Pitch accent variation and realization in interactive discourse in Australian English

Janet Fletcher and Debbie Loakes

School of Languages and Linguistics
University of Melbourne
Centre of Excellence for the Dynamics of Language
janetf;dlloakes@unimelb.edu.au

Abstract
This paper examines pre-nuclear and nuclear pitch accent variation and realization in map task interactions for Standard Australian English. An analysis of pitch accent choice in accented words reveals that pitch accent type is not always critical to the realization of information structure categories; for instance in the case of informational focus participants often use a range of pitch accent shapes. Nevertheless nuclear bitonal accents (e.g. L+H* pitch accents) are scaled higher than simple H* pitch accents across the board in line with previous findings for other English varieties. Furthermore there is evidence of an interaction between dialog act, pitch accent choice and realization with more rising and/or higher scaled accents used in particular kinds of interactions.

1. Introduction
The majority of research on Australian English intonation has focused on the prevalence and functions of Uptalk in interactive discourse [see 1 for a comprehensive overview]. Uptalk refers to the use of rising (often high rising) pitch at the end of syntactic declarative statements. Conversely in previous studies of Australian English intonation there has been relatively little focus on post-lexical pitch accent variation, particularly with regard to pitch accent type and pitch accent scaling, compared to other varieties of English. Earlier corpus studies for British English varieties [2] and experimental studies of American English [e.g. 3] have revealed that there is strong evidence of variation in terms of pitch accent type and pitch accent realization across different speech communities. Moreover, it has generally been assumed for Australian English that, as in Southern British English and General American English, pitch accent choice and realization play a crucial role in signaling information structure categories (e.g. focus) in spoken communication. With regard to the realization of neutral, informational, and contrastive focus, studies of American English have shown that a high target pitch accent with a strong lead tone, e.g. L+H*, tends to be used in narrow and contrastive focus contexts rather than simple H* pitch accent [3, 4] although there is some evidence of variation among different American varieties [e.g. 3]. Moreover the H* tone target of the rising L+H* accent is generally scaled higher in a speaker’s range in these narrow and contrastive focus contexts. Similar effects of focus have been found in many different languages [e.g. see 5, 6, 7, 8 for an overview]. Preliminary acoustic studies of Australian English also suggest that rising accents play an important role in signaling contrastive focus in carefully scripted laboratory phonology type experiments [e.g. 9].

An example of a potential contrastive focus context is shown below. In this example (utterance b), MANLY would potentially be realized with a rising L+H* rather than simple H* accent (both shown schematically in Figure 1) to emphasise that the interlocutor went to Manly and not Bondi. The simple H* shown on the left does not have the strong rising lead tone that characterizes the L+H* pitch accent on the right (located on the main stressed syllable of Manly). According to the standard ToBI descriptions of these accents for other varieties of English, the H* target of the bitonal L+H* accent is also likely to be scaled higher in the speaker’s pitch range as shown in Figure 1.

a) Did you go to Bondi baths yesterday?
b) No I went to MANLY baths.

Figure 1. Schematic representation of two major pitch accent f0 shapes: H* and L+H*.

It remains to be seen whether rising pitch accents are widely used in less scripted data in Australian English in general, and/or whether they are used frequently to signal contrastive focus in spoken interaction compared to what has been observed in controlled laboratory-style speech. It also remains to be seen whether pitch accent variation constitutes another source of potential differentiation between Standard Australian English and other so called “mainstream” varieties of English. It may be that rising accents, and not only high-rising boundaries, contribute to the general perception of Australian English as a ‘rising’ variety but this remains to be quantitatively verified, e.g. [1].

The main focus of this study is to investigate pitch accent variation and realization in a corpus of map task data (Australian National Database for Spoken Language, henceforth ANDOSL) for Australian English [10]. The additional aim of this study is to provide a “reference” dataset for future comparison with the AusTalk data collection which also includes a map task [11]. It should be noted that the ANDOSL Map Task used here was collected in Sydney in the ’90’s and is indicative of Standard Australian English spoken in the Sydney/NSW region at that time, although it should also be noted there is no indication that prosody has changed over time in Australian English. This study will give a further
understanding of (rising) prosodic patterns in spontaneous Australian English.

1.1 ANDOSL Map Task

The Map Task is an exercise in controlled quasi-spontaneous talk between two participants and therefore constitutes a rich resource for the investigation of situated interaction and in particular the interaction between intonational categories like pitch accents and information structure (analysed in the current study). The structure of a map task can be summarized as follows: participants work in pairs, each with a map in front of them that the other participant cannot see. The maps contain a number of landmarks, which are not all identical. One participant (the ‘instruction-giver’ IG) has a route marked on his/her map and is required to instruct the other participant (the ‘instruction-follower’ IF) to draw the correct route onto their own map. The mismatching landmarks on each map ensure that a range of queries, checks and negotiation talk will be elicited. This is one reason why map tasks are an effective tool to examine intonational variation given the known relationship of intonation to utterance modality and discourse segmentation (e.g. [12]).

2. Method and Materials

2.1 Participants

Eight dialogues (4 speaker pairs) from the Australian Map Task corpus of the Australian National Database [10] formed the dataset for this study. This was a subset of the corpus that was analyzed in earlier studies of Uptalk and high rising tunes in Australian English (e.g. [13]) and studies of discourse modeling (e.g.[14]). Each speaker pair consisted of a male and female who took it in turns to be Instruction Giver (IG) or Instruction follower (IF). The dialogues were chosen randomly and the speakers all belong to the standard Australian English dialectal grouping. The dialogues were between 485.93 sec and 810.24 sec in duration and were digitized at 22,500 kHz.

2.2 Word and prosodic labeling

The map data were annotated according to ToBI (Tones and Break Indices) conventions that have been adapted for Australian English ([15]). Fundamental frequency contours (F0) were annotated for pitch accents. Major pitch targets or movements corresponding to pitch accents (e.g. H*, L+H) intermediate phrase boundaries (e.g. L-H) and intonational phrase boundaries (e.g. L%H%) were labelled using the F0 signal and auditory analysis by four expert ToBI transcribers. There were 2831 intermediate or intonational phrases in the dataset with the equivalent number of nuclear accents (recall that a nuclear accent is the head of an intermediate phrase in the AM model that underpins AusE ToBI), and 1964 pre-nuclear accents (7626 tokens in total). Word boundaries were manually identified from speech waveforms and spectrograms and orthographically annotated using ESPS/xwaves and emulabeller [16]. Pitch accent categories were chosen from the main set of tonal categories proposed for AusE ToBI: L* (low tone target realized in the lowest part of a speaker’s pitch range), H* (high tone target corresponding to a shallow peak in the higher part of a speaker’s pitch range), L+H* (high tone target with strong rising lead tone), L+H tone target (low tone target on accented syllable with late rising tone), and downstepped variants including H* (lowered pitch peak relative to a pre-nuclear accent in the same intermediate phrase).

Of particular interest in this study is variation in simple (H*, !H*) versus bitonal (L+H*, L*+H) pitch accent choice, and scaling of the L and H targets in pre-nuclear and nuclear contexts. An example of intonational annotation illustrating two different pitch accent types, H* and L+H* respectively is shown in Figures 2 and 3. Both are examples of contrastive focus constructions which were of additional interest in our study. Due to the mismatch between the two maps used by the participants in the task, the IF’s often produced these kinds of constructions in response to a request by the IG. The illustration in Figure 2 shows the second part of an exchange between an IG and an IP pair which is shown below.

IG “Do you have a galah open cut mine?”
IF “I’ve got a DINGO.”

The IF responds to an information question posed by the IG about whether or not he has a particular landmark on his map, in this case “the galah open cut mine”. The IF has a “mine” with a different name on his map and so in his response, the word “DINGO” is realized with a simple H* nuclear accent scaled in the upper part of this speakers’ pitch range, rather than a L+H* bitonal accent (there is no strong lead tone evident on the accented first syllable of “Dingo”).

Figure 2. F0 contour showing an example of a nuclear H* pitch accent produced by a male participant (Instruction follower) on the accented first syllable of “Dingo” in a map task dialog.

Figure 3. F0 contour showing an example of a nuclear L+H* pitch accent produced by a male participant (Instruction Giver) on the accented first syllable of “bottom” in a map task dialog.

Figure 3 by contrast shows a L+H* pitch accent on the nuclear accented word “bottom” in the utterance “down to the bottom of the spruce trees”. In both cases the boundary tones also differ. This kind of pitch accent variation suggests that we may find more variation in these kinds of quasi-spontaneous interactions than has been found in more scripted laboratory tasks.

2.3. Data Analysis

F0 was calculated using the Schaeffer-Vincent algorithm in the Wrassp library in emuIt [18] and values extracted for pitch accent categories at the location of the annotated starred tone. For H*, !H* and L+H* pitch accents this was usually late in the rhyme of the accented syllable and for L* and L+H accents this was located in the middle of the (low) pitch elbow
associated with the accented syllable rhyme. The f0 values extracted at the * tone target were converted to semitones (benchmark 50). A linear mixed effects analysis was then performed in R (lmerTest and step [19]) to compare pitch accent scaling across the corpus. A maximally specified mixed model was implemented with fixed factors and interactions for PITCH ACCENT, SEX, POSITION (nuclear versus pre-nuclear) and map task ROLE (Instruction Giver [IG], Instruction Follower [IF]) with random factor SPEAKER. Post-hoc comparisons using Bonferroni correction were undertaken to investigate any significant interactions. It is well-attested that females have higher mean pitch than male speakers and it was assumed that this would be a significant factor. The main point of this analysis was to explore any simple main effects and potential interaction between speaker role and pitch accent scaling differences, with a particular focus on rising versus simple high pitch accents in nuclear versus pre-nuclear contexts.

3. Results

3.1. Nuclear and pre-nuclear pitch accent distribution

The distribution of major simple and bitonal accents in the eight map tasks is shown in Figure 4 for speakers in the IG role, and in Figure 5 for speakers taking the IF role.

![Figure 4. Distribution of simple and bitonal (two tone) pitch accents in nuclear (black) and pre-nuclear (grey) position - Instruction Giver (IG) role.](image)

There is similarity in nuclear pitch accent distribution across the two roles with some exceptions. Simple H* accents are the most frequent accent-type across the corpus (more than 50% of all accents produced by IGs, and 62% of all accents produced by IFs). L+H* pitch accents also significantly outnumber the instances of L*+H pitch accents. The higher proportion of simple high accents in the Instruction Follower sections of the task can be related to the lower number of downstepped !H* accents. Speaker turns and utterances tended to be shorter in IF contexts [see 17] with a high number of IPs and dialog acts consisting of a single pitch accent and boundary tone configuration. With regard to rising (L+H*, L*+H) versus simple H* or !H* accents, around 13% of nuclear accents are rising in this corpus. The effect of speaker role (IG or IF) was not significant ($\chi^2=3.3351$, $p>0.05$) in contrast to speaker SEX which was ($\chi^2=9.3775$, $p<0.001$). Females produced more bitonal rising accents (L+H*) in nuclear position than males who also produced fewer L*+H accents than females. Pre-nuclear accents show a broadly similar distribution - 60% of pitch accents are simple H* and around 15% are rising L+H* accents and there was no significant effect of speaker SEX ($p>0.05$) in terms of pitch accent selection.

![Figure 5. Distribution of simple and bitonal (two tone) pitch accents in nuclear (black) and pre-nuclear (grey) position - Instruction Follower (IF) role.](image)

3.2. Pitch scaling of nuclear and pre-nuclear accents

Pitch level (scaling) values in semitones (ST) for the different pitch accent types (simple and bitonal) in nuclear position are shown in Figure 6 for female participants and in Figure 7 for male participants in the Instruction Giver (IG) and Instruction Follower (IF) roles. Notwithstanding the predictable effect of speaker SEX (all pitch accents are scaled higher for females compared to males), there were statistically significant main effects for ROLE, POSITION (nuclear versus pre-nuclear) and PITCH ACCENT, with interactions between POSITION and PITCH ACCENT and ACCENT and ROLE.

![Figure 6. Pitch level (semitones) of the *tone of pitch accents produced by female participants in the Instruction Follower role (left panel) and Instruction Giver role (right panel).](image)

Post-hoc tests (Bonferroni) confirm that the differences in scaling between rising accents (L+H*) and simple H* accents is significant in both speaker roles for males and females (M: $t=14.73$, $p<0.0001$; $t=-9.95, p<0.0001$). It should also be noted that there is a degree of overlap between pitch values for H* pitch accents and the H tone of L+H* pitch accents. In terms of speaker ROLE, H* accents are scaled higher for IGs versus IFs ($t=-5.78$, $p<0.0001$), but there is no significant scaling difference for L+H* accents according to ROLE. Nuclear H*
and L+H* accents are scaled higher than pre-nuclear high tone accents (H*; t=-5.77, p<0.001; L+H*; t=-13.41, p<0.0001) in both roles with nuclear L+H* accents consistently scaled higher than H* accents. Pre-nuclear !H* or L* accents are not scaled higher or lower than their nuclear counterparts.

4. Discussion

Australian English speakers use a variety of pitch accents in these map task interactions. By far the most common accent type is the simple H* accent which reflects a pattern observed in corpus studies of other varieties including Mainstream American English, although proportionally fewer rising accents are actually observed in our corpus compared to [20]. This suggests that the observed 'rising' nature of Australian English is actually due to the high incidence of intonational phrase-final rises, i.e. Uptalk [see 1, 17], rather than frequency of rising pitch accents within phrases. The contrastive status of the L+H* versus H* accent has also often been a source of debate in intonational phonology [e.g. 21] and it could be that transcribers were conservative in their use of this accent category in this corpus given its well-known ambiguity with H* accents [20]. Nevertheless, the H tone of transcribed nuclear and pre-nuclear rising accents (L+H*) is consistently scaled higher than simple H* accents in these Australian English map task interactions, in keeping with earlier findings for other English varieties [3, 20], although there is unsurprisingly a degree of overlap observed in accent realizations.

These findings also support observations from scripted laboratory-style speech for younger Australian English speakers from Melbourne [9], although in this study (using unscripted speech) it is not always the case that rising accents are used exclusively in contrastive focus or informational focus constructions. Higher scaled H* accents are more commonly used, as well as the occasional use of H+!H*. Rising accents also tend to be used in informational focus contexts where one participant wishes to draw the other participant’s attention to pertinent information in relation to the task. Rising accents also occur in specific types of dialog acts including action directives to influence the action of the addressee i.e. the IF, and responses to the IG. A preliminary analysis of dialog acts which include contrastive focus contexts (e.g. cases of negative responses to instructions by the IG), suggests that speakers produce both single high tone and rising pitch accents in these focus contexts, although both accent types are typically scaled higher than pre-nuclear accents in the same intonational/intermediate phrase. This suggests that the normal downturn trends patterns often observed in an intonational phrase are not apparent in these cases. Female participants also tend to use more rising accents than male participants. It may be that the expanded pitch range associated with female speech makes it easier for transcribers to annotate a rising versus simple high pitch accent, or we may be taping into pragmatic and/or sociophonetic variation amongst Australian English speakers, but this requires further investigation. In summary, intonational analysis of situated interaction (even in quasi-spontaneous speech tasks) produces significant patterns of variation that should be captured by our intonational models, and gives a greater understanding of the nuanced ways speakers use prosody in spoken communication.

5. References

Author/s:
Fletcher, J; Loakes, D

Title:
Pitch accent variation and realization in interactive discourse in Australian English

Date:
2018

Citation:

Persistent Link:
http://hdl.handle.net/11343/219323

File Description:
Published version