Interpreting rising intonation in Australian English

Janet Fletcher, Deborah Loakes

School of Languages and Linguistics, University of Melbourne
janetf@unimelb.edu.au, dloakes@unimelb.edu.au

Abstract

Australian English is referred to widely as a rising variety of English due to the prevalence of rising tunes in interactive discourse. Australian English subjects were required to listen to a series of rising stimuli that varied in terms of pitch level and pitch span and were asked whether they heard a question or statement. The results showed that both rise span and pitch level of the rise elbow influenced the pattern of responses. If both were relatively high, subjects were most likely to interpret the rise as a question, with fewer question responses when the rise elbow was relatively low and the pitch span narrow. The results provide limited evidence for two simple rises in Australian English, but also confirm a high level of phonetic gradience amongst rising tunes in this variety.

Index Terms: intonation, uptalk, phonetic gradience

1. Introduction

The most frequently observed intonational phenomenon in Australian English is the rising statement tune, whereby utterances that might be typically produced with falling intonation in one variety are often realized with a high rising tune in this variety. This phenomenon is remarked on frequently in public discourse in Australia and elsewhere. While it has been argued in [1] that there are potential phonetic realizational differences in statement high rises versus polar question rises, strong arguments have been put forward that it should be treated as a "coherent" discourse or sociophonetic phenomenon ([2,3]). It has also been argued that any observed variation in high rising intonation is low-level and not necessarily categorical. This gives rise to a more general question concerning the extent of phonetic and phonological variation amongst rising tunes in Australian English.

Traditional descriptions of Australian English intonation are closely based on models of Standard British English intonation and the most widely-used model of analysis before the 'nineties was the nuclear tone model developed by [4]. In this model, two nuclear tones Tone 2 and Tone 3 are posited: a high rise and low rise, respectively. Since the beginning of the 'nineties, most of the recent research on Australian intonation has been couched within an autosegmental-metrical framework (e.g. [2], [5,8]). In particular, the ToBI (Tones and Break Indices) intonational framework (e.g. [7]) has been used extensively. The annotation framework for Australian English as currently conceived, is more or less identical to MAE_ToBI for General American English. Like other autosegmental-metrical models, it is a highly componential model of intonational analysis where contours are decomposed into strings of low and high tonal events or tone targets aligned with different elements in the prosodic structure of an utterance. One implication of a componential model is that tone targets can be combined in a number of ways to account for a variety of potential tunes. Table 1 gives a summary of some the rising tunes that are theoretically permissible within the ToBI system as applied to Australian English. It is loosely based on a summary provided in [2]. This list can be further expanded to take into account downstepped !H* pitch accents or! H- phrase accents. Also included are the equivalent English Tone categories adapted from Halliday's [3] model of English intonation.

Table 1 **Summary of rising contours, Halliday's English Tones, and ToBI annotation categories**

	High Rise, Tone 2	Low Rise, Tone 3	High/Mid-level	Fall-Rise, Rise-Fall-Rise Tone <u>2</u> Tone 4
Rises with low or "scooped" rising pitch accents (pitch level of rise elbow)		L*L-H% L* H-H% L*+H H-H%		L*+H L-H%
Rises with high, or rising pitch accents (pitch level of rise elbow)	H*H-H% L+H* H-H%		H* H-L%	H* L-H%

Whereas most contour-based analyses of intonation (i.e. [3]) have two simple rises that vary in pitch span and pitch level, at *least* three kinds of simple rise can be transcribed within ToBI. These rises combine an L* or L*+H pitch accent, or an H* or L+H* pitch accent, with two different phrase-accent and boundary tone combinations: L-H% and H-H%. This means that tunes transcribed as L* L-H% (low rise with low phrase accent) are more or less equivalent to Halliday's Tone 3, and H* H-H% tunes (high rise) are equivalent to Halliday's Tone 2. A third potential rise is L* H-H% (low rise, expanded range with high phrase accent). Schematic illustrations of these three rises are shown in Figure 1. Rises a) and c) rise relatively high into a speaker's range, whereas rise b) terminates at roughly mid-pitch level.

It has been suggested (e.g. [2]), that the differences between these ToBI-annotated rises may not be categorical even though the componential nature of the model (potentially) allows for this to be the case. In particular, the difference between nuclear rises that are regularly transcribed in current intonational analyses of Australian English as L*H-H% and H*H-H% may merely reflect "paralinguistic" variation of pitch level at the start of the rise. A similar argument could be made for L* H-H% and L* L-H% tunes. In other words the variation between rise a) and rise b) shown in Figure 1 (which terminates relatively high in the speakers range) may also reflect "paralinguistic gradience". This raises a valid

point about whether there are categorical differences between the three simple rises (illustrated in Figure 1).
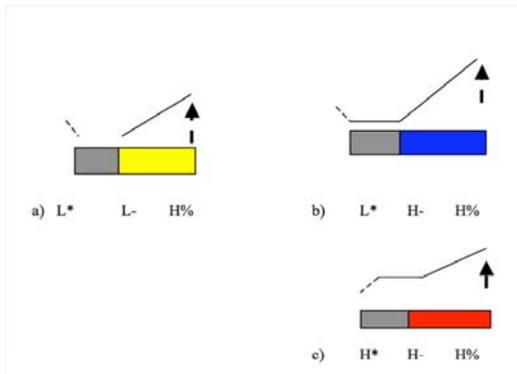

Figure 1: *Schematic illustration of three simple rises for Australian English for two syllable words, transcribed using ToBI annotation conventions: a) L* L-H% (low rise, narrow range); b) L* H-H% (low rise, expanded range); c) H* H-H% (high rise). The grey "block" corresponds to the nuclear accented syllable and the association point of the pitch accent. The coloured "block" corresponds to the post-nuclear "tail" of the tune.*

In previous corpus work on Australian English [1,5,6,8], simple rises were transcribed using the ToBI annotations listed in Table 1, and illustrated in Figure 1. In [8], we also examined the correspondence between these three broad rise-types and dialog acts in ToBI-labelled map task interactions and found that 97% of rises that were labelled H* H-H% were associated with information requests (i.e. polar questions), whereas rises that were labelled L*L-H% corresponded to acknowledgment/answer and acceptance dialog acts, or back channels. Rises labelled L* H-H% were rarely correlated with information requests, and were more likely to conclude statement directives or instruction dialog acts. Conversely, few (3%) H* H-H% tunes were associated with non-question acts. The componential nature of the model therefore allowed us to capture important differences in the way these different rises were exploited in these interactive dialogues.

While these results suggest that both pitch level and pitch span of rises may be exploited in interactive discourse in Australian English, it is still remains to be shown whether listeners (beyond the trained transcriber) are sensitive to the kinds of phonetic and potential phonological variation that we have observed (and assumed) so far in our corpus work on Australian English. In the experiment reported in this paper, it was hypothesized that listeners will interpret rises that would normally be transcribed as H-H% tunes as polar questions. We further hypothesized that rises with a narrow pitch span which rise from a low elbow or pitch level (e.g. L* L-H%) are less likely to be interpreted as question-like in view of earlier corpus-based results. It remains to be seen how listeners interpret rises with a wide pitch span that start from a relatively low elbow (i.e. L* H-H%). Earlier studies (e.g. [1]) suggest there may be some effect of pitch level of the nuclear pitch accent (i.e. L* versus H*), but it has been suggested in [2] that this might be merely phonetic variation. Our goal, therefore was to see whether listeners are sensitive to manipulations of pitch level and pitch span that would normally correspond to the three different ToBI-annotated tunes shown in Figure 1.

2. Method and Materials

Listeners were required to listen to a set of isolated utterances and identify them as either a question or a statement. They were also required to indicate the level of confidence of their answer on a scale of 1 to 5. Some of the sentence materials are shown in (1). The word in bold font represents the nuclear accented word in each phrase. A short and longer utterance were used to see whether there was any influence of utterance length on rise interpretation.

(1)
It was **Miranda**.
It was definitely Anne **Manning**.
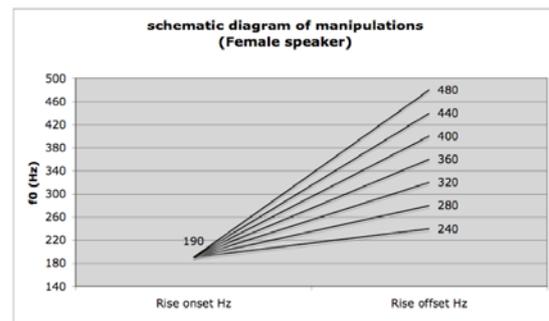It was definitely Melanie **Maloney.**

Figure 2. *A schematic illustration of rising stimuli with manipulated pitch span (female speaker). Rise onset value corresponds to one of the three 'elbow' values of the rise.*

Two thirty year old speakers of Australian English (male and female) recorded the original sentence materials and a range of manipulations were then performed using Praat. The original f_0 information for each sentence was replaced by a set of pre-determined rising contours that were manipulated to change pitch span, and the pitch level of the elbow or starting point and end point of the nuclear rise. This involved manipulating the f_0 value of the nuclear accented syllable in each utterance from relatively low pitch to high pitch in three pitch steps (70 Hz to 112 Hz for the male speaker; 140 Hz to 240 Hz for the female speaker), and the f_0 value of the terminal point of the rise in seven pitch steps (110 Hz to 190 Hz for the male speaker; 240 Hz to 480 Hz for the female speaker).

Figure 2 shows some of the rise stimuli for the female speaker. The rise turning point or elbow of the rise (the rise "onset" shown in Figure 2) was temporally fixed at 200 ms from the onset of the main stressed syllable based on the results of an earlier corpus analysis of nuclear accent alignment in Australian English ([9]). The pitch level of the rise end point was temporally aligned with the last voiced element of the utterance final word. The pitch level of the nuclear accented word was held constant at the same pitch level of the elbow point. In longer utterances the pre-nuclear portion of the contour was held constant with relatively high pre-nuclear pitch on the first content word ("definitely") of the utterance.

The full stimulus set consisted of 160 stimuli and a small number of foils. The stimuli were presented in random order to 40 listeners who were undergraduates at the University of Melbourne. There was a 4s gap between each successive utterance. Subjects were required to indicate on an answer-

sheet whether they heard the utterance as a statement or question and the relative confidence of their response using a scale from 1 to 5. They were paid for their participation. All responses were tabulated and a series of analyses of variance (ANOVA) were performed on the data. In all analyses the dependent variable was proportion of question responses and independent variables were rise-starting point, rise endpoint, utterance length, and speaker.

3. Results

Figure 3 plots the % of question responses for the different rise stimuli for the male speaker. There was a significant overall effect of pitch level of the rise onset (Male: F=29.91; p<0.0001; Female: F=32.17; p<0.0001) and rise endpoint (Male: F=22.22; p<0.0001; Female: F=34.87; p<0.000). As the rise endpoint increased in pitch value (from steps 1 – 7 in Figure 3), there was an increase in the proportion of question responses. The highest proportion of question responses occurred for rises that started relatively high in pitch (step 3) and continued to rise to a higher pitch value (steps 5 to 7). These stimuli were most similar to tunes that would normally be transcribed as H* H-H% in previous corpus studies. In other words, rises that occurred in a high pitch key were most likely to be interpreted as questions. It is also apparent from Figure 3 that as pitch span increased for rising stimuli that had a relatively low pitch elbow (steps 1 and 2), more question responses were also assigned. These were rises that would normally be transcribed as L* H-H%.

Figure 4 shows a factor plot summarizing the number of statement responses for the different rise stimuli. As one would expect, there were fewer statement responses compared to question responses overall. Nevertheless, there was a significant overall effect of pitch level of the rise onset or rise elbow (Male: F=30.11; p<0.0001; Female: F=35.59; p<0.0001) and rise offset level (Male: F=20.72; p<0.0001; Female: F=38.4; p<0.000). The highest number of statement responses occurred for stimuli with rise elbows that started at pitch steps 2 and 1, and ended with pitch steps 1,2 and 3. In other words rises with the narrowest pitch spans that also had relatively low pitch elbows were more likely to be heard as statement-like. These were stimuli that are closest to tunes that would normally be transcribed as L* L-H% according to ToBI conventions.
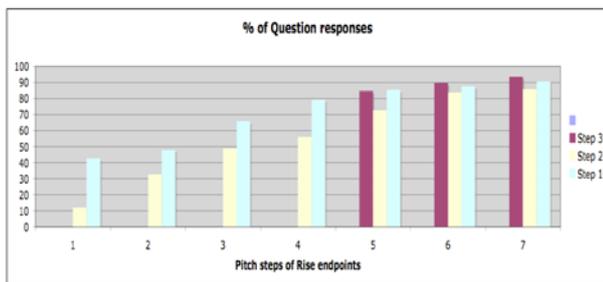

Figure 3. *% Question responses plotted according to two factors: pitch level of elbow (step 1 lowest pitch level, step 2 next lowest, step 3 highest pitch level of rise endpoints with 1 having the lowest Hz value and 7 the highest.*

A further exploration of the data also revealed that the female speaker elicited more question responses than the male speaker, although the overall response trends were similar. Moreover, utterance length also proved to be a significant

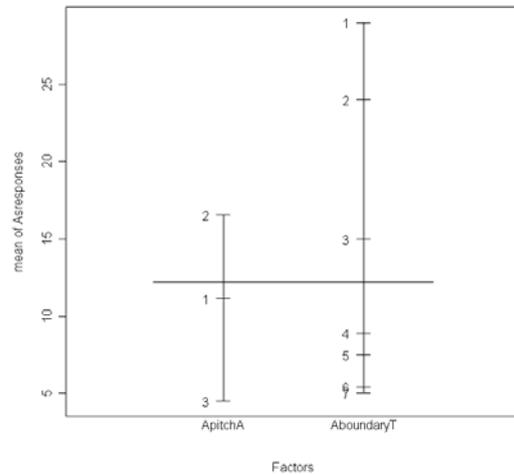factor for both speakers (Male: F=4.5, p<0.05; Female:F=19.2, p<0.0001).

Figure 4. *A factor plot showing the pattern of statement responses plotted according to two factors: pitch level of elbow (three values lowest (1), next lowest (2), highest (3)), and boundary tone (values 1 to 7) with 1 being the lowest value in Hz and 7 the highest value.*

In longer utterances, rises that had a high pitch elbow and continued to rise were more likely to be interpreted as questions and rarely as statements, particularly at step 5 of the rise endpoint continuum. At the same pitch step in shorter utterances, fewer question responses were elicited. Once again, those rises that started from a low to mid pitch level and rose to a high level were also perceived as questions. However, as the rise offset got higher in pitch level (steps 6, 7 & 8), more question responses were elicited regardless of the pitch height of the rise elbow in longer utterances. In shorter utterances, the results followed a similar trend. Very few question responses (10 across the corpus) were elicited for rises with the lower rise elbows. However, the differences between the two low pitch elbows (steps 1 and 2) were less evident in shorter utterances compared to longer utterances.

In terms of certainty of response, an interesting pattern was evident in the data. Speakers were most certain of their question responses for rises that consisted of step 3 rise elbows (i.e. that would be analysed as H* pitch accents) and higher pitch level endpoints. Conversely, they were most certain of their statement responses with step 1 and 2 rise elbows and the lowest rise endpoint (F=5.7, p<0.05). This suggests that tunes that could be transcribed as L* H-H% fell somewhere in the middle. The largest number of "less certain" responses were for tunes that had lower pitch elbows (steps 1 and 2) and mid pitch range endpoints (steps 2,3,4). The response patterns were the same for each speaker (Male: F=9.0, p<0.05; Female: F=12.8, p<0.03). The stimuli at the extreme ends of the rise continuum were the most confidently interpreted as either questions or statements.

4. Discussion and Conclusions

The results of this experiment confirm that pitch level and pitch span were both important factors in determining rise

interpretation in this small experiment. Subjects were more likely to respond that they heard a question if the f_0 values associated with the nuclear accented syllable, i.e. pitch level of the rise elbow, and rise endpoint values were relatively high, with fewer question responses when the rise elbow was relatively low in the speakers' pitch range. Statement responses were most likely when the rise was realised with a narrow pitch span starting from a relatively low pitch level on the nuclear accented syllable. Listeners were most confident of their question responses for rises that would normally be analyzed as H* H-H% rises. The fact that the different combinations of pitch levels that correspond to potential pitch accent and boundary combinations elicited different certainty of response patterns also suggests that tunes that would normally be annotated as the three rises, L* L-H%, L* H-H% and H* H-H% were not always interpreted identically in this experiment. Nevertheless high rises that would normally be transcribed as H* H-H% and L* H-H% were still more likely to be interpreted as questions versus statements in this task, whereas low rises were least likely to elicit question responses. We interpret this as evidence that if there is "gradience" across as well as within these categories, it is more likely that stimuli that are closest to L* L-H% are interpreted differently from those that are closer to L* H-H% or H* H-H% tunes, at least in this experiment. Interestingly, listeners were most certain of their question response if the rise was realised entirely in high pitch key. They were least certain of a question response for a shallow rise in low to mid-pitch key. The highest number of question responses (96%) were for short utterances with a high rise that would be transcribed as H* H-H%. The most confident statement responses were for shallow rises that started from a relatively low pitch elbow and rose to mid-level, i.e. a L* L-H% tune.

These results suggest that listeners (in this experiment) were not overly sensitive to pitch level differences of the two high rise possibilities provided by ToBI: L* H-H% and H* H-H% even though these potentially gradient differences in pitch range have been shown to be meaningful in earlier corpus analysis of Australian English interactive discourse [8]. However, listeners *were* sensitive to other differences in pitch span so that rises that would normally be transcribed as L* H-H% were not interpreted in the same way as L* L-H%. On the other hand, the rise stimuli that were most like L* H-H% tunes were less confidently interpreted as question rises compared to H* H-H% rises, and were marginally more likely to attract statement responses in certain utterance length conditions. This seems to concur with earlier results from corpus studies ([8]). It might also reflect the kind of variation outlined in other corpus studies of Australian English (e.g. [6]), where adolescents produced both L* and H* high rises with questions and statements. It also should be noted that we altered the pitch level of the rise elbow and rise offset, and did not modify the temporal alignment of the rise, unlike earlier studies based on New Zealand English (e.g. [10]). Listeners in the New Zealand English experiments were more likely to hear a high rise as a statement the later and sharper the terminal rise (i.e. rise speed was an important factor). It may be that rise timing could also be a crucial parameter, but as argued in [2], this might still reflect phonetic gradience rather than providing any evidence of a categorical distinction. The task was probably not sensitive enough to fully address this question.

Given that Australian English is one of the so-called "rising varieties" of English, it is perhaps surprising that so few statement versus question interpretations were elicited in this experiment. On the other hand, one could argue it is surprising that there were always some statement responses for all rises in this corpus. This could be just a by-product of the task with some listeners being uncertain whether an utterance was "questioning" in spite of its high pitch key. In any case, it has become widely accepted that rising statements are more likely to be used in interactive discourse or in particular types of complex narrative (e.g. [1,2,5, 6], so it could be that a simple fixed choice task like this was not going to elicit a lot of statement responses due to the restrained discourse context. Another factor to take into account is text frequency of H-H% tunes with syntactic declarative utterances. In corpus studies of Australian English intonation ([5,6]), it has been shown that L* H-H% or H* H-H% tunes with non-questioning utterances only account for between 11-19% of overall tune usage. Australian English speakers typically use a wide variety of tunes depending on the discourse context or utterance genre. Our results may merely reflect this distributional trend.

It is apparent that componential models of intonation like the ToBI model currently implemented for Australian English give the analyst a wide range of tune choices. It is not clear that another model that subsumes these tune choices into fewer categories would be more insightful until further detailed perception experiments are undertaken. This experiment was merely a small step towards refining our understanding of meaningful intonational variation in Australian English.

5. Acknowledgements

6. References

[1] Fletcher, J, Warren, P. & Grabe, E. "Intonational variation in four dialects of English: the high rising tune". In Sun-Ah Jun (ed.). *Prosody and Typology – The Phonology of Intonation and Phrasing* (pp 390-409*)*. Oxford University Press: Oxford2005.

[2] Ladd, D.R. "Intonational Phonology". 2nd. Edition. Cambridge: CUP. 2008

[3] Guy, G. & Vonwiller, J. "The High Rise Tones in Australian English". In P. Collins & D.Blair, Eds. *Australian English*,. St Lucia: University of Queensland Press, 21-34. 1989.

[4] Halliday, M.A.K. *Intonation and Grammar in British English.* The Hague:Mouton. 1967.

[5] Fletcher, J. & Loakes, D. "Patterns of rising and falling in Australian English". In P. Warren and C.I. Watson Eds. *Proceedings of the 11th Australian International Conference on Speech Science and Technology 2006*. 42-72. 2006

[6] McGregor, J. & Palethorpe, S. "High rising tunes in Australian English: Communicative functions of L* and H* pitch accent onsets". *Australian Journal of Linguistics*, 28, 171-93.2008

[7] Beckman, M.E., Hirschberg, J. and Shattuck-Hufnagel, S. "The original ToBI system and the evolution of the ToBI framework". In S-A. Jun (ed.). *Prosodic Typology*. Oxford: OUP. 9-54. 2005

[8] Fletcher, J. Wales, R, Stirling, L. & Mushin, I. "Intonational rises and dialog acts in the Australian English Map Task". *Language and Speech* 45: 229-253, 2002.

[9] Fletcher, J. & Harrington, J. Timing of intonational events in Australian English. In. P. McCormack and A. Russell (eds.) *Proceedings of SST2006*, 611-15.

[10] Warren, P. "Patterns of late rising in New Zealand English: intonational variation or intonational change?" *Language Variation and Change* 17: 209-30. 2005