OOPS, I'VE DONE A FUTT:

QUALITY AND QUANTITY IN A NEW ZEALAND VOWEL CONTRAST

Paul Warren: *School of Linguistics and Applied Language Studies, Victoria University of Wellington, P O Box 600, Wellington < paul.warren@vuw.ac.nz>*

Abstract

New Zealand English (NZE) is one of a number of varieties of English that exhibit a qualitative merger of the START and STRUT[1] vowels. Such a qualitative merger might lead to an expectation that quantitative differences are crucial to the identification of stimuli containing these vowels. To the extent that the qualitative merger is variably complete for different speaker groups, we might also expect listeners' sensitivity to quantitative (as opposed to qualititative) distinctions to depend on characteristics of the speaker, as well as on their own speaker-group membership. These questions are addressed through a lexical decision experiment, which confirms the importance of the quantity distinction of these vowels in NZE, and indicates that there is some sensitivity in participants' responses to social variables. Importantly, the results suggest that this sensitivity is located within the lexicon, in a way that would seem to support an exemplar-based approach to word recognition.

1. Introduction

The first part of the title of this paper derives from an exchange heard between two brothers. The younger boy (born in New Zealand) produced this statement on passing wind, whereupon the older (born in England, and still employing largely a Southern British vowel system) exclaimed that there was no such thing as a *futt*, although from the interaction it was clear that he knew what his brother intended. The point being illustrated here is that odd outcomes can arise when two vowel systems are confused. That is, for most speakers of New Zealand English (NZE) there is a qualitative merger of the START and STRUT vowels, both are fronted and START is raised relative to Southern British English, while the latter variety maintains both a qualitative and a quantitative distinction. If a speaker of Southern British English relies primarily on the qualitative distinction, then a START vowel might well be confused with a STRUT vowel. A further factor here might arise if the younger child is still establishing a length difference in his production of the two vowels.

While a dialectal difference between quality- and quantity-based distinctions for the same vowel contrast may result in misunderstanding, it is probably a rare occurrence, the likelihood of ambiguity or miscommunication being mitigated by contextual constraints on interpretation. In the current paper, however, the focus is on the importance of the length distinction in the NZE START-STRUT contrast. Specifically, this paper presents the results of a lexical decision experiment using materials containing vowels in the START-STRUT formant space but with manipulated lengths. The analysis demonstrates that for NZE participants not only does changing the duration of a vowel result in nonwords being perceived as words and vice versa, but also responses to START-STRUT stimuli are affected by demographic variables which might be expected to influence sound change in language, such as sex, age and socio-economic status.

2. Quality vs. quantity in the START-STRUT contrast in English

The qualititative overlap of START and STRUT in Australasian varieties of English has been noted for some time (see for instance Bernard 1967, for Australian English). This is reflected in common symbols for the two vowels in a range of transcriptions. Thus for Australian English (AusE) Clark (1989: 209) suggests [ɐː] and [ɐ], and refers to both vowels as 'very stable' (see also Horvath 2004). Clark notes that [ɑ:] and [ɑ] may also be found, resulting from 'co-articulation with a following /l/', and he comments further that [ʌ] may be found for STRUT, but is 'far more likely in New Zealand.' Watson et al. (1998: 203) suggest the transcription symbol [ɐ] for the quality of the two vowels in both AusE and NZE, so that START is [ɐː] and STRUT is [ɐ]. This transcription is also recommended in their description of NZE by Bauer and Warren (2004), who comment (588) that '[i]f vowels are to be paired in terms of length/tension, then in New Zealand English the STRUT vowel should be paired with the START vowel, with which it is virtually identical in terms of formant structure, resulting in a distinction primarily of length between *cut* and *cart*.' The consensus in these transcription systems appears to be both vowels are fronted in NZE and AusE when compared with RP. Although Watson et al. claim that in the two varieties these vowels 'are positioned in very similar places in the F1/F2 plane' (1998: 192), Easton and Bauer (2000: 104) have argued that they are both fronter in NZE than in AusE. The fronting in particular of START in NZE is documented in quite early comments on the variety (see discussion in Gordon et al. 2004).

A search through the descriptions of language varieties in the *Mouton Handbook of Varieties of English* (ed. Kortmann, Schneider, Burridge, Mesthrie and Upton 2004) indicates that START and STRUT have qualitative overlap in a number of other varieties, though the variability in the level of transcription in the contributions to that volume makes detailed comparison difficult - it is for instance not always clear whether a front or central quality is intended with an [a] transcription, particularly for contact varieties with only one open vowel.

Since many of the languages described are contact varieties, much of the overlap may derive from the vowel systems of the other substrate languages involved. In the Pacific region we find that in Fijian English both START and STRUT approximate to [a] (Tent and Mugler 2004: 759), with Indo-Fijian speakers realising STRUT as 'a retracted [a], often extra short …' (op. cit., 769). There is both qualitative and quantitative overlap in Australian Creoles (Malcolm 2004: 659) and in Hawai'ian Creole (Sakoda and Siegel 2004: 740), though the latter maintains a possible qualitative distinction, in that while it has [ɑ] for START, both [ʌ] and [ɑ] are found in free variation for STRUT. Crowley (2004: 676-7) similarly indicates that there is no length distinction in Bislama, with [a] for both vowels. For Solomon Islands Pijin (Jourdan and Selbach 2004: 694-5) an [a] transcription is also suggested, and the length difference is noted as inconsistent.

In South East Asia, there is overlap predominantly on a back form [ɑ] in both Singapore English (Wee 2004) and Malaysian English (Baskaran 2004), in both cases presumably as a result of the influence of substrate languages, and in both cases with no or at least inconsistent length differences.

Elsewhere, we find a range of African varieties of English that appear to have a qualitative overlap, usually on a central open variant, and often again with overlap in quantity: Ghanaian (Huber 2004a; including Ghanaian Pidgin: Huber 2004b), East African (Schmied 2004) and Black South African English (van Rooy 2004). In addition, Cameroon English (Bobda 2004), Cameroon Pidgin English (Menang 2004), and Cape Flats English (Finn 2004) share a feature also found in the English West Midlands (Clark 2004), namely a merger with START for a restricted lexical set of STRUT words including the word *one* (for which an [a] transcription is indicated for the African varieties mentioned and an [ɒ] form for the English West Midlands). Finally, Penhallurick (2004: 103) notes (after Parry 1999) that in some Welsh English varieties around the north and mid Wales border with England, STRUT 'strays into [a] territory'.

In summary, the overlap of START and STRUT is attested in a range of English varieties. The suggested transcriptions appear to indicate that there is no overall preference for back, central or front realisations. Since in many cases – particularly with the contact varieties – there is both qualitative and quantitative merger, AusE and NZE would seem to present a special case, with these vowels forming a tense-lax pair. However, recent research by Torgersen et al. (2006) suggests that such a pairing is now evident also in London varieties of English. In the case of London English, however, this results from backing of STRUT to the vowel-space of START, rather than the START-fronting found in NZE and AusE. Torgersen et al. also indicate a role of ethnicity in STRUT-backing in London, with speakers of Afro-Carribean origin seeming to lead the change.

3. Start and strut in NZE

The above survey shows that while the START-STRUT qualitative overlap is not unique to NZE, the extent and direction of the overlap differs from one variety of English to another. For NZE, the overlap seems to be pretty well established on a fronted vowel space. That it is an established overlap appears to be supported by the absence of any clear claims in the literature that it is more or less advanced for any particular social group. Indeed, an analysis of word-list /hVd/[2] data from the New Zealand Spoken English Database (NZSED; see Warren 2002) suggests that on the whole the two vowels show qualitative overlap for speakers of both sexes and all ages. However, detailed analysis of this small sample suggests a recent change that may be undoing the effects of the quality overlap, though in a new direction.

The analysis included the distributions of START and STRUT, as well as TRAP as a reference vowel, for old (45-60), mid-age (31-44), and young (18-30) speakers from NZSED, separated for males and females. Measurements were taken of first (F1) and second (F2) formants at a stable midpoint of each vowel token, for a single repetition of each /hVd/ word from 12 or 13 speakers in each of the six speaker groups. Hertz values were converted to the perceptually-relevant Bark scale, and statistical comparisons (Multivariate Analysis of Variance with F1 and F2 values as the dependent variables and vowel as the independent variable) were carried out of the distributions for each of the 6 groups of speakers. These analyses showed that both START and STRUT are significantly distinct from TRAP for all groups (all comparisons are significant at $p<0.001$), and that START and STRUT do not differ significantly from one another except in the case of the young females ($p<0.05$). Comparisons between the three age groups for the women indicated that distributions within the sets of both TRAP and START vowels do not differ according to age, but that the young women's distribution of STRUT differs from that of the old women ($p<0.01$) and marginally from that of the mid-age women ($p<0.09$). The same pattern was not found for the male speakers, which was confirmed also by an Analysis of Variance of the Euclidean distance between the START and STRUT vowels for each speaker, with speaker sex and age as independent variables. This produced a significant main effect of speaker age ($F[2,67] = 3.37, p<0.05$) and a significant interaction of age with sex ($F[2,67] = 3.77, p<0.03$). The interaction effect results from the finding that the average Euclidean distance between START and STRUT did not change across the age groups for men, but became progressively larger for women, from old to mid-age to young.

The pattern of these differences suggests that the STRUT vowel is raising away from START for the younger women. This raising of STRUT can be accommodated within the NZE vowel space because the raising of the short front vowels for which NZE is well-known has left space for such a closer realisation. It might have been triggered by the fronting of START and the resultant potential confusion of the two vowels. As the next analysis shows, there is also an indication that the quantitative contrast between these vowels is becoming less reliable, which may add to this confusion.

The analysis of the NZSED /hVd/ data also included durational measurements of the vowels. For this START/STRUT set, the vowel durations were entered into Analysis of Variance with speaker sex, age and vowel as independent variables. The main effect of vowel was highly significant (F[1,67]=977.43, p<0.001) – STRUT vowels had on average 40% of the duration of START vowels. In addition, the interaction of vowel and speaker age approached significance (p<0.09) – younger speakers tended to show a smaller durational difference (for old speakers the STRUT/START ratio was 0.37, for mid-age it was 0.40 for young speakers it was 0.44). Note though that this pattern was not different for the male and female speakers, so that the durational pattern is not related in any simple way to the pattern of qualitative differences between the vowels. (That is, we might predict that the durational difference would be less important for the younger females only, since they make a stronger qualitative distinction.)

In summary, and on the basis of a relatively small amount of word list data, it would appear that the START and STRUT vowels are distinguished in NZE primarily by duration, but that younger female speakers make a small but significant qualitative distinction, with a closer STRUT vowel. The following sections present data from a perceptual experiment examining the effects of a durational manipulation of words and nonwords containing these vowels, a manipulation that explores the impact of the qualititative overlap and whether it is affected by demographic variables associated with either the speaker or the participant.

4. Perceptual study of START and STRUT

4.1. Method

Bernard's (1967) early observation of the quality overlap of STRUT and START in AusE arose during his acoustic analyses of /hVd/ materials containing the full range of AusE vowels. For a subsequent listening test Bernard asked AusE speakers to hold the vowel part of short-vowel /hVd/ words, from which he then excised the vowel for identification. For another test he excised the vowel from unheld long-vowel /hVd/ recordings and presented shortened versions of the vowel for identification. Most confusions involved STRUT and START; in particular the shortened versions of START vowels (taken from *hard*) were reported in 96.2% of cases as STRUT, while the held STRUT vowel (from *hud*) was identified as START 54.4% of the time, with 33.5% STRUT and 8.7% NURSE responses.

Since the motivating interest behind the current research is with the consequences of sound mergers (in this case qualitative overlap without quantitative overlap) for the automatic processes of word recognition and speech comprehension, the present experiment replaces the vowel identification task used by Bernard with a timed lexical decision task. That is, participants heard a series of isolated word-length stimuli (with a range of vowels including but not limited to START and STRUT), for each of which they had to make a binary forced-choice under time pressure: was this stimulus a word of English or not a word?

Simulus materials

The test stimuli were created from twenty source tokens with START vowels. Half of these source tokens were real words (e.g. *flask*) and half were nonwords (e.g. **brarsh*). Within each set there were five monosyllabic tokens and five bisyllabic tokens with the START vowel in the first syllable (e.g. *basket*, **marther*). The vowels in these source tokens were shortened to produce stimuli with STRUT vowels. Shortening the vowel in the real word items produced a new set of nonwords (e.g. **flusk* from *flask*), while shortening the vowel in the nonword items produced a new set of real words (e.g. *brush* from **brarsh*). The shortening was proportional to the START/STRUT differences observed in the NZSED production data (i.e. to 40% of the source vowel duration). The entire test stimulus set thus consisted of 10 real words with the START vowel, 10 real words with STRUT, 10 nonwords with START and 10 nonwords with STRUT.

To gauge whether the perceptual effects of the qualitative overlap of the two vowels are influenced by aspects of the speaker's identity, two speakers were used, one female and one male, and for each of these speakers an attempt was made to associate them with two different age groups. Photographs were used to indicate to participants whether they were about to hear a token from the female or from the male voice, and the age of the speaker was prompted by using different photographs, depicting different ages, for different participant groups. In a pre-test with a separate group of participants, the age ratings given to the photographs corresponded to the old and young speaker groups in NZSED – the old and young female photographs were given average age ratings of 54.7 and 23.5 years respectively, while male photographs were given average age ratings of 48.2 and 22.9 years. The speakers themselves were both in the mid-age range (the female was 36 years old and the male was 44).

Procedure

The 20 participants were all NZE speakers; 10 were female and 10 were male. They were tested individually and were allocated to one of four groups (*1a*, *1b*, *2a*, *2b*), on a random basis as they arrived for the experiment. The stimulus sets for groups *1a* and *1b* always used the older male photograph as a prompt for the male voice, and the younger female photograph as a prompt for the female voice, while the stimulus sets for groups *2a* and *2b* always used the younger male photograph and the older female photograph. All participants responded to both the long-vowel and the shortened-vowel version of each source token, but these were placed into separate halves of the test session, with equal numbers of long and short vowel stimuli in each half. The difference between the *a* and *b* groups was that the order of these two halves was reversed. The entire stimulus list, which included filler words and nonwords, as well as items exploring other vowel contrasts (and which served as additional fillers for the materials reported here) contained 322 items. To give participants an opportunity to rest and to help them maintain their concentration, the two halves of the experiment were separated by a short interval. In addition, the stimulus list was preceded by a short practice block of 12 filler items, so that participants could become familiar with the equipment and procedure.

Participants sat in a quiet room, wearing headphones over which the stimulus lists were played at a comfortable listening level. For each stimulus, the participant was required to press one of two buttons: 'word' for a real word response and 'nonword' for a nonword response. The buttons were configured so that each participant used their dominant hand for the 'word' response. Participants had two seconds in which to make their responses, after which the next stimulus was presented. Between stimuli they were required to rest their index fingers on the two buttons. Both the speed of the participants' responses and the choices they made were recorded by the software that controlled the experiment (E-Prime: Schneider, Eschman and Zuccolotto 2002).

After the lexical decision task, participants completed a questionnaire covering demographic information. Since the participants all had the same occupation (students), a socio-economic score was derived from information concerning the occupations of the participants' parents (Elley and Irving 1985). Although participant age was collected as part of the demographic information recorded in the questionnaire, it was not included in the analysis of the lexical decision task because the 20 participants covered a narrow age range (19-27), with three-quarters of the participants covering the four years from 20-23. Note that all participants would be included in the 'young' age group in NZSED.

The experiment, including the questionnaire, took approximately 30 minutes to administer. Participants received a NZ$10 music or book voucher as recompense for their time.

Research questions

The research questions that were addressed through the experiment, with the above design features, were:

1. because of the quality overlap in START and STRUT, is it more likely that a token with a shortened START vowel will be heard as a token with a STRUT vowel than as a token with a START vowel?
2. if the overlap of START/STRUT is dependent on demographic differences between speakers (cf. the suggestion from the above analysis of NZSED data that there may be such effects), will participants be sensitive to this?
3. if the overlap of START/STRUT is dependent on demographic differences between speakers, will this be reflected in differences between participant performance, according to their own demographic characteristics?

The measures taken to assess these questions were i) the proportion of "word" responses (and its inverse, the proportion of "nonword" responses) to the experimental stimuli and ii) participants' response times (RTs). The independent variables in the experiment belonged in three groups, as follows:

A. Stimulus related *Source*: The START-vowel source tokens were real words or nonwords. *Vowel Length*: Stimulus vowels were long (as recorded in the source tokens) or shortened. *Syllables*: One or two syllables.

B. Speaker related *Speaker Sex*: The voice was female or male. *Photo Age*: The face in the prompting photograph was young or old.

C. Participant related *Participant Sex*: Participants were female or male. *Participant Socio-Economic Score*: based on the occupational categories of the participants' parents.

4.2. Results

i) Proportion of "word" responses

Overall, there was a high proportion of correct responses (98% of items intended as words received a 'word' response, 84% of items intended as nonwords received a 'nonword' response). Such a high level of overall accuracy suggests that the length manipulation produces a good STRUT percept. It also shows that the long vowel tokens are accurately perceived as having a START vowel. The higher correct response rate for items intended as real words reflects a general bias in lexical tasks towards interpreting ambiguous stimuli as real words rather than as nonwords (e.g. Ganong 1980).

Because the 98% correct response rate for the word targets (*flask*, *brush*, etc.) is near to ceiling performance, there is little value in assessing the impact of the independent variables on the response choice for these word targets, though we will return in the next section to a consideration of response times for this set. The higher error rate for the intended nonwords, at 16%, does however give some scope for exploring the influence of the experimental variables on performance with stimuli such as **brarsh* and **flusk*. A logistic regression model was fitted to the response choice data with the independent variables listed in the preceding section, except that Source of the stimulus was not included, since the levels of Source and Vowel Length represent the same groupings (i.e. all long vowel nonwords have a nonword source and all short vowel nonwords have a real word source). This model indicated significant main effects for the number of Syllables in the stimulus ($p<0.001$), for the Vowel Length of the stimulus ($p<0.001$) and for Participant Sex ($p<0.01$). In addition there were significant interactions of the number of Syllables with Vowel Length ($p<0.001$) and of Participant Sex with Participant Socio-Economic Score ($p<0.03$).

The interaction of the two stimulus-related variables reflects the finding that erroneous responses are more likely to unmanipulated long-vowel nonword stimuli if they are bisyllabic (e.g. more 'word' responses to **barket* than to **brarsh*), but are more likely to shortened vowel stimuli if they are monosyllabic (e.g. more 'word' responses to **flusk* [from *flask*] than to **busket* [from *basket*]). In the case of the long-vowel stimuli, bisyllabic items like **barket* might have received more word responses because a) there is greater segmental/syllabic overlap with the real word, b) the actual vowel length difference between short and long vowels in the first syllable of a bisyllabic form may be smaller, because of general shortening effects of the following syllable (Lehiste 1972; Warren 1985: 32-38)[3], and thus the long vowel in this first syllable may be misperceived as a short vowel (and **barket* heard as *bucket*). For the materials with shortened vowels it may be the case that the short vowel in the bisyllabic stimuli is now so short that it is no longer compatible with a long vowel, and so the error (**busket* heard as *basket*) is unlikely.

The overall main effect of number of Syllables results from a higher 'word' response rate to bisyllabic nonword stimuli, which may simply reflect the greater segmental and syllabic overlap that such stimuli have with the corresponding real word. The main effect of Vowel Length reflects the finding that the short vowel stimuli are more likely to be erroneously reported as real words. This may reflect the relative likelihood of shortening vs. lengthening effects in speech, but may also indicate that the shortened START vowels are not entirely convincing as STRUT.

It is interesting that neither of the speaker-related demographic variables showed an effect in the analysis, but that the sex of the participants had a significant impact. Female participants were more likely to respond to a nonword as though it were a word. This suggests that the young women participating in this experiment are less likely than the men to use the quantitative distinction between START and STRUT, which might in turn imply that they rely more on the qualitative distinction. Note in this context that it was the young women in the NZSED production data who were found to be making a greater qualitative distinction. The fact that the error rate for nonwords was higher for women with lower socio-economic scores (hence the interaction of Sex and Socio-Economic Score) is entirely in line with the general tendency for young women from lower socio-economic groupings to be at the forefront of sound change in NZE (cf. Holmes 1997).

ii) Response times

Responses were measured by the experimental software from the onset of each stimulus. Since in a lexical decision task respondents may delay initiation of their responses till the end of the stimulus, in case what at first appears to be a real word actually achieves nonword status on the basis of a segment late in the stimulus, the response times were adjusted to provide times from the offset of the word, giving *offset RTs*. It should be noted that analyses were also conducted on the *onset RTs*. The pattern of significant effects in the onset RT analyses differed from the analyses reported below only in that longer words (i.e. those with longer vowels or those that are in some other respect longer) had longer RTs, such differences being largely in proportion to the durational differences between the items.

Separate analyses were carried out for correct ('word') responses to real words and for correct ('nonword') responses to nonwords. This was because 'word' and 'nonword' responses use different hands (dominant and non-dominant respectively) as well as being different outcomes.

Real Words

An ordinary-least-squares linear regression analysis was conducted of 'word' response times to stimuli intended as real words (i.e. original real words such as *flask* and words produced by shortening the vowel in nonwords such as *brush* from **brarsh*). The independent variables were as listed above. The analysis of the offset RTs showed a significant main effect of Participant Sex ($p<0.001$), while the interaction of Speaker Sex with Photo Age approached conventional levels of significance ($p<0.06$). No other main effects or interactions were significant. The significant effect for Participant Sex reflects the fact that the female participants responded more rapidly to the stimuli than the male participants (by an average of 77 ms.). Since this factor did not interact with any of the other experimental variables, we can assume that this is a simple difference in response speeds of the two participant groups that is not of further interest to the current discussion.

The interaction of Speaker Sex and Photo Age comes close to conventional levels of significance. The male voice is responded to more rapidly after the older male photograph, and the female voice more rapidly after the younger female photograph. The patterning of this interaction is interesting, since it seems to reflect the extent of the match of photograph with speaker. That is, the male speaker, at 44 years of age at time of recording, is closer to the age indicated by the old male photograph (given an average age rating of 48.2 years by an independent group of participants) than to that of the young photograph (22.9), while the female speaker, at 36 years is closer to the young photograph (23.5) than to the old photograph (54.7).

The fact that none of the stimulus variables were significant, as well as the lack of significant interactions involving these variables, shows that all of these stimuli, regardless of their source (i.e. whether they have full-length START vowels or are constructed to have STRUT vowels by shortening of a START vowel) are responded to at a fairly uniform latency from stimulus offset. It does not appear that responses were influenced by any differences in the extent of the START-STRUT overlap for different speaker groups. Instead, the indications are that shortening the START vowel in tokens like **brarsh* creates convincing real words with STRUT vowels.

Nonwords

'Nonword' response times to stimuli intended as nonwords (i.e. original nonwords such as **brarsh* and nonwords derived from real words such as **flusk* from *flask*) were also subjected to ordinary-least-squares linear regression analyses, with the same independent variables as above. The analysis of offset RTs showed a significant main effect of number of Syllables ($p<0.001$). In addition, both of the participant variables produced significant main effects (Participant Sex at $p<0.04$ and Socio-Economic Score at $p<0.001$), and there was also a significant interaction of Speaker Sex and Photo Age ($p<0.05$). No other main effects or interactions were significant.

The effect of number of Syllables is such that the monosyllabic items receive faster responses than bisyllabics. It may be the case that the START-STRUT vowel length difference is more discriminable in the monosyllabic stimuli, because of the compression of the first syllable in bisyllabic forms like **barket* and **busket*, and that it is therefore easier to reject the monosyllabic nonword stimuli. Recall from the error analysis above that the bisyllabic nonwords were also more likely to be reported as real words.

As was the case with the real word response times, the Participant Sex effect is a faster mean response time from the female participants than from the men (in this case by 49ms). The effect of Socio-Economic Score was that participants with higher scores in this measure took longer to reject the nonwords as not being words known to them. It may be that participants with a higher score by this measure (which, recall, is dependent on the occupational status of their parents) have a larger vocabulary, or a heightened expectation that word-like tokens that they hear are likely to be words, and that this is reflected in the longer time taken to reject a nonword.[4]

The interaction of Speaker Sex and Photo Age shows a different pattern from the interaction of these factors reported above for the real words. For the words, the effect was that responses to the male voice were faster after the older photograph and for the female voice they were faster after the younger photograph, a result which was argued to reflect a match between speaker age and photograph age. For the nonwords, however, responses are slower for both voices after the older photographs, with a stronger effect for the male voice. For both voices, then, the nonwords take longer to reject when associated with the older photograph. This interesting difference between the word and nonword results will be examined in more detail in the following Discussion section.

5. Discussion

The experiment reported in this paper used a lexical decision task with length-manipulated stimuli to determine whether listeners rely on quantitative distinctions between START and STRUT for word identification because of the qualitative overlap of these vowels in New Zealand English. The results indicated that the length of the stimulus vowel is indeed a good predictor of vowel identity. This is shown by the high degree of accuracy in the task and by the fact that response times to both real words and nonwords were unaffected by whether the vowel was an unmanipulated START vowel or a shortened STRUT vowel.

A clear but unsurprising effect in the results was the higher accuracy rate with the tokens intended as real words than with the nonword tokens. This is in line with the general research finding of a bias towards real word responses in lexical tasks.

The effect of syllable count on this error rate with the nonword stimuli shows that the greater the overlap of the stimulus with an existing real word, the more likely it is that the real word response will be given. This increased uncertainty about the nature of the stimulus is reflected also in the longer response times for the correct nonword responses to the bisyllabic stimuli.

The second and third research questions focussed on whether the slight differences in qualitative overlap of START and STRUT observed in the NZSED study would be reflected in performance in the lexical decision task. We might expect this to be revealed in the patterning of response times by participant variables (e.g. the age, sex, or social status of the participants) or by speaker variables (i.e. the sex of the speakers providing the stimuli, and the age implied by the prompting photographs). The results indicated a possible participant effect here. First, the error rate for nonwords was higher for women with lower socio-economic scores (as reflected in an interaction of Sex and Socio-Economic Score), which is in line with a general tendency for young women from lower socio-economic groupings to be at the forefront of sound change in NZE. Second, there was an effect of Socio-Economic Score on the response times for the nonword stimuli, though this did not interact with participant sex. There was also a main effect in the response time analysis of the sex of the participants, with the women being significantly faster in their responses. In addition, the error rates showed that the women participants were less likely than the men to use a quantitative distinction between START and STRUT in determining the lexical status of the stimuli, which might in turn imply that they rely more on the qualitative distinction. Note in this context that these participants were in the same age range as the ‘young’ speakers in NZSED, and that it was the young women in NZSED who were found to be producing a greater qualitative distinction.

Turning to the speaker-related demographic variables, we observed that while these had no effect on participant accuracy as reflected in the error rates, they did produce an interesting pattern of effects on response speeds. For the real word responses, the interaction of speaker sex and speaker age (as indicated by the prompting photograph) seems to reflect an age match between the speaker’s actual age (which presumably is indicated in some way by their voice, e.g. through voice quality features) and the age of the photograph. That is, fastest responses were when the voice followed the most age-appropriate photograph. The nonword response times showed a different interaction pattern, since both male and female voices were responded to more slowly after the older photograph (and the interaction effect reflects the finding that this difference is greater for the male speaker). The following is offered as a possible interpretation of this pattern of response times. In the case of the real word stimuli, response times are dependent on the characteristics of the available mental representations of words against which the stimuli need to be matched in order for the positive response (‘yes, this is a word of English’) to result. In line with previous discussion of experimental results from other studies with New Zealand English vowels (Hay, Warren and Drager 2006; Warren and Hay 2006; Warren, Hay and Thomas submitted), these word representations can be seen as phonetically detailed exemplars with associated social information about the kinds of speakers who – in the listener’s experience – might be likely to produce these tokens. If this social information includes details which relate to the type of STRUT and START distinction being maintained by males and females across different age groups, then we would predict the age-match result found for the real words in this experiment. However, the same pattern is not likely for the nonwords, because these have no lexical representation, i.e. there are no exemplars with associated social indexical information. Rather, the participants’ experience with the nonword stimuli may be that although they fail to find a matching lexical representation for the spoken stimulus both when the speaker is cued as an older speaker and when they are cued as a younger speaker, they are nevertheless faster in rejecting it when the cued speaker is from their own age group, and slower – more conservative perhaps – in rejecting it for an older speaker. This suggests that participants are sensitive to phonetic differences between the age groups, or at least that they are more efficient in dealing with the word/nonword response for tokens that they are assessing against some more heavily populated distribution of exemplars, i.e. the exemplars from their own age group, to whom they presumably have greatest exposure. The fact that this difference is greater for the male voice may be because both the voice and the old photo suggest an older speaker, and the combined effect here is stronger than with the female, where at least the age of the voice was closer to the age group of the participants.

In summary, then, the experiment reported in this paper has shown that a quantitative distinction between the START and STRUT vowels is sufficient to cue words and nonwords containing these vowels in New Zealand English stimuli heard by young NZE-speaking participants. It has shown that responses are only to a minor extent sensitive to demographic variables that influence the pattern of phonetic distinction between these two vowels in production data, a result which is not entirely surprising given the relatively small qualitative difference between productions of these vowels. An interesting additional result is that participants appear to be more sensitive to speaker-related demographic variables for real word stimuli than for nonword stimuli, a result that is compatible with the indexing of social information to mental representations (exemplars) for real words.

Notes

1 This paper uses START and STRUT (and later in this paper also TRAP) to refer to the lexical sets containing the relevant vowels, following Wells (1982). *Cart* and *cut* would be a relevant minimal pair for the contrast in question.

2 The /hVd/ context is frequently used in word-list recordings (e.g. producing words like ‘hid, heed, had, hard’) in order to maintain the same consonantal contexts for the vowels.

3 The START vowel of the bisyllabic source items used in the current experiment had a mean duration of 182 ms., while that of the monosyllabic items had a mean vowel duration of 261 ms. This difference was significantly different (in ANOVA, at $p<0.001$).

4 A reviewer has suggested that this effect of Socio-Economic Score might also result from a greater hesitancy among the speakers from the higher groups, reflecting their greater reluctance to appear ‘in error’. While plausible, this possibility cannot easily be evaluated on the basis of the current data.

References

Baskaran, Loga. 2004. 'Singapore English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 1034–1046.

Bauer, Laurie and Paul Warren. 2004. 'New Zealand English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 580–602.

Bernard, J.R. 1967. 'Length and the Identification of Australian English Vowels.' *AUMLA (Journal of the Australian Universities Language and Literature Association)* 27: 37–58.

Bobda, Augustin Simo. 2004. 'Cameroon English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 885–901.

Clark, John. 1989. 'Some proposals for a revised phonetic transcription of Australian English.' In Peter Collins and David Blair (eds) *Australian English: The Language of a New Society*. 205–213.

Clark, Urszula. 2004. 'The English West Midlands: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 134–162.

Crowley, Terry. 2004. 'Bislama: phonetics and phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 671–689.

Easton, Anita and Laurie Bauer. 2000. 'An acoustic study of the vowels of New Zealand English.' *Australian Journal of Linguistics* 20(2): 93-117.

Elley, Warwick B. and James C. Irving. 1985. 'The Elley-Irving socio-economic index: 1981 census revision.' *New Zealand Journal of Educational Studies* 20: 115–128.

Finn, Peter. 2004. 'Cape Flats English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 964–984.

Ganong, W. F. 1980. 'Phonetic categorization in auditory word perception.' *Journal of Experimental Psychology: Human Perception and Performance* 6: 110–125.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill. 2004. *New Zealand English: its origins and evolution*. Cambridge, UK: Cambridge University Press.

Hay, Jennifer, Paul Warren and Katie Drager. 2006. ‘Factors influencing speech perception in the context of a merger-in-progress.’ *Journal of Phonetics* 34: 458–484.

Holmes, Janet. 1997. ‘Setting new standards: sound changes and gender in New Zealand English.’ *English World-Wide* 18(1): 107–142.

Horvath, Barbara. 2004. ‘Australian English: phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 625–644.

Huber, Magnus. 2004a. ‘Ghanaian English: phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 842–865.

Huber, Magnus. 2004b. ‘Ghanaian Pidgin English: phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 866–873.

Jourdan, Christine and Rachel Selbach. 2004. ‘Solomon Islands Pijin: phonetics and phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 690–709.

Kortmann, Bernd, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds). 2004. *A Handbook of Varieties of English: A Multimedia Reference Tool*. Berlin: Mouton de Gruyter.

Lehiste, Ilse. 1972. ‘The timing of utterances and linguistic boundaries.’ *Journal of the Acoustical Society of America* 51: 2018–2024.

Malcolm, Ian G. 2004. ‘Australian creoles and Aboriginal English: phonetics and phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 656–670.

Menang, Thaddeus. 2004. ‘Cameroon Pidgin English (Kamtok): phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 902–917.

Parry, David. 1999. *A Grammar and Glossary of the Conservative Anglo-Welsh Dialects of Rural Wales*. Sheffield: National Centre for English Cultural Tradition.

Penhallurick, Robert. 2004. ‘Welsh English: phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 98–112.

Sakoda, Kent and Jeff Siegel. 2004. ‘Hawai’i Creole: phonology.’ In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 729–749.

Schmied, Josef. 2004. 'East African English (Kenya, Uganda, Tanzania): phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 918–930.

Schneider, Walter, Amy Eschman and Anthony Zuccolotto. 2002. *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Tent, Jan and France Mugler. 2004. 'Fiji English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 750–779.

Torgersen, Eivind, Paul Kerswill and Susan Fox. 2006. 'Ethnicity as a source of changes in the London vowel system.' In F. Hinskens (ed) *Papers from the Third International Conference on Language Variation in Europe, Amsterdam, June 2005*: Amsterdam: John Benjamins.

van Rooy, Bertus. 2004. 'Black South African English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 943–952.

Warren, Paul. 1985. *The Temporal Organisation and Perception of Speech*. Unpublished PhD, Cambridge: University of Cambridge.

Warren, Paul. 2002. 'NZSED: building and using a speech database for New Zealand English.' *New Zealand English Journal* 16: 53–58.

Warren, Paul and Jen Hay. 2006. 'Using sound change to explore the mental lexicon.' In Claire Fletcher-Flinn and Gus Haberman (eds) *Cognition and language: Perspectives from New Zealand*. Bowen Hills, Queensland: Australian Academic Press. 105–125.

Warren, Paul, Jen Hay and Brynmor Thomas. submitted. 'The loci of sound change effects in recognition and perception.' In José Ignacio Hualde and Jennifer Cole (eds) *Laboratory Phonology 9*.

Watson, Catherine I., Jonathan Harrington and Zoe Evans. 1998. 'An acoustic comparison between New Zealand and Australian English vowels.' *Australian Journal of Linguistics* 18(2): 185–207.

Wee, Lionel. 2004. 'Singapore English: phonology.' In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (Vol. 1). Berlin: Mouton de Gruyter. 1017–1033.

Wells, John C. 1982. *Accents of English*. Cambridge, England: Cambridge University Press.