

JOURNAL ARTICLE

The Interpretation of Prosodic Variability in the Context of Accompanying Sociophonetic Cues

Paul Warren

School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, NZ
paul.warren@vuw.ac.nz

Production data have shown that one of the features distinguishing uptalk rises from question rises in New Zealand English (NZE) is the alignment point of the rise start, which is earlier in question utterances realized by younger speakers. Previous research has indicated that listeners are sensitive to this distinction in making a forced-choice decision as to whether an utterance is a statement or a question. NZE is also characterized by an ongoing merger of the NEAR and SQUARE diphthongs, with younger speakers more likely to realize the vowel in a word such as *care* with a closer starting point (as in [iə]), overlapping with their realization of the NEAR vowel), whereas older speakers would have more open starting point (as in [eə]). The current study uses the mouse-tracking paradigm to provide evidence that the realization of SQUARE with an innovative vs. a conservative variant in a word early in an utterance affects NZE listeners' sensitivity firstly to a rise as a potential signal of an uptalked statement and secondly to the early alignment of the rise as a signal of a question. This finding indicates that the interpretation of prosodic variability can depend on speaker characteristics imputed from other sociophonetic cues.

Keywords: uptalk; rising intonation; New Zealand English; sociophonetics; speech perception; mousetracking

1 Introduction

This study considers whether listeners' interpretation of prosodic variables, namely the use of an utterance-final rise to indicate an uptalked statement and the alignment of the starting point of that rise to indicate whether the utterance is indeed uptalk or a question, depends on socially-conditioned expectations linked to a segmental phonetic cue earlier in the utterance. The response required of participants in the experiment reported below is whether an utterance is intended as a question or as a statement, a distinction that has been linked in New Zealand English (NZE) to the difference between early and late starting points respectively for the final rise. The segmental sociophonetic cue used in the experiment as a potential indicator of social grouping is the realization of a single vowel in a word that occurs earlier in each test utterance, prior to the intonational cue. This is the realization of the SQUARE diphthong,¹ which for many speakers of the same dialect is merged with the NEAR diphthong. For non-merging speakers, the SQUARE and NEAR vowels might be transcribed phonemically as /eə/ and /iə/ respectively. For merging speakers, these two centering diphthongs both have a closer starting point that is more typical of the realization of NEAR for non-merging speakers, i.e., [iə]. Uptalk use, the distinction between early and late rises for questions and statements and the NEAR-SQUARE merger are all linked in this language variety with younger speakers (predominantly

¹ The lexical-set labels SQUARE and NEAR are used in this paper, following Wells (1982), to refer to the relevant centering diphthongs.

teenagers and those in their twenties), especially females, although not all young New Zealand women regularly use uptalk or merge their NEAR and SQUARE diphthongs. The research question is whether manipulation of the vowel causes a change in listeners' expectations about the speaker, resulting in a shift in the interpretation of the intonation. Background information on both the intonational pattern and the vowel cue will now be presented, in order to contextualize the experimental design.

Uptalk has been documented for New Zealand English (largely under the label HRT, for high-rising terminal) since the mid-1960s (Benton, 1965). Early systematic studies indicated that it was most prevalent amongst the young, and particularly young women (Bell & Johnson, 1997; Britain, 1992), and there remains a strong association of uptalk with younger speakers. An important factor for the experiment reported below is the extent of the overlap of uptalk with question intonation. Since intonational rises can also mark questions, there is a frequent lay perception that uptalk indicates a questioning nature, and that uptalkers are therefore insecure or lacking in confidence (see discussion in Warren, 2016). This association is not entirely surprising, given that it has been claimed that uptalk rises and question rises may be phonetically indistinguishable (Guy et al., 1986; Ladd, 1996; Lakoff, 1973). However, more recent research indicates emerging differences between these rise types in a number of varieties of English, with the nature of the differences dependent on variety (Warren & Fletcher, 2016). In Australian English, for instance, uptalk rises tend to start from a lower pitch, possibly as part of a fall-rise contour (Fletcher & Harrington, 2001; McGregor, 2005). Other research has suggested that question rises have earlier onsets than statement rises, in particular in NZE (Fletcher et al., 2005; Warren, 2005; Warren & Daly, 2005), South African English (Dorrington, 2010), and Southern Californian (Ritchart & Arvaniti, 2014). It should be stressed that early and late are relative concepts, and refer to the alignment of the rise onset with regard to the nuclear accent and following material, rather than in the utterance as a whole. In NZE, this alignment difference has been shown to be related to speaker age, with younger speakers more likely to distinguish between an early rise start for questions and a later start for uptalk utterances, while older speakers use later onsets for both sentence types (Fletcher et al., 2005; Warren, 2005).

A number of perceptual studies have investigated the cueing value of such differences. In one study, Fletcher and Loakes (2010) found that the most statement-like utterances in a set of rising Australian English utterances were those with a low onset, while questions were signalled more reliably by a high onset, reflecting the production data for that variety. In a small-scale study in NZE (Warren, 2005; Zwartz & Warren, 2003), participants were asked to classify variants of a single utterance. The utterance in question had a six-syllable nucleus + tail sequence (*basketball stadium*), and the final rise was resynthesized firstly so that its onset was temporally aligned at a range of five equally distant points from the initial accented syllable (*ba-*) through to the utterance-final syllable (*-um*), and secondly so that rises either progressed linearly towards a final high point at the end of the utterance, or involved a sharp rise followed by a high plateau. The results indicated that questions were more clearly signalled by the sharp early rise, and statements by the sharp late rise, reflecting respectively the convex and concave contours found in production data from NZE speakers (Warren, 2005).

The current study aims to build on these findings by additionally investigating whether the listeners' perceptions of uptalk, and of early and late rises in NZE as signals of questions and statements respectively, are also dependent on whether the speaker is a likely 'uptalker.' As indicated above, it does so indirectly by manipulating a segmental variable that has a social distribution similar to that of uptalk, i.e., the realization of the SQUARE diphthong. In a longitudinal study of the NZE NEAR-SQUARE merger, Gordon and

Maclagan (2001) examined production data based on words containing these vowels, read in sentence and word-list contexts by adolescents from the same demographic, sampled every five years. While the diphthongs were still both widely present in their first sample from 1983, they showed significant overlap in adolescent speech by 1998. The merger has been towards a closer starting point for SQUARE, i.e., a merger-by-approximation towards NEAR. In the current paper, this innovative realization of SQUARE will be given the label [iə], which is the transcription recommended for the NEAR diphthong in NZE by Bauer and Warren (2004). The more conservative variant will be labelled [eə], the transcription they give for SQUARE, although it should be noted that single labels cannot do justice to the variation that is found in both the closer and the more open realization of the diphthong. Like the use of uptalk intonation, the use of a variant of SQUARE with a closer starting point continues to be commented on in the media and in letters-to-the-editor, particularly as a possible source of ambiguity (in this case resulting in perceived homophony of words like *beer* and *bare* or *cheer* and *chair*). It also continues to be associated with the speech of younger speakers, with the diphthongs remaining unmerged for many, especially older speakers.

Based on a series of experiments involving the production and perception of NEAR and SQUARE, Warren et al. (2007) concluded that the social indexing of the merger in NZE meant that lexical items containing a diphthong with the closer starting point of [iə] were least likely to be identified as containing the SQUARE vowel when the speaker was perceived to be older and male. This was demonstrated in a study employing four voices (two of each sex) that were independently rated for probable speaker age. In another study Hay et al. (2006b) manipulated age through photographs and showed that perceived age of the speaker influenced identification accuracy of NEAR and SQUARE word pairs. The photographs were presented as though they were pictures of the speakers being listened to, and accuracy was greatest after photographs of older individuals, even though the same male and female speech tokens were presented after each of the male and female photographs respectively. The consequences of the merger for lexical access have also been investigated in semantic priming experiments (Warren et al., 2007; Warren et al., 2003), which showed that there was an asymmetry in the priming exhibited by young NZE-speaking participants, but that listeners were again sensitive to the age of the speaker. When the stimuli were from a younger voice, the form [tʃiə] primed both *shout* (an associate of *cheer*) and *sit* (an associate of *chair*), but the form [tʃeə] only primed *sit* (the associate of *chair*). When an older speaker was used, the priming was not asymmetrical—[tʃiə] primed only *shout* and [tʃeə] primed only *sit*.

The brief summaries above of research on uptalk and on the NEAR-SQUARE merger in NZE indicate some changes-in-progress that have been occurring over a similar time-frame, and which are similarly socially stratified. Younger speakers, particularly but not exclusively women, are more likely to use uptalk, to distinguish uptalk rises and question rises through the alignment of the start of the rise, and to have a closer starting point for the SQUARE diphthong.

The results reported above from the studies of NEAR and SQUARE conducted by Hay et al. (2006b) and by Warren et al. (2007) are part of a growing body of research that shows that social characteristics associated with speakers can affect the interpretation of phonetic information. In some of the early perceptual work in this area, Strand and Johnson (1996) found that participants' categorization of fricatives on a [ʃ]-[s] continuum was influenced by the putative sex of the speaker as indicated by video clips with which the audio signals were aligned. Subsequently, Johnson et al. (1999) showed that listeners' categorizations of vowels on a continuum from [u] to [ʌ] were affected not only by a visually presented face (male or female), but also, in a separate experiment, by the imagined

sex of the speaker (participants in one group were told that the speaker was female and asked to imagine a female speaker while doing the experiment, while the other group were told to do the same for a male speaker).

In other research, the speaker's putative dialect origin has been shown to affect perceptual responses. Niedzielski (1999) asked participants, all residents of Detroit, to indicate which of a set of resynthesized vowels best matched a vowel in a sentence they had heard. The results showed that participants who were led to believe that the speaker was from Canada chose a different resynthesized vowel from those who were led to believe that the speaker was from Detroit, despite the fact that both groups of participants heard precisely the same sentences and resynthesized vowels. Hay et al. (2006a) asked their participants to identify which of a series of resynthesized vowels best matched an /i/ target vowel. Again, the experimental manipulation was the supposed dialect origin of the speaker, but in this case this was signalled not explicitly through instructions to the participants, but by means of the appearance of the words 'Australian' or 'New Zealander' at the top of the response sheets. While the results for male participants were inconclusive, female participants were more likely to select a raised, more Australian token from the continuum when they were in the 'Australian' condition. In a subsequent study, Hay and Drager (2010) found that the mere presence of a stuffed toy kangaroo (indicating Australia) or kiwi (indicating New Zealand) could influence listeners' responses.

As well as the age effects shown in the studies of NZE NEAR and SQUARE by Hay et al. (2006b) and Warren et al. (2007), Drager (2011) found that the speaker's perceived age influenced the categorization of vowels on a DRESS-TRAP continuum in NZE, a variety which has a well-established pattern of raising of the short front vowels. However, she found this effect only for older participants, which she conjectures may be linked to their greater experience of a range of speakers from different generations as well as to their greater exposure to the progression of the sound change.

The studies reviewed above have exploited the potential social indexicality of phonetic cues and have shown that segmental phonetic perception can be affected by the perceived characteristics of the speaker (as prompted by photographs, movie clips, dialect region labels, or even stuffed toys). The current study builds on these links between speaker characteristics and phonetic properties at the segmental level, and investigates whether segmental differences can impact on the interpretation of cues at the suprasegmental level. The interaction of segmental and suprasegmental cues in linguistic indexicality has previously been explored by Levon (2007), who manipulated sibilant duration and pitch range in a read passage. Levon found that pitch range only affected judgements of the speaker on an effeminate-masculine scale when the sibilants were short, and that sibilant duration only affected such judgements when pitch range was narrow. The current study takes a different approach to the interaction of segmental and suprasegmental cues. Rather than asking participants to judge characteristics of the speaker on the basis of phonetic cues, it exploits the fact that the realization of the SQUARE vowel, the use of uptalk, and the alignment differences in statement and question pitch rises are typically co-indexical of younger speakers,² and examines whether as a consequence variation in the segmental cue will affect the interpretation of the suprasegmental cues. That is, if a SQUARE diphthong in an utterance has an [iə] realization and signals a speaker who is more likely to produce uptalk, then a subsequent rising intonation is more likely to be interpreted as uptalk. In addition, if the type of speaker signalled by the [iə] realization

² All stimuli were prepared using source utterances from the same female speaker, so the experiment reported here cannot explicitly test the differences between female and male speakers in the realization of either of the phonetic variables.

of SQUARE is also the type of speaker who makes a phonetic distinction between question and statement rises, then a subsequent early rise should be more likely to signal a question and a late rise a statement, relative to a situation where the diphthong has a more conservative [eə] realization.

The two speech features indicated above—the realization of the SQUARE diphthong as [iə] or [eə] and the realization of a rise with an early or late onset—were manipulated on utterances with declarative word order which were then used in a forced-choice task, where participants had to select between two sentence types—question and statement. An example sentence is given in (1).

(1) John's mother cared for stray animals.

The word that has the SQUARE diphthong in this example is *cared*. All relevant words in the test utterances were words which would have an [eə] realization of the SQUARE diphthong in the speech of more conservative speakers, and for which there was no minimal pair word that differed only in having the NEAR vowel. This vowel was manipulated (see below) so that it had either the conservative [eə] realization or the innovative [iə] realization.

The rise was on the final nuclear accented word, *animals*, and was manipulated (see below) so that it started either at the end of the accented syllable (i.e., the first syllable) or at the beginning of the final syllable, which was also the last syllable in the utterance. All final words were three syllables long. These rise alignments are compatible with those found in the production research described above.

A further aspect of the experiment reported below that differs from the previous forced-choice studies of sentence type is that in addition to collecting response choice and reaction time, the experiment tracks mouse movements made by participants as they make their selection. The mouse-tracking technique records the (x, y) pixel coordinates of the trajectory of cursor movements on a computer screen as participants use the computer mouse to move the cursor from its starting point to a decision target. Studies using this technique have shown that details of the mouse trajectory, such as its curvature and complexity, co-vary with a range of cognitive processes involved for instance in auditory lexical decision (Spivey et al., 2005), visual lexical decision (Barca & Pezzulo, 2012, 2015), decisions as to the truth or falsity of negated sentences (Dale & Duran, 2011), memory strength (Papesh & Goldinger, 2012), the automatic activation of phonological information during visual word processing (Barca et al., 2016), and social categorization (Freeman, 2014; Freeman et al., 2008; Freeman et al., 2011). Analyzing mouse trajectories can provide additional insights into decision processes that cannot be measured simply from the outcome decision, such as the attraction strength of competing responses over time. In addition, it has been demonstrated that movement trajectories reflect the confidence of a participant's response, with more direct trajectories to the target correlating with higher confidence scores reported by the participants after their decisions (Papesh & Goldinger, 2012).³

The experiment tests the following hypotheses, which arise from consideration of the literature reviewed above. Firstly, there will be more 'question' responses (which will also be faster and more direct) following early rises than following late rises. Secondly, 'statement' responses will be more likely (and faster and more direct) in the context of

³ Although an eye-tracking paradigm might have been able to reveal similar patterns of behaviour, the mouse-tracking technique has the advantage of being considerably cheaper in terms both of equipment and of analysis costs.

an [iə] realization of the SQUARE diphthong. This would be a new finding, and would indicate that a sociophonetic cue at the segmental level influences the interpretation of a potentially ambiguous cue at the suprasegmental level (i.e., whether a rise indicates a question or uptalk). Thirdly, an [iə] realization of SQUARE, indicating a younger speaker, will show greater compatibility with the use of early and late rises to indicate questions vs. statements respectively. This again would be a new finding, but one which is compatible with a ‘gestalt-like understanding of indexicality’ (Levon, 2007: 546).

2 Experiment

2.1 Method

2.1.1 Materials

The experimental materials consisted of 20 short utterances (see Appendix), each of which contained a word with a SQUARE vowel, followed on average 5.2 syllables later (*SD* 1.77, range 2–8 syllables) by the word bearing the nuclear accent. The nuclear accented word was in all cases three syllables long and had initial stress. In addition, there were 40 filler items and 12 practice items. The fillers consisted of 10 questions involving inversion (e.g., “Are they leaving tomorrow?”), 10 wh-questions (“What is the best way of getting red wine stains out of clothes?”), and 20 statements with final falls (“They were disappointed that the concert ended so early.”). The practice items consisted of 6 statements with final falls, 2 wh-questions, 2 inversion questions, and 2 declarative utterances with final rises. So that participants would not be influenced by the realizations of any SQUARE or NEAR vowels outside of the test utterances, none of the filler or practice items contained either of these diphthongs.

A 25-year old female native speaker of NZE was recorded reading all 20 test items both as questions and as uptalk sentences. She was not asked specifically to produce a particular variant of the SQUARE vowel in the test utterances, since the intention was to resynthesize the vowel using average vowel formants from three further female speakers who consistently distinguished SQUARE and NEAR (see below). In the same recording session she also read out all the fillers and practice items. From the set of recorded test items, 10 question recordings and 10 uptalk recordings were selected as the source utterances for manipulations. Pitch manipulation was carried out using PSOLA implemented in Praat (Boersma & Weenink, 2012). Pitch values for the final rise were based on those of the source utterance. The average starting value for the pitch rise across the 20 test utterances was 184 Hz (*SD* 6.9 Hz). The average pitch at this point was marginally higher for utterances sourced from questions (185 Hz, *SD* 8.7 Hz) than for those sourced from statements (183 Hz, *SD* 4.8 Hz). This difference was not significant by *t*-test ($p = 0.64$). The pitch level at the end of the final rise was similarly based on that of the source utterance and was at an average of 423 Hz (*SD* 24.0 Hz). The final pitch value of utterances sourced from questions (422 Hz, *SD* 19 Hz) was marginally lower than that of utterances sourced from statements (425 Hz, *SD* 29 Hz). Again, this difference was not significant ($p = 0.80$). The fact that these beginning and end values did not differ is further confirmation that in this variety the distinction between question and uptalk rises does not equate to a difference in pitch height (unlike, say, Australian English).

In the early rise condition, the pitch level of the start of the rise was kept constant across the accented syllable, after which it rose linearly to the end of the word, i.e., at the end of the third syllable. In the late rise condition, the pitch level of the start of the rise was maintained across both the accented syllable and the following unaccented syllable, and then rose linearly and sharply across the final syllable to the end of the word. These pitch shapes are based on the previous production studies of NZE intonation noted in the Introduction. The mean duration of the early rise was 349 ms (*SD* 106 ms). It was slightly

longer for utterances taken from a statement source utterance (350 ms, *SD* 116 ms) than for those taken from questions (348 ms, *SD* 101 ms), a difference that was not significant ($p = 0.96$). The mean duration of the late rise was 174 ms (*SD* 84 ms), and was longer for utterances sourced from questions (182 ms, *SD* 84 ms) than for those sourced from statements (165 ms, *SD* 87 ms). This difference was also not significant ($p = 0.66$). That these duration values are so similar within each set is probably a reflection of the fact that all nuclei were realized across three-syllable words with stress on the first syllable. Examples of early and late rise stimuli are shown in **Figure 1** for the utterance “John’s mother cared for stray animals.”

The formants of the beginning portion of the SQUARE diphthong were manipulated and resynthesized using linear predictive coding in Praat, via a purpose-designed script written by the author, in order to produce variants of this diphthong with (relatively) closer and more open starting points. As indicated above, these variants will be referred to as the [iə] and [eə] variants of SQUARE. The [iə] variant is the innovative variant and would be in the NEAR vowel space of more conservative speakers. The script requested the first and second formant values to be used in the resynthesis and then prompted the user to specify, via mouse-clicks on the Praat display of speech wave and spectrogram, the beginnings and ends of the area to be manipulated. The script imposed logistic functions to create smooth transitions for the formants from the source wave before the manipulation area into the manipulation area and back again from the manipulation area to the source wave. The script was used to modify F1 and F2 of the first target of the diphthong, stretching over the first third of the diphthong, with the final target of the diphthong left

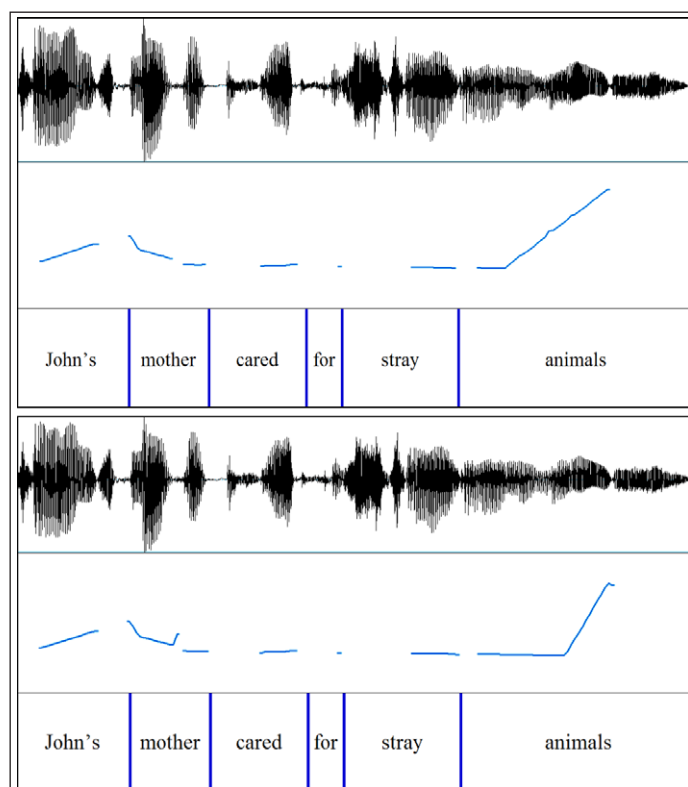


Figure 1: Examples of early rise (top) and late rise (bottom) utterances used in the experiment, showing waveforms, pitch contours, and word-level TextGrids. The pitch range of the display is 75Hz to 500Hz. For each such pair, two examples of each stimulus were used, with [iə] and [eə] variants of the SQUARE vowel, in this case in the word ‘cared.’ Examples of the four stimuli based on the sentence illustrated in this figure can be accessed at DOI: <https://doi.org/10.5334/labphon.92.s1>

as in the original recording and varying naturally depending on the following sound. The F1 and F2 values used for the first target of the resynthesized diphthongs were based on average formant values from minimal pair word-list recordings from a group of 3 female NZE speakers who distinguish NEAR and SQUARE in their speech. For NEAR the average F1 for these speakers was 331 Hz and F2 was 2711 Hz, while for SQUARE their average F1 and F2 were 657 Hz and 2170 Hz respectively. These values formed the basis for the resynthesis of the [iə] and [eə] variants of SQUARE.

2.1.2 Design

Two test lists were produced; in one the test sentences contained the more open conservative [eə] version of the SQUARE diphthong, and in the other they contained the closer innovative [iə] version. Participants were randomly allocated to one of these lists. In each list, the 20 test items occurred twice, once with the early rise, and once with the late rise. The two rise versions were allocated to two separate blocks, and each block had an equal number of early and late rise test items. Thus all participants heard each item with both the early- and late-rise intonation, but they only heard one version of the diphthong. So that not all repeated utterances were test utterances, half of the 40 fillers were repeated in the course of the experiment, giving 60 filler items in total, and a total experimental list (excluding practice items) of 100 utterances. The two instances of repeated fillers were assigned to separate blocks. The 60 filler items were evenly and pseudo-randomly distributed throughout the test list, such that sequences of more than two items of the same type were avoided. The position on the screen of the choice targets ('question' and 'statement,' presented in capitals; see below) remained constant for each participant, but was switched for half the participants in each list.

2.1.3 Participants

Data from 36 native speakers of NZE (27 females) were included in the analysis below. Their age range was 18–32 (mean 22.4, *SD* 3.5). Three further participants were replaced, 2 because of equipment failure and 1 because he failed to select a response before the time-out of 5 seconds for a high proportion of test items (52.5%; no other participant had more than 7.5% of test data missing for that reason).

2.1.4 Procedure

The experiment was run in E-Prime 2.0 (Psychology Software Tools, 2012), using a script developed by the author. The screen resolution was 1920 × 1080. Mouse positions were tracked every 10 milliseconds. The E-Prime script first presented an instruction screen that contained the text below. Note that the text points out that some utterances have declarative word order but rising intonation and thus have the potential of being either questions or statements, but does not draw attention to how these might differ from one another.

In this task, you will hear utterances that could be questions or statements. Your task is to decide whether each one is a question or a statement. Some of them will be easier than others because they start with question words like 'when.' Others may only be marked by intonation. Note, though, that some intonation patterns, such as rising intonation, are often found on statements too, so you will still need to decide whether the utterance was intended as a statement or question.

For each utterance you will first see a START box at the bottom of the page. Click on START. You will then hear the utterance and at the top of the page you will see the words QUESTION and STATEMENT. Click on the word that corresponds to the utterance type for that utterance.

Once participants had indicated that they understood these instructions, the practice items were presented. First, a screen with three clickable areas was presented, as per the instructions above. The ‘start’ point was centered at the bottom, taking up 11% of the screen’s width and 6% of its height. Response targets were top left and top right, each taking 25% of the screen’s width and 10% of its height (see **Figure 2** for a not-to-scale representation of the layout). The presentation of each audio stimulus over headphones commenced as soon as the participant clicked in the region marked by ‘start’ and mouse coordinates were continuously recorded until the participant clicked in one of the two target regions (‘question’ and ‘statement’), or until 5 seconds elapsed, whichever was sooner. Immediately after this, the screen was blanked briefly and then the cycle began for the next stimulus.

At the conclusion of the practice set, participants were encouraged to seek any advice on procedural issues, and then the two main blocks of test and filler items were run without a further break.

2.2 Mouse-tracking measures

The (x, y) coordinates from the mouse-tracking data can be analyzed in a number of ways. The most common analyses are measures of the displacement of the mouse trajectory from a straight-line response from the starting position to the target for the choice being made (see for instance Freeman, 2014; Freeman & Ambady, 2010; Hehman et al., 2014; Papesh & Goldinger, 2012). Two such measures are shown in **Figure 2**. One of these is Maximum Deviation (MD), which is the largest perpendicular distance from the straight-line trajectory to the actual trajectory. The other is the Area Under the Curve (AUC), which is the area defined by the actual trajectory and the straight-line trajectory. The greater the value of MD or AUC, the more the trajectory has deviated from the straight line and moved towards the alternative response. Negative values of AUC and MD can also exist, reflecting a path that goes below the idealized straight-line trajectory. In addition, (x, y) coordinate data can be analyzed for changes in direction, speed, or acceleration, in either the horizontal or vertical dimension, or both. Changes in direction can either be simple reversals (x-flips, y-flips, or both) or more complex measures such as sample entropy (for details of these and other measures, see Hehman et al., 2014).

MD is the measure selected for analysis in the current paper. It has been claimed to index “the partial, simultaneous activation of a competing representation of the opposite category” (Freeman, 2014: 87). In a test of recognition of visually-presented words as old (they had previously been seen in a training set) or new (they had not), stronger memories were associated with fast and linear mouse-tracking responses, while weaker memories had tracks that were slower and curvilinear (Papesh & Goldinger, 2012). The same researchers

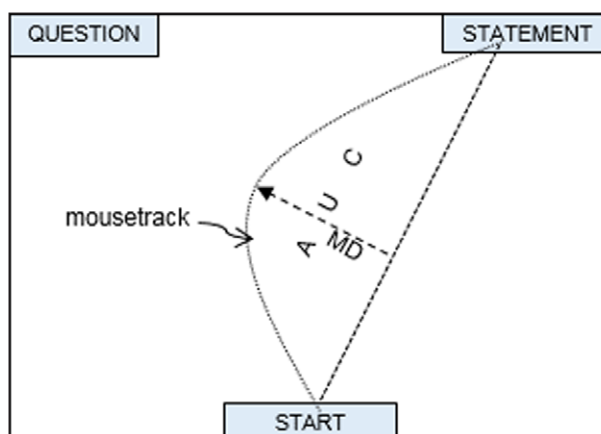


Figure 2: Screen layout (not to scale) with example mouse trajectory, showing start point, choice target points, and AUC and MD measures (see text).

argued that “movement trajectories revealed underlying response confidence” (p. 906). Competition effects were found by Spivey et al. (2005) in a task where participants listened to a word and had to match this to one of two pictures, one of the object corresponding to the word and one of a different object. They found less curvature of the mouse-track towards the alternative response when the word corresponding to the competing picture was a phonologically unrelated word (e.g., *jacket*) than when it was a member of the same phonological cohort (e.g., *candle*, for the auditory word *candy*). The MD measure, then, should provide additional information about the competition between question and statement responses in the context of the experimental manipulations. It should add an index of confidence to data involving response choice and response times, and it might reveal subtleties in participants’ response behaviours that do not show up in the other measures.

2.3 Statistical analysis

In addition to the participant exclusions described above, individual data that exceeded the 5 second time out were excluded from analysis. This amounted to 2.0% of the test data (29 responses). In addition, because our interest is in the impact of the rise alignment on decisions, responses that were made earlier than the point at which the rise started were also excluded. This was a further 2.6% (38 responses). A total of 1373 mouse tracks remained for the test items.

Statistical analysis of the response choices, response times, and mouse-tracking data for test items was by means of mixed effects models in R, using the *lme4* package (Bates et al., 2015). Logistic models (using *glmer*) were applied to binary data such as response choices, and linear models (using *lmer*) to continuous data such as response times and MD. The statistical significance of including a factor or interaction in a model was assessed using the *mixed* command from the *afex* package (Singmann et al., 2015). Fixed effects included Vowel ([iə] or [eə]), Rise (early or late), Source (whether the stimulus was derived from an original question recording or uptalk statement recording—see ‘Preparation of test stimuli’ above), the Serial Position of the stimulus in the experiment, and, where appropriate, the Choice (question or statement) made by the participant. Source was included as a test of whether there are other aspects of the utterances beyond Vowel and Rise that might affect responses. This seemed a sensible addition given that half of the source recordings were questions and half were uptalk statements. Items and participants were included as random effects, together with random slopes by participant for the Serial Position of a stimulus in the experiment. These random slopes were included to account for between-participant variation in how response behaviour (e.g., speeding up in making responses) changes over the course of the experiment.

2.4 Predictions

The hypotheses set out in the Introduction lead to the following predictions for the forced choice binary selection between question and statement responses:

1. There will be a significant effect of rise alignment on response selection, with more question responses for early rises than for late rises. Note that this augments the previous perceptual result reported by Zwartz and Warren (2003), which was based on a longer terminal sequence (a two-word sequence of 6 syllables, compared with a single 3-syllable word in the current study), which might be expected to result in a clearer contrast between early and late rises. In addition, the selection of the question response will be more rapid and involve a more direct mouse movement to the target in the early rise condition, while statement responses will be made more rapidly and with more direct mouse trajectories in the late rise condition.

2. There will be a significant effect of the realization of SQUARE on response selection. Since the realization of SQUARE as [iə] is more likely in the speech of younger speakers, who are also more likely to produce uptalk, statement responses are predicted to be more likely (and faster and more direct) following an [iə] realization of SQUARE than after an [eə] realization.
3. There will be a significant interaction of rise alignment and SQUARE realization on response selection. Since speakers who merge NEAR and SQUARE onto an [iə] pronunciation are members of the same social group (younger speakers) for whom there is evidence that question and statement rises are becoming distinguished through an earlier rise on questions than on statements, it should follow that an [iə] pronunciation of SQUARE will signal a speaker who is likely to have early question rises. Therefore when an early rise is heard in combination with an [iə] pronunciation of SQUARE, question responses will be more likely and will be made both quickly and with a direct mouse trajectory. Conversely, if an [iə] pronunciation of SQUARE is followed by a late rise, then this will be more likely to result in a statement response. However, since the [eə] realization of SQUARE is the more conservative variant, not only (as per prediction 2) will uptalk be unexpected after [eə], but also the difference between early and late rises will be less reliable as a cue to question and statement respectively.

3 Results

3.1 Response choice

The logistic regression model for response choice (question or statement) as the dependent variable included the random effect structure outlined above, together with Vowel, Rise, Source, and Serial Position as fixed effects, as well as the interactions of Rise with Vowel (to test whether an [iə] realization of the SQUARE diphthong would make it more likely that an early rise would be interpreted as marking a question) and of Rise with Source (to test whether there are other properties of the original question and uptalk utterances that operated in conjunction with rise alignment to signal the intended utterance). The model produced a significant effect of Rise ($\chi^2(1) = 43.47, p < 0.0001$). The factors Vowel, Source, and Serial Position all failed to produce significant simple effects, and there were no significant interactions. The effect of Rise is shown in **Figure 3**. Prediction 1 above is therefore supported by the finding of significantly more question responses after early rises than after late rises, but the lack of an effect of Vowel means that the response choice data do not support prediction 2. In addition, prediction 3 is not supported, since there are no differences in the proportion of question responses that depend on the interaction of the vowel with the alignment of the rise.

A further observation from this initial analysis is that overall there were more question responses than statement responses (73% across all test items and conditions). One likely reason for the low number of statement (uptalk) responses is that uptalk occurs more typically in narrative structures with connected utterances than in isolated sentences such as those presented in the experiment. This low overall count of statement responses will prove to be important in later analyses of subgroups of data.

3.2 Response times

An initial inspection of response times (RTs) indicated that, as is commonly found, they were not normally distributed. A comparison of a selection of typical transformations indicated that the logarithm of RTs produced the best fit to a normal distribution ($r = 0.999$, compared with $r = 0.983$ for untransformed RTs). The statistical tests for response times were therefore conducted on log-transformed RTs. For clarity of presentation, however, raw RTs will be graphed.

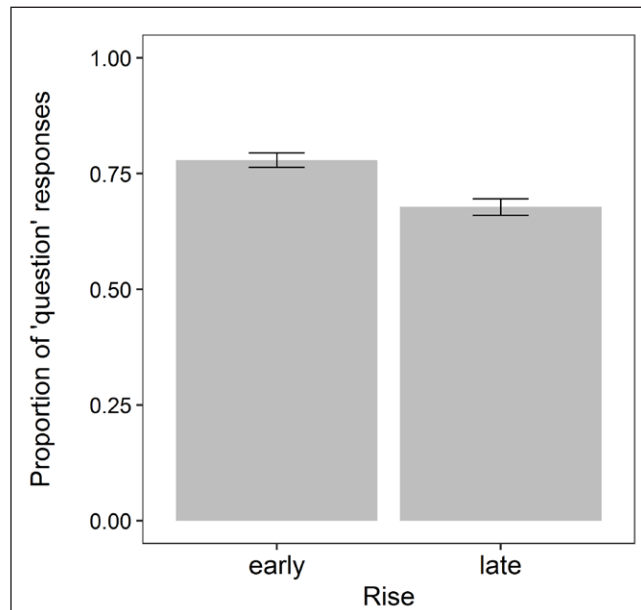


Figure 3: Effect of Rise alignment on the selection of question (vs. statement) responses.

In addition to the factors examined in the analysis of response choices, the RT analysis included Choice (i.e., whether the participant clicked on the question or statement response box). The analysis therefore considered the simple effects of Rise, Vowel, Source and Choice, as well as their interactions, and Serial Position. There was no simple effect (or interaction) related to whether the source of the manipulated stimuli was a question or a statement. Significant interactions were found between Choice and Rise ($\chi^2(1) = 16.52$, $p < 0.0001$) and between Choice and Vowel ($\chi^2(1) = 4.31$, $p < 0.05$). There was also a significant simple effect of Serial Position ($\chi^2(1) = 11.61$, $p < 0.001$) – participants made their response selection more quickly as the experiment progressed.

The interaction of Choice and Rise (**Figure 4**) partially supports prediction 1. That is, after early rises, question responses are faster than statement responses. However, after late rises the speed of statement responses is not any different to that of question responses. This suggests that although participants take an early rise as an indicator of a question, they do not show any preference to interpret a late rise as indicating a statement. Note, however, that this is not entirely surprising, since the trend over apparent time reported in the Introduction is for question rises to move to earlier alignment, meaning that older speakers (who our participants will still be listening to) will have late rises for questions as well as for statements, although rises on statements will be rather rare, since these older speakers are less likely to use uptalk.

The two-way interaction of Choice and Vowel (**Figure 5**) provides support for prediction 2. The interaction is due to an increase in latencies for statement responses following stimuli containing the more conservative [eə] realization of the SQUARE diphthong, compared to the other factor combinations shown in the figure. The finding that statement decisions take longer than question decisions after the [eə] realization is compatible with the observation that as the conservative variant [eə] signals speakers who are less likely to produce statements with rising intonation.

3.3 Maximum Deviation

It has been pointed out (e.g., Freeman & Ambady, 2010; Hehman et al., 2014) that if the mean trajectory for an experimental condition shows a moderate deviation from a straight line, and therefore also a moderate average MD value, then it might result from averag-

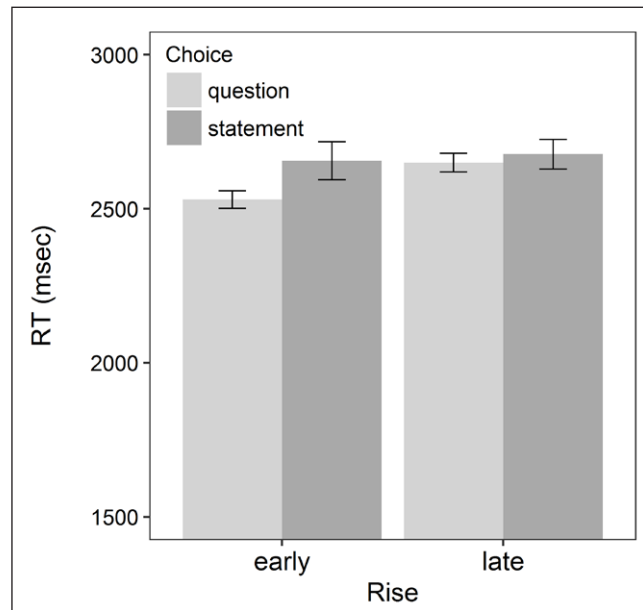


Figure 4: Effect of Rise alignment on response times (RTs) in the selection of question and statement responses.

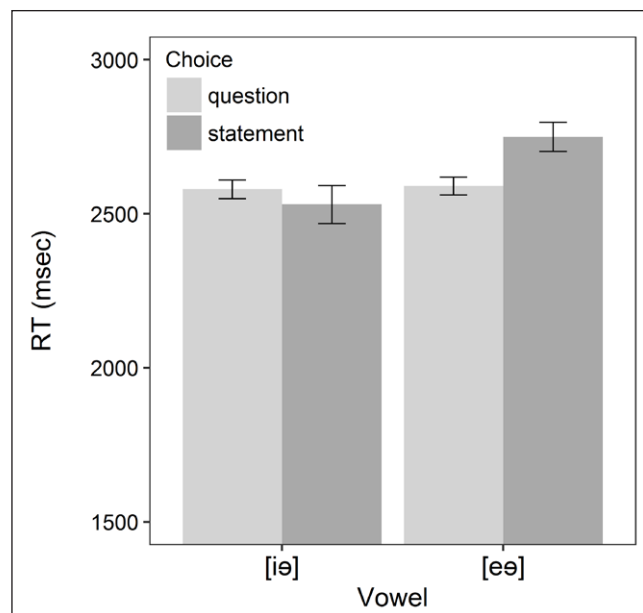


Figure 5: Effect of Vowel on response times (RTs) in the selection of question and statement responses.

ing two trajectory distributions—one that reflects marked attractions to the alternative response before resolving onto the selected response, and one that consists of more-or-less direct movement to the selected response. In at least one study using mouse-tracking there has been an explicit prediction of such a pattern. Dale and Duran (2011) asked participants to judge whether statements were true. One of the main experimental parameters of interest was whether the statement sentence contained a negation, since it has long been attested that the presence of a negation slows down readers during verification tasks. The researchers predicted that verification trials with high MD values, reflecting a strong but ultimately resisted temptation to respond that the sentence is not true, would be found when the sentence contained a negation. Their results supported this prediction. In another study, in which faces were categorized for sex, Freeman (2014) found that the

abrupt trajectory reversals (i.e., changes of mind) that are reflected in high MD values were more likely to occur with more ambiguous stimuli. In a second experiment these reversals were most likely when atypical faces (male faces with long hair or females with short hair) were seen in a normative context (i.e., where the majority of the male faces had short hair and the females long hair), but were also more likely for typical faces in counter-normative contexts (where the majority of male faces were seen with long hair, for instance) than in normative contexts.

The interpretation of MD data therefore crucially depends on whether mouse trajectories belong to a unimodal or bimodal distribution. While bimodality would mean that the usual statistical assumptions concerning normal distributions are challenged for the data-set as a whole, the presence of a bimodal distribution can itself be informative. As can be seen from **Figure 6**, the trajectories for test items in the current experiment are bimodally distributed. Hartigan’s dip statistic for unimodality (Hartigan & Hartigan, 1985) confirms that the distribution is significantly bimodal ($D = 0.032, p < 0.0001$). (For validation of the use of this statistic in the context of mouse-tracking data, see Freeman & Dale, 2013.) The vertical line indicates the empirically-derived cutoff value between the two distributions. Note that this is lower than the 0.9 value reported by Freeman (2014) for his study of sex categorization of faces. As explained earlier, negative MD values reflect trajectories that go below the idealized straight line from start to finish.

To explore factors that might influence the bimodal distribution evidenced in **Figure 6**, responses were categorized into low and high MD groups, and this classification was used as the dependent variable in further regression analysis. This analysis will be reported for the question responses only, since an initial model using all responses failed to converge once the low number of statement responses (see **Figure 3**) was further subdivided into low and high MD groups. As is apparent from the inset in **Figure 6**, the question responses showed a similar bimodal distribution to the overall pattern. The dip test returned the same result as that for the complete set ($D = 0.032, p < 0.0001$); the empirically-derived cutoff value used for the categorization was slightly higher (at 0.781). The high MD group had a mean MD value of 1.18 ($SD\ 0.20$) and the low MD group had a mean MD value of

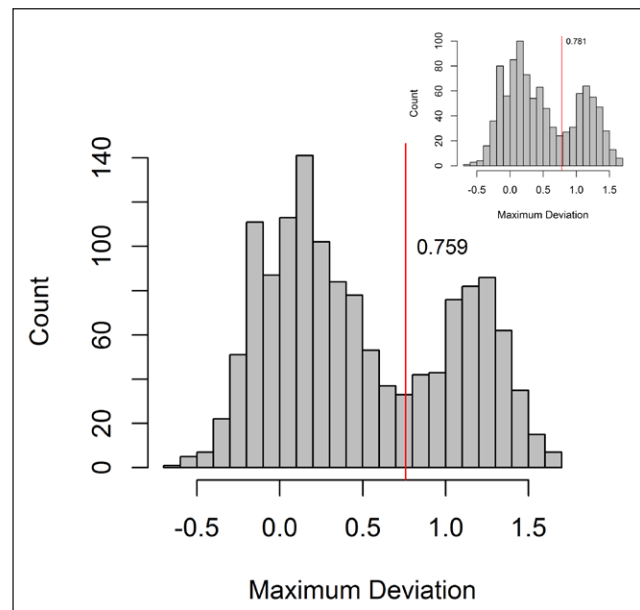


Figure 6: Distribution of Maximum Deviation values. (Inset is for question responses only. See text).

0.17 (*SD* 0.28). An analysis of RTs for question responses with MD group as one of the predictors returned an unsurprising result, with the more direct low MD tracks significantly faster than the high MD tracks ($\chi^2(1) = 46.46, p < 0.0001$).

Logistic regression analysis of the question response set with MD group as the dependent variable and the fixed effects of Vowel, Rise, Source, and Serial Position, as well as the interactions of Rise with Vowel and with Source, produced a single significant effect, namely that of Rise. Stimuli with late rises were significantly more likely to exhibit a trajectory reversal (early: 0.283, late 0.399; $\chi^2(1) = 21.82, p < 0.0001$). In other words, as anticipated by prediction 1, participants were significantly more likely to track towards the statement response box before making a question response in the context of late rises than in the context of early rises. If greater likelihood of a trajectory reversal is symptomatic of ambiguous stimuli, as claimed by Freeman (2014) in connection with his results for sex categorization of faces, then the late rise stimuli are more ambiguous than the early rise stimuli. This is further confirmation of the asymmetry in the signal value of the two rise types that was reflected in the response time data reported in connection with **Figure 4**.

However, an alternative explanation of this result is that in the case of an early rise, the information that indicates a question, i.e., the rise, becomes available earlier, and as a consequence there is less opportunity for the alternative response to compete. In the late condition, on the other hand, increased attraction to the statement response might be a result of the rise information becoming available later, with the information before the rise remaining compatible with the statement response. That is, the finding that the move to the question response is more direct with early rises (and potentially also the overall effect in response choice shown in **Figure 3**) may not be due to a functional distinction between early and late rises, but might simply be a consequence of when the high pitch information becomes available in the task being performed by participants for this experiment.

Such an explanation would seem to be discounted by a closer analysis of the MD measure in the question responses. Of particular interest is whether there are measurable influences on MD of factors other than rise alignment within the high MD distribution, i.e., in the set where the competition effects of the alternative response are evident. A further regression model was therefore performed on the high MD group, with MD as the dependent variable. The factors explored were Vowel, Rise, Source, and Serial Position, as well as the interactions of Rise with Vowel and with Source. Source had no significant impact, neither as a simple effect nor in interaction, indicating that trajectory movements towards the competing responses were not influenced by uncontrolled properties distinguishing the original sets of statement and question utterances from which the test items were derived. There was however a significant interaction of Vowel and Rise ($\chi^2(1) = 8.79, p < 0.005$), as shown in **Figure 7**. Stimuli with early rises exhibited less attraction to the alternative statement response (i.e., had smaller MD values) when the rise followed an [iə] vowel than when it followed an [eə] vowel. In addition, **Figure 7** indicates that after [iə], question responses to early rises show less attraction to the statement response than question responses to late rises. These interaction effects involving Vowel and Rise were anticipated by prediction 3.

4 Discussion

The experiment reported in this paper was designed to test whether the nature of a segmental phonetic variable influences listeners' interpretation of prosodic variation. The experiment exploited the co-variation, predominantly among younger speakers, of a

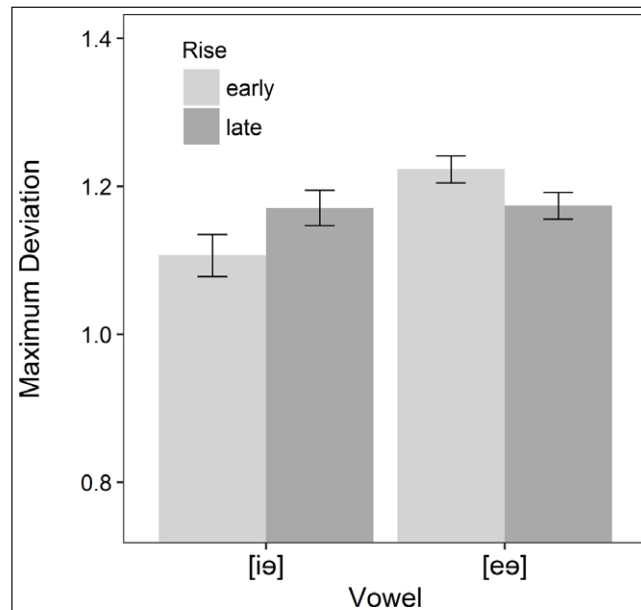


Figure 7: Maximum Deviation by Rise and Vowel, for question responses in the high MD set.

closer articulation of the SQUARE diphthong (approximating the NEAR vowel, i.e., [iə]) and final statement rises (uptalk) in NZE, and in particular the recent finding that earlier rises may provide a possible means of distinguishing questions from uptalk statements (Warren, 2005; Warren & Daly, 2005; Warren & Fletcher, 2016). Therefore two aspects of prosodic variability are involved—the variable use of final rises to signal either questions or statements in the same speech community, and the variability in the alignment of the final rise that is trending towards a marker that distinguishes these sentence types. Such variability becomes informative once it can be readily disentangled, so that there is greater clarity over whether a final rise indicates a question or a statement. The current study aimed to see whether the segmental phonetic cue in the test utterances would indicate whether or not the speaker is from a group of speakers that is likely firstly to produce uptalk and secondly to distinguish uptalk from question rises through alignment difference in the rise starting point.

The predictions set out in section 2.4 received good support. In line with previous perceptual research (Zwartz & Warren, 2003), the experiment provided further evidence that early rises are more clearly associated with questions and late rises with statements. This was reflected in the analysis of the response choice data, which showed a clear effect of rise type on the proportion of question responses. Question responses after early rises were also faster and showed a more direct mouse trajectory than question responses after late rises, and trajectory reversals were more likely after late rises, indicating increased competition from the statement response in this condition.

In addition, the manipulation of the SQUARE vowel successfully shifted performance in the forced-choice response task. The slower statement responses to items containing an [eə] version of this vowel than to items containing an [iə] version is an indication that although uptalk rises from more conservative speakers are not ruled out, they are less expected. The mouse-tracking data for question responses made after a trajectory reversal show that there is less competition from the statement response for an early rise after the [iə] realization of SQUARE than for the same rise after the [eə] realization. The trajectories also show that in the context of the [iə] realization there is less competition from the statement response when the rise is early than when it is late. These findings of significant results in the mouse-tracking data, particularly in the absence of significant differences

in overall response choice, support the value of the technique in adding a more qualitative aspect to the interpretation of participants' responses. In particular, these trajectory reversal data linked to the MD measure allow us to say more about the competition effects that exist between alternative response choices.

The results of this study indicate that variability in intonational rise alignment in NZE is meaningful, and is more closely tied to certain speaker groups than to others. Listener performance in the experiment reflects the parallel development in this variety of a segment-level merger and changes in the use and shape of an intonation pattern. The vowel merger has been able to happen largely because the consequences are not extreme—although letters to newspaper editors might suggest otherwise (e.g., complaints about hearing of people “crossing on the Cook Strait ‘Fairy’ and flying ‘Ear’ New Zealand” [Bravery, 2001]), confusion between words with NEAR or SQUARE tokens is not frequent, partly because the vowel sounds are relatively rare (17th and 18th most frequent out of 20 English vowels, Gimson, 1963), but also because there are few relevant minimal pairs, and because utterance contexts usually serve to disambiguate, just as they do for other homophones. Nevertheless, the merger is still strongly age-graded, and so the use of the [iə] or [eə] variant of the SQUARE vowel has potential as a marker of social grouping.

On the other hand, the prosodic variability considered here has the potential to cause confusion between two radically different sentence types—questions and statements. This is reflected again in complaints from the public, but also in the frequent definition of uptalk using terms such as ‘question-like intonation on a declarative utterance.’ Such a definition is problematic on a number of counts. The notion of ‘question’ is ambiguous between function and form, and the belief that there is any single type of ‘question intonation’ is misplaced, as is the assumption that questions have to have rising intonation. Even if we accept that ‘question-like intonation’ stands for the type of rising intonation often found on yes-no questions, there is lack of agreement about whether that type of rising intonation is identical or even similar to uptalk. Some of this disagreement may result from the presence of different styles and uses of uptalk in different English varieties (Warren, 2016). It is worth remembering also that different varieties may have reached different stages in the development and use of uptalk, with some varieties (e.g., NZE) beginning to show novel distinctions, while others (such as British English, Shobbrook & House, 2003) do not.

Ladd (2008: 126) acknowledges that uptalk and question rises may differ, but that “the differences are subtle, and arguably gradient,” and that it may be plausible to “analyse HRT statements as having the same phonological representations as high-rising question contours.” The results of the current study show that the ‘arguably gradient’ distinction between early and late rise in NZE does have the potential to signal a distinction in meaning, and that uptalk and question contours may have different phonological representations. As pointed out by House (2006), the development of a new phonological distinction is a natural resolution of a potentially confusing situation in which one pattern can have multiple functions. Note though that other factors may make such a split unlikely to happen precipitously. As Guy et al. (1986) pointed out in their discussion of Australian English, even without a phonetic difference between uptalk rises and question rises, ambiguity is unlikely, as contexts will clarify the intended meaning. For instance, true questions usually include some anaphoric reference to a previous utterance, and will usually be at the end of a turn, while uptalk tends to provide new information, with the speaker typically continuing to hold the floor. “If these clear contextual and textual differences were not sufficient to disambiguate between the two meanings, we might expect structural change to occur: For example, phonetic differentiation of the contours or disuse of

the contour for one of the meanings” (Guy et al., 1986: 27). It may be that recent phonetic differences between uptalk and question intonation that have been noted for Australian and New Zealand English indicate that a level of potential confusion has been reached that requires such differentiation.

Additional File

The additional file for this article can be found as follows:

- **File 1.** Examples of the four stimuli based on the sentence illustrated Figure 1. DOI: <https://doi.org/10.5334/labphon.92.s1>
- **File 2.** List of test items used in the experiment. DOI: <https://doi.org/10.5334/labphon.92.s2>

Acknowledgements

I am grateful to Chigusa Kurumada and to three anonymous reviewers for their comments on earlier versions of this article. My thanks also go to participants at *Variation and Language Processing 2* and at the *15th Australasian International Conference on Speech Science and Technology* for feedback on presentations of the research reported here.

Competing Interests

The author has no competing interests to declare.

References

- Barca, L., Benedetti, F. and Pezzulo, G. 2016. The effects of phonological similarity on the semantic categorisation of pictorial and lexical stimuli: Evidence from continuous behavioural measures. *Journal of Cognitive Psychology*, 28(2), 159–170. DOI: <https://doi.org/10.1080/20445911.2015.1101117>
- Barca, L. and Pezzulo, G. 2012. Unfolding visual lexical decision in time. *PLoS ONE*, 7(4), e35932. DOI: <https://doi.org/10.1371/journal.pone.0035932>
- Barca, L. and Pezzulo, G. 2015. Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, 10(2), e0116193. DOI: <https://doi.org/10.1371/journal.pone.0116193>
- Bates, D., Maechler, M., Bolker, B. and Walker, S. 2015. lme4 (Version 1.1–8). Retrieved from <http://CRAN.R-project.org/package=lme4>.
- Bauer, L. and Warren, P. 2004. New Zealand English: Phonology. In: Kortmann, B., Schneider, E. W., Burrige, K., Mesthrie, R. and Upton, C. (eds.), *A handbook of varieties of English: A multimedia reference tool* (Vol. 1, pp. 580–602). Berlin: Mouton de Gruyter.
- Bell, A. and Johnson, G. 1997. Towards a sociolinguistics of style. *University of Pennsylvania Working Papers in Linguistics*, 4, 1–21.
- Benton, R. A. 1965. *Research into the English language difficulties of Maori school children 1963–1964*. Wellington: Māori Education Foundation.
- Boersma, P. and Weenink, D. 2012. Praat: Doing phonetics by computer (Version 5.3.06).
- Bravery, L. 2001. Letter to the Editor. *New Zealand Listener*, June 9th.
- Britain, D. 1992. Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, 4, 77–104. DOI: <https://doi.org/10.1017/S0954394500000661>
- Dale, R. and Duran, N. D. 2011. The cognitive dynamics of negated sentence verification. *Cognitive Science*, 35(5), 983–996. DOI: <https://doi.org/10.1111/j.1551-6709.2010.01164.x>
- Dorrington, N. 2010. *‘Speaking up’: A comparative investigation into the onset of uptalk in General South African English* (BA[Honours] thesis). Rhodes University.

- Drager, K. 2011. Speaker age and vowel perception. *Language & Speech*, 54(1), 99–121. DOI: <https://doi.org/10.1177/0023830910388017>
- Fletcher, J., Grabe, E. and Warren, P. 2005. Intonational variation in four dialects of English: The high rising tone. In: Jun, S.-A. (ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 390–409). Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199249633.003.0014>
- Fletcher, J. and Harrington, J. 2001. High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, 58(4), 215–229. DOI: <https://doi.org/10.1159/000046176>
- Fletcher, J. and Loakes, D. 2010. *Interpreting rising intonation in Australian English*. *Speech Prosody 2010*, 100124: 1–4.
- Freeman, J. B. 2014. Abrupt category shifts during real-time person perception. *Psychonomic Bulletin & Review*, 21(1), 85–92. DOI: <https://doi.org/10.3758/s13423-013-0470-8>
- Freeman, J. B. and Ambady, N. 2010. MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241. DOI: <https://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., Ambady, N., Rule, N. O. and Johnson, K. L. 2008. Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, 137(4), 673–690. DOI: <https://doi.org/10.1037/a0013875>
- Freeman, J. B. and Dale, R. 2013. Assessing bimodality to detect the presence of a dualcognitive process. *Behavior Research Methods*, 45(1), 83–97. DOI: <https://doi.org/10.3758/s13428-012-0225-x>
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M. and Ambady, N. 2011. Looking the part: Social status cues shape race perception. *PLoS ONE*, 6, e25107. DOI: <https://doi.org/10.1371/journal.pone.0025107>
- Gimson, A. C. 1963. *An Introduction to the Pronunciation of English*. London: Edward Arnold.
- Gordon, E. and Maclagan, M. 2001. Capturing a sound change: A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of Linguistics*, 21(2), 215–238. DOI: <https://doi.org/10.1080/07268600120080578>
- Guy, G., Horvath, B., Vonwiller, J., Daisley, E. and Rogers, I. 1986. An intonational change in progress in Australian English. *Language in Society*, 15(1), 23–51. DOI: <https://doi.org/10.1017/S0047404500011635>
- Hartigan, J. A. and Hartigan, P. M. 1985. The dip test of unimodality. *The Annals of Statistics*, 13, 70–84. DOI: <https://doi.org/10.1214/aos/1176346577>
- Hay, J. and Drager, K. 2010. Stuffed toys and speech perception. *Linguistics*, 48(4), 865–892. DOI: <https://doi.org/10.1515/ling.2010.027>
- Hay, J., Nolan, A. and Drager, K. 2006a. From fush to feesh: Exemplar priming in speech perception. *Linguistic Review*, 23(3), 351–379. DOI: <https://doi.org/10.1515/TLR.2006.014>
- Hay, J., Warren, P. and Drager, K. 2006b. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4), 458–484. DOI: <https://doi.org/10.1016/j.wocn.2005.10.001>
- Hehman, E., Stoller, R. M. and Freeman, J. B. 2014. Advance mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations*, 18(3), 384–401. DOI: <https://doi.org/10.1177/1368430214538325>
- House, J. 2006. Constructing a context with intonation. *Journal of Pragmatics*, 38(10), 1542–1558. DOI: <https://doi.org/10.1016/j.pragma.2005.07.005>

- Johnson, K., Strand, E. A. and D'Imperio, M. 1999. Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384. DOI: <https://doi.org/10.1006/jpho.1999.0100>
- Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R. 2008. *Intonational Phonology* (2nd ed.), Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511808814>
- Lakoff, R. 1973. Language and woman's place. *Language in Society*, 2, 45–79. DOI: <https://doi.org/10.1017/S0047404500000051>
- Levon, E. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 36(4), 533–554. DOI: <https://doi.org/10.1017/S0047404507070431>
- McGregor, J. 2005. *High rising tunes in Australian English* (Doctoral dissertation). Macquarie University, Sydney, Australia.
- Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85. DOI: <https://doi.org/10.1177/0261927X99018001005>
- Papesh, M. H. and Goldinger, S. D. 2012. Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, 19(5), 906–913. DOI: <https://doi.org/10.3758/s13423-012-0281-3>
- Psychology Software Tools. 2012. E-Prime (Version 2.0). Retrieved from <http://www.pstnet.com>.
- Ritchart, A. and Arvaniti, A. 2014. The form and use of uptalk in Southern Californian English. In: Campbell, N., Gibbon, D. and Hirst, D. (eds.), *Social and Linguistic Speech Prosody (Proceedings of the 7th international conference on Speech Prosody)* (pp. 331–335). Dublin, Eire.
- Shobbrook, K. and House, J. 2003. High rising tones in Southern British English. In: Solé, M. J., Recasens, D. and Romero, J. (eds.), *15th International Congress of Phonetic Sciences* (pp. 1273–1276). Barcelona: Universitat Autònoma de Barcelona.
- Singmann, H., Bolker, B. and Westfall, J. 2015. afex: Analysis of Factorial Experiments (Version 0.13–145). Retrieved from <https://cran.r-project.org/web/packages/afex/index.html>.
- Spivey, M. J., Grosjean, M. and Knoblich, G. 2005. Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398. DOI: <https://doi.org/10.1073/pnas.0503903102>
- Strand, E. A. and Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D. (ed.), *Natural Language Processing and Speech Technology* (pp. 14–26). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110821895-003>
- Warren, P. 2005. Patterns of late rising in New Zealand English: Intonation variation or intonational change? *Language Variation and Change*, 17, 209–230. DOI: <https://doi.org/10.1017/S095439450505009X>
- Warren, P. 2016. *Uptalk: The phenomenon of rising intonation*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781316403570>
- Warren, P. and Daly, N. 2005. Characterising New Zealand intonation: Broad and narrow analysis. In: Bell, A., Harlow, R. and Starks, D. (eds.), *Languages of New Zealand*. Wellington: Victoria University Press.
- Warren, P. and Fletcher, J. 2016. Phonetic differences between uptalk and question rises in two Antipodean English varieties. In: Barnes, J., Brugos, A., Shattuck-Hufnagel, S.

- and Veilleux, N. (eds.), *Speech Prosody 2016* (pp. 148–152). Boston, MA. DOI: <https://doi.org/10.21437/SpeechProsody.2016-31>
- Warren, P., Hay, J. and Thomas, B. 2007. The loci of sound change effects in recognition and perception. In: Cole, J. and Hualde, J. I. (eds.), *Laboratory Phonology 9* (pp. 87–112). Berlin: Mouton de Gruyter.
- Warren, P., Rae, M. and Hay, J. 2003. *Word recognition and sound merger: The case of the front-centering diphthongs in NZ English*. In Solé, M. J., Recasens, D., and Romero, J. (eds.), *15th International Congress of Phonetic Sciences* (pp. 2989–2992). Barcelona: Universitat Autònoma de Barcelona.
- Wells, J. C. 1982. *Accents of English*. Cambridge, England: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511611759>
- Zwartz, J. and Warren, P. 2003. This is a statement? Lateness of rise as a factor in listener interpretation of HRTs. *Wellington Working Papers in Linguistics*, 15, 51–62.

How to cite this article: Warren, P 2017 The Interpretation of Prosodic Variability in the Context of Accompanying Sociophonetic Cues. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 11, pp.1–21, DOI: <https://doi.org/10.5334/labphon.92>

Published: 04 May 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 