The effect of listener and speaker gender on the perception of rises in AusE

[1]Elaine Schmidt, [2]Brechtje Post, [1]Carmen Kung, [1]Ivan Yuen & [1]Katherine Demuth

[1]ARC Centre of Excellence in Cognition and its Disorders (CCD), Centre of Language Sciences (CLaS), Department of Linguistics, Macquarie University; [2]University of Cambridge
Elaine.Schmidt@mq.edu.au

ABSTRACT

Australian English (AusE) uses High Rising Tunes at the end of questions and statements. However, it remains unclear whether listeners can distinguish between them perceptually. This study analyses the identification of question- and statement-rises in the absence of contextual information.

Results suggest that identification is strongly influenced by speaker and listener gender. Specifically, it appears that male listeners use pitch differences in pitch accents for perceptual discrimination, just as they do in production, while female listeners rely on the speaker gender: female utterances are perceived as questions, male utterances as statements.

Contrastingly, listener gender did not affect the interpretation of boundary tones: the highest tones are associated with questions, the lowest with statements. However, the middle step shows the bias for questions of female and statements of male speakers again.

These results are important for L2 learners of AusE, and hearing impaired populations where subtle pitch differences are lost.

Keywords: intonation, perception experiment, High Rising Tunes (HRTs), Australian English

1. INTRODUCTION

Australian English intonation is characterised by rises (HRTs) which occur both at the end of statements (S) and at the end of questions (Q). While early sociolinguistic research claimed that this is a phenomenon exclusive to female adolescents, more recent research suggests that HRTs are now observable to the same extent for male and female speakers alike [5, but also 3]. However, the question is whether declarative/interrogative HRTs are phonetically the same and whether they are distinguishable without context (like in noisy environments) where top-down cues cannot be used.

1.1 Production of HRTs

Previous research has suggested that statement- and question-rises are phonetically different [2]. However, it is unclear whether they differ as a result of discourse function (Q vs. S) only, or whether the difference is purely a result of speaker gender. McGregor [5] suggests that questions generally have higher boundary tones than statements, and that questions are often realised with H* pitch accents. However, even though she observed this as a trend, statistical analyses did not reach significance.

We reanalysed McGregor's production data excluding data points that present creaky voice (pitch values lower than 75 Hz for male speakers, 100 Hz for female speakers) or are unnaturally high (more than 300 Hz for male, 500 Hz for female speakers). Additionally, we teased apart contours with L* and H* pitch accents as this might affect the output. Results now showed significant differences in the realisation of S and Q for male and female speakers. Specifically, male speakers showed significant differences in pitch accents depending on discourse type (both for L* and for H* accents) but did not differ in the realisation of the boundary tone. In contrast, female speakers had significantly different (higher) boundary tones for Q than for S, both for tunes with L* and H* pitch accents.

The goal of this paper was to determine if these differences in production of statement and question HRTs might also be reflected in perception.

1.2 Perception of HRTs

Perception experiments testing the ability to discriminate between question and statement HRTs in Australasian varieties of English are scarce. Warren [6] tested the perception of HRTs in New Zealand English by manipulating the location of the rise. Consequently, earlier rises became flatter and later rises were steeper. Listeners consistently associated early rises with Q and late rises with S.

Fletcher et al.'s [4] perception experiment on AusE also included length and steepness of the rise. However, in contrast to Warren [6], the pitch height of the pitch accent and boundary tone were also manipulated simultaneously. Listeners consistently associated higher rises with Q and lower rises with S. Additionally, a gender bias was visible that manifested itself in more Q responses for the female than the male speaker.

However, stimuli were manipulated at pitch accent and boundary simultaneously. Since our production reanalysis suggests male and female speakers adopt different strategies in the production of HRTs, it is conceivable that listeners might also attune to different cues in the perception. The current study addresses these issues by teasing apart *pitch accent* and *boundary* cues and testing them independently in separate blocks. This enables us to establish whether the gender of the listener affects the type of cue (*pitch accent* or *boundary)* that is used to discriminate Q from S in absence of top-down contextual information.

2. HYPOTHESES

Based on the reanalysed production data, we hypothesise that male and female listeners use different cues to discriminate between questions and statements. Specifically, we hypothesise that:

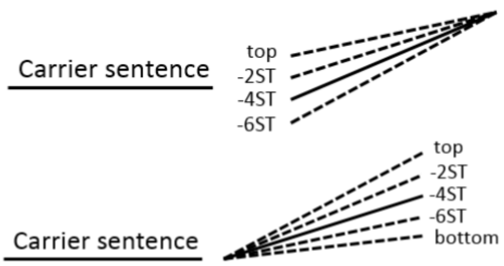
H1. Female listeners use *boundary tone*, male listeners the *pitch accent* differences to discriminate between Q and S.

3. METHODOLOGY

16 male and 16 female monolingual Australian-English speakers (M age: 31 yrs.) participated in the perception experiment. All reported normal hearing.

The carrier sentences and target words were recorded by 1 male and 1 female native speaker of Standard AusE (as judged by a trained phonetician and AusE native speaker). The carrier sentences ("they often played with…") had neutral almost monotone intonation. Only the target words were resynthesised; the carrier sentence remained the same in all conditions.

Figure 1: Schematic representation of *pitch accent* manipulations (top) and *boundary* tone manipulations (bottom), in 2 semitone (ST) steps



The target words were trisyllabic pseudowords with a CVCVCV syllable structure consisting of nasals/ liquids and vowels (e.g. maluna) and initial stress. The pitch was manipulated using PSOLA in Praat [1] either at the *pitch accent*, i.e. the first syllable of the target word, or at the *boundary*, i.e. the last syllable of the target word (see Fig. 1). The pitch was shifted upward/downward in 2 semitone (ST) steps starting from a medial rise (labelled as -4ST) to make a total of 5 steps for *boundary* and 4 steps for *pitch accent* manipulations. All *pitch accent* manipulations ended in the same boundary tone (-4ST), and the *boundary* manipulations all used the -4ST step of the pitch accent continuum.

Four trained phoneticians judged independently whether the resynthesised sentences sounded natural. Cases in which there was disagreement were discarded with all associated pitch steps for both conditions. After agreement, 8 target words (for male/female speaker) were resynthesised to make 152 items.

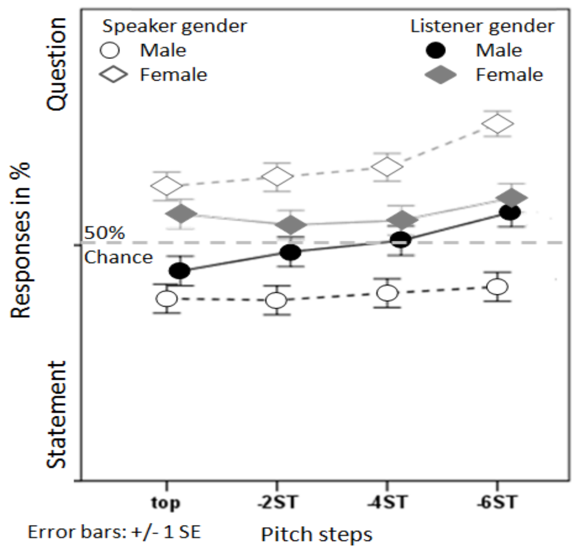
4. JUDGEMENT RESPONSES

Listeners were presented with one target sentence at a time and were instructed to decide as fast as possible whether the sentence was a question or a statement. *Pitch accent* and *boundary* manipulations occurred independently of one another and were presented in separate blocks, as were stimuli produced by the male and female speaker.

4.1 Results Judgement Responses

A univariate ANOVA with factors listener gender (LiG), speaker gender (SpG), position (accent vs. boundary), pitch steps, and dependent variable Q vs. S revealed a significant effect of LiG ($F(1, 4811)= 14.77$, $p<.001$), SpG ($F(1, 4811)= 212.29$, $p<.001$) and pitch step ($F(3, 4811)= 140.60$, $p<.001$).

Figure 2: Percentage of Q/S responses for *pitch accent* manipulations according to pitch step and listener-/speaker gender



Additionally, there are significant interactions between SpG and pitch steps ($F(3, 4811)= 3.78$,

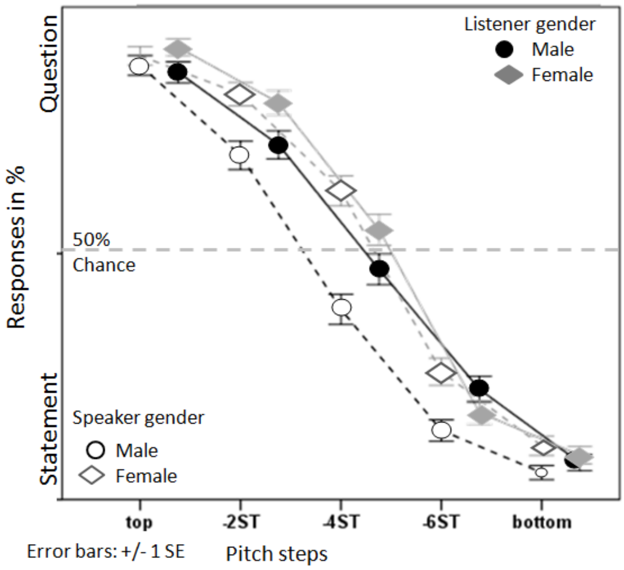
p<.01) and position (F(1, 4811)= 31.54, p<.001), as well as between pitch steps and position (F(3, 4811)= 154.99, p<.001), and a three-way interaction between SpG, step and position (F(3, 4811)= 2.96, p<.05).

Since the pitch steps differed in the *pitch accent* and *boundary* conditions, we ran further ANOVAs separately in each condition.

Results in the *accent* condition showed significant effects for LiG (F(1, 2026)= 8.85, p<.01), SpG (F(1, 2026) = 172.89, p<.001) and pitch steps (F(3, 2026)= 2.83, p<.05) (see Fig. 2). However, Scheffe Posthoc comparisons do not show significant differences between pitch steps and only a trend between the top and -6ST steps (p= .071).

Results in the *boundary* condition also showed significant differences for the factors LiG (F(1, 2531)= 5.91 p<.05), SpG (F(1, 2531)= 53.89 p< .001) and pitch steps (F(4, 2531)= 419.41 p<.001), as well as significant two-way interactions between LiG (F(4, 2531)= 2.58 p< .05) and SpG (F(4, 2531)= 6.16 p<.001) with pitch steps respectively. Scheffe Posthoc comparisons showed significant differences between each pitch step (all at p<.001) (Fig. 3).

Figure 3: Percentage of Q/S responses for boundary manipulations according to pitch step and listener-/speaker gender



4.2 Discussion Judgement Responses

The significant differences found for speaker and listener gender in both conditions suggest that the discrimination of declarative-/interrogative-rises in AusE is subject to gender biases. Specifically, female utterances were more likely to be interpreted as Q by male and female listeners alike. Contrastingly, sentences by the male speaker were more likely to be identified as S.

Interestingly though, prosodic cues are weighted differently in the two conditions. In *pitch accent* manipulations, pitch height between steps does not seem to affect the interpretation of an utterance, instead it is only speaker gender that determines the identification: female utterances are perceived as Q, male as S (Fig. 2). However, listener gender shows an interesting trend: Female listeners provide Q/S responses in roughly equal distribution while male listeners seem to use pitch height in the accent to discriminate between Q and S with the top step eliciting more clear S responses but the -6ST step more clear Q responses. The steps in between appear to show a continuum. This would suggest that H1 can be partly confirmed: male listeners use pitch accent differences to discriminate between different discourse functions, female listeners do not. The overall interaction between listener gender and pitch steps fails to reach significance indicating that this is only a trend. However, it is likely that this pattern will reach significance with more data points.

In contrast, results of *boundary* manipulations demonstrate that pitch height plays a crucial role in the discrimination of Q and S: the two highest boundary tone steps (top, -2ST) are unequivocally identified as Q whereas the bottom two steps (-6ST, bottom) are clearly identified as S by all listeners (see Fig. 3), and the S-shape of the identification function suggests that the distinction is perceived categorically, with wavering judgements limited to the middle step of the continuum. However, gender biases also come into play. While female utterances are more likely to be identified as questions at all pitch steps, the gender bias properly manifests itself in the interpretation of the middle (-4ST) step. Here the identification curve is skewed since female utterances are identified as Q, while the same step in the male speaker is interpreted as S. The same pattern is also apparent in listener responses. Thus, H1 is only partly confirmed since *both* male and female listeners use boundaries to discriminate between Q and S, but only male listeners use pitch accents to distinguish the two.

In summary, the discrimination of Q and S in boundary manipulations appears to rely more heavily on pitch height than it does in pitch accents. However, if the pitch height appears to be between categories, gender biases of the listener and the speaker determine listeners' responses.

5. REACTION TIMES

Reaction times serve as confirmation and validation of our data since they reflect speed of processing, and hence difficulty, when listeners placed stimuli with different pitch steps into the binary categories Q and S. Here, we hypothesise that

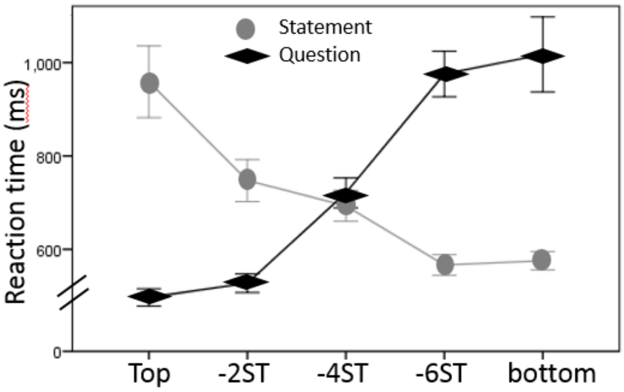
> H2. Reaction times should be faster for items that were clearly identified as belonging to a category (i.e. top two steps for Q and bottom two steps for S in boundary condition) but significantly slower for items that are between categories (-4ST step).

For the analysis of reaction times we have removed outliers that were more than 2 standard deviations from the mean (about 5% of the data).

5.1 Results Reaction Time

In order to test whether the reaction times for *pitch accent* manipulations differed as a result of gender biases, pitch steps or Q vs. S responses we carried out a univariate ANOVA with the same factors as before. Results revealed a significant effect of LiG ($F(1, 2095)= 24.87$, $p<.001$) as well as a significant interaction between Listener Gender (LiG) and Q vs. S ($F(1, 2095)= 4.59$, $p<.05$).

Figure 4: Mean reaction time for Q/S responses for boundary manipulations



We carried out a further ANOVA with the same factors for *boundary* manipulations to establish whether the gender biases that were apparent in the judgement responses and the different pitch steps are reflected in reaction times as well. Results show a significant effect for pitch steps ($F(4, 2380)= 5.43$, $p<.001$) as well as significant interactions between LiG and SpG ($F(1, 2380)= 5.29$, $p<.05$), pitch step and Q vs. S ($F(4, 2380)= 36.03$, $p<.001$), three-way interactions between LiG, SpG and pitch steps ($F(4, 2380)= 3.97$, $p<.01$) as well as between LiG, SpG and Q vs. S ($F(1, 2380)= 10.22$, $p<.001$) and between SpG, pitch steps and Q vs S ($F(4, 2380)= 2.87$, $p<.05$). Planned Scheffe Posthoc comparisons show significant differences between top and -4ST / -6ST (at $p<.001$ and $p<.05$ respectively), -2ST and -4ST (at $p<.001$) and bottom and -4ST (at $p<.05$).

5.2 Discussion Reaction Time

Our reaction-time results *validate* our stimuli and *confirm* our judgement-response data. Reaction times for pitch accents show that only listener gender is significant, with male listeners responding slower than females. The fact that speaker gender and pitch are not significant overall is an important *validation* for the naturalness of our stimuli. Since accent manipulations were consistently associated with Q for the female speaker and S for the male speaker regardless of pitch height (Fig. 2), any big differences in reaction time for pitch step would have indicated processing difficulties for those stimuli, which could have indicated that they sounded unnatural.

Reaction times for the *boundary* condition fully confirm our judgement data since they clearly indicate that some pitch steps can be more easily identified as Q/S while others (most notable -4ST) are straddling category boundaries. As such the data confirm H2 which expected reaction times for Q responses for the top two steps to be faster, as are S responses for the bottom two steps (Fig. 4). In contrast, the -4ST step was hypothesised to reveal slower reaction times since it is more difficult to assign to either Q or S (Fig. 3). Our reaction times show that this was indeed the case.

6. CONCLUSION

The results of this perception experiment suggest that in the absence of contextual top-down information, the identification of question- and statement-rises in AusE is strongly influenced by gender biases. These gender biases work two-ways: they influence discrimination abilities depending on the gender of the speaker, as well as on the gender of the listener. Specifically, it seems as if male listeners indeed use pitch height cues in pitch accent position for perception, just as they do in production, while female listeners only rely on the gender of the speaker: female utterances are more likely to be interpreted as Q, male utterances as S.

In contrast, pitch height differences in boundary tones are interpreted independent of the gender of the listener: the highest tones are associated with questions, the lowest with statements. It is only for steps in the middle that gender bias plays a role again, with the aforementioned bias for questions by female speakers and statements by male speakers.

These results are important in the context of L2 learners of AusE, especially if the L1 uses falling pitch to mark statements. These findings also raise many questions as to the discrimination abilities of hearing impaired populations since pitch differences are only poorly translated in cochlear implants.

ACKNOWLEDGEMENTS

We are very grateful to Jeannette McGregor for kindly providing us with her data set, and to the Child Language Lab for their helpful comments and suggestions. This project is funded by the ARC Centre of Excellence in Cognition and its Disorders (ARC CE110001021) and the ARC Laureate Fellowship (ARC FL130100014) to Katherine Demuth.

REFERENCES

[1] Boersma, P., & Weenink, D. 2007. *Praat: Doing phonetics by computer (Version 5.1.12) [Computer program]*. Retrieved November 18, 2009, from http://www.praat.org/

[2] Fletcher, J., & Harrington, J. 2001. High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, *58*(4), 215-229.

[3] Fletcher, J., Grabe, E., & Warren, P. 2004. Intonational variation in four dialects of English: the high rising tune. *Prosodic Typology: an Approach through Tone and Break Indices*.

[4] Fletcher, J., & Loakes, D. 2010. Interpreting rising intonation in Australian English. In M. e. Hasegawa-Johnson (Ed.), *Speech Prosody 2010* (pp. 100124:1-4). United States: University of Illinois Press.

[5] McGregor, Jeannette. *High rising tunes in Australian English*. Unpublished Doctoral dissertation. Macquarie University, Sydney

[6] Warren, P. 2005. Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change* , *17*, 209-230.