

**Wellington Working Papers in
Linguistics**

Volume 15, 2003

edited by Paul Warren

**School of Linguistics and Applied Language Studies
Victoria University of Wellington
P.O. Box 600
Wellington
New Zealand**

Published 2004

Printed for
Victoria University of Wellington
P.O. Box 600
Wellington
New Zealand

ISSN 1170-1978

Wellington Working Papers in Linguistics 15, 2003 Contents

Helen Ainsworth	How she says it and how he says it - differences in the intonation of dairy farming women and men in South Taranaki	1
Patricia Vermillion	The ups and downs of Kiwis: An experiment investigating tonal cues used to identify NZE intonation	17
Paul Warren, Shari Speer and Amy Schafer	Wanna-contraction and prosodic disambiguation in US and NZ English	31
Joel Zwartz and Paul Warren	This is a statement? Lateness of rise as a factor in listener interpretation of HRTs.	51
	Contents of previous volumes	63

Editorial

Once again the publication and volume dates of the Wellington Working Papers in Linguistics are out of synchrony, for which the editor apologises.

The theme to this volume is prosody in New Zealand English (NZE). The papers are contributed by staff and students working in this area. Helen Ainsworth has previously published a number of short pieces on aspects of the rhythm of NZE, and has for the last few years been carrying out intonational analysis of conversational data collected in the Taranaki region, towards a PhD. Her paper in this volume covers a part of that analysis. Patricia Vermillion is also working towards a PhD, exploring the characteristics of NZE intonation. In her paper in this volume she reports results from a judgement task that aims to determine specific traits that distinguish the intonation of NZE from that of British English. The paper by myself with Shari Speer and Amy Schafer is a small part of a larger collaborative study (the SPOT project – see references in that paper) investigating the use of prosody in marking syntactic structure and in resolving structural ambiguity. Finally, Joel Zwartz's contribution is based on a study he conducted for a Linguistics Honours class, and explores the perceptual effects of differences in the starting points of intonational rises in utterances that are ambiguous between statement and question interpretations.

Together, these papers give an introduction to some of the current research on intonation in NZE that is being conducted in the School of Linguistics and Applied Language Studies at Victoria University of Wellington.

Paul Warren

May 2004

How she says it and how he says it - differences in the intonation of dairy farming women and men in South Taranaki

Helen Ainsworth

Abstract

This paper explores gender differences in the intonation of a group of New Zealanders. The pitch dynamism of speakers was compared, the measure of pitch dynamism being the number of times on average speakers changed pitch direction within intonation groups. The kinds of pitch accents favoured were also analysed. The women in the sample had fewer changes in pitch direction than the men and a larger percentage of their pitch accents were level as opposed to moving. In other words, the intonation of the women was characterised by less pitch dynamism than the men's. This runs counter to commonly held beliefs about men's and women's intonation patterns. Explanations for the differences were explored, and geolinguistic, social network and sociolinguistic theories provided possible clues as to the source of the gender differences.

Introduction

Stereotypically, female intonation has been described in rather negative terms such as 'high-pitched', 'shrill', 'over-emotional' and 'swoopy' (Henton 1989: 299). It has also been characterised as 'more dynamic than men's, displaying wider ranges of pitches, more frequent and rapid shifts in pitch, and more frequently ending with a non-falling terminal than men's' (McConnell-Ginet 1983:555). McConnell-Ginet questioned the validity of such global characterisations, saying that they were based on 'unsystematic observation' and 'a few experimental studies in laboratory settings involving oral reading' (ibid.). She called for systematic investigation to determine the extent to which these male/female stereotypes were based on fact.

In New Zealand Warren and Daly took up the challenge and compared Wellington men's and women's pitch range and dynamism (Warren & Daly 2000, 2004, Daly & Warren 2001). Using an Equivalent Rectangular Bandwidth scale (which normalises for absolute differences in male and female pitch ranges) they measured speakers' pitch range in unscripted map task dialogues as well as in scripted sentences and stories which included instances of direct speech. Their findings confirmed speaker-sex stereotypes. The female subjects had greater pitch ranges than the men in the study. Using the same material they measured pitch dynamism by looking at the speed of changes in a speaker's pitch range from high to low points and low to high points. The female speakers were found to use more dynamic pitch than the males. They also examined the size and rate of change of rises and the alignment of pitch movement against segmental aspects of utterances in questions realised only by intonational means in the

speech of young Pakeha women and men. The women were found to use larger and more rapid rises which were aligned later with respect to text, tending to confirm once again the male/female stereotypes listed above.

I examined a sample of the speech of young dairy farmers, both male and female, in South Taranaki within the context of a larger project (Ainsworth, in prep.) in which I compared the intonation patterns of Taranaki speakers with those of speakers from Wellington. It was my impression that female intonation in my corpus was not markedly 'high pitched' or 'shrill', to use Henton's words, but there did seem to be a lot of pitch movement in the speech of the elderly women of Taranaki, prompting the description 'sing songy' by one observer (Priestley, personal communication). I decided therefore to focus on pitch dynamism as measured by frequency of shifts in pitch. I was also interested in what kinds of pitch accents speakers favoured. Described below are the methodology used, my findings, and discussion of possible reasons why the younger men and women in my survey appeared to have intonation patterns which did not conform to stereotypical male/female patterns.

The data

Data for the present paper was taken from tape recorded conversational interviews with five male and five female Pakeha dairy farmers resident in South Taranaki, aged between 25 and 35 and matched for educational level.

Following Ladefoged's recommendation (1997:140) that short narratives or other material containing longer stretches of speech be used in the description of the intonation of as yet undescribed languages or dialects, I selected for analysis narrative speech samples. Whenever possible these sections conformed with Labov's definition of narrative, where a minimal narrative is 'a sequence of two clauses which are *temporally ordered*' (Labov 1972a: 360). Fully-formed oral narratives in Labov's framework contain the following six elements: an abstract, orientation detail, complicating action recounting what happened, an evaluation which gives an indication as to what makes the story interesting, a resolution saying what happened in the end, and a coda signalling the end of the story. Not every narrative in my corpus contained all six elements but most contained more than the minimal two temporally ordered clauses.

Where fluent narrative sections of speech were difficult to find I occasionally used sections which might be termed 'factual reporting' such as descriptions of a typical day's work. Although not strictly speaking narratives, the passages selected were at least fluent and contained more than one clause. For the most part, however, narratives were abundantly available. The majority of speakers had interesting stories to tell in response to the classic Labovian (1966) 'danger of death' question. The positioning of the question immediately after a reading passage which told the story of a woman who got into trouble trying to get her sailing dinghy back to shore, prompted the narration of numerous stories from people's pasts featuring lakes, rivers and the sea. The recreational preferences of many of the subjects were centred around aqua-sports - fishing, diving,

swimming, waterskiing, sailing and surfing - a watery focus which reflects the island status of New Zealand. The corpus contains harrowing tales of swimmers and surfboard riders being swept out to sea on rip tides and others being dumped by freak waves. In one a diver surfaced to see his support boat disappearing over the horizon. There is a handful of stories which concern the hazards of farming and one might well wonder at the standard of driving in New Zealand given the number of stories involving motor accidents. One of my favourite stories concerns a novice skier standing petrified at the top of a ski run who then opts to walk down the mountain rather than risk life and limb by skiing down.

Analysis of the data

Five minute sections of narrative text from each recording were orthographically transcribed and intonation groups, the basic units of intonational structure, were identified (following Cruttenden 1997:29-37). The boundaries of these units were indicated by some or all of the following characteristics: a pause, lengthening of the final syllable with concomitant slowing of tempo, increase or decrease in volume, pitch discontinuity, terminal falls or rises in pitch, and anacrusis (speeding up of unstressed syllables at the beginning of an intonational group). Very often syntactic cues indicated intonation group boundaries and many boundaries coincided with changes in speaker turn.

Syllables which were rhythmically prominent, i.e. stressed, and pitch prominent (accented) were auditorially identified. Pitch prominence could be expressed as moving pitch, as a step up or step down in pitch, or as a change in pitch trend (eg. from falling to rising, or from level to rising). 'Accent unit' is the term for the domain of a pitch accent which is linked to a prominent syllable. The realisation of a pitch accent is not limited to the prominent syllable itself, but can spread over syllables up to, but not including, the next pitch accent.

Accent units within the intonation groups were marked according to whether they were rising, falling, or level in pitch. The most prominently accented syllable in each intonation group, i.e. the one with the most marked pitch level or pitch movement and normally the last accented syllable in the group, was identified as the nuclear accent unit. Some nuclear accent units were characterised by a combination of both rising and falling pitch movement and were marked accordingly.

Further auditory assessment was made of accented syllables which did not sit on the declination line, the default pattern for RP English intonation. If they started at a markedly higher pitch than the preceding accented syllable they were considered to show "up-step", which was marked by an up arrow. If they started at a lower pitch than the preceding accented syllable they showed "down-step", marked by a down arrow. If they started at the same pitch as the preceding accented syllable they showed "continuation" of the pitch level, which was marked by a horizontal arrow. Used alone, on an unstressed syllable, the arrows marked a pitch pattern which was significantly above or below the

expected pitch level. It is somewhat debatable whether or not the declination line typical of RP English should be regarded as the default pattern for New Zealand English. It was clearly present for some speakers in my corpus but less clearly so for others. Where the declination line appeared to be absent and the base-line was level in pitch I used the arrow symbols to mark pitch level above or below what might be expected, given my auditory assessment of what might be the speaker's normal pitch base-line.

A sample of approximately 50% of my auditory assessments was checked by Paul Warren and a high degree of consensus was reached. Blind checks were also undertaken on sections of the transcriptions. My assessments did not vary greatly from one listening to another. A few adjustments were necessary but overall there was a high level of consistency.

A minimum of 200 intonation groups per speaker were analysed. The number of syllables per intonation group was averaged for each sample group. Women averaged 4.39 syllables per intonation group and men averaged 4.75. The number of accent units per intonation group was averaged for each group and again there was little difference - women averaging 1.93 accent units per intonation group and men averaging 1.90.

The first step in the analysis was to find a method of globally measuring changes in pitch direction. By counting changes in pitch direction in each intonation group and then in each accent unit, I constructed an index by which I could measure the difference between the sets of sample texts. My count of changes in pitch direction included changes of direction within complex nuclei, changes from rising accent units to falling accent units in the head and from falling accent units to rising accent units in the head. Where there was a step up to a level or falling pitch accent, the step up was counted as a change in pitch direction and where there was a step down to a level or rising pitch accent, the step down was counted as a change in pitch direction. In passages of continuous speech, resetting of pitch upwards or downwards at the beginning of a new intonation group after a fall or rise at the end of the previous intonation group counted as a change of direction. Resetting of pitch at the beginning of a new intonation 'paragraph' was not counted as a change of pitch direction. Paragraphs in writing have analogues in speech. There are prosodic markers, particularly in organised or rehearsed speech styles such as news-reading, or story telling, which indicate the introduction of new subject matter. Tench (1996:23) describes a typical intonation paragraph in RP English as follows: 'the first onset syllable of the initial intonation unit of the new item is high and then comes a gradual descent in the general pitch level of the following units until the lowest point is reached with the final unit.' The descent in pitch level in New Zealand English may not be as pronounced as in RP but new subjects are nevertheless clearly marked intonationally. Speakers tend to expand their pitch range at the onset of a new topic and contract the range at its conclusion.

Applying this measure of pitch dynamism, i.e. calculating the average number of changes in pitch direction per intonation group and then in each

accent unit, the men in the sample demonstrated more pitch dynamism than the women (see Table 1).

	CODs per intonation group	CODs per accent unit
women	1.72	0.91
men	2.68	1.41

Table 1 Changes in pitch direction (CODs)

With changes in pitch direction patterning in this way one could expect that listeners might be able to detect a difference in tone between the speech of the two groups. And indeed when I played samples of a Taranaki dairy farming young man and a Taranaki dairy farming young woman to a class of third year students of linguistics at Victoria University of Wellington, the students confidently claimed that one speaker had more 'bouncy' intonation and one had 'flatter' intonation than the other. The 'bouncy' intonation belonged to a young man and the 'flatter' intonation to a young woman. The young man had on average 3.02 CODs per intonation group and 1.59 CODs per accent unit while the young woman had on average 1.25 CODs per intonation group and 0.68 CODs per accent unit.

My attention turned next to the analysis of nuclear accents. The nucleus is the most prominent and salient accent in the intonation group and could therefore be expected to make a major contribution to the perception of intonation. Pitch movements associated with nuclear accents largely determine the tune of the intonation group (Cruttenden 1997). Analysis of broad tune types gave me an indication of the kind of tunes favoured by speakers.

Consistent with the overall measure of pitch dynamism outlined above was the finding that women's nuclei were more likely to be level than were men's (see Table 2). Chi-square analysis of the counts of items making up the percentages of level vs. non-level nuclei revealed a significant difference between women and men in the distribution of their nuclear accent types ($\chi^2=77.43$, $df:1$, $p<0.001$).

	% level nuclei
women	28.61
men	12.84

Table 2 Level nuclear accents

Table 3 shows the patterns for moving pitch. The description of moving nuclei was very similar for men and women except that men used more falling nuclear accents than women.

Nucleus type:	falling	rising	complex
women	50.43	7.17	13.77
men	63.83	9.68	13.63

Table 3 Nuclear accents with moving pitch

Women's preference for level pitch extended to non-nuclear accents (see Table 4). Again, the counts on which these percentage values are based were analysed by chi-square, and showed a significant difference in female and male use of level vs. non-level non-nuclear accents ($\chi^2=209.74$, $df:1$, $p<0.001$)

	% level non-nuclear accents
women	70.75
men	38.27

Table 4 Level non-nuclear accents

When I narrowed the focus still further and looked at the manner in which each falling accent was approached or 'attacked' I found further evidence of gender differences in pitch movement. I focussed my attention on falling accents for the simple reason that there were more falling accents (especially nuclear ones) than there were other moving accents for both groups of speakers. The count of falling accents included nuclear and non-nuclear falling accents plus the first part of complex fall-rise accents. The manner in which speakers approach falling accents varies. The starting point of downward pitch movement is usually slightly above that of the preceding unaccented syllable, but sometimes there is a noticeable step up to it, sometimes the onset is level with the previous unaccented syllable and occasionally there is a noticeable downwards step to the onset of the fall in pitch. When I looked at the manner in which speakers approached falling accents I found that men were more likely to use stepped up onsets than were women. Table 5 shows the number of falling accents with and without up-step onsets, and the proportion of all falling accents with up-step. Chi-square analysis of the counts in Table 5 reveals a significant difference in the incidence of up-step in women and men ($\chi^2=93.67$, $df:1$, $p<0.001$).

	# falling accents with up-step	# falling accents without up-step	% of falling accents with up-step
women	49	873	5.31
men	261	1071	19.59

Table 5 Falling accents with up-step

Finally, I looked at how speakers approached level accents, both nuclear and non-nuclear, again considering specifically the incidence of up-step. Table 6 shows the number of level accents with and without up-step onsets, and the proportion of all level accents with up-step. Chi-square analysis of the counts in

Table 6 reveals a significant difference in the incidence of up-step in women and men ($\chi^2=6.59$, $df:1$, $p=0.01$), this time with more up-step from the women.

	# level accents with up-step	# level accents without up-step	% of level accents with up-step
women	135	885	13.23
men	43	451	8.70

Table 6 Level accents with up-step

Discussion

It would seem that, for the most part, the intonation of South Taranaki dairy farming young women does not conform with the stereotypical female patterning of intonation as outlined by McConnell-Ginet (1983:555). It most certainly cannot be accused of being 'swoopy', one of the descriptors which features in Henton's list of stereotypically female intonation traits (Henton 1989:299). Using my operational definition of pitch dynamism, i.e. the average number of times a speaker changes the direction of their pitch per intonation group and per accent unit, the young men in my sample were shown to have more pitch dynamism than the young women in the sample. The young women were shown to have a greater preference for level accent units, both nuclear and non-nuclear, than the men, adding to the effect of an overall flatter and less dynamic intonation.

In one respect, however, these women conformed to McConnell-Ginet's stereotypes. She stated that women were said to have intonation which 'more frequently end[s] with a non-falling terminal than men's' (McConnell-Ginet 1983:555). This was true of the women in my survey. They had 13% more non-falling terminals than the men. (For 'terminal' read 'nuclear accent'). Not only did the women have more level nuclear accents, they also had a higher percentage of level accents with stepped up onsets. The latter amounted to upwards change in pitch direction and often seems to serve a similar function to the High Rising Terminal (HRT), a feature of Australasian intonation (and increasingly of the intonation of other varieties) that has received much attention in the literature. It is my impression that in my data HRTs tend to be preceded by level sections of intonation, and falling terminals by falling accents. The latter pattern is characterised by more changes in pitch direction as each fall is preceded by an upwards movement in pitch. This suggests that the finding that the women in my sample had fewer non-falling terminals than the men is not inconsistent with the finding that the women's intonation was characterised by less pitch dynamism than the men's.

The research reported here and that in Warren and Daly (2001) have picked out slightly different measures of pitch dynamism, each of which feature in McConnell-Ginet's (1983:555) list of female stereotypes. Where I focussed on

'more frequent ... shifts in pitch' they looked at 'displaying wider ranges of pitches' and 'more ...rapid shifts in pitch'. For their sample, Warren and Daly found that women had greater pitch ranges than men and that the speed of change in a speaker's pitch range from high points to low points and from low points to high points was greater for the women than for the men. My results are not necessarily inconsistent with these findings. It is possible that the Taranaki women had greater pitch ranges than the men in the sample and that the women's changes from high points in their pitch range to low points and vice versa may be faster than those of the men. I did not measure speakers' pitch ranges nor the speed of shifts in pitch.

The question I address in this paper is why the women and the men in my sample do not appear to conform to gender stereotypes with respect to my measure of pitch dynamism (i.e. frequency of changes in pitch direction). In order to understand the differences I believe it is necessary to look at gender in conjunction with other social factors. As already mentioned the analysis was undertaken as part of a much larger project. In the wider study, speaker location and age were considered as well as gender. When the effect of these additional factors is studied, the gender differences revealed are perhaps less surprising. Analysis of all samples in the study showed that as a rule:

- (1) Speakers in Taranaki locations (the South Taranaki dairy farming community and New Plymouth) showed more pitch dynamism than did Wellingtonians.
- (2) Older speakers (aged from 70 years) in Taranaki locations (the South Taranaki dairy farming community and New Plymouth) and Wellington showed more pitch dynamism than did younger speakers (aged 20-30 years) in these areas.

Exceptions to the first rule were the young dairy farming women in South Taranaki and the young women of New Plymouth who had pitch patterns more like Wellingtonians than like their fellow Taranakians, and exceptions to the second rule were the young dairy farming men who shared the pitch patterns of older speakers, particularly those from Taranaki. The young New Plymouth men were mid-way along the continuum, having more pitch movement than the young women and the young Wellington men, but less than the elderly and the young dairy farming men.

For explanation of these exceptions to the rules I turn to geolinguistic, social network and sociolinguistic theories.

In designing the broader study the decision to investigate the intonation of speakers from Taranaki communities and Wellington city was motivated by a desire to address the issue of whether or not regional variation, aside from Southland/Otago English, exists in New Zealand. The finding that there was more pitch dynamism in the speech of most Taranaki speakers surveyed than in the speech of most Wellingtonians surveyed suggested the possible existence of regional variation. But the fact that not all Taranaki speakers shared the same

preference for pitch dynamism suggested that a change with respect to this linguistic feature might be underway. The change could be the result of 'hierarchical' diffusion (Callary 1975, Trudgill 1983) whereby a linguistic innovation begins in a larger population centre and then trickles down to smaller centres. In this case the linguistic innovation consists of a comparative lack of pitch dynamism. According to the theory of hierarchical diffusion the tendency towards flatter intonation would have started in the centre with the biggest population, Wellington, then moved to New Plymouth, the next most populous centre, and thence to the more sparsely populated South Taranaki dairy farming community. The results described above suggest that such a pattern of diffusion has taken place.

Proponents of geolinguistic theory suggest that the effects of 'place' as opposed to 'space' need to be examined in order to understand geographic distribution of linguistic features (Britain 1991, Horvath & Horvath 2002).

'Place effects refer to the ensemble of sociolinguistic conditions within a speech locality, whereas space effects refer to the relationship between speech localities. Space focuses attention on distance, proximity, or location; place addresses the effects of linguistic and social conditioning within particular speech localities...' (Horvath & Horvath 2002:63).

Social space or 'place' as defined by the Horvaths is an important aspect of social network theory developed within sociolinguistics by Milroy and Milroy (1985). Rural communities have traditionally been characterised by the 'dense' and 'multiplex' social networks which the Milroys identified amongst working class people of Belfast. In such communities people were related to each other, lived close by and worked together. There was often a strong sense of solidarity and pressure exerted on members to conform to social norms. 'As a consequence of this inter-group distinctiveness is emphasised and the members of close-knit networks are insulated from outside influences, including mainstream linguistic values' (Bortoni-Ricardo 1985:81). Scales have been devised for measuring the degree of density and multiplexity in speakers' personal networks on the basis of indicators of network attributes such as kinship, workmates and friends. Network scores and the linguistic features of individuals have been compared. The men in the Belfast communities studied by the Milroys, if employed, worked locally and were found to have high network scores and a preference for maintaining local vernacular norms. The women in the study tended to be employed outside their immediate neighbourhood and consequently had more open networks and were more influenced by mainstream linguistic values (Milroy 1987:76).

If we look at the close-knit South Taranaki dairy farming community under scrutiny here in the same light it may seem surprising that gender differences in the patterns of pitch movement exist. Husbands and wives typically worked together as business partners on their farms. Childcare was shared by both partners. One young couple described how their preschool children played in a caravan parked alongside the cowshed, freeing up both parents at milking time.

Several of the young men commented that the amount of time they spent with their young families was one of the advantages of their rural lifestyle over life in town. Most of the speakers in the sample mixed with each other socially and one of the men and one of the women were married to each other. According to social network theory these close network ties should correlate with similar scores on the measure of pitch dynamism, but this was not the case.

Careful scrutiny of the speakers' employment histories provided clues as to the possible source of the gender differences identified. In the biographical section of the conversational interviews I conducted, interviewees were asked for details of their schooling and their employment, historical as well as current. All interviewees had been brought up on dairy farms in South Taranaki, had attended primary schools in the region followed by attendance at Opunake High School. All had spent the early years of their adulthood away from the area and had returned to take up dairy farming in their mid-twenties. On leaving high school the men attended polytechnic in New Plymouth, or in one instance Massey University in Palmerston North, to obtain agricultural and/or trade certificates/diplomas and they worked in New Plymouth as tradesmen for a few years. The women on the other hand moved to New Plymouth and worked in offices, banks or hair dressing salons. The women's jobs could be described as service jobs involving a lot of contact with people from all walks of life whereas the men's jobs could be described as practical, and everyday contact would probably have been predominantly with men in similar employment. These years of early adulthood may have been critical in the development of differing intonational patterns. The young women may have accommodated their pitch patterns to match those of the young women they were in contact with in New Plymouth. The intonation of the young New Plymouth women in the larger sample was notable for a similar lack of pitch dynamism. The young men, mixing mostly with each other or with men like themselves, maintained the intonation patterns they had grown up with.

Analysis of data from two widely separated age groups provided me with a kind of apparent-time window onto the diachronic development of the change in intonation. We know that intonation characterised by a comparative lack of pitch dynamism is the more innovative form because it is favoured by the young speakers in my larger sample and not by the elderly speakers. This would appear to conflict with the fact that the HRT, a particularly dynamic intonational feature, is also favoured by younger speakers (Britain 1992). (But note my comments regarding a possible connection between HRTs and level intonation on p7.) It appears that the young women in the Taranaki sample have adopted the more innovative level intonation and the young men have not. Analysing a very wide range of research, Labov (1990) argues that research in dialectology indicates that while women have been shown to lead in linguistic changes which are new and dynamic, they have also been shown to lag behind men in the use of variables representing older changes.

This generalisation was illustrated by Woods (1997) in a study of gender-related variation and linguistic change using real-time evidence of the

development of New Zealand English. She compared the speech, recorded in 1948, of men and women who had lived most of their lives in the Otago region of New Zealand, with the speech of some of their present day descendants who continued to live in Otago, and who were recorded in the early 1990's. The speakers in both sets of recordings were aged between 60 and 87. She analysed 30 tokens each of the MOUTH, TRAP and DRESS vowel sets for both generations of speakers. For the purposes of the study she termed the earlier generation of speakers 'first' generation and the subsequent generation 'second' generation. I shall do the same. She found that the MOUTH diphthong had become closer over time and that the first generation women were leaders in the change, but that fifty years later, when the feature had become a more established variant of New Zealand English, the second generation women were using a more open, less advanced form than the men. The vowels in TRAP and DRESS had also been raised over time with women leading the change in both cases. In the case of TRAP women of the first generation had closer articulation than the first generation men, but by the second generation the gap between the women and the men had closed. The raising of the vowel in DRESS was shown to be a more recent change and the women of the second generation were clearly leading this change.

The above analysis illustrates three distinct stages in the diachronic development of New Zealand English as spoken in Otago with women being the first to adopt innovative features which are then adopted by men over time, at which point women may have a tendency to revert to earlier forms. Woods hypothesises that women's preferred discourse strategies may provide an explanation as to why women tend to introduce new phonological features into New Zealand English. She cites research which shows that women have been found to be more co-operative and listener-orientated in face-to-face interaction than men and that they are more accommodative to their interlocutors' needs (Woods 1997:117). She suggests that in the levelling stage of New Zealand English dialect formation her first generation women would have accommodated their speech to the linguistic features of people they were in contact with, in this case, quite possibly Australians. Australians were present in Otago, especially during the time of the gold rushes, and their dialect was characterised by raised MOUTH and TRAP vowels. She also suggests that, by way of contrast, where it is found that men are leading linguistic change it is in situations of geographic restriction and comparative lack of contact with speakers of different varieties. Labov's (1972b) analysis of the centralisation of /ay/ and /aw/ in Martha's Vineyard is a case in point. Men were in the lead in this change and it was clearly not the result of dialect mixing but rather the adoption of a form associated with conservative usage on the island. Woods also compares her findings to those of Milroy and Milroy (1993) who found that 'strictly localised' variants were most frequently used by men in their Belfast study and 'supralocal' forms were favoured by the women in their sample.

The situation of the young dairy farming men in South Taranaki could in some regards be compared with that of the men in Labov's study of Martha's Vineyard (Labov 1972b). The middle-aged fishermen of Martha's Vineyard

showed a strong tendency to maintain the linguistic features typical of the older generation of fishermen in the region. This was interpreted by Labov as a linguistic response to socio-economic changes in the community. On Martha's Vineyard there had been a downturn in the fishing industry and a decline in farming and dairying with a consequent increasing reliance on tourism and summer holidaymakers for income. The linguistic response amongst the men to these changes was to demonstrate their pride in being different to the incoming flood of socio-economically privileged mainlanders by maintaining the linguistic features typical of the older generation of fishermen. The socio-economic history of the young dairy farmers in Taranaki in no way resembles that of the Martha's Vineyard fishermen in Labov's study. I draw the comparison because of the similar social isolation of the young Taranaki men and because of the pride they expressed in being from the 'Naki'. All had had a taste of the world outside South Taranaki and all had subsequently chosen to return home. All said they enjoyed their work and life-style. None had any desire to live elsewhere. These young men may be subconsciously using older, local intonation patterns as a badge of local identity.

The young women's response to a seemingly similar situation was to adopt the linguistic norms of society outside the farming community of South Taranaki, accommodating to the speech patterns of the young women they mixed with in New Plymouth. The intonation of the young women in New Plymouth was in turn very different to that of their grandmothers who had very dynamic pitch movement. The intonation of the young women in Taranaki more closely resembled that of their urban sisters in Wellington. The gender pattern of the younger age group in Wellington showed a reversal of the situation present in Taranaki. In Wellington it was the young men who favoured less dynamic pitch movement. In fact they had the most level intonation of all groups measured. It is tempting to interpret the pattern of change from more dynamic to less dynamic pitch movement as being similar to that demonstrated by Woods (1997) in Otago with regards to the MOUTH diphthong. Women may have led the change from dynamic pitch to flatter intonation in Wellington and they continue to do so in Taranaki. The pattern in Wellington may have become a more established variant and the men have adopted it, the women meantime having reverted in part to an earlier norm. With no hard evidence on this issue to date, however, this needs to be a subject for future research.

A case can be made that the young women in Taranaki are acting out the role which New Zealand women have long been playing in adopting innovative features which lead to changes in the standard accent. It has been observed that 'the female gender role of creating language standards has been operating for at least a hundred years within New Zealand' (Maclagan 2000:88). Reasons for the role of women in New Zealand in language change have been explored by Holmes: 'Whatever their origins, the success of particular innovations (as opposed to linguistic variants which never acquire the status of changes in the standard accent) appears to depend on their being adopted and endorsed by women' (1997: 131). She argued that this may be because of the greater influence Pakeha women have traditionally had in matters of social etiquette since the

early days of the colony, with women being 'expected to provide models of correct and refined linguistic usage' and that 'issues of language usage have long been regarded as more appropriately the concern of women than men in New Zealand' (ibid.). A second possible reason given for women's linguistic variants being treated as models by New Zealand society and becoming established as components of standard New Zealand English is that 'women typically interact with a wider range of social contacts than men' and 'since they are more responsive to the speech of others than men are, they are thus likely to acquire a number of linguistic variants from which innovations may develop' (ibid.: 133). These innovations, being favoured by women, then become standard forms over time and by looking at women's usage it may be possible to predict what New Zealand English will sound like in future. This being the case it would be reasonable to assume that the future standard for New Zealand English will be, if it is not already, the flattish intonation preferred by the young women of Taranaki. The distinctiveness of Taranaki's intonation currently rests with the young rural men who, for now at least, express their staunchness and pride in being from the Naki via their dynamic pitch movement.

To state the obvious, language is constantly changing. Four decades on from Labov's study of the speech of Martha's Vineyard, researchers have returned to the region and undertaken a synchronic analysis of the /ay/ diphthong studied by Labov (Blake & Josey 2003). Blake and Josey found that the earlier linguistic pattern had undergone changes in the intervening years and argued that the changes were linked to socio-economic restructuring and resulting ideological changes taking place on the island. Fishermen no longer had the status they once had, and young people did not have the same incentive to choose fishing as a profession. 'The economy of the island [had] been maintained primarily by tourism, with locals going from an oppositional to cooperative stance towards tourists' (ibid.: 481). 'The sociolinguistic findings for the present-day data base suggest that, with a change in the socio-economic structure of the Vineyard, locals' allegiance to a traditional way of life has diminished. As a consequence, there has been a decline in the linguistic marking of opposition to non-local populations' (ibid.).

Changes in the socio-economic and cultural environment of rural South Taranaki will undoubtedly occur over the coming years just as they have done in Martha's Vineyard. The linguistic response to the impact of these changes is as yet unknown. But it is not impossible that the current difference between the intonation patterns of the men and women of the area will be less obvious in future. The young men of Taranaki may follow the lead of the women of the area and decide it is cool to adopt the supralocal level pitch patterns of their urban counterparts in Wellington.

A combination of informal but close observation and the range of measures reported in this paper suggests that rural Taranaki young women's intonation is not 'high-pitched', 'shrill', 'over-emotional' or 'swoopy'. Nor, according to my operational definition, is it 'more dynamic than men's'. The analysis indicated that the women in the sample have pitch patterns which more frequently ended

with non-falling terminals than men's. But overall, the gender stereotypes quoted at the beginning of this paper appear to be unfounded with respect to this particular speech community.

References

- Ainsworth, Helen. (in prep.) *Regional variation in New Zealand English intonation*. Unpublished PhD, Victoria University of Wellington.
- Blake, Renee & Meredith Josey. 2003. The /ay/ diphthong in a Martha's Vineyard community: What can we say 40 years after Labov. *Language in Society* 32, 451-485.
- Bortoni-Ricardo, Stella Maris. 1985. *The Urbanisation of Rural Dialect Speakers*. Cambridge: Cambridge University Press.
- Britain, David. 1991. *Dialect and space: a geolinguistic study of speech variables in the Fens*. PhD dissertation, Essex University, Colchester.
- Callary, Robert. 1975. Phonological change and the development of an urban dialect in Illinois. *Language in Society* 4, 155-169.
- Cruttenden, Alan. 1997. *Intonation*. Cambridge: Cambridge University Press.
- Daly, Nicola & Paul Warren. 2001. Pitching it differently in New Zealand English: speaker sex and intonation patterns. *Journal of Sociolinguistics* 5(1), 85-96.
- Henton, Caroline. 1989. Pitch dynamism in male and female speech. *Language and Communication* 15(1), 43-61.
- Holmes, Janet. 1997. Setting standards: sound changes and gender in New Zealand English. *English World-Wide* 18(1), 107-142.
- Horvath, Barbara M. & Ronald J. Horvath. 2002. The geolinguistics of /I/ vocalization in Australia and New Zealand. *Journal of Sociolinguistics* 6(3), 319-346.
- Labov, William. 1966. *The Social Stratification of English in New York City*. Washington DC: Centre for Applied Linguistics.
- Labov, William. 1972a. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1972b. The social motivation of sound change. In William Labov *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press. 1-42.
- Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2, 205-254.
- Ladefoged, Peter. 1997. Instrumental techniques for linguistic phonetic fieldwork. In William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell Publishers. 137-166.
- Maclagan, Margaret A. 2000. How long have women been leading language change? In Janet Holmes (ed.), *Gendered Speech in Social Context: Perspectives from Gown and Town*. Wellington: Victoria University Press. 87-98.

- McConnell-Ginet, Sally. 1983. Intonation in a man's world. In Barrie Thorne, Cheri Kramarae and Nancy Henley (eds). *Language, Gender and Society*. Massachusetts: Newbury House Publishers. 69-88.
- Milroy, James & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 25, 339-384.
- Milroy, James & Lesley Milroy. 1993. Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics* 3, 57-77.
- Milroy, Lesley. 1987. *Observing and Analysing Natural Language. A Critical Account of Sociolinguistic Method*. Oxford: Basil Blackwell, Inc.
- Tench, Paul. 1996. *The Intonation Systems of English*. London: Cassell.
- Trudgill, Peter. 1983. *On Dialect: Social and Geographical Perspectives*. Oxford: Basil Blackwell.
- Warren, Paul & Nicola Daly, 2000. Sex as a factor in rises in New Zealand English. In Janet Holmes (ed.) *Gendered Speech in Social Context: Perspectives from Gown and Town*. Wellington: Victoria University Press 99-113.
- Warren, Paul & Nicola Daly, 2004. Characterizing New Zealand English intonation: broad and narrow analysis. To appear in Allan Bell, Ray Harlow & Donna Starks (eds.) *Languages of New Zealand*. Wellington: Victoria University Press.
- Woods, Nicola J. 1997. The formation and development of New Zealand English: interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics* 1(1), 95-125.

The ups and downs of Kiwis: An experiment investigating tonal cues used to identify NZE intonation

Patricia Vermillion

Abstract

The aim of the experiment reported in this paper was to investigate which tonal cues are indicative of New Zealand English (NZE) intonation in comparison to British English (BrE). Fourteen potential tonal cues were chosen for manipulation. Results from a rating task demonstrated that listeners used several tonal characteristics when differentiating NZE from BrE. The principal cues were the relative height of the final intonation phrase boundary and the height of the phrase-medial low values. Specifically, the lower of the two different values of a low phrase boundary (L%) was interpreted as sounding more like NZE; conversely if the final target was high (H%), the higher target was believed to sound more like a NZE speaker. For phrase-medial low values (L), it was found that higher L values indicated NZE. From these results, we assert that native NZE listeners use several tonal cues, including the height of H%, L% and L values, when determining NZE as distinctive from BrE. This indicates NZE as having relatively high pitch values within an intonation phrase with marked departures from this to either low L% or high H% phrase-final boundaries.

Introduction

A number of phonetic cues help us determine which variety an English speaker is using. These include the phonetic realisations of different segments, such as the use of a postvocalic /r/ in *car*, as a marker of General American English in comparison to the English spoken in many parts of New Zealand. There is also semantic variation of certain lexical items; *bush*, for example, refers to a "native forest" in New Zealand English (NZE) and a "shrub" in British English (BrE). In addition, listeners may also use the intonational pattern of a word or phrase, such as a rising terminal contour on a statement which has been claimed to be widely used in NZE (amongst other varieties) rather than a falling tune, which has been claimed to be a "universal" marker of declaratives (Gussenhoven forthcoming).

Past research conducted on NZE intonation has shown several prosodic features to be indicative of this variety. The most noted feature is the High Rising Terminal Contour (HRT), or the high rising pitch at the end of a phrase which is realised on statements (Allan 1990; Britain 1992; Britain & Newman

1992; Warren & Britain 2000). Timing also appears to be a factor; Ainsworth (1993) and Warren (1998, 1999) indicate that there is a rhythmic distinction between NZE and BrE. Ainsworth examined strong vs. weak vowel productions in function words, comparing news readings from several radio stations. She found that Maori NZE news-readers produced more full vowels than Pakeha¹ NZE news-readers, who produced more full vowels than BBC news-readers. Warren (1999) confirmed these findings in acoustic measurements of the same newsreader data.

Although there has been research conducted on NZE intonation, it is relatively sparse. The aim of this experiment is to serve as a starting point for future examinations, looking at various features of intonation as possible factors which may be indicative of NZE intonation. Using native speakers as expert judges, this study examines explicitly different tonal cues which listeners use when distinguishing NZE intonation from BrE.

As an exploratory experiment, a range of possible factors was examined. The factors examined can be described with reference to the example pitch contour in Figure 1. The assessed potential indicators are as follows: [1] High pitch accent values (pitch points 2 and 5), [2] pitch range, [3] Low pitch values between pitch accent peaks (pitch points 3 and 4), [4] phrase-initial intonation phrase boundaries (pitch point 1), including Low values, and [5] High values, [6] phrase-final phrase boundaries (pitch point 6), including Low values, and [7] High values, [8] pitch level, [9] number of peaks, [10] Peak alignment on earlier peaks (i.e. high peak occurring on the first syllable of the utterance, at pitch point 2), [11] Peak alignment on later peaks (i.e. high peak occurring on the penultimate syllable in the utterance, at pitch point 5). A final set of manipulations involve the relative steepness of the pitch slope, or *dynamism* at three locations within the phrase: [12] dynamism on the High phrase-final boundary (pitch point 6), [13] dynamism on the High phrase-initial boundary (pitch point 1) and [14] dynamism on a High phrase-internal target.

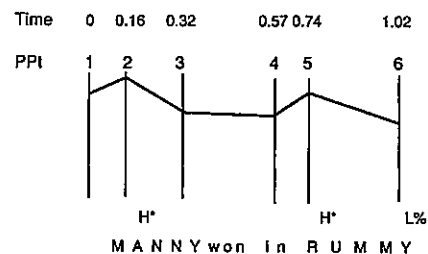


Figure 1. Schematic representation of source utterance (Time = sec, PPI = Pitch point)

Method

The notation system adopted for this experiment is similar to that of the Tone and Break Indices (ToBI) system (Pierrehumbert 1980; Beckman & Ayers 1994) which describes intonation according to tone targets, such as High 'H' pitch targets and Low 'L' pitch targets. Boundary specifications note the end of an intonation phrase (IP) with the diacritic '%' appearing after the tone target (i.e. L% indicates a Low IP-final boundary specification) whereas '%' before the L or H tone indicates the beginning of the phrase (i.e. %L indicates a Low IP-initial boundary). Pitch accents, or pitch movements focused on the perceptually salient tone are indicated with a '*' (e.g. H*).

The source utterance used in this study was taken from an adult male speaker of RP British English. He was instructed to read aloud from a provided script, "as naturally as possible" whilst being recorded onto a Maxell 74 minidisk using a Sony MDS-JB930 minidisk recorder and a Sony unidirectional microphone. The speaker was recorded in a recording studio in the Language Learning Centre at Victoria University of Wellington.

The segmental makeup of an utterance invariably affects the F0 trace. Voiceless obstruents, for instance, are not evident on the F0 trace due to the absence of voicing whereas plosives cause the F0 trace to rise after the stop is released. Such segmental effects often prove problematic for synthesising speech though they are easily and unconsciously normalised by listeners' perceptual system. The sonorant utterance *Manny won in rummy* was selected as the base utterance for the experimental materials, in order to avoid such complications. This utterance was digitised at a sampling rate of 16 kHz in mono. Using the signal processing system Praat (Boersma & Weenink 1996), the source utterance was stylised and the six pitch points (PPI) were located at tonal movement positions (see Figure 1).

Pitch Points	1	2	3	4	5	6
Male	3.80	4.52	3.90	3.84	4.15	3.11
Female	5.69	6.41	5.78	5.73	6.04	4.99

Table 1. Source Pitch Points (in ERBS) for 'male' and 'female' stimuli

To compare male and female pitch levels, the Total Rescaling method (Ladd & Morton 1997), or an up/down-scaling of all pitch points within the contour, was adopted. The six pitch points from the source utterance were adjusted to one of two levels using pitch point values as shown in Table 1.

¹ New Zealander of European descent.

This produced two pitch levels (PLs) reflecting the averages of ten male (3.98 ERB)² and ten female (5.83 ERB) NZE speakers reading aloud a story.

The intonation of the source utterance was further manipulated in order to assess which intonational variations may be used by listeners to identify NZE. Being an exploratory experiment, a wide range of different tonal characteristics was assessed, as indicated above and detailed here below.

First, eight sets of manipulations involved setting pitch targets to high (H) or low (L) values, relative to the source PPTs indicated in Table 1. The range between these high and low values was set at 1.25 ERB. As noted earlier, studies examining and defining NZE intonation are relatively sparse. Due to this lack of information, the two contrasting high and low values were not based on actual values used by NZE speakers, but were selected according to what appeared natural to the author, which was later confirmed by two native NZE speakers. Specifically, the adjustments were as follows (illustrated in Figure 2).

First, the height of the H* targets were compared (Fig 2a). The two H* targets (PPT 2 and PPT 5, Table 1) were adjusted to levels 0.25 ERB below and 1.00 ERB above the source PPT values, creating low and high H* targets, respectively. We also examined whether the values to which pitch is allowed to fall between two peaks are used as indicators of NZE; thus L values at PPT 3 and PPT 4 were lowered by 0.75 ERB or raised by 0.50 ERB to create low and high L values (Fig 2b). The height of the final boundary tone was also assessed as a possible determiner of NZE (PPT 6 in Fig 1). A final boundary tone can be either L or H. The low final boundary tone (L%) was set to 0.75 ERB below or 0.50 ERB above the source PPT values, creating a low L% and a high L% (Fig 2c). In addition, the relative height of a high final boundary tone was assessed (Fig 2d) by setting the boundary tone (now designated H%) to 1.50 ERB or 2.75 ERB above the source value, creating a low and a high H%, respectively. Similar to the phrase-final boundary, the height of the phrase-initial boundary target (PPT 1 in Fig 1) was examined as a possible factor. A low phrase-initial boundary (%L) value was adjusted to one of two values relative to the source PPT: creating a high and low %L, the PPT was set 0.25 above and 1.0 below the source PPT 1, respectively (Fig 2e). An initial high boundary tone (%H) was set at 0.75 and 2.00 ERB above the source PPT 1, creating both a relatively low and a high %H (Fig 2f).

Adjusting the heights of tonal values inevitably affects the mean F0 over the utterance as a whole. Therefore, we also examined explicitly how Pitch Level affects judgments of the stimuli as being perceived as NZE or BrE (Fig 2g). Pitch levels (PL) were adjusted either to a low level by lowering all PPTs by 0.625 ERB or to a high level by raising the PPTs values 0.625 ERB above the

² The ERB (Equivalent Rectangular Bandwidth) scale is used as it has been shown to be perceptually more relevant than a physical Hertz scale.

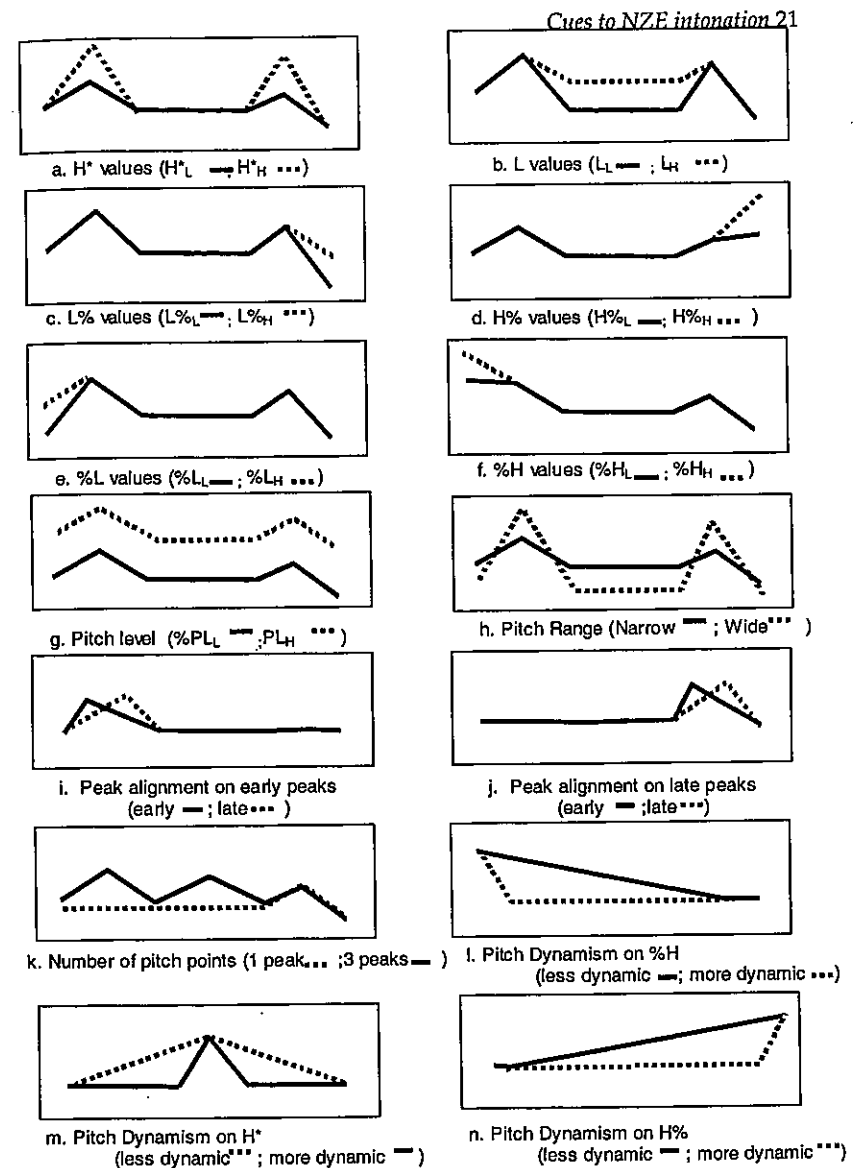


Figure 2. Schematic representations of the 14 manipulations used in the experiment.

source PPTs. The resulting average F0s for male and female PLs were 3.35 and 5.23 for PL_L, and 4.57 and 6.48 for PL_H, respectively.

Pitch range may also be an intonational feature which listeners use when distinguishing NZE. Effects of pitch range can be explored while keeping

pitch level (relatively) constant if high and low values are adjusted around a midpoint. The H and L values were raised and lowered, respectively, by 0.625 ERBs from the source utterance PPTs, creating a wide range (Figure 2h). It was determined by the experimenter that further raising and lowering both L and H values, respectively, made pitch movements difficult to hear. Therefore, High and Low values in the source utterance were not adjusted for a narrow range in order to maintain perceivable pitch movement in the contour.

In addition to adjusting PPT values, the timing of pitch movements and the number of peaks were also manipulated. Previous studies have suggested that alignment of the pitch movement may affect the meaning of an utterance. One characterization of the difference between a fall and a rise fall is as a delay in the peak of a nuclear pitch accent, conveying non-routineness (Gussenhoven 1983). Similarly a delayed pitch movement on a rise, contrasting an early rise and a late rise may indicate whether a rising IP-final contour is interpreted as a statement or a question (Zwartz, this volume). In the current study, peak alignment was also assessed as to whether the NZE variety of English is distinguishable from BrE according to the alignment of tonal material with the text. In an attempt to make variations in peak alignment easier to perceive, the original pitch movement positions (Figure 1) were adjusted for each manipulation (c.f. Fig 2i and 2j). The location of each PPT is explained here as seconds after the onset of the source utterance voicing (0 seconds).

Peak alignment was assessed in two locations. First, the F0 peak occurring early in the phrase was examined (Fig 2i). Specifically on the first vowel in the utterance *Manny* (PPT 2, Figure 1), the F0 peak was adjusted to the either the onset of the vowel, located at 0.06 sec after the voicing onset of the utterance, or the end of the vowel, found at 0.16 sec after the voicing onset. The F0 peak occurring later in the IP, specifically on the first vowel of the utterance *rummy* (PPT 5, Figure 1), was also assessed (Fig 2j); the peak was aligned to the onset of the vowel, located at 0.76 sec. after the onset of voicing of the utterance, or at the end of the vowel, found at 0.83 sec. In total, the manipulation of peak alignment created 4 separate stimuli.

It has also been claimed that NZE is more "melodic" than other varieties of English (Holmes and Ainsworth 1996). One factor which may result in such claims is the number of pitch movements in an intonational phrase (see Ainsworth 2003). For example, if more contrasting H and L targets are produced in an utterance, the utterance may appear to be more dynamic and musical than a more monotone utterance with fewer contrasting targets. In order to assess such an assumption, the incidences of peaks were examined as a possible indicator of NZE (Fig 2k). Accordingly, the IP was manipulated to include either 1 F0 peak by lowering the first H value (PPT 2) to the following L value (PPT 3), which reduced the number of H* targets from 2 in the source to 1. Relative the source PPTs given in Table 1, an utterance with 3 H peaks was established by creating an additional H peak located in between PPT3 and

PPT 4, specifically situated at 0.45 sec. The natural effects of declination, or the reduction of subglottal air pressure over a breath-unit, and subsequent downstep of H values were considered when determining the height of this further H peak; therefore the height of this peak was set to a value half-way between the two original H* values given in Table 1 (PPT 2 and PPT 5), creating a downstep effect on the three peaks.

Dynamic pitch has been noted as a marker of social dialects, in particular women's speech (Austin 1965; Brend 1975, see also Ainsworth 2003) and homosexual male speech (Barrett 1997). In the current study, pitch dynamism is examined as a possible distinguishing feature of NZE, distinctive from BrE according to the steepness of pitch movement. Dynamism at three locations within the phrase was subsequently examined, including phrase-initial, phrase-internal and phrase-final locations (Figures 2l, 2m and 2n, respectively). In an attempt to make variations in dynamism greater and subsequently easier to perceive, the original pitch movement positions (Figure 1) were adjusted for each manipulation and PPTs different from Table 1 were adopted. Two PPT values were used for all H and L targets in assessing dynamism. Specifically, the height of any H target PPT was set to 4.52 and 6.41 ERB for male and female speakers, respectively, and the height of any L target PPT was set to 3.80 and 5.69 ERB for male and female speakers. The location of each PPT is explained here as seconds after the onset of the utterance voicing (0 seconds).

To assess dynamism at IP-initial position (Fig 2l), three PPTs were used; the first PPT (H target) was set at 0 sec, followed by an L target producing either a more dynamic pitch movement (the target was at 0.3 sec) or a less dynamic pitch movement (the target was at 1.0 sec), which was trailed by a final L target located at 1.02 sec. Three PPTs were also used to create a more and less dynamic stimuli at phrase-final locations (Fig 2n); the first PPT (L target) was set at 0 sec, the second L target was set to create either a less dynamic pitch movement (starting at 0.02 sec) or a more dynamic pitch movement (starting at 0.72 sec) and the final H target was set at 1.02 sec. For a more and less dynamic pitch movement at IP-internal position, five PPTs were used (Fig 2m). For a less dynamic H*, two L targets, located at 0 and 0.09 sec (respectively), were followed by a H target set at 0.49, and two L targets located at 0.89 and 1.02 seconds after the onset of the utterance. Contrasting the less dynamic H*, a more dynamic H* was created: two L targets, located at 0 and 0.39, were followed by the H target (again located at 0.49 sec), and two L targets set at 0.59 and 1.02 sec. Collectively, dynamism was assessed as relatively more dynamic and less dynamic in three locations within an IP, including phrase-final, phrase-initial and phrase internal.

Each of the 14 manipulation types shown in Figure 2 had two settings, giving 28 stimuli for each of the two voices (male and female), creating 56 stimuli in all. Two native NZE speakers listened to each manipulation and confirmed that they created "natural" sounding utterances. The manipulated stimuli were then Low-Pass Filtered (LPF) using Cool Edit (cut-off frequency

110 Hz, filter order 3.5). The filter used in the current experiment was previously used in an experiment by Rietveld and Vermillion (2003), who confirmed that listeners are able to distinguish the height of the H values at the different pitch levels associated with male and female voices.³ After filtering the stimuli were amplified by 500%, to ensure the pitch information was audible.

The "female" and "male" stimuli were randomised separately; separating the sexes and indicating the sex of the speaker at the top of each answer sheet ensures that the listeners are rating the appropriate sex (in case they perceive a different sex than that one intended). One half of the completed recordings had "female" speakers played first while the other played "male" speakers first. Prior to every fifth stimuli, a 0.5 second beep was inserted in order for the participants to determine whether they were keeping pace with the recording. Each stimulus was separated with an interval of 3.5 seconds. The first five stimuli were played a second time within the experiment and the answers given for the first set were discarded from the included data; this was done to discount erroneous answers often made at the beginning of experiments. In total, the experiment extended over a period of approximately 9 minutes, plus a 5-minute introduction.

Participants

56 participants took part in this experiment on a voluntary basis. A questionnaire was used to determine social variation amongst the participants. The data provided by 30 native NZE speakers were included in this experiment. Participants' datasheets were discarded if they were not a native NZE speaker. In addition, they were not included in the dataset if the participant did not complete both this experiment and a related attitude experiment (not discussed in this paper). Six participants in the selected group were male and the other twenty-four were female. All were students at Victoria University of Wellington. 27 of the selected group were aged between 18-22 years and three were 23-29 years old. None of the listeners had any known hearing impairments.

Procedure

Participants were told that they would hear a number of speakers saying one sentence, which were played over a loudspeaker. The participants were told to mark on a given scale "how much the speaker sounds like a New Zealand English speaker" with the left side of the scale indicating that "the speaker does not sound like a NZE speaker" and the right side of the scale indicating that "the speaker sounds very much like a NZE speaker".

³ Although the filter has a "cut-off" of 110Hz, the slope of the filter used ensures that pitch-relevant information above the cut-off frequency, including the range for female voices, is still present.

To assess the listeners' responses, we used a Magnitude Estimation Scale. This was provided by a 100mm line on the answer sheets, between the two labels for the scale ends. This type of scale does not assume equal intervals of perception (Zraick and Liss 2000). The listeners were asked to mark on the line their impressions of each stimulus. The introduction to the experiment, which was read aloud by the author, identified the variety of NZE that we were investigating. Specifically, the listeners were asked to identify a variety of NZE that is spoken by Pakeha speakers with a 3-year University degree and from the Wellington area. We asked them to compare this variety to one other variety of English (RP British English) in order to give all of the listeners a similar comparative model. That is to say, we wanted the listeners to indicate what makes NZE distinct from one specific variety rather than having different listeners judging NZE against different comparison varieties. We avoided indicating a specific city or town for the RP comparator in the introduction, as listeners may know a person from this place who may speak a different variety from the one which we are attempting to locate. The listeners were provided with samples of speech from a standard BrE speaker (educated speaker from South England) and a NZE speaker (educated speaker from Wellington, NZ) in order to illustrate the varieties of English we were referring to. Both samples of speech were taken from female speakers, however speech from neither one was used as source data for this experiment. These example speech samples were not altered or filtered in any way, as we wished to indicate clear examples of both varieties of English.

The introduction was also used to provide motivation for the listeners. In particular, the listeners were told that we wanted to determine what a NZE accent is in order to synthesise a realistic sounding NZE accent on the computer. The participants were told that each utterance had been read aloud by one of two British English speakers: one male and one female. The purpose of using British speakers, rather than NZE speakers was to reduce the danger of all stimuli being regarded as having a "NZE accent", which might have resulted had we said they had been derived from the speech of a NZE speaker. By using a British speaker as our source utterance rather than a speaker of some other variety, we are also priming a comparison to be made between NZE and BrE, which is the comparative aim of this experiment. The participants were also told that each utterance was then put onto a computer and changed in some way. Their task was to determine which of these manipulated stimuli sounded like a NZE speaker and which did not.

Scoring

Scores were measured to the nearest whole millimetre from the 100mm response scale. Scores could range from 0 (for "not at all like a Kiwi") to 100 (for "very much like Kiwi"). These scores were then entered into separate Analyses of Variance (repeated measures design, Huynh-Feldt corrected) for each pair of stimuli (as determined by the manipulations listed above, e.g. the H_H and H_L versions of the H* height manipulation formed one such pair). In each case the independent variables were the two within-subject factors: Sex

of Speaker (2 levels: male and female, based on the initial manipulation of pitch level of the source utterance, as indicated in Table 1) and Manipulation (2 levels: high and low, as described in the Method section for each manipulation).

Note also that there are three points of comparison that involve similar manipulations. Two contrasting pairs are manipulations at the same location within the phrase; they are the phrase-initial boundary (high and low values of %L and %H) and the phrase-final boundary (high and low values of L% and H%). Another contrasting pair is a mirrored change in pitch movement at two opposite ends of the phrase, specifically pitch dynamism at %H and H% (less and more dynamic pitch movements on %H and H%). For each of these points of comparison an additional three-way ANOVA (Huyhn-Feldt corrected) was carried out, with Sex of speaker (2 levels), Manipulation (2 levels) and Tune (2 levels). Specifically, the two compared levels of Tune were L vs. H for the initial and final boundary and %H and H% for pitch dynamism. It should be noted that statistics were not performed repeatedly on all of the responses. Specifically, statistics were performed first on the contrasted pairs (once per pair), and secondly on the three *a-priori*, or planned three-way comparisons.

Results

When analysing the fourteen separate phonetic variables, two factors turned out to be significant: H% values and L values. Of these two sets of results, those for the height of the L values ($F_{1,29} = 5.489$, $p = 0.026$) showed the larger effect size ($\eta^2 = 0.053$). Comparing the values for the low and high L values, the response averages were 36.75 and 44.82 respectively, illustrating that the listeners, on average, believed that higher L values are more indicative of NZE than lower L values (see Figure 3).

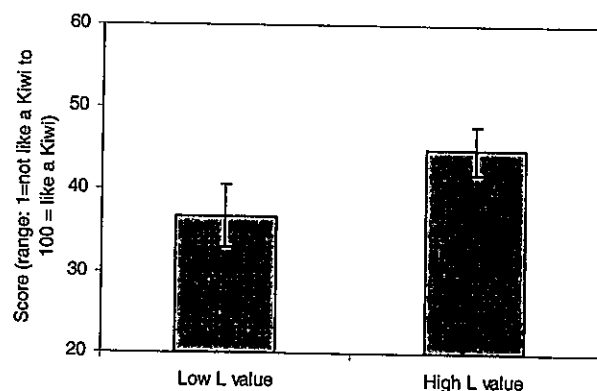


Figure 3. Mean scores for low and high L values (higher score = more NZE-like)

The second largest effect size ($\eta^2 = 0.013$) was the height of the H% values ($F_{1,29} = 5.075$, $p = 0.032$), illustrating a response average of 41.73 for low H% and 52.27 for high H% (Figure 4). In other words, the higher the H% values, the more likely the listener would be to interpret the speaker as a NZE speaker.

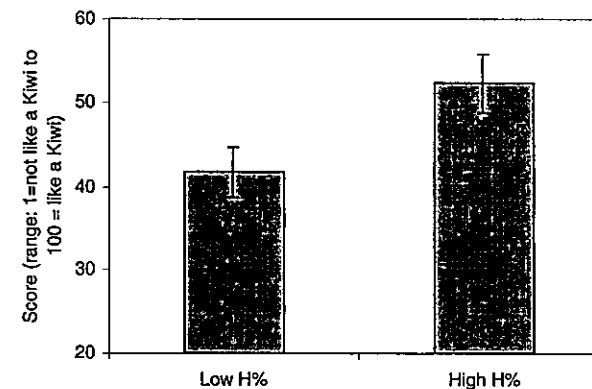


Figure 4. Mean scores for low and high H% values (higher score = more NZE-like)

In the three separate "contrasting pairs" comparisons, there was one significant interaction, which involved Tune and Manipulation for phrase final targets, with an effect size of 0.043 ($F_{1,29} = 6.190$, $p = 0.019$, see Figure 5).

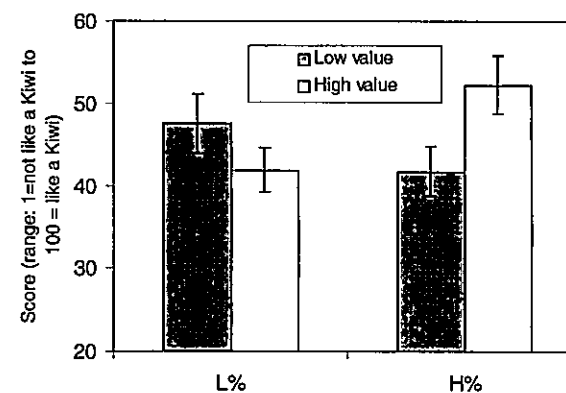


Figure 5. Mean scores for low and high L% and H% values (higher score = more NZE-like)

As shown in Figure 5, listeners were more likely to believe the stimuli was a NZE speaker if the phrase final low target (L%) was lower and conversely, if the phrase final high target (H%) was higher.

Discussion and conclusions

From these results, we assert that native NZE listeners use several tonal cues, including the heights of H%, L% and L values when determining NZE as distinctive from BrE. Specifically, a higher value for a high final boundary target (H%) was more characteristic of a NZE speaker. This is compatible with several studies that have asserted that high rising terminal contours (HRTs) are characteristic of NZE (Ainsworth 1994; Allan 1990; Britain 1992; Warren and Britain 2000). In addition, the lower value for a final low boundary target (L%) was interpreted as sounding more like NZE. These two results may be due to the rate of change at IP-final positions; the steepness of the F0 change is greater the higher the H% and the lower the L%. However, pitch dynamism as measured separately (see Figures 2l, 2m and 2n) was found not to be a significant factor in this experiment.

The results may also be illustrating that pitch range may be a factor; rather than the steepness of the pitch change, the important factor may be the pitch span of the final target, which is considerably larger for lower L% and higher H% than the alternatives. Similarly, Warren and Daly (2004, p20 of manuscript) assert that "...the extent of pitch movement is considerably larger for NZE speakers [than for either Cambridge or Leeds]". However, two results appear to conflict with the supposition that pitch range is a factor. First, pitch range was examined as a possible main effect in this experiment by adjusting all of the pitch values in the phrase to either a wide or a narrow position. The results indicated that global pitch range was not used by listeners in determining NZE. Second, the height of the phrase-medial low values (L) was also a significant factor; higher L values were found to indicate NZE. As evident in Figure 2b, higher L values illustrate a more shallow pitch movement within the phrase.

It may be that the two contrasting variables, specifically wide and narrow pitch range, are bound to the location within a phrase; although wide pitch movements at phrase-final locations were found to be indicative of a NZE speaker, within the phrase high L values and subsequent narrow movements were the expected norm.

It should also be noted that the method used in this experiment may have unduly influenced the results. As noted in the method section, the two contrasted values selected for the manipulated test stimuli were not based on actual values commonly reached by NZE speakers. These values were selected according to what appeared natural by the author. As a result, specific features used to distinguish NZE intonation, such as a wider pitch range than that created, may not have been available. Examining intonation features using numerous successive steps rather than two contrasting variables may illustrate other features which are used to indicate NZE intonation. This experiment was exploratory in nature, and future comparative investigations will further investigate the findings reported here.

Also, in measuring listeners' perceptions, their beliefs about the speaker group may have affected the outcome. For instance, if a listener believes that NZE speakers illustrate more communicative support in conversation than BrE speakers do, their response indicating that NZE speakers realise higher H% values may be due to this belief; higher H% has been found to indicate more concern and greater interest in the conversation than lower H% values (Vermillion, in prep). In order to examine such an effect, listeners' beliefs about NZE speakers in comparison to BrE were assessed in separate questionnaire given to the same participants. An analysis of the data illustrated that listeners' beliefs may have had influence on the results reported here (Vermillion, in prep.).

Nonetheless, the results illustrate which features of intonation native NZE speakers believe to be indicative of NZE. Examining these intonation features further may prove beneficial to the understanding of NZE intonation and the meanings conveyed by such use. In order to assess the validity of these perceptions, it would be necessary to compare BrE and NZE intonation use, specifically examining whether NZE speakers realise higher L values and wider pitch movements at phrase-final boundaries. Further examining the features which speakers believe are indicative of their variety may also be valuable in understanding the perceived meanings of intonation, investigating specifically what is interpreted by such use of intonation, and whether these meanings are the same for speakers of a different variety of English, such as BrE.

References

- Ainsworth, Helen 1993. *Rhythm in New Zealand English*. Unpublished BA (Honours) Dissertation, Wellington: Victoria University.
- Ainsworth, Helen 1994. The emergence of the High Rising Terminal contour in the speech of New Zealand children. *Te Reo* 37, 3-20.
- Ainsworth, Helen 2003. How she says it and how he says it - differences in the intonation of dairy farming women and men in South Taranaki. *Wellington Working Papers in Linguistics* 15, 1-15.
- Allan, Scott 1990. The rise of New Zealand intonation. In Allan Bell and Janet Holmes (eds), *New Zealand Ways of Speaking English*. Clevedon: Multilingual Matters, 115-128.
- Austin, William 1965. Some social aspects of paralanguage. *Canadian Journal of Linguistics* 2, 31-39.
- Barrett, Rusty 1997. The Homo-genius Speech Community. In Anna Livia and Kira Hall (eds), *Queerly Phrased: Language, Gender & Sexuality*. Oxford: OUP, 181-201.
- Beckman, Mary & Gayle Ayers 1994. *Guidelines for ToBI labeling (Vol 2)*. Columbus: Ohio State University Linguistics Dept.

- Boersma, Paul & David Weenink 1996. *PRAAT: A system for doing phonetics by computer*. Amsterdam: University of Amsterdam.
- Brend, Ruth 1975. Male-Female Intonation Patterns in American English. In Barrie Thorne & Nancy Henley (eds), *Language and Sex: Difference and Dominance*. MA: Newbury House, 84-87.
- Britain, David 1992. Linguistic change in intonation: the use of HR terminals in New Zealand English. *Language Change and Variation* 4, 77-104.
- Britain, David & John Newman 1992. High rising terminals in New Zealand English. *Journal of the International Phonetic Association* 22(1/2), 1-11.
- Gussenhoven, Carlos 1983. *On the Grammar and Semantics of Sentence Accent*. Dordrecht: Forris.
- Gussenhoven, Carlos, forthcoming. Why question intonations rise and why they sometimes fall. In Michael Hoey and N'elia Scott (eds), *Questions: Multiple Perspectives on a Common Phenomenon, Language and Discourse Series*. Liverpool: University of Liverpool Press.
- Holmes, Janet & Helen Ainsworth 1996. Syllable-timing and Maori English. *Te Reo* 39, 75-84.
- Ladd, Robert & Rachel Morton 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25, 313-342.
- Pierrehumbert, Janet 1980. *The Phonology and Phonetics of English Intonation*. PhD Thesis: MIT, Cambridge, USA.
- Rietveld, Toni & Patricia Vermillion 2003. Cues for Perceived Pitch Register. *Phonetica* 60, 261-272.
- Vermillion, Patricia 2001. *The Perception and Production of Intonational Meaning by British Men and Women*. MPhil Thesis: University of London, UK.
- Vermillion, Patricia, in prep. *The Meaning of New Zealand English Intonation*. PhD Thesis: Victoria University of Wellington, New Zealand.
- Warren, Paul 1998. Timing patterns in New Zealand English rhythm. *Te Reo* 41, 80-93.
- Warren, Paul 1999. Timing properties of New Zealand English rhythm. *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1843-1848*.
- Warren, Paul & David Britain 2000. Intonation and Prosody in New Zealand English. In Allan Bell and Koenraad Kuiper (eds), *Varieties of English around the World: New Zealand English*. Amsterdam/Philadelphia: John Benjamins, 146-172.
- Warren, Paul & Nicola Daly 2004. Characterizing New Zealand English intonation: broad and narrow analysis. To appear in Allan Bell, Ray Harlow & Donna Starks (eds.) *Languages of New Zealand*. Wellington: Victoria University Press.
- Zraick, Richard & Julie M. Liss 2000. A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language and Hearing Research* 43, 979-988.
- Zwartz, Joel 2003. This is a statement? Lateness of rise as a factor in listener interpretation of HRTs. *Wellington Working Papers in Linguistics* 15, pp-pp.

Wanna-contraction and prosodic disambiguation in US and NZ English

Paul Warren (Victoria University of Wellington), Shari Speer (Ohio State University) and Amy Schafer (University of Hawai'i)

Abstract

This paper presents results from a game-based study, conducted in both New Zealand and the United States, demonstrating how naïve speakers use prosodic phrasing to disambiguate syntactically ambiguous sentences. The particular set of results discussed in this paper concerns how speakers differentiate two types of *wh*-questions, as in (i) and (ii).

(i) Which triangle₁ do you want ____₁ to change the position of the square?

(ii) Which triangle₁ do you want to change the position of ____₁ this time?

Our results show that *wanna*-contraction was more frequent for Midwestern American English speakers than for New Zealand English speakers, and more likely in both data sets when there was no syntactic gap between *want* and *to*, as in (ii). Phonetic analyses indicated that both groups of speakers consistently lengthened the word preceding the relevant gap location and the following silence in both (i) and (ii). ToBI transcriptions showed prosodic constituent breaks most often patterning with the location of major syntactic breaks. While our results need not imply that syntactic gaps are directly pronounced in the phonetic structure of an utterance, it seems likely that prosody-syntax correspondence rules may result in phonological regularities in naïve speakers' productions of gap sentences, and that these regularities are available to help comprehenders locate gaps in *wh*-questions.

Wh-gaps and prosodic marking

The ambiguous question in (1) has two readings, indicated by (a) and (b). A commonly accepted syntactic analysis of the ambiguity (e.g. Chomsky, 1977) is that the different readings correspond to movement operations of *wh*-constituents from different positions in the underlying form of the sentence. Thus in the (a) reading, a fronting operation moves the *wh*-word from an underlying subject position before *to leave*, while in the (b) reading it is moved from an object position after *leave*.

(1) Who does John want to leave?

(a) John wants *who* to leave?

(b) John wants to leave *who*?

It has also been argued that the movement operation leaves a "trace" (or *wh*-gap) at the position of the *wh*-word in the underlying structure, which will block the operation of phonological rules that require adjacency of the words either side of the trace. One such phonological rule might result in the contraction of *want to* to *wanna*. Note however that fast speech processes such as *wanna*-contraction are optional (Fodor, 1979), and cannot therefore provide

determinate cues to a gap between *want* and *to*. That is, the absence of contraction need not mean that there is a gap. Nevertheless, it has been widely argued that the presence of contracted forms in utterances of (1) would indicate that *want* and *to* are in the same constituent, with no intervening trace, thus disambiguating the utterance towards the (b) reading indicated above (Baker & Brame, 1972; Lakoff, 1970). However, it also appears that the presence of contraction does not necessarily signal a gap-less interpretation. For instance, Karins and Nagy (1993) tested listeners' interpretation of *wanna*-contraction. They presented a short story followed by a question with or without *wanna*-contraction (*Who do you (want to / wanna) help?*). In the context of the story, the *wh*-constituent in the question was compatible with both subject and object roles, though the materials were designed to show a preference for the subject role. The experiment showed that listeners were statistically more likely to select the subject interpretation (*you want X to help*) in the absence of contraction and the object interpretation (*you want to help X*) if there had been contraction, showing that the presence or absence of contraction does have a strong cueing function. However, the contracted form was given a subject interpretation in some 32% of cases, demonstrating that the relationship between the syntactic and phonetic forms is not categorical.

Pullum (1997) argues for a non-syntactic explanation for the incidence of various types of contraction, including *wanna*-contraction, and posits that forms such as *wanna*, *gonna*, *hafta*, etc. are separate lexemes synchronically related to *want to*, *going to*, *have to* etc. Pullum suggests that there are 'liberal dialects' that have a wider distribution of contracted forms than other dialects, including contexts in which more traditional analyses have argued that contraction should not be found. He argues that contraction in these dialects is constrained by phonological phrasing rather than by syntactic factors. It is possible therefore that the non-categorical responses reported by Karins and Nagy reflect such dialect (or idiolectal) differences between individual listeners. This is an explanation offered by Bley-Vroman and Kweon (2002) for over-general use of *wanna* by the native speakers tested in their study of learner knowledge of *wanna*-contraction.

In addition to contraction, a gapping site may also be marked by further phonetic properties of the utterance, such as may be expected in the presence of some (perhaps low-level) phonological or prosodic boundary between *want* and *to*. This was indeed found in an early study by Danly (1980), in which participants read ambiguities similar to that in (1), following disambiguating paragraphs intended to make the meaning clear. For non-reduced tokens in Danly's data set, the sequence *want to* was some 25msec longer in the (a) reading, where the syntactic analysis argues for a trace between these two words.

Recent research has disputed the conditions under which *wh*-gaps might be marked prosodically. A study by Nagel et al (1994) analysed words in the position of *call* in sentences like (2), where the gap locations in each sentence are indicated by Δ . Like Danly, they found longer durations of these key words in the (a) versions. They also found greater pitch movements over these words.

(2a) Which doctor did the supervisor call Δ to get help for his youngest daughter?

(2b) Which doctor did the supervisor call to get help for Δ during the crisis?

Subsequent studies by Straub et al (2001) argued that the prosodic differences reported by Nagel et al for sentence pairs such as (2) derive not from gapping, but from other factors that are known to affect intonational phrasing. They pointed out that the infinitival clause (*to get help for his youngest daughter*) in the (a) version is an adjunct to the verb *call*, whereas that in (b) (*to get help for*) is a complement to *call*. This difference can be demonstrated by replacing the infinitival clause with a purposive clause (*so that he could ...*), which is acceptable for (2a) but not for (2b), as shown in (3a) and (3b).

(3a) Which doctor did the supervisor call Δ so that he could get help for his youngest daughter?

(3b) *Which doctor did the supervisor call so that he could get help for Δ during the crisis?

Straub and colleagues pointed out that the critical potential gap location (*call / to*) is at the boundary between a main clause and an adjunct phrase in (2a) but within a phrase in (2b). They argued that the different prosodic phrasings observed by Nagel et al reflect this difference in argument structure rather than gapping *per se*, and that the prosodic consequences of this difference can be accounted for in terms of Selkirk's (1984) *Sense Unit Condition on Intonational Phrasing*. This condition states that the immediate constituents (C_v , C_i) of an intonational unit should form a sense unit, with C_i either modifying or being an argument of C_v (a head). This means that the intonational groupings in (4) are allowed, where an intonational break also coincides with the gap location in each case, but those in (5) are ruled out.

(4a) (Which doctor did the supervisor call Δ) (to get help for his youngest daughter)?

(4b) (Which doctor did the supervisor call to get help for Δ) (during the crisis)?

(5a) *(Which doctor did the supervisor call Δ to get help for) (his youngest daughter)?

(5b) *(Which doctor did the supervisor call) (to get help for Δ during the crisis)?

To confirm their re-interpretation of Nagel et al's experiment, i.e. that the finding of a prosodic effect at a gap location is in fact a prosodic effect due to

clause structure differences, Straub et al ran a production experiment of their own using materials such as those in (6):

(6a) *Phrase-Final Gap:*

What did you return Δ to make sure you would get a full refund?

(6b) *Phrase-Medial Gap:*

What did you return Δ to the store when you didn't expect to get a full refund?

(6c) *Phrase-Medial Control:*

Who did you return to the store with Δ when you wanted to get a full refund?

Straub et al argued that if prosodic contrasts reflect differences in gap location, then the critical region (*return to*) should show a difference between (6a) and (6b) on the one hand and (6c) on the other. However, if the prosodic contrast depends primarily on clause structure, then (6a) should differ at this region from both (6b) and (6c), since only in (6a) is there a clause boundary at *return*. Straub et al's production data support this second prediction – both the duration of *return* with its following pause and the pitch excursion across the critical region were significantly greater in (6a) versions than in (6b) and (6c), which did not differ significantly from one another (though there was a non-significant difference in the direction predicted by the gapping account).

While Straub et al's study points out the dangers of confounding gap location and structural boundary location, the earlier work of Danly (using examples with the same structure as (1) above) and related studies of ellipsis sites by Cooper and Paccia-Cooper (1980) showed that clause-internal gaps are likely to be marked by lengthening and pausing relative to comparable non-gap locations. The remaining sections of this paper present some new data relating to this issue from our SPOT speech production task.

Gaps and prosody in the SPOT project

The data to be reported in this paper come from the SPOT project, and involve the gap sentences illustrated in (7) below. Our study differs from those discussed above in two ways: first, both of the gap sentences used in our game task have complement *to* clauses (see further below); second, our analysis involves a combination of measures: durations, transcriptions, and the incidence of *wanna*-contraction across the potential gap site.

(7a) Which triangle, do you want Δ , to change the position of the square?

(7b) Which triangle, do you want to change the position of Δ , this time?

The SPOT project is a collaborative research programme involving researchers in New Zealand and the United States of America. The programme goals include the collection of speech production data under more naturalistic conditions than those typically used in speech production studies of prosodic

disambiguation. For instance, the gap studies referred to above use reading tasks, yet readers and talkers have different pragmatic goals, and different processing demands – unsurprisingly a number of researchers have commented on the resulting differences in the prosody of read and spontaneous speech (Ayers, 1994; Blaauw, 1994; Howell & Kadi-Hanifi, 1991). Ideally, prosodic distinctions between minimally contrasting syntactic structures would be studied using spontaneous speech. This is however hardly practicable, since it is unlikely that speakers will produce these minimally contrasting utterances without prompting. Some degree of spontaneity might be achieved using role play situations or map tasks (Anderson & colleagues, 1991). Here, though, there are also typically few constraints placed on the syntactic forms of expression. Other techniques, such as descriptions of node networks (Levelt & Cutler, 1983) or tangram shapes (Clark & Schober, 1992; Clark & Wilkes-Gibbs, 1986), have revealed much about the use of prosody in the planning and repair of utterances and in the construction of collaborative discourse, but these tasks are again not designed for eliciting specific syntactic contrasts.

In our own attempts at resolving or at least side-stepping this "laboratory speech" issue, we have constructed a task in which pairs of speakers negotiate the movement of objects around a game board. The task requires participants to use a fixed set of sentence frames and object names to construct instructions, requests and acknowledgements, thus producing multiple instances of a set of syntactic structures of relevance to our research. The two participants have slightly differing boards, and different roles. The "Driver" knows the ultimate goals for the objects, and issues instructions accordingly. The "Slider" moves the objects in response to these instructions, and knows the locations of bonuses and dangers, but not the goals for the objects. The Slider asks questions about which object to move and provides confirmation that the object has been moved. Both Driver and Slider are aware of basic game rules governing the types of movement for game board objects. We have found that participants quickly become familiar with the utterances available for use under the rules of the game, and learn to produce them fluently and without recourse to printed sentence lists. We believe that this game task produces a rich source of data for the study of syntactic ambiguity resolution in a more spontaneous speech situation than that used in traditional sentence reading tasks. Since the utterances include a number of syntactic ambiguities frequently studied in comprehension studies, we can also use our recordings to examine the use of prosodic information in ambiguity resolution during comprehension. What is more, by careful planning of our gameboard layouts, we can construct game situations in which potentially ambiguous utterances are contextually disambiguated and other situations where there is a contextual bias towards one of the meanings.

As well as the gap structures in (7), the syntactic contrasts that we have investigated include early (8a) vs. late (8b) closure of constituents (Schafer, Speer, Warren, & White, 2000) and high vs. low attachment of prepositional phrases (PPs; e.g. in (9) the triangle could be an instrument used to move the

square or it could be part of a description of a house-like piece, the "square with the triangle") (Schafer, Speer, & Warren, forthcoming).

- (8a) When that moves the square will...
 (8b) When that moves the square it...
 (9) I want to change the position of the square with the triangle

We claimed above that both utterances (7a) and (7b) involve infinitive clauses as complements. Recall that Straub et al (2001) argued that Nagel et al's (1994) phonetic differences were a reflection not of a gapping difference but of the contrast between adjunct and complement roles of the infinitive clause in (2), as revealed by the purposive test in (3). While it would seem that our sentence in (7a) would allow a purposive, as in (10a) below (and contrasting with (10b)), it is clear from the context of our game task that a different sense of *want* is involved in our materials, one which is compatible with a complement interpretation of the infinitive clause rather than with an adjunct interpretation. This different meaning might be paraphrased as "I wish to change the position of the square with the triangle", which does not allow the purposive (as shown by the awkwardness of a version of (10a) with "wish" in place of "want").¹

- (10a) Which triangle do you want in order to change the position of the square?
 (10b) *Which triangle do you want in order to change the position of this time?

Data

Our project data include recordings of eight pairs of speakers of each of Midwestern American English (MAE) and New Zealand English (NZE). Each pair completed as many as six games (not including practice games) during a session of approximately two hours, swapping Driver and Slider roles between games. Data for one speaker have been removed for each dialect, in both cases because of a high level of disfluency. The remaining data have been analysed both acoustically (duration, pausing, F0 patterns) and

¹ Note that the Slider sentence in (7a) is produced in response to the instrumental meaning of the Driver sentence illustrated in (9) (i.e. "I want to use the triangle to move the square" - the rules of the game are such that the square cannot move on its own, but must be pushed by another object). The situation of the game prior to the Driver's utterance of (9) is that both game boards (the Driver's and the Slider's) have a triangle on each of two sides of the square, and either of these triangles may legitimately move the square. The Driver's utterance of (9) with the instrumental meaning thus presupposes that there is at least one triangle capable of moving the square, but the utterance remains unclear as to which triangle should do the moving, which is why the Slider produces the question in (7a). Notice though that the purposive in (10a) means something like "of the available triangles, which one do you want to have so that you can use it to move the square?", which is distinct from the meaning intended in our game task.

auditorily (using Pierrehumbert's Autosegmental Metrical framework encapsulated in the ToBI transcription system (Beckman & Ayers, 1997; Pierrehumbert, 1980)).

Across the different sentence types of the SPOT project, and for both MAE and NZE speakers, our phonetic measures and transcription data have shown that naïve speakers produce patterns of prosodic phrasing that indicate the syntactic constituent boundaries they intend to communicate. This paper presents detail of our analysis of the gap structures in (7). Utterances containing disfluencies (hesitation pauses, repeats, false starts, mispronunciations, etc.) were excluded from the analysis (this accounted for approximately 9% of the total gap structures available). The remaining data for this analysis include, for MAE, 61 early gap utterances (as in (7a), repeated below) and 193 late gap utterances (i.e. where the gap is after *of*, as in (7b)). For NZE there are 41 early and 194 late gap utterances. Both of the early and late gap utterances are Slider utterances, i.e. they are requests for more information from the participant in the role of Slider. (7a) and (7b) respectively are uttered by Sliders in response to the instructions from the Driver given as (11a) and (11b) below. (The intended meaning of the ambiguous utterance in (11a) was tracked by an experimenter, who noted the object being moved by the Driver.) As stated earlier, the rules of the game are such that the square piece cannot be moved on its own, but must be pushed by another piece. Often there is a choice of pieces that can be used to push the square, including a choice between two triangles (of differing colours), and this will lead to the Slider utterance in (7a). In some gameboard configurations, the objects that can push the square might be shapes other than triangles (e.g. cylinders), so the total count of early gap utterances includes a relatively small number in which there would be a different object name in the position of *triangle* in (7a). The total numbers of fluent triangle utterances of (7a) are 53 for MAE and 38 for NZE. The instruction in (7b) refers to the movement of objects by themselves (rather pushing a square), and can similarly relate to the movement of shapes other than triangles. The counts of fluent triangle utterances of (7b) are 155 for MAE and 151 for NZE.

- (7a) Which triangle, do you want Δ_1 to change the position of the square?
 (7b) Which triangle, do you want to change the position of Δ , this time?
 (11a) I want to change the position of the square with the triangle. [with the intended meaning of "use the triangle to push the square"]
 (11b) I want to change the position of the triangle.

The following sections present summaries of our results for these materials, first for the incidence of wanna-contractations and then for phonetic (duration) and finally phonological prosodic analyses. Each section presents both the MAE and NZE data.

Wanna contraction

Earlier research referred to above (Baker & Brame, 1972; Lakoff, 1970) claimed that the presence of a syntactic gap is likely to block the contraction of *want to* to *wanna*, but because wanna-contraction is optional (Fodor, 1979), the absence of contraction need not indicate the presence of a gap. Likewise, we noted earlier that both perceptual and more theoretical studies (Karins & Nagy, 1993; Pullum, 1997) have suggested that the presence of a contracted form may not unambiguously indicate that there is no intervening gap. However, it is clear that much of the literature in theoretical syntax that has used contracted forms such as *wanna* as evidence has espoused a strong belief that contraction will be more frequent in no-gap contexts. In line with this, and with the overall preference in Karins and Nagy's (1993) perceptual study, we predicted a greater incidence of contraction in the late gap versions (7b) of our temporary ambiguity than in the early gap versions (7a).

The contraction literature has been centred largely on US English, and research on the phenomenon in New Zealand English is virtually non-existent. Researchers at the University of Canterbury in Christchurch (Quinn, p.c., 2004) report that contractions (including wanna-contraction) are more frequent in the speech of the younger speakers in their corpora than in that of the older speakers, but that this finding may be confounded by a difference in the degree of formality of the recordings across age groups. But even in the youngest group (i.e. the most progressive group in terms of contraction) the uncontracted form is far more common than the contracted. Although there is no empirical data to support a difference between the varieties, general informal comments from informants in New Zealand suggests that wanna-contraction is more typically associated with American than with New Zealand varieties of English. We therefore also predicted that wanna-contraction would be more frequent in the MAE recordings than in the NZE recordings of our SPOT task.

Figure 1 gives the percentage incidence of wanna-contraction (i.e. contraction at the potential early gap site) in the early (7a) and late (7b) gap conditions of our experiment, separately for MAE and NZE. The figure shows quite clearly that contraction was indeed more frequent in MAE than in NZE, and that in both varieties contractions are more likely when there is no syntactic gap between *want* and *to* (i.e. in the late gap condition). The counts from which Figure 1 was derived were analysed in the following manner. Counts of tokens with and without contraction were entered into a hierarchical loglinear analysis (Hays, 1981: 561-6), with Variety (MAE vs. NZE), Type (early gap vs. late gap) and Contraction (present vs. absent) as factors. A parsimonious model for the data was built by backwards elimination of factor combinations. The tests of this model that are of interest are the interactions of Contraction with the other factors, since these interactions indicate an effect of these other factors on the incidence of contraction. The analysis shows that both of these other factors are significant (as measured by the likelihood ratio χ^2 changes, which for the interaction of

Contraction with Variety was 11.89, $p < 0.001$, and for the interaction of Contraction with Type was 18.88, $p < 0.001$). There was no significant interaction involving both Variety and Type.

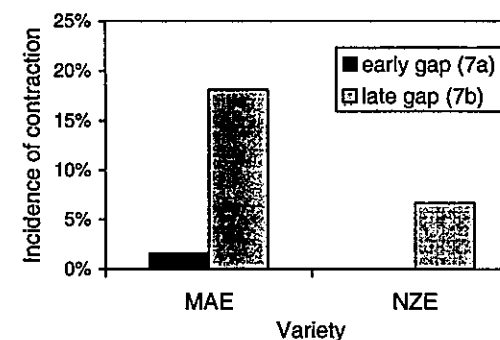


Figure 1. Incidence of wanna-contraction in early and late gap sentences (see examples (7a) and (7b) respectively) in Midwestern American English and New Zealand English. Data are from 15 speakers of each variety.

Despite the fact that contraction is – as predicted – much more likely when there is no gap, the low overall level of contraction indicates that it is unlikely to be a reliable indicator of the difference between (7a) and (7b).

Wanna-contraction is 2.5 times more likely in the MAE data (13.8% overall) than in the NZE data (5.5% overall). This would seem to reflect the informal observations that such contractions are more widespread in colloquial American English than in colloquial New Zealand English.² Interestingly though, the incidence of contraction is due in each case to high levels of *wanna* from a few speakers (one NZE speaker and three MAE speakers have incidence levels of 35% or more, and these account for 85% or more of wanna-contraction in the data for each variety), suggesting that the use of *wanna* is something of an individual characteristic, used by Pullum's (1997) "liberal" dialect speakers, while still more likely in one variety than the other. Note however that Pullum suggests that such liberal dialect speakers are likely to contract across the board, and not just when there is no intervening syntactic gap. But this general pattern of wanna-contraction does not seem to be the case here, as shown by the difference between the early and late gap conditions. Pullum argues that contraction is constrained for liberal dialect speakers by phonological factors, such as whether the relevant

² This pattern of greater wanna incidence in MAE than in NZE is confirmed in an analysis of the Driver utterances exemplified in (11) that precede utterances such as (7). For these, there was an overall contraction rate of just under 5% for MAE, but less than 1% for NZE.

words are in the same intonational phrase. For this reason, we will return to discussion of wanna-contraction after our discussion of other phonological distinctions between utterances (7a) and (7b).

Durational analyses

Recall that the research discussed above was equivocal as to whether durational differences would provide a reliable indicator of the presence of a syntactic gap. On the one hand, Danly (1980) and subsequently Nagel et al (1994) found significant durational differences. On the other, Straub et al (2001) argued that Nagel et al (1994) confused gapping and other syntactic differences; Straub et al went on to present data that showed non-significant differences. We predicted for our data that we will find similar differences to those reported by Danly for similar materials, and that these differences would obtain despite the lack of overt instruction to our participants to disambiguate the structures they produced.

We measured key durations at both the early gap site and the late gap site. At the early gap we measured the duration of *want* and *to*, and any pause between these words. In the cases of contraction, we took a single measure of the duration of *wanna*. For the following presentation we consider the duration of the whole sequence *want+pause+to*, which is the measure used by Danly (1980). At the late gap site we measured the preposition *of* and any following silence. Again, the presentation below gives the values of the entire *of+pause* sequence. For each variety, speakers' data were included only if they had two or more fluent tokens of each of the early and late gap utterances (7a) and (7b). This means that the data analysed below comes from 13 MAE speakers and 11 NZE speakers.

Durations at the early gap site

Two analyses were carried out on the data from the early gap site. One included all data from the 24 speakers in the two varieties, while the other excluded contracted (*wanna*) tokens, reflecting Danly's (1980) analysis. Figure 2 displays the results for the larger dataset.

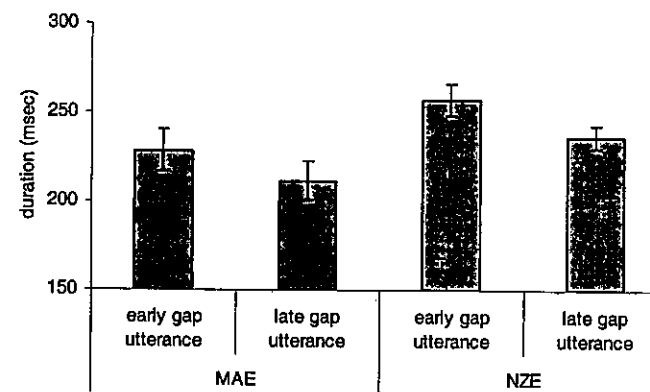


Figure 2. Duration of wanna sequence (*want+pause+to*, or contracted *wanna*) in early and late gap utterances (see examples (7a) and (7b) respectively) in Midwestern American English and New Zealand English. Data are from 13 speakers of MAE and 11 speakers of NZE.

Analysis of Variance was conducted on the subject mean values, with utterance type (early gap vs. late gap) and variety (MAE vs. NZE) as factors. The ANOVA showed a significant main effect of utterance type ($F[1,22]=12.92, p<0.01$) and a strong near-significant main effect of variety ($F[1,22]=4.087, p=0.056$). The averages of the subject means showed a 20 msec difference in the duration of the wanna sequence in early and late gap utterances (242 vs 222 msec). The difference between the means of the complete datasets (i.e. not first averaged for each subject) was slightly larger at 27 msec (246 vs 219 msec). The differences were slightly smaller when contracted *wannas* were excluded, but the difference between early and late gap utterances remained highly significant. The difference between the varieties is consistent for both early and late gap utterances (there was no significant interaction of the two factors in the analysis). While this may indicate a general difference in speech rate between the two groups, it may also reflect differences in phrasing (see Discussion below).

Durations at the late gap site

The averages of the subject mean durations of the sequence *of+pause* are shown in Figure 3. Analysis of Variance revealed a significant main effect of utterance type ($F[1,22]=25.497, p<0.001$), and no other effects. The difference between the two utterance types was 79 msec (123 msec vs. 202 msec), indicating that this late gap position is clearly marked by durational differences. Note that there is a tendency for the NZE data to show shorter durations here, in contrast to a tendency in the opposite direction at the early gap position (compare Figure 2). This may also reflect differences in phrasing between the two varieties (again, see Discussion below).

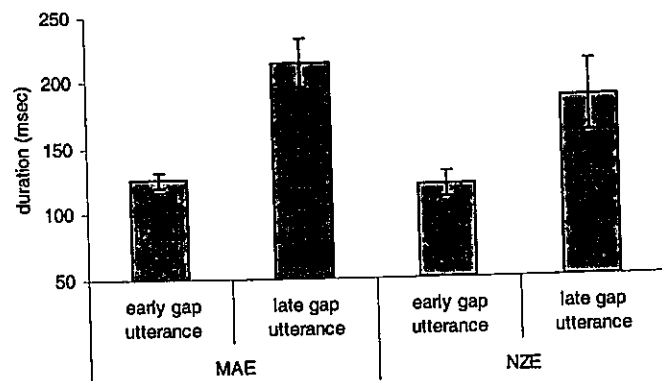


Figure 3. Duration of the sequence of +pause in early and late gap utterances (see examples (7a) and (7b) respectively) in Midwestern American English and New Zealand English. Data are from 13 speakers of MAE and 11 speakers of NZE.

Overall, the differences in duration at the two sites are similar for the two varieties – there was no suggestion of an interaction of variety and utterance type at either site. The difference at the early gap site is similar in magnitude to that reported by Danly (1980) (25msec). The difference at the later gap site is clearly larger, as might be expected given the difference between the structures (in (7a) the preposition is followed by its complement, in (7b) by a new phrase, serving as an adjunct to the verb phrase).

The data in Figures 2 and 3 focus on the gap sites in utterances (7a) and (7b). Further measurements show that there are also differences at other sites in the utterances, particularly at *change* and *position*. These two sites are summarised in Figure 4. The clearest differences are for NZE, though Analysis of Variance shows an interaction of variety and utterance type only for the *change+pause* data ($F[1,22]=5.156$, $p<0.05$). In analyses for both *change+pause* and *position+pause* there was a significant main effect of utterance type (with longer durations in early gap than in late gap utterances in each case: *change*: $F[1,22]=11.236$, $p<0.01$; *position*: $F[1,22]=5.255$, $p<0.05$).

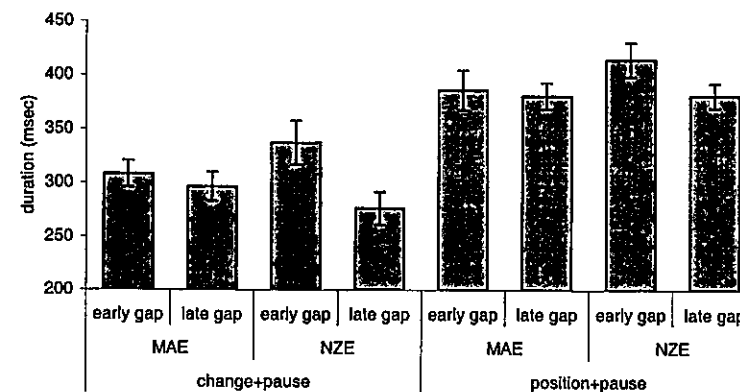


Figure 4. Duration of the sequences *change+pause* and *position+pause* in early and late gap utterances (see examples (7a) and (7b) respectively) in Midwestern American English and New Zealand English. Data are from 13 speakers of MAE and 11 speakers of NZE.

At issue here is whether this pattern of differences at *want*, *change* and *position* is indicative of a general lower rate of speech in early gap vs. late gap utterances, or whether it reflects a tendency to place intonational phrase breaks at these positions in the early gap utterances. Note that this need not mean that speakers break at each of these positions in any one utterance, but that there are breaks at these three locations in various combinations across the utterance set as a whole. An explanation in terms of speech-rate could in fact argue for some general speeding-up through "want to change the position of" in late gap utterances, while an explanation based on boundary locations may relate to how a constraint such as the Sense Unit Condition may operate in the case of early and late gap utterances. The two explanations are of course not mutually exclusive. The phonological analysis discussed in the next section is relevant both to this issue and to the question of whether possibilities for wanna-contraction reflect intonational phrasing.

Phonological analysis

The phonological analysis of our wanna-materials in (7) consists of ToBI-style transcriptions of the utterances. These transcriptions are available for 159 of the fluent MAE recordings (49 early gap utterances and 110 late gap utterances), and for a subset of 40 of the NZE recordings (a random selection of 20 early and 20 late gap utterances). Each transcription was carried out by at least one expert ToBI transcriber. Utterance files were cut after "of" so that transcribers did not know from the lexical content whether the early or late gap utterance had been produced. Transcribers marked pitch accents (H^* , L^* or bitonal e.g. $L+H^*$), phrase accents (at the ends of intermediate phrases, i.e. $L-$ or $H-$) and boundary tones (at the ends of full intonational phrases, i.e. $L\%$ or $H\%$), and indicated break index values at each word boundary. (Possible

break index values are 0 for where words are run together, as in *wanna*, 1 for phrase-internal word boundaries, 3 for intermediate phrase boundaries and 4 for intonational phrase boundaries. 2 is used to indicate an ambiguous boundary that carries perhaps tonal but not rhythmic phrase boundary markings.)

The following presentation focuses on the locations of phrase boundaries in the utterances, i.e. corresponding to break index values of 3 or 4 and the tonal markings of intonational and/or intermediate phrase boundaries. Panels (a) through (d) in Figure 5 show, for each word location from *which* through to *of*, the proportion of utterances that contain such a phrase boundary at that location, for MAE early and late gap, and NZE early and late gap utterances.

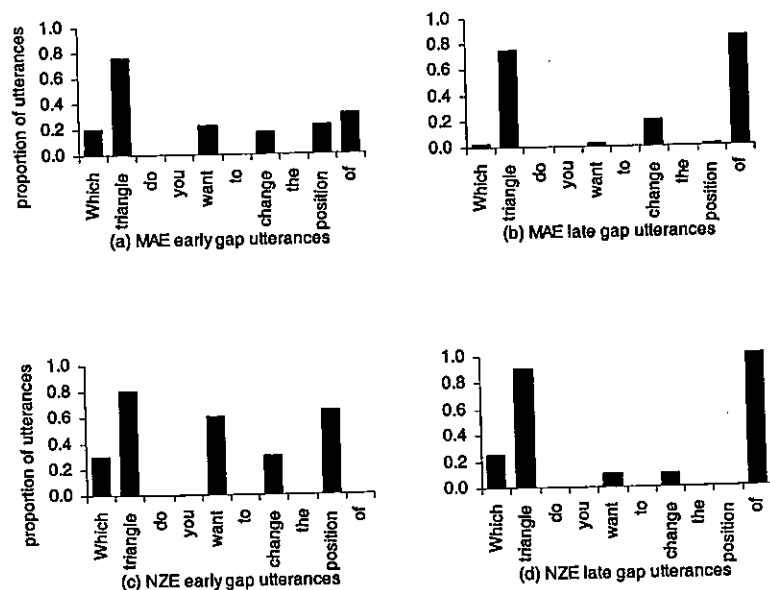


Figure 5. Incidence of phrase breaks (intonational and/or intermediate) across the sequence *Which triangle do you want to change the position of*, in early and late gap utterances (see examples (7a) and (7b) respectively) in MAE (panels (a) and (b)) and NZE (panels (c) and (d)).

The counts from which Figure 5 was derived were analysed in the following manner. For each location where there were sufficient data (i.e. excluding *do*, *you*, *to*, *the*) counts of tokens with and without breaks were entered into a hierarchical loglinear analysis, with Variety (MAE vs. NZE), Type (early gap vs. late gap) and Break (present vs absent) as factors. For each analysis a parsimonious model was built by backwards elimination of factor combinations. The tests of this model that are of interest are the interactions of

Break with the other factors, since these interactions indicate an effect of these other factors on the incidence of intonational breaks. Table 1 shows the significant interactions for each location. The differences between breaks reported for MAE and NZE occur at *which*, *want* and *position*, with, in each case, more breaks in NZE. At *which* and *of* there was a main effect of Type, with more breaks at *which* in the early gap utterance, and more at *of* in the late gap utterance. The interactions of Type with Variety at these locations simply reflect differences in the size of these effects – the Type difference at *which* is greater for MAE, the Type difference at *of* is greater for NZE. Two further main effects of Type, at *want* and *position*, reflect the greater likelihood of breaks at these locations in the early gap utterances.

	Variety	Type	Variety*Type
which	**	**	*
triangle			
do			
you			
want	**	**	
to			
change			
the			
position	**	**	
of		**	**

Table 1. Results of statistical analysis of the incidences of phrase breaks shown in Figure 5. For each factor (combination) shown by the column headings, significant interactions with the Break factor at each word location are indicated by ** (for $p < 0.01$) or * (for $p < 0.05$), blanks indicate that no significant difference was found, shaded cells show comparisons that were not possible because of small datasets. (See text for further details of the analysis.)

Thus, while there are some differences between MAE and NZE in the level of incidence of breaks, both varieties show similar differences between the early and late gap utterances in how these are phrased. The analyses and the pattern of breaks shown in Figure 5 indicate that late gap utterances are likely to be marked by two main breaks, after *triangle* and *of*. Further, while early gap utterances are just as likely as late gap utterances to have a break after *triangle*, they are more likely to also have breaks at *want* and *position*, and are unlikely to have one at *of*.

Discussion

The above analyses have shown that patterns of prosodic marking of gap constructions demonstrated in a range of previous studies are found also in data from the SPOT project. Phonological (transcription) and phonetic (duration) measures reflect a greater tendency to break at *want* in the early gap utterance (7a) than in the late gap utterance (7b). In addition, contraction of *want to* to *wanna* was more likely in the late gap utterances, i.e. when there is no gap site at *want*.

The differences between the two varieties at the early gap site (i.e. at *want to*) are consistent. NZE has fewer contractions (see Figure 1), which is in agreement with general informal observations on the two varieties. At the same time, however, NZE has greater durations across the *want+pause+to* sequence (Figure 2). It is unlikely that this difference is a simple result of slower speech in the NZE because the opposite pattern (shorter durations in NZE) is found at the late gap site (Figure 3). It is also the case that these early gap sites are more frequently marked by prosodic breaks in NZE, both in early gap utterances and in late gap utterances (Figure 5 and Table 1). There may therefore be a general difference between the two varieties in the extent to which these utterances are broken into intonational and intermediate phrases. The greater incidence of such phrasing at the early gap site in NZE would then correlate with the lower incidence of wanna-contraction.

We also found that the prosodic phrasing of the utterances is highly variable in both varieties. The distribution of breaks at different locations in Figure 5 shows little evidence that any particular word boundary is consistently marked by an intermediate or intonational phrase break, with the exception of *triangle* in both utterance types and *of* in the late gap utterances. The variability of phrasing is revealed in the counts of different phrasings for the transcribed materials. For instance, from the 20 transcriptions completed for each of the utterance types in the NZE data we obtained 10 different phrasings (i.e. different organisations into intermediate and/or intonational phrases) for early gap and 7 different phrasings for late gap utterances. From the MAE transcriptions, we get 20 distinct phrasings for 49 early gap utterances, and 12 phrasings for the 110 late gap utterances. Despite this degree of variability, there are some areas of consistency. The break after *triangle*, for instance, is found overall in approximately 77% of cases (across both varieties). This break appears to place the phrase *which triangle* into focus as the element for which further information is being sought. This is appropriate to the context of the game, in which a Slider utterance from (7) follows one of the Driver utterances in (11).

What is of interest is whether the subsequent phrasing of the Slider utterances is constrained in any way, in particular in order to indicate the possible location of a gap to which the filler *which triangle* is linked. Constraints from such phrasing might operate over and above any explicit marking of a gap site through the prosodic and other phonological means investigated in our analyses of the gap locations. Research into the analysis of filler-gap dependencies has considered a range of factors that might affect the linking of fillers with potential gap sites (Clifton & Frazier, 1990). In one study of prosodic constraints, Broderick (1996) reports that listeners prefer to link a filler to a gap within the same intonational phrase. As Straub et al suggest, in connection with Broderick's work, "it is conceivable that the comprehension and the production of filler-gap dependencies interact with the comprehension and production of prosodic structure, even in the absence

of prosodic cues that exist solely to mark the location of syntactic gaps" (Straub et al., 2001: 392). If speakers are sensitive to the apparent listener preferences indicated by Broderick (1996), then this would suggest that the structure for the utterances would be as in (12), where (/) indicates (possible) break locations. Linking of the filler *which triangle* to the gap after *want* is possible with any of the phrasings in (12a) (as well of course as with the phrasing in (12b)), since the filler will attach in each case to a gap within the phrase started at or before *do*. On the other hand, the presence of such a phrase break before *of* would make the late gap interpretation less likely, since the gap location would no longer be in the phrase started with or immediately after the filler.

(12a) Which triangle (/) do you want (/) to change (/) the position (/) of the square?

(12b) Which triangle (/) do you want to change the position of (/) this time?

In the transcribed data from our NZE speakers, full intonational phrase breaks are found after *triangle* in 6 (of 20) early gap utterances, and in 3 (of 20) late gap utterances. All of the remaining utterances have no internal intonational phrase breaks, except for a single early gap utterance, which has a break after *position* (and is therefore compatible with the postulated break locations in (12a) above). The majority of the transcribed NZE utterances do not therefore distinguish early and late gap structures on the basis of the location of full intonational phrase breaks. If we include intermediate phrase breaks (while noting that Straub et al's discussion of Broderick's work refers to intonational phrase breaks.), then we find that nineteen of the twenty transcribed early gap utterances have a break at either *want*, *change* or *position* (as predicted in (12a)), while only 5 of the late gap utterances have breaks before *of* (3 at *want* and 2 at *change*).

In the transcribed MAE data, the locations of full intonational phrase breaks again do not fully distinguish early and late gap structures – for example, 40 (of 49) early gap utterances have no intonational phrase break between *triangle* and the end of the utterance, and this is true of 87 (of 110) late gap utterances. If we include intermediate phrase breaks, then there are still some that have no break later than *triangle* in both gap types (12 early and 7 late). There are nevertheless some differences in phrasing – some 55% (27) of early gap utterances have a break at one or more of *want*, *change* or *position*, while only 25% (27) of the late gap utterances have breaks at one or more of these positions. Conversely only 20% (10) of the early gap utterances have a break at *of* without an earlier break at *want*, *change* or *position*, while the figure for late gap utterances is 69% (76). These analyses show that there are clear tendencies towards the patterns of phrasing suggested in (12) for early and late gap utterances, but also some exceptions. Closer inspection of pitch accent and boundary tone patterns over the utterances might reveal the use of tonal cues to structure not exposed by this analysis.

In summary, the data reported in this paper provide support for intonational distinction of the utterances in (7a) and (7b). However, they also reveal considerable variation in intonational realizations and in the incidence of wanna-contraction. While some of this variation may be due to differences between English varieties, it is clear that much more of the variation is a result of individual choice in how to phrase the utterances. Importantly, however, the choices are not unconstrained, and an important constraint is the distinction between early and late gap structures. Further experimentation will determine the significance of the observed distinctions for the comprehension of these utterances, and the immediacy with which prosodic cues are interpreted as indicators of the different structures.

References

- Anderson, A. H., & colleagues. (1991). The HCRC map task corpus. *Language and Speech*, 34, 351-366.
- Ayers, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, 44, 1-49.
- Baker, C. L., & Brame, M. K. (1972). Global rules: a rejoinder. *Language*, 48, 51-75.
- Beckman, M. E., & Ayers, G. (1997). *Guidelines for ToBI labelling*. Unpublished manuscript, Columbus, OH: Ohio State University
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14, 359-375.
- Bley-Vroman, R., & Kweon, S.-O. (2002). *Acquisition of the constraints on wanna contraction by advanced second language learners: Universal grammar and imperfect knowledge*. Unpublished manuscript. <http://www.sls.hawaii.edu/bley-vroman/> (accessed 9th Feb 2004)
- Broderick, K. (1996). The role of prosody in Wh-movement: Intonational phrasing and adjunct construal. Poster presented to the *Ninth Annual CUNY Conference on Human Sentence Processing*, New York, March 21-23, 1996.
- Chomsky, N. (1977). On wh-movement. In A. Akmajian, P. Culicover & T. Wasow (Eds.), *Formal Syntax*. New York: Academic Press.
- Clark, H. H., & Schober, M. F. (1992). Understanding by addressees and overhearers. In H. H. Clark (Ed.), *Arenas of Language Use* (pp. 176-197). Chicago: University of Chicago Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Clifton, C., & Frazier, L. (1990). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273-318). Dordrecht: Kluwer.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.
- Danly, M. (1980). *The wh-movement transformation and speech timing*. Unpublished manuscript
- Fodor, J. D. (1979). Superstrategy. In E. Walker & W. Cooper (Eds.), *Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.
- Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech. *Speech Communication*, 10, 163-169.
- Karins, K., & Nagy, N. (1993). Developing an experimental basis for determining grammaticality. *The Penn Review of Linguistics*, 17, 93-100.
- Lakoff, G. (1970). Global rules. *Language*, 46, 627-639.
- Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, 2, 205-217.
- Nagel, H. N., Shapiro, L. P., & Nawy, R. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research*, 23, 473-485.
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Unpublished Doctoral dissertation, MIT.
- Pullum, G. K. (1997). The morpholexical nature of English to-contraction. *Language*, 73(1), 79-102.
- Quinn, H. (p.c., 2004, p.c.). Modal contractions in the ONZE copora.
- Schafer, A. J., Speer, S. R., & Warren, P. (forthcoming). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to Studying World-Situated Language Use: Bridging the Action and Product Traditions*.
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.
- Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.
- Straub, K., Wilson, C., McCollum, C., & Badecker, W. (2001). Prosodic structure and wh-questions. *Journal of Psycholinguistic Research*, 30, 379-394.

This is a statement? Lateness of rise as a factor in listener interpretation of HRTs.¹

Joel Zwartz and Paul Warren

Abstract

High Rising Terminal intonation contours (HRTs) have been popularly viewed as a "questioning" intonation, a view that suggests the neutralisation of a phonological distinction between HRT and question rises. Recent production studies have however indicated phonetic differences between questions and HRTs in New Zealand English, specifically in the alignment of the F0 contour of the final rise. The pilot perception study reported in this paper shows that differences in alignment similar to those found in the production studies are perceptible, and that they affect the listener's interpretation of whether the utterance is a statement or a question. The study additionally shows that it is not just the alignment of the rise that is important, but also its shape. These results suggest that the observed phonetic difference may be the basis of a phonological distinction between statement and question rises.

Introduction

High Rising Terminal contours (HRTs) are a rising intonation pattern found at the end of statement utterances in the English speech of young people in New Zealand, Australia, the United States, Canada and Britain (Ainsworth 1994, Ladd 1996, Warren & Britain 2000, Cruttenden 1995). As Warren (ms) notes, the number of papers written by linguists about HRTs is not really warranted by the relative infrequency of these contours. Instead, it seems that HRTs have come under scrutiny because of the negative interpretation they have received as indicating uncertainty or hesitancy, an interpretation that is reflected in comments by popular columnists and correspondents to newspapers. This negative perception of HRTs is due to the fact that rises at the ends of utterances are more commonly associated with questions than with statements in many varieties of English, particularly perhaps those varieties spoken by these commentators. More generally across languages, it seems that there is a universal tendency for rises to indicate openness or

¹ The experiment reported here was conducted by the first author, under the supervision of the second author, in partial fulfilment of the requirements of a BA Honours course in Psycholinguistics at Victoria University of Wellington. Both authors have contributed equally to the preparation of this paper.

incompleteness, as opposed to finality (Ladd 1996:113ff, Hirst & Di Cristo 1998: 27). The pragmatics of HRTs in New Zealand English (NZE) could arguably fall under this universal tendency, since studies of NZE HRTs have shown that they are widely used to include the listener in the discourse, which may well be interpreted as indicating openness, rather than uncertainty (Warren & Britain, 2000). However, the current paper is not concerned with attitudes towards HRTs, but instead focuses on differences that have been found between the phonetic forms of question rises and HRT rises, and addresses whether these differences correspond to a nascent phonological distinction between question and statement rises.

Typically, as the above paragraph has indicated, HRT rises are seen as having the same shape as question rises. For this reason, the same phonological labels are frequently used for the two, even though there is a semantic difference between the question and statement functions being conveyed. Of course, if there is genuinely no difference in the form of these two types of rise, then we may well be dealing with the neutralisation of a contrast. The phonological labels reported here are based on the conventions of the Autosegmental Metrical (AM) approach to intonation (Pierrehumbert, 1980), and the associated ToBI transcription system (Beckman and Ayers Elam, 1997). The AM approach analyses the intonation of utterances into sequences of high and low tonal targets. Within an intonational phrase in English there will be one or more pitch prominent syllables with high (H*) or low (L*) pitch accents, or bitonal pitch accents (e.g. a rising L+H*). Each intonational phrase contains at least one intermediate phrase, which is bounded at its right edge by a phrase accent (H- or L-), while the intonational phrase itself ends with a boundary tone, again either high (H%) or low (L%).

Using this system, Fletcher, Grabe and Warren (2004) classified as possible HRTs a range of tunes in Australian and New Zealand English that might be transcribed as L* H-H%, L*+H H-H%, and H* H-H%. Ladd (1996: 121) also refers to the "high-rising 'question' intonation" used on statements in North American and Antipodean English as probably including the first and third of these tunes. He also indicates (1996:82, Table 3.1) that the first two tunes can be characterised as low rises. Clearly, then, these descriptions indicate that High Rising Terminals do not correspond to high nuclear tones that rise, but rather more generally to tunes that rise, from either a high or a low starting point, to a marked high endpoint.

Turning now to the phonetic description of HRTs, we find a range of definitions. Britain and Newman note that the physical realisation of HRTs can vary dramatically, corresponding to rises in fundamental frequency (F0) that may be less than 50Hz or more than 100Hz, and stretching over a period of speech that may be as small as a single syllable or greater than an entire word (Britain & Newman 1992: 4-8). Some of the variation reported by Britain and Newman may of course be due to differences between male and female speech – a 50Hz rise from a typical average male pitch would be roughly equivalent (at around 7 semitones) to a 100 Hz rise from a typical

average female pitch. This demonstrates that it is more useful to describe the magnitude of a rise using a perceptual measure, e.g. number of semitones, or possibly by means of an indication of the proportionate change in pitch. Guy and Vonwiller (1989: 23) in their analysis of Australian HRTs described the final pitch as approximately 40% higher than the starting point of the nuclear accent, a difference that corresponds to 6 semitones.

Ladd (1996:123) says that "[p]honetically, the American/Australian rises [i.e. HRT rises] begin high on the accented syllable and keep rising to the end of the phrase". Note that this phonetic description here seems to be consistent primarily with a H* H-H% tune, though it may also match a L*+H H-H% tune with a bitonal pitch accent. It is less consistent with the L* H-H% tune also mentioned by Ladd as an example of the statement rise.

The phonological tune types mentioned above, as well as the phonetic descriptions of HRTs, would also be appropriate characterisations of questions in English. Ladd (1996: 121-2) states that

there is to my knowledge no phonetic difference that depends on whether the high-rising contour is used with syntactically marked questions or with statements. It therefore seems reasonable to say that we are dealing with 'the same' tune across varieties of English. The only difference is in the meaning or function of the tune.

However, recent production studies by Fletcher, Grabe and Warren (2004), and Warren and Daly (2004) would seem to have found evidence that goes against Ladd's observation. Fletcher et al. found that some Australian speakers start their rises with a lower pitch for statements (HRTs) than for questions and Warren and Daly (2004) found that some New Zealand speakers start their rises earlier for questions than for statements. Warren and Daly (2004) point out that both the Australian and the New Zealand trends result in a more dramatic rise for HRTs than for questions. In the first example this is because there is a greater total change of frequency; in the second example it is because the frequency change is compressed into a shorter time period, resulting in a greater rate of change. It is worth noting that both the above trends could also be partly due to differences in the pre-nuclear contour of statements and questions, which may be gradually falling from a high level for statements, but rising from an initial low for yes-no questions. The difference between the Australian and New Zealand English patterns described above, with Australian statement rises starting lower but NZE statement rises starting later, may well be related to a finding by Vermillion (2003) from a perceptual comparison with British English that NZE is characterized by higher valleys between pitch peaks. That is, NZE statement rises may be distinguished from question rises by the relative timing of the rise, because compared with Australian English, NZE allows less scope for lower rise onsets within the intonation phrase.

As just noted, Warren and Daly (2004) observed a relationship between whether an utterance was intended as a statement or as a question and the lateness of the rise produced by the talker. In terms of the AM framework,

this amounts to a difference in the *alignment* of the rise, relative to the segmental material with which it is *associated*. That is, in Warren and Daly's data a tune that in current descriptions has the same AM label (e.g. a sequence of H* H-H%) has different phonetic realisations according to whether the tune is indicating a question or a statement. If this is a consistent relationship, then it would be surprising to find that it is not exploited by listeners in comprehension. Such a finding would support the view that a difference in the alignment of the rise corresponds to a phonological difference, marking the utterance as a question or a statement. This would therefore require a revision of the phonological labelling of these rises in order to convey the distinction between them.

The question and statement rises analysed by Warren and Daly were realized on different lexical material, and may also have carried unmeasured distinctions in the pre-nuclear stretch of the tunes. The perception experiment described below includes control both of the words used in the utterances and of the earlier parts of the contour, and so allows a closer examination of the relationship between the alignment of the final rise and the statement/question status of the utterance.

The aim of this study was therefore to investigate the perceptual consequences of the types of phonetic difference found by Warren and Daly (2004) in the production data of young New Zealanders. The experiment uses listeners from the same demographic group as Warren and Daly's speakers.

Experiment

The current experiment investigates listeners' interpretation of different rise alignments in a binary choice task (deciding whether a stimulus is a statement or a question). If the phonetic differences observed by Warren and Daly (2004) are indicative of a phonological distinction between question and statement rises, then we would expect to find that later rises are more frequently reported as being statements. As well as indicating whether the stimulus was a statement or a question, participants were required to give a confidence rating for their decision, allowing a more qualitative look at response patterns.

Materials

The stimuli for this perception study were designed to reflect the distribution of lateness of rise found in Warren and Daly's (2004) production data. Their data consisted of more than two hundred utterances obtained from a map task, i.e. from a spontaneous unscripted interaction. For each final rising utterance, Warren and Daly recorded the starting point of the final rise, as a proportion of the distance (in syllables) from the nuclear accent to the end of the utterance. They report rises with a range of starting points between 0.2 and 0.9 of the total distance through the accent unit. To allow a similar range

of starting points in my perceptual study, this experiment uses an accent unit with 6 syllables from the accent-bearing syllable to the end of the unit. This utterance was *He'll be at the BASKetball stadium*¹.

As an exploratory first study of the perceptual consequences of rise alignment, this study employed just this single utterance for generating the experimental stimuli. This also allowed us more easily to ensure that listeners were basing their judgements solely on intonation. The experimental stimuli were derived from single utterances of the target sentence from both a female speaker and a male speaker. Each produced the sentence with an intonation contour that matched the definitions of HRT given above, i.e. it contained a rise that started at a high level and went even higher. (In fact the rise of 9 semitones, corresponding to the natural magnitude of rise for the female speaker, was slightly greater than that described by Britain and Newman, 1992.) The inclusion of materials from both a female and a male speaker is motivated by Warren and Daly's (2004) finding of differences between female and male speakers in their realisation patterns for statement and question rises.

The source utterances were produced with the pre-nuclear material (*He'll be at the*) on a low, level pitch contour, and with an H* H-H% tune on *basketball stadium*. The female speaker was recorded first, and then the male speaker (the first author) was recorded, taking care to mimic the tune, rhythm and duration of the female speaker. The F0 contour of the accent unit (*BASKetball stadium*) in each of the source utterances was then manipulated to produce stimuli with a range of starting points for the rise, using the software package *Praat* (Boersma & Weenink 2001). First the duration of the accent unit, i.e. the segmental material starting at the release of the /b/ in *basketball* and finishing at the end of the /m/ in *stadium* was divided into five equal time segments. For each speaker, two sets of five stimuli were then synthesised, with rises starting at the beginning of each of these segments. One set (gradient stimuli, see Figure 1) had rises of five different gradients, each one linear from its respective starting point to the end of the utterance. The other set (plateau stimuli, see Figure 2) had a "rise-and-plateau" contour, with various durations of plateau time. Note that the level of the plateau is that of the final H%. The purpose of having the two sets was to investigate whether the crucial factor in variable alignment of the rise is in fact the actual timing of the start of the rise, or the amount of pitch information at the higher level.

¹ The capitals in *BASKetball* denote the nuclear stress, i.e. the first syllable of the accent unit under consideration.

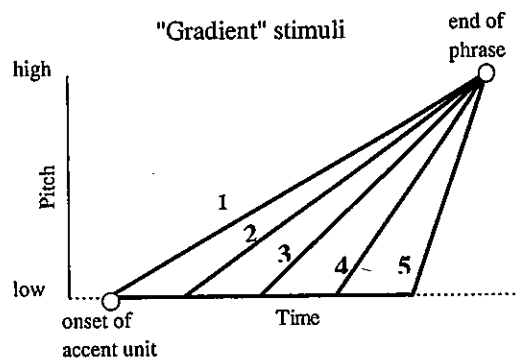


Figure 1. Schematic representation of the "gradient" stimuli used in the perceptual experiment.

For the purposes of the experiment, it is clearly important to try to ensure that listeners are not influenced in their decision by the lexical make-up of the stimuli. Two main ways of reducing such an influence were considered prior to the preparation of the materials. One was to low-pass filter the utterances so that the lexical material was unintelligible. Low-pass filtering is a common technique in intonation studies to mask lexical and grammatical information (see Vermillion, 2003). Removing the source of such information from the listener allows us to be surer that listeners are basing their judgements on the intonation of the stimuli, since lexical and grammatical information commonly indicates whether something is a question or a statement. However, we could not be sure that listeners wouldn't imagine a sequence of words to fit the rhythmic structure of the low-pass filtered utterance. Their decision as to whether the utterance was a statement or a question might then be influenced by this imagined sentence. In addition, note that the chosen utterance contained the stressed (but unaccented) post-tonic syllable *sta*. There was a danger that if the utterance were filtered, tokens with late rise onsets might lead to the percept of a nuclear accent on this syllable, since

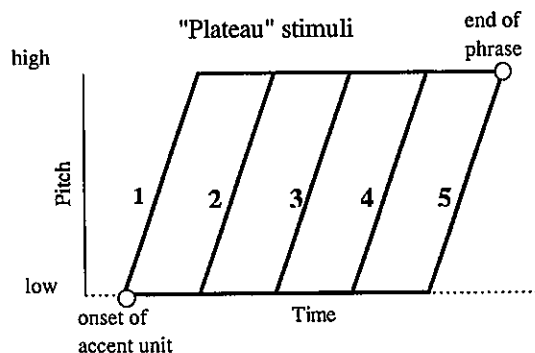


Figure 2. Schematic representation of the "plateau" stimuli used in the perceptual experiment.

there is no lexical/grammatical information suggesting it should be earlier. Since a sequence of five or more post-tonic syllables will almost certainly include a stressed syllable, low-pass filtering would severely limit the possible length of the accent unit, and compromise the range of starting points for the rise.

The other means by which the influence of lexical material can be reduced, and the method chosen for this experiment, is to provide contexts in which both the statement and the question interpretation are possible. This entailed providing two dialogue context passages for the utterance *He'll be at the BASKetball stadium*, as shown in 1) and 2) below. Dialogue 1) indicates how the words of the stimuli could be interpreted as a question, while dialogue 2) shows how they could be a statement. Note also how the dialogue contexts render unlikely the placement of a nuclear accent on *stadium*.

1) Question dialogue

A: Jim's just finished a job. That arts show was doing a feature on the architecture of stadiums, and he was getting footage of the basketball one. Can you give him a lift home?

B: *He'll be at the basketball stadium?*

A: Yeah.

B: Sweet as. That's on my way back from the workshop.

2) Statement dialogue

A: Jim's just finished a job. That arts show was doing a feature on the architecture of stadiums, and he was getting footage of the basketball one. Can you give him a lift home?

He'll be at the basketball stadium.

B: Sweet as. That's on my way back from the workshop.

Procedure

Ideally, the experimental stimuli would be presented to each participant separately, allowing us to present different orders for each participant, in order to control for any sequencing effects caused by one particular order. However, because this study was run in tandem with another unrelated experiment, having multiple separate orders was not possible. Consequently, only two orders were used. For each source utterance (i.e. female and male), the ten stimuli were arranged in two blocks of five, preceded by a practice block of five stimuli. This practice block exposed participants to the extremes of both types of stimulus. Within each of the female and male lists the stimuli were placed in random order, but the male and female sets were kept separate.

Before hearing the stimuli, the participants were asked to read silently the two dialogues in 1) and 2) above. They were asked to re-read the dialogues between hearing the male and female stimuli, and to imagine the speaker of the stimuli as being of the appropriate sex. In addition to the forced choice between statement and question, participants were asked to give a confidence rating, from one to five, for each of their judgements.

Subjects

The participants were native speakers of New Zealand English and mostly aged younger than 20, and therefore in a similar demographic category to the informants recorded by Warren and Daly (2004). Most of them were students of linguistics, but care was taken to ensure that they were naïve with respect to research on HRTs. Time constraints and a technical fault in the language laboratory equipment meant that the experiment was run on only 22 participants. Three of these gave "trivial" responses, that is all or nearly all "statement" or all "question", and were thus discarded. Therefore the data analysis was done on the responses from 19 participants.

Hypothesis

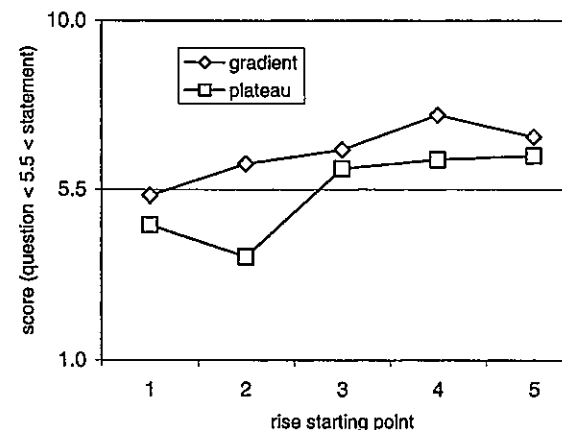
If the phonetic differences observed by Warren and Daly (2004) correspond to a nascent phonological distinction between question and statement rises, then we would expect to see a link between the lateness of the rise and whether the stimulus is perceived by the listener as a statement or a question. Specifically, we would expect to see more "question" responses for stimuli with earlier rises, and more "statement" responses for stimuli with later rises. If we do see such a pattern, indicating that there is a phonological distinction, and if the sole parameter for that distinction is the timing of the start of the final rise, then we would expect to see no significant difference in the pattern of responses from the two series of stimuli. If, however, the shape of the final rise is important as well as the start point, we would expect to see a different pattern of responses for each of the series of stimuli.

Results

The results presented here are based on a reanalysis of the original data from this perception study (see further Warren, ms). Each response was assigned a score on an integer scale from 1 to 10, ranging from "very question-like" (i.e. a "question" response with a maximum confidence rating of 5, which was re-scored as 1) through to "very statement-like" (a "statement" response with a maximum confidence rating of 5, re-scored as 10). An average score of 5.5 would therefore represent bias towards neither question nor statement responses. Average scores lower than 5.5 indicate a bias towards question responses, while average scores higher than 5.5 represent statement bias. These data were then entered into an ANOVA with speaker sex (male vs. female), continuum type (gradient vs. plateau) and rise starting point (1 through 5) as factors. The main effects of continuum type and rise starting point were significant ($F[1,360]=12.909, p<0.001$ and $F[4,360]=7.581, p<0.001$). There was no significant difference for speaker sex, nor did speaker sex interact significantly with either of the other factors. The overall results, collapsing male and female, are shown in Figure 3. Note that negative values

indicate that, on average, participants thought that the stimulus was a question, while positive values favour a statement. Note also that the participants were fairly conservative in their use of the confidence rating, so the average values cluster round the neutral 5.5 score. The simple counts of "statement" v "question" responses have a similar pattern to that shown in Figure 3, but with steeper curves.

The main effect of continuum type reflects the higher "statement" response for the gradient stimuli (average over all 5 starting points: 6.49) than for the plateau stimuli (5.40). The main effect of starting point arises because later starting points attract higher "statement" scores. Post-hoc analyses of the starting point data show that response scores to stimuli with starting points 1 and 2 are significantly lower (more "question-like") than those with starting points 3, 4 and 5.



Discussion

The results of the forced-choice perception experiment presented in this paper clearly show that early rises are more likely to result in question judgements, and later rises in statement judgements, reflecting the pattern found by Warren and Daly (2004) in their production data. It also shows that in contrast to Warren and Daly's observation from production data, this tendency is not more strongly associated with male speakers, since no significant effect was found for speaker sex in the current study. The significant difference between the gradient and plateau continuum types shows that it is not just the alignment of the final rise start that is important, but also the shape of the rise.

A common feature of the two significant results in this study (rise starting point and continuum type) is that in each case more high pitch information leads to a more "question-like" response. As an indication of how this varies, consider how the areas below the respective graphs in Figures 1 and 2 vary according to the amount of high pitch information. This area measured for the first stimulus in the gradient series is equivalent to that for the third stimulus in the plateau series.

It would appear that a feature for distinguishing question utterances is therefore that there is a good amount of high pitch in the accent unit. Conversely, the property that distinguishes HRT rises on statements is that they are late and dramatic, again matching observations from Warren and Daly's production data. In terms of the response scores shown in Figure 3, it would appear that the best "question" rise would be one with a quick early rise (somewhere in the first two fifths of the accent unit) to a sustained high level, while the best "statement" rise is a late rise to the endpoint of the contour.

In the description of HRT tunes presented by Ladd (1996, see above for details), both statement HRTs and question rises are categorized as having the same possible tunes. Evidence from production studies by Fletcher et al. (2004) and Warren and Daly (2004), and further evidence now from the perception data reported in the current study all suggest that there is a difference in the realization of these rise types. If – as the perception study suggests – this difference has phonological significance, in that it signals a meaning difference between statements and questions, then an intonational transcription system for NZE needs to be able to represent the contrast. Since the tunes otherwise seem similar in AM terms (the tunes used in the perceptual experiment both being H* H-H%), the difference might indeed be best represented in terms of the alignment of part of this tune (see similar treatment of question and statement differences in terms of alignment in Neapolitan Italian: D'Imperio & House, 1997). We suspect that the difference may be in the alignment of the phrase accent H-. The H* pitch accent is both associated and aligned with the nuclear syllable (the first syllable in *BASKETBALL* in the perception experiment), and the H% boundary tone is aligned and associated with the end of the utterance. With respect to phrase accents, the guidelines for ToBI labelling (Beckman & Ayers Elam, 1997: 17) indicate that "when the nuclear accent is far from the end of the intermediate phrase, the phrase accent fills in the space in between it and the phrase edge, creating [...] a long plateau-like region for H- realized over a long stretch." The conventional transcription of a tune such as H* H-H% places the H- label adjacent to the H% at the end of the intonational phrase (Beckman & Ayers Elam, 1997: 18), and assumes that the phrase accent spreads back to the previous pitch accent. A modification of this AM treatment for our NZE rises would, we maintain, involve differences in the alignment of our H- phrase accents in statements or questions, indicated in transcription e.g. by using a < symbol to indicate early alignment (similar to the use of < and > to indicate alignment of the phonetic peak of pitch accents relative to segmental

material). Further detailed study of the realization and understanding of different contour shapes in NZE intonation will be necessary for determining the status of alignment differences in the context of the intonational system as a whole, as well as a consideration of further issues not discussed in this paper, such as whether phrase accents in NZE might have secondary associations to other material in the utterances (cf. Grice, Ladd & Arvaniti, 2000; Gussenhoven, 2000), which might in turn vary according to utterance types.

References

- Ainsworth, Helen. 1994. The emergence of the high rising terminal contour in the speech of New Zealand children. *Te Reo* 37, 3-20.
- Beckman, Mary E. and Gayle Ayers Elam. 1997. *Guidelines for ToBI Labelling Version 3*, March 1997. (http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/)
- Boersma, Paul and David Weenink. 2001. *Praat 4.0 – a software package for doing phonetics by computer* (<http://www.praat.org>)
- Britain, David and John Newman. 1992. High rising terminals in New Zealand English. *Journal of the International Phonetic Association*, 22, 1-11.
- Cruttenden, Alan. 1995. Rises in English. In J. Windsor-Lewis (Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor*. London: Routledge. Pp. 155-173
- D'Imperio, Mariapaolo and David House, 1997. Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech '97*, volume 1, 251-254, Rhodes, Greece.
- Fletcher, Janet, Esther Grabe and Paul Warren. 2004. 'Intonational variation in four dialects of English: the high rising tone.' In Sun-Ah Jun (Ed.), *Prosodic typology and transcription: a unified approach*.
- Hirst, Daniel and Albert Di Cristo (Eds.) 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Grice, Martine, D. Robert Ladd and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17, 143-185.
- Gussenhoven, Carlos. 2000. The boundary tones are coming: on the nonperipheral realization of boundary tones. In Michael Broe and Janet Pierrehumbert (Eds.), *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press. Pp. 132-151.
- Guy, Gregory and Vonwiller, Julie 1989. The High rising tone in Australian English. In Peter Collins and David Blair (Eds.), *Australian English. The Language of a New Society*. St. Lucia: University of Queensland Press. Pp. 21-34.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: CUP
- Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation. Cambridge, Massachusetts, USA: MIT
- Vermillion, Patricia. 2003. The ups and downs of Kiwis: An experiment investigating tonal cues used to identify NZE intonation. *Wellington Working Papers in Linguistics* 15, 17-31.
- Warren, Paul and David Britain. 2000. 'Intonation and prosody in New Zealand English.' In Allan Bell and Koenraad Kuiper (eds), *New Zealand English*. Wellington: Victoria University Press. Pp. 146-172.

Warren, Paul and Nicola Daly 2004. Characterising New Zealand English Intonation: broad and narrow analysis. In Allan Bell, Ray Harlow and Donna Starks (Eds.), *Languages of New Zealand*. Wellington: Victoria University Press.

Warren, Paul (ms). Patterns of late rising in New Zealand – intonational variation or intonational change?

Previous volumes

CONTENTS: Volume 1, 1990

'Anahina 'Aipolo and Janet Holmes: The Wellington Tongan community: prospects for language maintenance

Nicola Daly: Sri Lankans and Sinhala language maintenance in New Zealand

Jane Pilkington: Mixed marriages and language maintenance: some Samoan data

Wendy Thompson: Attitudes to Maori and the use of Maori lexical items in English

CONTENTS: Number 2, 1990

Laurie Bauer: English vocabulary during the Twentieth century: an experimental approach

Lisa Matthewson: The negative clitic in Middle High German

Elizabeth Pearce: An analysis of negated infinitivals in Middle French

Xuezhong Zhang: The structure of Chinese dative/benefactive and double object sentences

CONTENTS: Number 3, 1991

Janet Holmes: Women's verbal contributions in public settings

Caroline McGhie: The jargon of Wellington taxi dispatchers

Catherine Neazor: Language maintenance and shift in the Wellington Polish community

Michael Williams: Post-vocalic (r) in the urban speech of the Isle of Wight

CONTENTS: Number 4, 1992

Laurie Bauer: Review of J. Kelly and J. Local *Doing Phonology*

Gavin Carr: On Heads in Morphology and Syntax

Lisa Matthewson: Prosodic Morphology in Alabama Negation

Matthew Scott: An Assimilatory Neutralization in New Zealand English

CONTENTS: Number 5, 1992

Chris Lane: Repetitive questioning in courtroom interaction and interlanguage communication

Lisa Matthewson: Talking to the TV. Conversational behaviour of television viewers

Jane Pilkington: 'Don't try and make out that I'm nice!' The different strategies women and men use when gossiping

Maria Stubbe: What's the score? Qualitative analysis in gender research.

CONTENTS: Number 6, 1994

Anita Easton: Talk and laughter in New Zealand women's and men's speech

Jen Hay: Jocular abuse in mixed gender interaction

Janet Holmes and Allan Bell: Consonant cluster reduction in New Zealand English

Camille Plimmer: Language maintenance and shift in the Italian community in Wellington

CONTENTS: Number 7, 1995

Laurie Bauer: Is the morpheme dead?

Jen Hay: The Adjacency Condition and the Atom Condition: compatible morphological constraints

Elizabeth Pearce: V-movement and optionality

Paul Warren: The sound of desert trains: delay strategies and constraints in spoken sentence processing

CONTENTS: Number 8, 1996

- Jen Hay: No laughing matter: gender and humour support strategies
Gary Johnson: The management of interaction in the television interviews of Maggie Barry
Shelley Robertson: Maori English and the bus-driving listener: a study of ethnic identification
And phonetic cues
Ben Taylor: Gay men, femininity and /t/ in New Zealand English. Report on Research in
Progress
Janet Holmes and Chris Lane: Interactional Variation in Environmental Interviews
Book Review: Mary Crawford *Talking Difference*. Reviewed by Janet Holmes

CONTENTS: Number 9, 1997

- Diane Massam: Three Faces of Niuean *aki**
Elizabeth Pearce: DP Licensing and Spec roles in Maori
Elizabeth Pearce: Genitive Case in the Maori DP
Shizuka Torii: Two Types of Evaluation Time and subject Marking in Japanese

CONTENTS: Number 10, 1998

- Janet Holmes: The linguistic construction of gender identity
Jane Pilkington, Janet Holmes and Chris Lane: Gossip: its context and its boundaries
Sandra Shearn: Interviewing styles on New Zealand National Radio: is Mike Hosking really so
outrageous?
Robert Sigley: Interpreting social variation using stylistic continua: the strange case of
relativiser choice
Lynnette Sollitt-Morris: The paradox of silence in interaction: an indicator both of power and
powerlessness

CONTENTS: Number 11, 2000

- Sai Hui: Humour in a Cantonese family: an analysis in an asymmetrical small group setting
Jinnie Potter: Chartwell School: sixteen years of a Japanese-English bilingual venture
Maria Stubbe, Chris Lane, Jo Hilder, Elaine Vine, Bernadette Vine, Janet Holmes, Meredith
Marra and Ann Weatherall: Multiple Discourse Analyses of a Workplace Interaction

CONTENTS: Volume 12, 2000

- Angela Ford: A DP analysis of Iaa noun phrases
Elizabeth Pearce: Transitivity, Incorporation and Animacy in Iaa
Elizabeth Pearce and Samuel Ukewea Wadjeno: Uny me Cako: Iaa text
Shizuka Torii: What is tense?

CONTENTS: Volume 13, 2001

- Laurie Bauer: Whither the Thesaurus
Laurie Bauer: How and Why the Phonological Word and the Morphosyntactic Word do not Coincide
Hannah Buchanan: Neutralisation of DRESS and TRAP before /l/ in New Zealand English
Yono Sukarno: vP-aspect: Aspectual Base Properties of Indonesian Predicates

CONTENTS: Volume 14, 2002

- Laurie Bauer and Paul Warren: Affix-able words: morphological productivity as the basis for
(de)composition in processing
Gloria Hu: Phonological awareness in young Cantonese-Mandarin bilinguals
Megan Rae and Paul Warren: The asymmetrical change in progress of NEAR and SQUARE vowels in NZE:
psycholinguistic evidence
Jen Hay and Paul Warren: Experiments on /r/-intrusion