short vowel whereas *pan* contains a long vowel.

2.2. Participants

18 female and 3 male AusE-speaking university students participated in the perception task. 19 participants were between 18 and 23 years of age and two female participants were in their 40s. All were from NSW.

2.3. Analysis

Logistic function (i.e. a sigmoid curve) was fitted to the identification response data for individual participants to estimate the F1 categorical boundary at the 50% crossover point of perceived /æ/ in each condition. There were 168 estimated category boundary values obtained (8 conditions X 21 listeners). Eleven outliers were returned and removed. The outliers occurred in cases where categorical boundaries could not be established such as when listeners responded with a single category across the F1 continuum (4 cases), or responded with a single category interspersed with isolated random responses (7 cases: three for long nasal, three for short nasal and one for short oral contexts).

The remaining 157 estimated 50% boundary values were used as the dependent variable in a multilevel modeling analysis in SPSS. Nasality (oral vs. nasal), length (180msec vs 300 msec) and F2 (1900Hz vs 2100Hz) were included as fixed factors and listener was included as a random factor.

3. Results

The results show significant main effects for both nasality [$F(1,149)=19.592$, $p<0.0001$] and length [$F(1,149)=161.611$, $p<0.0001$] and a significant interaction between nasality and length [$F(1,149)=53.213$, $p<0.0001$]. There was no effect for

F2 or listener. In the oral case (i.e. *bad/bed*), the mean 50% estimated F1 category boundary was at 653.05 Hz (sd 66.02) for short vowels and 577.171 Hz (sd 40.19) for long vowels. In the nasal case (i.e. *ban/ben*) the category boundary was 693.03Hz (sd 121.7) for short vowels and 412.607 Hz (sd 101.96) for long vowels. Figure 2 illustrates the interaction showing extreme difference in the 50% cross-over values for long and short vowels in the nasal context compared to the oral context.

Figures 3 and 4 show the proportion of listeners who responded with either /æ/ or /e/ at each value of F1 in the oral and nasal contexts across vowel length conditions. In the oral context (Figure 3) the F1 boundary between /æ/ and /e/ was higher for short vowel tokens compared to long vowel tokens. This indicates that listeners interpret a greater proportion of vowels as /e/ when the vowel is short and that a short vowel requires a high F1 to be perceived as /æ/. Even in high F1 cases there was not a robust response to /æ/ with identification scores failing to reach 90% even at the highest F1 value. For long vowels there was a more even distribution of /æ/ and /e/ responses.

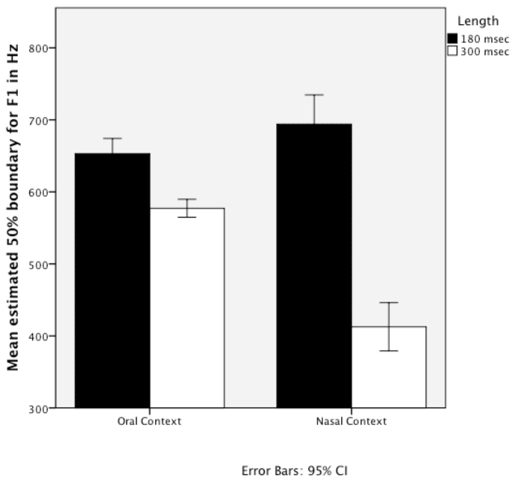

Figure 2: *Mean estimated 50% categorical boundary for oral and nasal contexts according to vowel length.*

In the nasal context (Figure 4) the short vowel condition produced a greater proportion of responses for /e/ than for /æ/ with the 50% cross-over point occurring for tokens with very high F1 (that is, phonetically low tokens). F1 values beyond the 50% cross-over point yielded responses near chance level. This result shows that robust identification only occurred for /e/ indicating that, regardless of F1, listeners were unlikely to identify the short nasalised vowel as /æ/.

In the long nasalised vowel case, contrary to the short vowel case, there were no reliable responses to /e/, showing that listeners are highly likely to interpret a long nasalised vowel as /æ/. For tokens with very low F1 values (that is phonetically high vowels) the response was at chance level.

4. Discussion

The results of the forced-choice identification tasks show that listeners are highly sensitive to F1 in the oral context when making decisions about vowel category but not so sensitive to this cue in the nasal context. For the oral stimuli we observed the expected continuous classification of vowels with boundaries mediated by F1. There was also a length effect for the oral context forcing a boundary shift in the anticipated direction. That is, short vowels were more likely to be perceived as /e/. As /æ/ is the longest of the inherently short vowels in AusE, and the word we chose to represent oral /æ/ *bad* contained a lexically specified long vowel, this length effect is not unexpected. The shift in boundary position according to length was small in the oral context compared to the nasal context. Our results confirm that listeners are highly sensitive to duration in the nasal context and that F1 is largely ignored. Listeners perceived long nasalised vowels as /æ/ and short nasalised vowels as /e/. When F1 was incompatible with listeners' expectations they behaved randomly. For example, listeners were unable to reliably identify a long nasalised vowel as either /e/ or /æ/ if the F1 was very low even though such an F1 value would be more indicative of /e/. Similarly if the F1 was very high, as would be expected for /æ/, a short nasalised vowel could not be reliably identified. Conflicting spectral and durational signals could not be resolved in the nasal context whereas in the oral context duration did not have a detrimental impact on identification.

Our results support phonologisation of length for nasalised vowels suggesting that a once phonetic effect (the increased length of a low vowel in a nasal context) has become phonologised to the extent that it is now the primary contrastive cue for the [æ̃] vs [ẽ] distinction.

These results provide a baseline for future work on sound change. An obvious extension to this work is to examine both

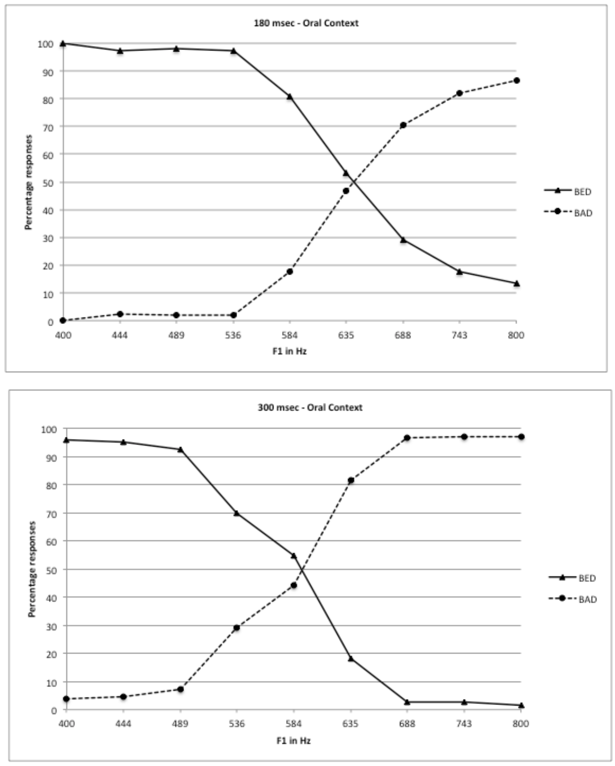

Figure 3: *Percentage of responses to 'bed' and 'bad' in the oral context condition. Top panel is the short vowel (180 msec) context and bottom panel is the long vowel (300 msec) context.*

production and perception of the same participants to determine the extent to which production mediates perception. According to non-functional approaches to sound change, listeners are considered the agents of change through a process of ‘innocent misapprehension’ (see [21], [22]). We might therefore expect listeners of different ages, who have varying degrees of experience with this sound change, to behave differently from each other. As this study focused on a sound change in progress, it is conceivable that older listeners may have responded differently to younger listeners along the lines of [22]. From a functional perspective Kirby’s Probabilistic Enhancement Hypothesis [5] prioritises the role of the speaker in sound change and considers phonologisation to result from adaptive cue enhancement to preserve contrast. A program of research to explore both perception and production is required to tease out the important mechanisms by which sound change occurs and the various roles played by speaker and listener. AusE sound change provides an exciting opportunity to explore these questions.

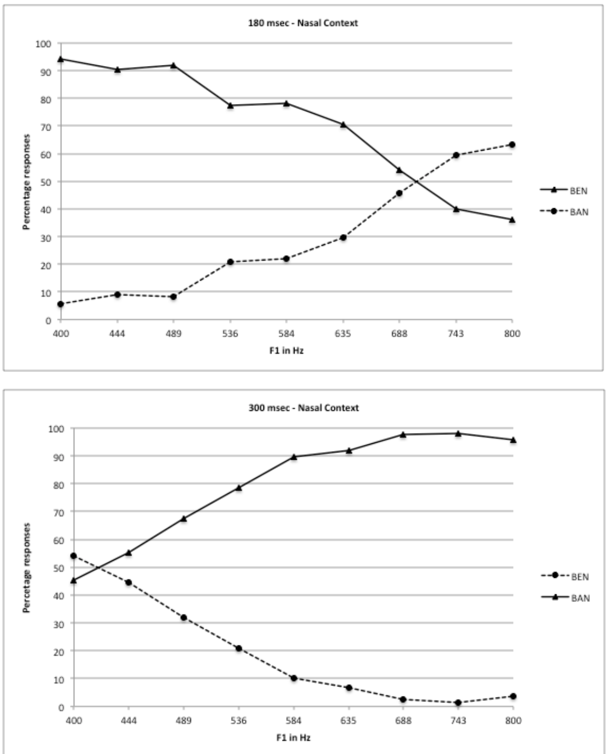

Figure 4: *Percentage of responses to ‘ben’ and ‘ban’ in the nasal context condition. Top panel is the short vowel (180 msec) context and bottom panel is the long vowel (300 msec) context.*

5. Conclusions

We have demonstrated that listeners behave differently when responding to oral compared to nasal stimuli in the identification of /e/ and /æ/ in AusE showing that F1 takes primacy in oral contexts whereas duration is the more important cue in nasal contexts. Our results support production studies that indicate an allophonic split between the nasalised and oral allophones of /æ/ and confirm phonologisation of length as a structural change for these vowels in AusE.

6. Acknowledgements

We would like to thank Jonathan Harrington and Ivan Yuen for insightful and helpful advice and Ulrich Reubold for creating the synthetic stimuli. This work was supported by MQRDG 9200900671 and ARC DP110102479.

7. References

[1] Cox, F. and Palethorpe, S., “Reversal of short front vowel raising in Australian English”, Proc Interspeech, Brisbane, 342-345, 2008.
[2] Cox, F. and Palethorpe, S., “Illustrations of the IPA: Australian English”, JIPA, 37: 341-350, 2007.
[3] Cox, F. and Palethorpe, S., “Probablilistic Enhancement and Australian English /æ/”, SCIHS, U. California, Berkeley, 2014.
[4] Hyman, L., Phonology: Theory and Analysis, Holt, Rinehart & Winston, New York, 1975.
[5] Kirby, J., “The role of probablilistic enhancement in phonologisation”, In Yu, A. (ed.) Origins of Sound Change: Approaches to Phonologisation, OUP, Oxford, 228-246, 2013.
[6] Cox, F., “The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers”, AJL, 26:147-179, 2006.
[7] Yuen, I., Cox, F. and Demuth, K., “Three-year-olds’ production of Australian English phonemic vowel length as a function of prosodic context”, JASA, 135: 1469-1479, 2014.
[8] Bernard, J., “Length and identification of Australian English vowels,” AUMLA 27: 37-58, 1967.
[9] Fletcher, J. and McVeigh, A., “Segment and syllable duration in Australian English”, Speech Commun. 13: 355-365, 1993.
[10] Wells, J., Accents of English: Beyond the British Isles, CUP, Cambridge, 1982.
[11] Durie, M. and Hajek, J., “Getting it right: More on an orthography for Australian English”, AJL, 15: 227-239, 1995.
[12] Watson, C. and Harrington, J., "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am., 106: 458-468, 1999.
[13] Chen H., Xu Rattanasone, N. and Cox, F., “Perception and production of phonemic vowel length in Australian English-learning 18 month-olds”, LabPhon, Tokyo, 2014.
[14] Labov, W., Principles of Linguistic Change: Internal Factors. Basil Blackwell, Oxford, 1994.
[15] Eckert, P. “California Vowels” accessed 1st June 2014. http://www.stanford.edu/~eckert/vowels.html
[16] McBurney, S., “Colonial Pronunciation”, In Ellis, A.J. (1889) On Early English Pronunciation. Part V: The Existing Dialectal as compared with West Saxon Pronunciation, Greenwood press, New York, 237-248, 1887.
[17] Mitchell, A. and Delbridge, A., The Speech of Australian Adolescents, Angus & Robertson, Sydney, 1965.
[18] Cox, F., Palethorpe, S. and Tsukada, K., “One hundred years of accent change in Australian English”. 10th SST, Sydney, 2004.
[19] Cutler A., Smits, R. and Cooper, N., “Vowel perception: Effects of non-native language vs. non-native dialect” Speech Commun. 47: 32–42, 2005.
[20] High Level Parameter Speech Synthesis System. http://sens.com/hlsyn/
[21] Ohala, J. Coarticulation and Phonology, Language and Speech, 36:155-170, 1993
[22] Harrington, J., Kleber, F., and Reubold, U., “Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study,” J. Acoust. Soc. Am., 123: 2825–2835, 2008.