PERCEPTION OF PITCH HEIGHT AND PROMINENCE IN SPEECH BY LISTENERS IN ENGLAND, BELFAST AND GLASGOW

Hae-Sung Jeon

School of Psychology and Humanities, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UNITED KINGDOM; hjeon1@uclan.ac.uk

Antje Heinrich

Manchester Centre for Audiology and Deafness, School of Health Sciences, The University of Manchester, Manchester, UNITED KINGDOM; antje.heinrich@manchester.ac.uk

Proc. Mtgs. Acoust. 54, 050001 (2024)

https://doi.org/10.1121/2.0001898

Acoustics Week in Canada

Joint Meeting

186th Meeting of the Acoustical Society of America

and the Canadian Acoustical Association

Ottawa, Ontario, Canada

13-17 May 2024

Psychological and Physiological Acoustics: Paper 4aPP6

1. INTRODUCTION

Previous research has shown a perceptual asymmetry between fundamental frequency (f_o) peaks and valleys in speech for both the psychoacoustic perception of pitch and linguistic judgement of prominence (Jeon and Heinrich, 2022a, b; Barnes et al., 2023). In general, rising or high pitch seems to be perceptually more salient than falling or low pitch (Evans, 2015; Hsu et al., 2015; Lialiou et al., 2024). Listeners showed heightened discrimination of the relative height between two successive f_o peaks associated with two stressed syllables in an utterance (e.g., 'does Néllie know Lénny?') compared to valleys when the peaks and valleys were created to be acoustically symmetrical (see Figure 1; Jeon and Heinrich, 2022a, b). Meanwhile, the experimental task type, i.e., whether listeners were asked to judge relative height or prominence, affected listeners' responses. For judging prominence (i.e., which of the two successive peaks or valleys sounded more prominent, standing out or emphatic), listeners needed a larger f_o excursion to perceive the two peaks or valleys equal compared to when they judged pitch height. Furthermore, listeners seemed reluctant to associate f_o valleys with prominence even when they could discriminate the valley height (Jeon and Heinrich, 2022b). These results corroborate the finding that the perception of prominence is complex (Ladd, 2008, Chap. 2; Baumann and Winter, 2018; Cole et al., 2010; Bishop et al., 2020).

The experimental participants in Jeon and Heinrich (2022a, b) were native English speakers living in England. Many varieties spoken in the United Kingdom follow the cross-linguistic tendency that high pitch and a pitch rise is associated with prominence (Liberman, 1967), and pitch falls at the end of declaratives (cf. Liberman, 1967; Gussenhoven, 2004; Evans, 2015). However, prosodic properties of listeners' native language(s) can affect their perception (Cumming, 2011; Barnes et al., 2023), and the association of high pitch and prominence may be the outcome of listeners' constant exposure to accentuation realized by high pitch. The present study examined the effect of linguistic experience by testing listeners who have been exposed to Belfast and Glaswegian varieties in which low but prominent syllables are frequent. In spontaneous and informal speech of Belfast and Glaswegian varieties, the default pitch accent has a low pitch followed by a rise (Cruttenden, 1997), and both declaratives and interrogatives have a frequent final rise (Jarman and Cruttenden, 1976; Grabe, 2002; Lowry, 2002; Smith and Rathcke, 2020).

2. BELFAST ENGLISH AND GLASWEGIAN ENGLISH

A. BELFAST ENGLISH

Four nuclear accent patterns were identified for Belfast English in Lowry (2002): the rise-plateau (the most frequent), the rise-plateau-slump (terms by Cruttenden, 1986), the high rise, and the fall. What is of interest here is that prominence is often associated with low pitch (e.g., in impressionistic illustrations in Rahilly, 1997, in yes-no interrogatives in Grabe, 2002, and Jeon and Nichols, 2022). Existing literature has little discussion on the acoustic properties of prominence in Belfast English, but it seems that pitch provides an important cue to prominence. Listeners impressionistically comment that Belfast English sounds monotonous, but the most prominent syllable in an utterance is preceded by a "relatively large amount of pitch obtrusion, compared to the obtrusion which separates the previous stressed syllable (p. 118, Rahilly, 1997)." In Kochanski et al. (2005), when automatic prominence classifiers were trained with speech in different English varieties (e.g., Leeds, London, Newcastle, etc.), loudness was always identified as the most important cue to prominence. But f_o was weighted more for Belfast compared to other varieties. Finally, the stressed syllables in Belfast English did not seem as acoustically strong as those in the Standard Southern British English (SSBE) when periodic energy mass was analyzed (Jeon and Nichols, 2022).

B. GLASWEGIAN ENGLISH

The nuclear rise in Glaswegian English often forms a 'rise-plateau-slump' (Mayo, 1996; Mayo et al., 1997). That is, the pitch rises after the lowest point in the accented syllable, stays high or slightly falls until near the end of the phrase where it falls. Glaswegian English has a shallower 'prominence gradient', i.e., relatively small contrasts between unaccented and accented, or unstressed and stressed syllables compared to SSBE with respect to intensity and duration (Smith and Rathcke, 2020). For instance, the Glaswegian short vowels that undergo the Scottish Vowel Length Rule (/i, u, aɪ/, Aitken, 1981; Scobbie et al., 1999) do not lengthen under accentual prominence, and the contrast between strong and weak syllables in intensity and duration in the post-nuclear final trochaic words is larger in SSBE than in Glaswegian English (Rathcke and Stuart-Smith, 2016). Glaswegian English seem to have a relatively weak concentration of prominence cues in the stressed syllable and some of the cues (e.g., the plateau-slump portion of the rise-plateau-slump) are delayed beyond the stressed syllable. That is, the unstressed syllable following the stressed syllable may have high pitch and long duration (Rathcke and Stuart-Smith, 2016).

3. EXPERIMENT

The experiment investigated whether listeners' linguistic experience affected their psychoacoustic (pitch height) and linguistic (prominence) judgement on the f_o movement. In the experimental materials, we varied the type of f_o extrema (peaks vs valleys) associated with two stressed syllables in an utterance, 'does Néllie know Lénny?', the second f_o extremum shape (25 ms vs 50 ms plateau) and its f_o. Listeners identified either the relative pitch height or prominence between the peaks or valleys.

A. HYPOTHESIS

Based on Jeon and Heinrich (2022a, b), we expected listeners to treat the height and prominence tasks differently. The longer f_o plateau was expected to make a peak sound higher, a valley sound lower, and make both a peak and a valley more prominent compared to extrema with a shorter plateau (Jeon and Heinrich, 2022a, b; Knight, 2008; Barnes et al., 2023). If the prosodic properties of listeners' language influence their perception, then there would be an interaction between language variety and task, and between language variety and extremum type (peaks vs valleys).

Specifically, the first hypothesis was that for judging peak height, which should be an easy task (Jeon and Heinrich, 2022a, b), there would be no effect of language variety. Second, for judging peak prominence, we expected reduced discrimination of Belfast and Glasgow listeners; the stressed syllables in Belfast English are weaker, and Glaswegian English shows shallow 'prominence gradients' compared to SSBE. Third, for judging valley height, we expected that compared to English listeners, Belfast and Glasgow listeners would show heightened discrimination, because their frequent exposure to low accents could improve their discrimination. Fourth, for judging valley prominence, we expected Belfast and Glasgow listeners to show heightened discrimination, because their languages often associate prominence with low pitch.

B. METHODS

I. PARTICIPANTS

The data of 32 listeners living in England (21–35 years old; 19 female, 13 male) are from Jeon and Heinrich (2022b). Twenty-six native speakers of Northern Irish and 26 native speakers of Scottish English were recruited online using Prolific (www.prolific.co, 23 November 2021–14 January 2022). All participants had normal hearing and vision, no history of cognitive impairments, and no professional music training. For the Belfast group (20–35 years old, 16 female, 10 male), participants were born and brought up in Northern Ireland, and they were currently living in Belfast (postcode BT). For the Glasgow group (19–34 years old; 15 female, 11 male), participants were born and brought up in Scotland, and they were currently living in Glasgow (postcode G). Participants were randomly assigned into two Task groups, Height (16 England, 16 Belfast and 14 Glasgow listeners) and Prominence (16 England, 10 Belfast and 12 Glasgow listeners).

II. EXPERIMENTAL DESIGN AND STIMULI

The experimental design was: 3 Areas (England, Belfast, Glasgow) × 2 Tasks (Height, Prominence) × 2 Extremum Types (Peak, Valley) × 2 Second Extremum Plateau Durations (25 ms, 50 ms) × 5 Steps (-2, -1, 0, 1, 2) for the second extremum height.

The stimuli were resynthesised utterances, 'does Néllie know Lénny?' spoken by a female native speaker of SSBE. (See Jeon and Heinrich, 2022a for details of the recording and resynthesis procedure.) The built-in 'manipulation' function of Praat ver. 6.1.16 (Boersma and Weenink, 2020) was used to resynthesize experimental stimuli from the recorded base stimulus. Half of the experimental stimuli had two peaks associated with the stressed syllables and the other half valleys (Fig. 1). The first extremum always formed a 25 ms f_o plateau; the first peak was set at a constant value of 260 Hz, valley at 200 Hz (Table 1). The second peak or valley varied in height in five 0.25 EBR_N (11–13 Hz) steps (Moore, 2012, p. 76) and formed an either 25 ms or 50 ms plateau. The f_o at Step -2 was the closest to the baseline. The right edge of the f_o plateau was aligned at the end of the stressed vowel. The flat baseline for Peak was at 200 Hz, for Valley at 260 Hz. This setup allowed all pitch events to occur in the normal speaking range for the speaker.

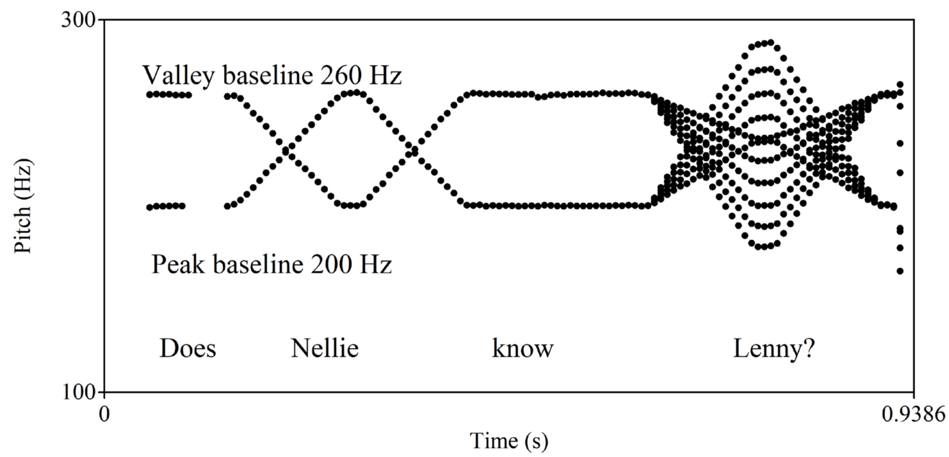

Figure 1. Sample f_o contours of the experimental materials. The stimuli in this figure have a 25 ms plateau at the peaks and valleys.

Table 1. Maximum and minimum f_o of the second extremum. The first extremum's height was always at Step 0. A negative Step indicates that the second extremum had a smaller f_o excursion size from the baseline than the first. The values in semitone were calibrated relative to 100 Hz (ST = 12 ln(Hz /100 /ln 2).

Step	**EBR_N**	**Hz**	**semitones**
Peak			
-2	6.55	234	14.75
-1	6.80	247	15.66
0	7.05	260	16.54
1	7.30	273	17.41
2	7.55	287	18.25
Valley			
-2	6.34	224	13.94
-1	6.09	212	12.98
0	5.84	200	12.00
1	5.59	189	10.99
2	5.34	178	9.94

III. EXPERIMENTAL PROCEDURE

The study was approved by the Business, Arts, Humanities and Social Science Ethics committee of the University of Central Lancashire (BAHSS2 0122). The experiment was hosted on Gorilla (Anwyl-Irvine, 2019). Participants completed a questionnaire on their native English variety, gender, age, and musical experience. Participants were instructed to wear headphones and passed the headphone screening test (Woods et al., 2017).

Experimental stimuli were divided into four blocks; two blocks included Peak stimuli and the other two blocks had Valley stimuli. The presentation order for Peak and Valley was counterbalanced; half of the participants completed two Peak blocks first and the other half completed two Valley blocks first. The participants were told that they would hear an utterance with either two high 'peaks' or two low 'valleys'. Participants in the 'Height' group were asked to judge which sounded higher for the peaks or lower for the valleys. Participants in the 'Prominence' group were asked to judge which peak or valley sounded more prominent, standing out or emphatic. They indicated their response by clicking either the 'Nellie' or 'Lenny' button on the screen. There was a practice session before the first block and also before the third block. For each practice session, there were 4 trials (2 Plateau Durations [25 ms, 50 ms] × 2 Steps [-3, 3]). In the main experiment, each listening stimulus was presented four times in total. Four catch trials (simple mathematical operations with the answer being either 1 or 2, e. g., 4 - 3 = ?) were presented visually to keep participants attentive. The stimulus presentation order was randomized for each participant for both the practice session and the main experiment within a block. Participants could repeat the practice session if they wished. No feedback was provided for both the practice session and the main experiment.

C. ANALYSES AND RESULTS

None of the participants reported to have absolute pitch. In total, 6,720 data points (2 Extrema × 2 Plateau Durations × 5 Steps × 4 repetitions × 84 participants) were collected. For data processing and statistical analysis, the package *tidyverse* Version 2.0.0 (Wickham et al., 2019) and the package *brms* Version 2.21.0 (Bürkner, 2017) were used respectively for R Ver. 4.3.1 (R Core Team, 2023) and R Studio 2022.07.1+554 (RStudio Team, 2020). The package *praatpicture* Version 1.2.0 (Puggaard-Rode, 2024) was used for plotting a spectrogram and f_o track together.

Figure 2 shows the frequency of 'second' (Lenny) response. Compared to the peak stimuli (Figure 2a), the response functions for the valley stimuli (Figure 2b) are flatter across all Task and Area conditions. In particular, the response functions for the prominence judgement task for valleys do not show a positive slope like others. Furthermore, listeners' 'second' responses were below the reference line marking the 0.5 frequency ratio on the y-axis, i.e., listeners were more likely to choose the 'first' response when they judged prominence for valleys.

To investigate the trends described above, we constructed Bayesian logistic models (binomial, links = logit). The categorical predictors were sum-coded; Area (contrast 1: Belfast [1]. England [0], Glasgow [-1]; contrast 2: Belfast [0], England [1] and Glasgow [-1]), Extremum Type (Peak [1], Valley [-1]), the second Plateau Duration (25 ms [1] vs 50 ms [-1]), and Task (Height [1], Prominence [-1]). This coding scheme allowed the model intercept to be in the middle of categories concerned. The five manipulation Steps were incorporated as a continuous predictor. The models estimated the maximum likelihood of the 'second' (Lenny) response.

We carried out the modelling in two stages. Firstly, we investigated the interactions involving Area using all data (Section 3. C. I). Secondly, we examine the hypothesized differences in discrimination across the variety groups by incorporating interaction terms with the Step predictor for the two extremum types, Peak, and Valley, respectively (Section 3. C. II). The step parameter shows the response function's slope; a steep and positive slope indicates heightened discrimination. For the intercepts, default priors were used. For coefficients, the weakly informative priors formed normal distributions centered at zero (SD = 0.5). Hamiltonian Markov Chain Monte Carlo (MCMC) sampling was conducted with four chains and 10,000 iterations (2,000 for warm-up). This resulted in a total of 32,000 posterior samples. There was no indication of convergence issues in all models (all Rhat values = 1.00).

I. ALL DATA: INTERACTIONS INVOLVING AREA

As we aimed to investigate the role of Area, all three-way interaction terms involving Area were incorporated in the model. We avoided incorporating any four-way or higher-order interaction terms which are not readily interpretable. Listener was incorporated as a random factor. The model structure is provided in (1). (See Bürkner et al., 2024 for notation conventions.)

$$brm(response|trials(4) \sim 1 + area + taskType + extremumType + plateau + step + area * taskType * extremumType + area * taskType * plateau + area * extremumType * plateau + (1 \,|\, ID)) \quad (1)$$

In the model summary (Table 2), a positive log odd coefficient ($\breve{\beta} > 0$) indicates that the relevant predictor is associated with an increase in the 'second' responses. The 95% credible interval (CI) not straddling zero is considered an indicator for the predictor's reliable effect. The posterior probability shows the estimated coefficients' posterior sample distribution. If the posterior probability of the coefficient being above or below zero is 1 ($Pr(\breve{\beta} > 0) = 1$ or $Pr(\breve{\beta} < 0) = 1$), this indicates a strong effect, i.e., not a single posterior sample for this coefficient was below or above zero. Below only the effects with strong evidence are discussed.

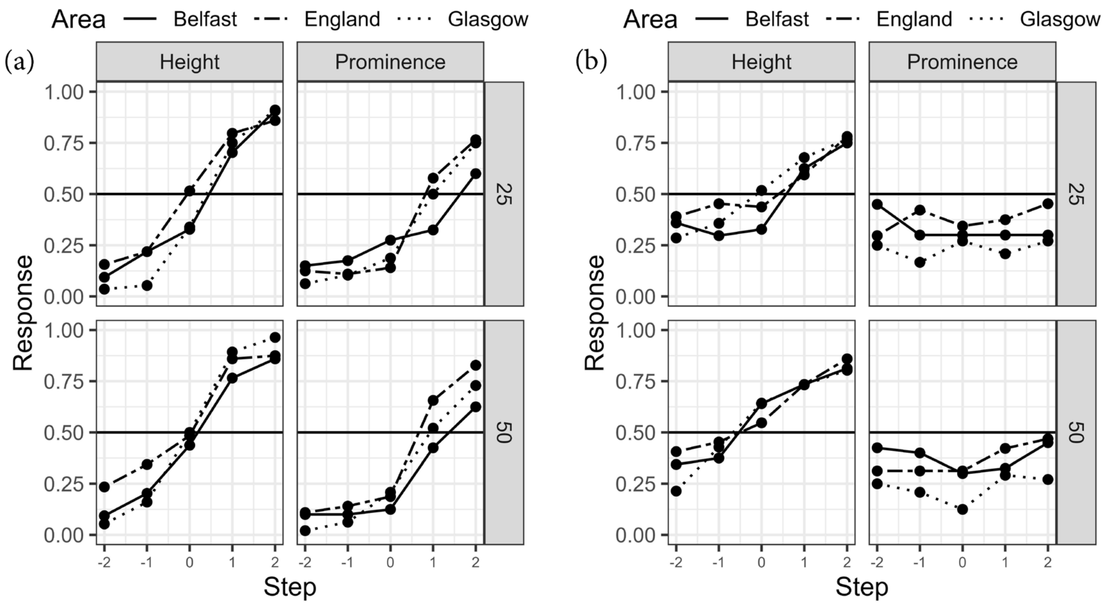

Figure 2. 'Second' response frequency (mean ratio) for (a) Peak and (b) Valley. Data were collapsed across participants and trials.

We first discuss the main predictors with a reliable effect (Table 2). First, there was strong evidence for the Area effect for the second contrast ($\breve{\beta}$ = 0.13, CI [0.01, 0.25], Pr ($\breve{\beta}$ > 0) = 0.98), but not for the first contrast ($\breve{\beta}$ = -0.01, CI [-0.14, 0.12], Pr($\breve{\beta}$ < 0) = 0.58). That is, listeners in England (n = 1175, 46%) were more likely to choose the 'second' response compared to those in Glasgow (n = 827, 40%) when Belfast (n = 890, 43%) was treated as the reference in the middle. Second, a positive coefficient for Task ($\breve{\beta}$ = 0.31) was reliable with the 95% credible interval away from zero (CI [0.22, 0.40]) and high posterior probability (Pr($\breve{\beta}$ > 0) = 1); the 'second' response probability was higher for Height (n = 1901, 52%) compared to Prominence (n = 991, 33%). Third, there was evidence for the Plateau Duration effect ($\breve{\beta}$ = -0.04, CI [-0.08, 0], Pr ($\breve{\beta}$ < 0) = 0.97), showing a lower 'second' response probability for 25 ms (n = 1382, 41%) compared to 50 ms (n = 1510, 45%). Finally, the Step effect was reliable ($\breve{\beta}$ = 0.37, CI [0.34, 0.40], Pr($\breve{\beta}$ > 0) = 1); a step increase was associated with an increase of the 'second' response probability.

For interactions, the Area × Extremum Type interaction effect was reliable (contrast 1, $\breve{\beta}$ = -0.07, CI [-0.13, -0.01], Pr($\breve{\beta}$ < 0) = 0.99). There was also evidence for the Task × Extremum Type interaction effect ($\breve{\beta}$ = -0.04, CI [-0.08, 0], Pr($\breve{\beta}$ < 0) = 0.97). Furthermore, there was indication that the Area effect was dependent on Task and Extremum (contrast 1, Area × Task × Extremum Type, $\breve{\beta}$ = 0.06, CI [-0.00, 0.12], Pr($\breve{\beta}$ > 0) = 0.97). These interactions are demonstrated in Figure 3; listeners in the England group were most likely to choose the 'second' response across all Task and Extremum Type conditions, but the ordering between the Belfast and Glasgow groups varied. For peak height discrimination, Belfast listeners were more likely to choose the 'second' response compared to Glasgow listeners; but for valley height discrimination, Belfast listeners were less likely to choose the 'second' responses. For prominence discrimination, Belfast listeners were more likely to choose the 'second' response compared to Glasgow listeners for both peaks and valleys. For prominence discrimination for valleys, the 'second' response probability was notably low for Glasgow listeners.

To summarize, the statistical analysis showed evidence for the effects of Area, Task, Plateau Duration and Step. There was also evidence for the interaction between Area, Task and Extremum Type.

II. PEAK AND VALLEY DATA: THE EFFECTS OF STEP

Bayesian logistic models were constructed for data split by Extremum Type (Peak, n = 3360; Valley, n = 3360). The model structure is provided in (2). (See Bürkner et al., 2024 for notation conventions.) The interaction term between Area, Task Type and Step was incorporated to examine how the response function slope varied depending on Area and Task Type. Because there was no evidence for interactions involving Plateau Duration (Section 3. C. I), they were not incorporated.

$$brm(response|trials(4) \sim 1 + area + taskType + extremumType + plateau + step + area * taskType * step + (1 \,|\, ID)) \tag{2}$$

The results are summarized in Table 3. (See Section 3. C. I for interpretations of the models.) Figure 4 shows response functions plotted with estimated probability.

Table 2. Output of the Bayesian regression model (LB: Lower Bound, UB: Upper Bound, c_1: contrast 1, c_2: contrast 2).

	$\breve{\beta}$	SE	LB	UB	Pr
Intercept	-2.34	0.05	-2.43	-2.25	1
Area_c1	-0.01	0.07	-0.14	0.12	0.58
Area_c2	0.13	0.06	0.01	0.25	0.98
Task	0.31	0.05	0.22	0.40	1
Extremum Type	-0.03	0.02	-0.07	0.02	0.88
Plateau Duration	-0.04	0.02	-0.08	0.00	0.97
Step	0.37	0.01	0.34	0.40	1
Area_c1 × Task	-0.03	0.07	-0.17	0.10	0.70
Area_c2 × Task	-0.06	0.06	-0.18	0.06	0.83
Area_c1 × Extremum Type	-0.07	0.03	-0.13	-0.01	0.99
Area_c2 × Extremum Type	0.00	0.03	-0.05	0.06	0.53
Task × Extremum Type	-0.04	0.02	-0.08	0.00	0.97
Area_c1 × Plateau	-0.00	0.03	-0.07	0.06	0.56
Area_c2 × Plateau	-0.00	0.03	-0.06	0.05	0.55
Task × Plateau	-0.03	0.02	-0.08	0.01	0.95
Extremum Type × Plateau	0.01	0.02	-0.03	0.05	0.62
Area_c1 × Task × Extremum Type	0.06	0.03	-0.00	0.12	0.97
Area_c2 × Task × Extremum Type	0.03	0.03	-0.03	0.09	0.86
Area_c1 × Task × Plateau	0.00	0.03	-0.06	0.06	0.54
Area_c2 × Task × Plateau	0.01	0.03	-0.04	0.07	0.68
Area_c1 × Extremum Type × Plateau	0.05	0.03	-0.01	0.10	0.95
Area_c2 × Extremum Type × Plateau	-0.02	0.03	-0.07	0.03	0.05

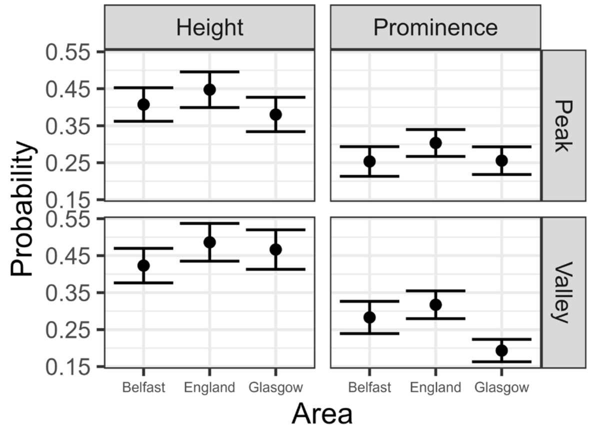

Figure 3. Estimated probability for the 'second' response from model (1) collapsed across Plateau Duration and Steps. Error bars show one standard error.

Table 3. Output of the Bayesian regression models for Peak and Valley data (LB: Lower Bound, UB: Upper Bound, c_1: contrast 1, c_2: contrast 2).

	Peak					**Valley**				
	$\breve{\beta}$	SE	LB	UB	Pr	$\breve{\beta}$	SE	LB	UB	Pr
Intercept	-2.54	0.06	-2.66	-2.43	1	-2.27	0.06	-2.40	-2.15	1
Area_c1	-0.03	0.08	-0.18	0.13	0.63	0.06	0.09	-0.12	0.23	0.73
Area_c2	0.19	0.07	0.04	0.33	0.99	0.14	0.08	-0.03	0.30	0.95
Task	0.28	0.06	0.17	0.39	1	0.33	0.06	0.21	0.46	1
Plateau Duration	-0.05	0.03	-0.10	0.01	0.94	-0.06	0.03	-0.11	-0.00	0.98
Step	0.61	0.03	0.56	0.66	1	0.15	0.02	0.11	0.19	1
Area_c1 × Task	-0.04	0.08	-0.20	0.12	0.68	-0.10	0.09	-0.28	0.07	0.88
Area_c2 × Task	0.03	0.07	-0.11	0.18	0.68	-0.08	0.08	-0.25	0.08	0.84
Area_c1 × Step	-0.07	0.04	-0.14	0.00	0.97	-0.04	0.03	-0.10	0.02	0.90
Area_c2 × Step	-0.06	0.03	-0.13	0.00	0.98	0.01	0.03	-0.04	0.07	0.67
Task × Step	-0.03	0.03	-0.08	0.02	0.86	0.12	0.02	0.07	0.16	1
Area_c1 × Task × Step	0.07	0.04	-0.00	0.14	0.97	0.05	0.03	-0.02	0.11	0.93
Area_c2 × Task × Step	-0.08	0.03	-0.14	-0.01	0.99	-0.06	0.03	-0.11	-0.00	0.98

For Peak, first, there was strong evidence for the Area effect for the second contrast ($\breve{\beta}$ = 0.19, CI [0.04, 0.33], Pr ($\breve{\beta} > 0$) = 0.99). That is, listeners in England were more likely to choose the 'second' response compared to those in Glasgow when Belfast was treated as the reference in the middle. Second, Task ($\breve{\beta}$ = 0.28) showed a reliable effect (CI [0.17, 0.39], (Pr($\breve{\beta} > 0$) = 1); the 'second' response probability was higher for Height compared to Prominence. Third, there was evidence for the strong Step effect ($\breve{\beta}$ = 0.61, CI [0.56, 0.66], Pr($\breve{\beta} > 0$) = 1); an increase in Step led to an increase in the 'second' responses. For interactions, there was evidence for the two-way interaction between Area and Step (contrast 1, $\breve{\beta}$ = -0.07, CI [-0.14, 0], Pr($\breve{\beta} < 0$) = 0.97; contrast 2, $\breve{\beta}$ = -0.06, CI [-0.13, 0], Pr($\breve{\beta} < 0$) = 0.98). The three-way interaction between Area, Task Type and Step was also reliable (contrast 1, $\breve{\beta}$ = 0.07, CI [0, 0.14], Pr($\breve{\beta} > 0$) = 0.97; contrast 2, $\breve{\beta}$ = -0.08, CI [-0.14, -0.01], Pr($\breve{\beta} < 0$) = 0.99). These interactions are shown in Figure 4(a); the England listener group showed the steepest response functions for both task types. In addition, they cross the 0.5 probability line on the y-axis at a point closer to the Step zero on the x-axis compared to the Belfast and Glasgow groups. If a response function crosses the 0.5

probability line at Step zero, this indicates that listeners perceived equal pitch height or prominence between the two extrema when they had acoustically the same *fo*. A positive Step value for the crossing point indicates that the two extrema were perceived as equal when the first extremum, compared to the second, was closer to the baseline with a smaller *fo* excursion. For height discrimination, the response functions for Belfast and Glasgow overlap. The response functions also indicate that the Belfast and Glasgow groups required a large *fo* excursion size of the second peak compared to the first peak for their perceptual equivalence in height. On the other hand, for prominence discrimination, Belfast listeners' response function is flatter, and they were more likely to choose the 'first' response compared to the other two groups. While all response functions cross the 0.5 probability line at a positive Step value, the Belfast group seems to have required a particularly large *fo* excursion size for the perceptual equivalence.

For Valley, first, there was a reliable positive coefficient for the Task ($\breve{\beta}$ = 0.33, CI [0.21, 0.46], Pr($\breve{\beta}$ > 0) = 1); the 'second' response probability was higher for Height compared to Prominence. Second, there was evidence for the Plateau effect ($\breve{\beta}$ = -0.06, CI [-0.11, 0], Pr($\breve{\beta}$ < 0) = 0.98). The shorter plateau (25 ms) associated with the second valley lowered the 'second' response probability overall. Third, the Step effect was reliable ($\breve{\beta}$ = 0.15, CI [0.11, 0.19], Pr($\breve{\beta}$ > 0) = 1). For interactions, there was evidence for the two-way interaction between Task and Step ($\breve{\beta}$ = 0.12, CI [0.07, 0.16], Pr($\breve{\beta}$ < 0) = 1). There was also evidence for the three-way interaction between Area, Task Type and Step (contrast 2, $\breve{\beta}$ = -0.06, CI [-0.11, 0], Pr($\breve{\beta}$ < 0) = 0.98). Figure 4(b) shows a striking difference between the height and prominence conditions. For height discrimination, the listeners in Belfast and Glasgow groups required a larger f_o excursion size of the second valley for the perceptual equivalence of the two valleys compared to the English. However, the response functions for the three area groups overlap to some extent. On the other hand, for prominence discrimination, the response functions for all area groups are flat. The Belfast group's response function shows a negative slope while the Glasgow group seemed to have been strongly biased to choose the 'first' response.

To summarize, for Peak, the statistical analysis showed evidence for the effects of Area, Task and Step. There was also evidence for the interaction between Area, Task and Step. For Valley, there was evidence for the effects of Area, Plateau Duration, Task and Step. There was also evidence for the interaction between Area, Task and Step.

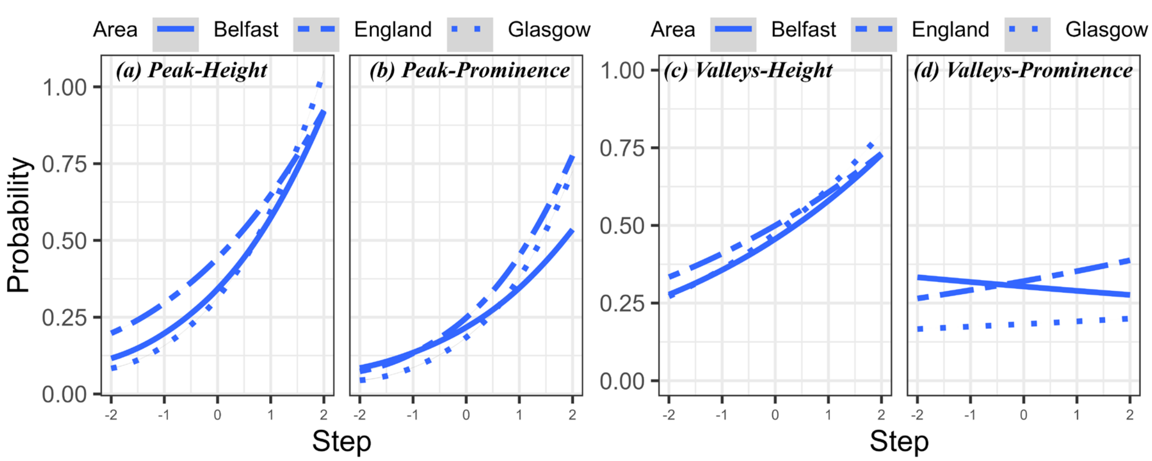

Figure 4. Probability for (a) Peak-Height, (b) Peak-Prominence, (c) Valleys-Height and (d) Valley-Prominence. Data collapsed over Plateau Duration.

4. DISCUSSION

As expected, a long f_o plateau had an effect of making a peak sound higher, a valley sound lower and increasing the extremum's perceived prominence. The type of the experimental task also had a reliable effect; listeners were less likely to choose the 'second' response for prominence discrimination compared to height discrimination, and they required a larger f_o excursion size for prominence discrimination. Importantly, the results highlight the effect of listeners' linguistic experience on judging prominence. Overall, Belfast and Glasgow listeners showed reduced prominence discrimination compared to English listeners.

The first hypothesis was that for judging peak height, there would be no cross-varietal differences. This hypothesis was not supported. While the response functions from the raw data in Fig. 2a for the three areas overlapped, the area predictor had a reliable effect in the statistical analysis, interacting with the step predictor (Section 3. C. II). Listeners in England showed heightened discrimination compared to Belfast and Glasgow listeners (Fig. 4a).

Second, for judging peak prominence, as hypothesized, the Belfast and Glasgow listeners showed relatively reduced discrimination. This was shown by their less steep response functions compared to listeners in England (Fig. 4b). In particular, Belfast listeners were least likely to identify the second peak as prominent. An unexpected finding is that listeners seem to have perceived the two peaks equal in prominence when the second peak was physically higher than the first; in both Figures 2(a) and 4(b), the response functions cross the 0.5 point on the y-axis at the step higher than zero. This contrasts with the well-known previous finding that two peaks are perceived equivalent in prominence when the second peak is acoustically lower than the first (Pierrehumbert, 1979). It is postulated that listeners' judgement is influenced by their expectation on declination, i.e., gradual lowering of f_o peaks over an utterance, which is commonly observed in speech (see Ladd, 2008, Chap. 2 for declination). However, listeners' expectations seem to be established based on the precise acoustic properties of the listening stimuli. Throughout the present experiment, the first peak's f_o height in the listening stimuli was kept constant with a normal f_o excursion size for the speaker's high accent. Therefore, the prominent first peak repeatedly presented in the experiment may have led listeners to establish a high threshold of the f_o excursion for perceiving the second peak to be equivalent or more prominent than the first, instead of expecting declination.

Third, for judging valley height, we expected the Belfast and Glasgow listeners to show heightened discrimination compared to English listeners. This expectation was not borne out (Fig. 4c).

Fourth, for judging valley prominence, the hypothesis that the Belfast and Glasgow listeners would show heightened discrimination was not supported. Listeners in all area groups were biased towards perceiving the first valley prominent (Fig. 4d). In particular, the listeners in Glasgow showed a strong bias towards perceiving the first valley prominent.

Taking the findings together, the area groups showed more notable differences for prominence judgement than for pitch height judgement. In general, Glasgow and Belfast listeners showed reduced prominence discrimination compared to English listeners. The reduced prominence discrimination may be related to the shallow 'prominence gradients' in the language varieties they have been exposed to. Belfast and Glaswegian English may not contrast stressed/accented and unstressed/unaccented syllables as in SSBE or some varieties spoken in England, so the Belfast and Glasgow listeners may not have interpreted the acoustic variations in the experiment in the same way as English listeners did.

For the unexpected findings for judging valley prominence, perhaps the valleys in the present study were not in the appropriate forms to be perceived as low prominence by Belfast and Glasgow listeners. In the experimental stimuli, the plateau was aligned at the end of the stressed vowel. That is, for the valleys, the stressed vowel contained a pitch fall followed by a relatively short plateau (25 ms or 50 ms). Belfast and Glasgow listeners may have needed more salient cues, such as a longer f_o plateau, a plateau aligned at the earlier point of the vowel, or a more salient rise from the f_o minimum in the stressed vowel. For instance, the early alignment of a f_o valley, e.g., before the stressed syllable, would bring the rise portion early in the stressed vowel, as it is often the case in Glaswegian English (Mayo, 1996; Mayo et al., 1997). Belfast listeners may need a strong f_o cue, given that f_o is a heavily weighted parameter in their production of prominent syllables. Another potential cue to low prominence is the acoustic events in the post-stressed syllable, such as a rise or a rise followed by a plateau. Finally, one reason for Belfast listeners' strong bias perceiving the first valley more prominent than the second could be related to how they enhance low prominence in their speech. In yes-no questions, Belfast listeners may interpret an utterance-final high-pitched, not low-pitched, valley as prominent or emphatic. This is possible if Belfast listeners expect declination of the successive f_o valleys in neutrally spoken utterances, while emphasis could be marked by raising the f_o valley. Figure 5 shows an example utterance 'are you growing limes or lemons?' spoken by a Belfast English speaker. Here, each low accent (marked by *) is followed by a salient rise f_o and gradual lowering of the valley-shaped accents is observed. The relationship between declination in speech, realization of valley accents and their perception requires further investigation.

5. CONCLUSIONS

There was no evidence that listeners' frequent exposure to low accents enhances their discrimination of pitch height or prominence of f_o valleys. However, for judging prominence, listeners seemed to interpret the acoustic cues in line with the prosodic properties of their language variety. The present findings show that even for English which is considered a typical stress-based language, the definition and perception of prominence diverges across varieties (also see Barnes et al., 2023, reporting absence of the 'plateau effect' for valleys for speakers of American English). The prominence-lending or -cueing function of pitch in listeners' native language may shape their interpretation of pitch movements. These findings form a useful basis for further investigation on how different phonetic and phonological aspects of native language affects perception of intonation.

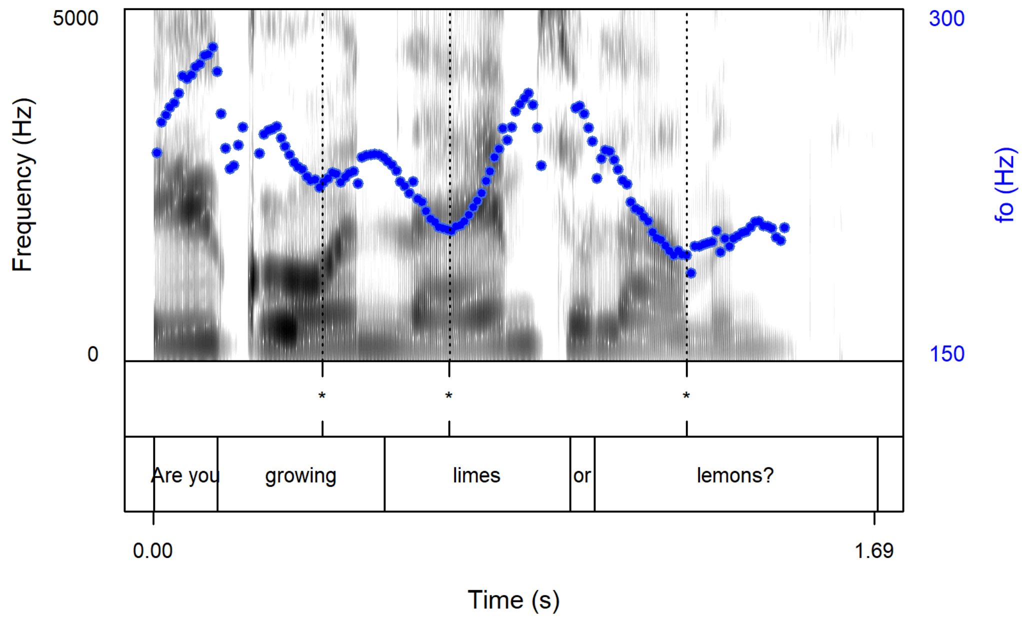

Figure 5. F_o track and spectrogram (0-8kHz) of a Belfast English utterance from the IViE corpus (Nolan and Post, 2013). The minimum f_o for a valley-shaped accent is marked by an asterisk (*).

ACKNOWLEDGMENTS

This research was supported by the Department for Business, Energy and Industrial Strategy, UK (the British Academy\Leverhulme Small Research Grant, SRG19\190109).

REFERENCES

Aitken, A. J. (**1981**). "The Scottish Vowel Length Rule," In M. Benskin and M. L. Samuels (Eds.), So Many People Languages and Tongues: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh, pp. 131–157, Edinburgh: Middle English Dialect Project.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (**2019**). "Gorilla in our midst: An online behavioral experiment builder," Behav. Res. Methods, **52**(1), 388–407.

Barnes, J., Lee, C. K., Brugos, A., ShattBtesuck-Hufnagel, S., and Veilleux, N. (**2023**). "Sometimes low really is just the opposite of high: perception of low F0 targets by tone and non-tone language speakers," in Proceedings of the 20th ICPhS, Guarant International, Prague, Czech Republic, pp.1295–1299.

Baumann, S., and Winter, B. (**2018**). "What makes a word prominent?: predicting untrained German listeners' perceptual judgments," J. Phon., **70**, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004

Bishop, J., Kuo, G., and Kim, B. (**2020**). "Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription," J. Phon, **82**, 100977. https://doi.org/10.1016/j.wocn.2020.100977

Boersma, P. and Weenink, D. (**2020**). Praat: Doing phonetics by computer [Computer program]. available at http://www.praat.org/.

Bürkner, P.-C. (**2017**). “brms: An R package for Bayesian multilevel models using Stan,” J. Stat. Softw., **80**(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H. S., Weber, F., Vehtari, A., Ben-Shachar, M. S., Rabel, H., Mills, S. C., Wild, S., and Popov, V. (**2024**). “Package ‘brms’ (March 20, 2024)”, available at: https://cran.r-project.org/web/packages/brms/brms.pdf (last access 31 May 2024).

Cole, J., Mo, Y. and Hasegawa-Johnson, M. (**2010**). “Signal-based and expectation-based factors in the perception of prosodic prominence,” Lab. Phon., **1**, 425–452. https://doi.org/10.1515/labphon.2010.022

Cruttenden, A., (**1986**). Intonation (1st ed.), Cambridge, UK: Cambridge University Press.

Cruttenden, A., (**1997**). Intonation (2nd ed.), Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139166973

Cumming, R. E., (**2011**). “The language-specific interdependence of tonal and durational cues in perceived rhythmicity,” Phonetica, **68**, 1–25. https://doi.org/10.1159/000327223

Evans, J. P., (**2015**). “High is not just the opposite of Low,” J. Phon,, **51**, 1–5. https://doi.org/10.1016/j.wocn.2015.05.001

Grabe, E., (**2002**). “Variation adds to prosodic typology,” Speech Prosody 2002, Aix-en-Provence, France, 2002, pp. 127–132.

Gussenhoven, G. (**2004**). The Phonology of Tone and Intonation, Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Hsu, C.-H., Evans, J. P., and Lee, C.-Y. (**2015**). “Brain responses to spoken F0 changes: Is H special?,” J. Phon., **51**, 82–92. https://doi.org/10.1016/j.wocn.2015.02.003

Jarman, E., and Cruttenden, A. (**1976**). “Belfast intonation and the myth of the fall,” JIPA, **6**, 4–12. https://doi.org/10.1017/S0025100300001432

Jeon, H.-S. and Heinrich, A. (**2022a**). “Perceptual asymmetry between pitch peaks and valleys,” Speech Commun., **140**, 109–127. https://doi.org/10.1016/j.specom.2022.04.001

Jeon, H.-S. and Heinrich, A. (**2022b**). “Perception of pitch height and prominence by old and young listeners,” Speech Prosody 2022, pp. 654–658, Lisbon, Portugal, 23–26 May 2022. https://doi.org/10.21437/SpeechProsody.2022-133

Jeon, H.-S. and Nichols, S. (**2022**). “Investigating prosodic variation in British English varieties using *ProPer*,” in Proceedings of Interspeech 2022, Incheon, Republic Korea, September 2022, pp. 1313–1317. https://doi.org/10.21437/Interspeech.2022-792

Knight, R.-A. (**2008**). “The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. plateaux,” Lang. Speech, **51**(3), pp. 223–244. https://doi.org/10.1177/0023830908098541

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). “Loudness predicts prominence: Fundamental frequency lends little,” J. Acoust. Soc. Am., **118**, 1038–1054. https://doi.org/10.1121/1.1923349

Ladd, D. R. (**2008**). Intonational Phonology (2nd Edition). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lieberman, P. (**1967**). Intonation, Perception and Language, Cambridge, MA: MIT Press.

Lialiou, M., Grice, M., Röhr, C., and Schumacher, P. B. (**2024**). “Auditory processing of intonational rises and falls in German: Rises are special in attention orienting,” J. Cogn. Neurosci, **36**(6), 1099–1122. https://doi.org/10.1162/jocn_a_02129

Lowry, O. (**2002**). “The stylistic variation of nuclear patterns in Belfast English,” JIPA, **32**, 33–42. https://doi.org/10.1017/S0025100302000130

Mayo, C. J. (**1996**). “Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI,” unpublished PhD thesis, University of Edinburgh.

Mayo, C., Aylett, M., and Ladd, D. R. (**1997**). “Prosodic transcription of Glasgow English: An evaluation study of GlaToBI,” In Botinis, G. Kouroupetroglou, and G. Carayiannis (Eds.), Proceedings of the ESCA Tutorial and Workshop on Intonation: Theory, Models and Applications, Athens, Greece, pp. 231–234.

Moore, B. C. J. (**2012**). An Introduction to the Psychology of Hearing, 6th edition, Bingley, UK: Emerald Group Publishing Limited.

Nolan, F. and Post, B. (**2013**). The IViE Corpus. In J. Durand, U. Gut and G. Kristoffersen (ed.), The Oxford Handbook of Corpus Phonology, pp. 475–485, Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.013.025

Puggaard-Rode, R. (**2024**), “praatpicture. A library for making flexible Praat Picture-style figures in R,” In Fougeron, C. and Perrier, P. (Eds.), Proceedings of the 13th International Seminar on Speech Production, Autrans, France, pp. 115–118.

Pierrehumbert, J. (**1979**). “The perception of fundamental frequency declination,” J. Acoust. Soc. Am., **66**(2), 363–369. https://doi.org/10.1121/1.383670

Rahilly, J. (**1997**). “Aspects of prosody in Hiberno-English: The case of Belfast,”, in J. L. Kallen (Ed.), Focus on Ireland, pp. 109–132, John Benjamins Publishing Company. https://doi.org/10.1075/veaw.g21.09rah

Rathcke, T. V. and Stuart-Smith, J. H. (**2016**). “On the tail of the Scottish Vowel Length Rule in Glasgow,” Lang. Speech, **59**, 404–430. ttps://doi.org/10.1177/0023830915611428

R Core Team. (**2023**). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team. (**2020**). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. URL http://www.rstudio.com/.

Scobbie, J. M., Hewlett, N., and Turk, A. (**1999**). "Standard English in Edinburgh and Glasgow: The Scottish Vowel Length Rule revisited," In P. Foulkes and G. Docherty (Eds.), Urban Voices: Accent Studies in the British Isles, pp. 230–245, London: Hodder.

Smith, R. and T. Rathcke, T. (**2020**). "Dialectal phonology constrains the phonetics of prominence," J. Phon., **78**, 100934. https://doi.org/10.1016/j.wocn.2019.100934

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (**2019**). "Welcome to the tidyverse," J. of Open Source Softw. **4**(43), 1686. https://doi.org/10.21105/joss.01686

Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (**2017**). "Headphone screening to facilitate web-based auditory experiments," Atten. Percept. Psychophys., **79**(7) 2064–2072. https://doi.org/10.3758/s13414-017-1361-2