Interspeech 2018
2-6 September 2018, Hyderabad

Weighting of Coda Voicing Cues: Glottalisation and Vowel Duration

Joshua Penney[1,2], Felicity Cox[1,2], Anita Szakay[1,2]

[1]Centre for Language Sciences, Department of Linguistics, Macquarie University, Australia
[2]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Australia
joshua.penney@mq.edu.au, felicity.cox@mq.edu.au, anita.szakay@mq.edu.au

Abstract

Recent research suggests that a trading relationship may exist in speech production between vowel duration and glottalisation as cues to coda stop voicing in Australian English. Younger speakers have been shown to use glottalisation to signal voicelessness more than older speakers who instead make greater use of vowel duration. This suggests a sound change in progress for the voicing cues. In addition, the vowel duration cue to voicing is greater in inherently long vowel contexts compared to inherently short vowel contexts. We report on a perceptual study designed to examine whether the weighting of these two cues found in production is replicated in perception.

Older and younger listeners were presented with audio stimuli co-varying in vowel duration and glottalisation. In accord with findings from production, the vowel duration cue was weaker for contexts containing inherently short vowels than for those containing inherently long vowels. Complementarily, glottalisation had a stronger effect on the perception of coda voicelessness in inherently short vowel contexts. Older and younger listeners did not differ in their use of glottalisation as a perceptual cue to voicelessness despite previously identified age differences in production. This finding raises questions about the link between perception and production in sound change.

Index Terms: Glottalisation, vowel duration, coda voicing, coda stops, cue weighting, Australian English, sound change.

1. Introduction

It is well known that the phonological distinction between voiced and voiceless stops in English is often not cued by the presence of phonetic voicing; while voiced stops may sometimes contain voicing in the closure period, voicing is often weak or absent [1, 2, 3]. Voiceless coda stops can be cued by a combination of shorter preceding vowel duration, longer closure duration, higher F0 at preceding vowel offset, and by the presence of glottalisation (i.e. the addition of glottal constriction to the oral coda stop resulting in laryngealised phonation on the preceding vowel [4]). Voiced coda stops, in contrast, may be cued by longer preceding vowel duration, shorter closure duration, lower F0, and by F1 cutback, in addition to the potential presence of voicing during the closure period [2, 5, 6, 7, 8].

Glottalisation functions as a cue to coda voicelessness in Australian English (AusE) [9], as is it does in a number of other English varieties [4, 10, 11, 12]. In an apparent time analysis of speech production, [9] found that younger AusE speakers utilised glottalisation to cue coda voicelessness more frequently than older speakers suggesting that the use of glottalisation is a recent change. Moreover, in addition to using glottalisation more than older speakers, younger speakers also appeared to make less use of preceding vowel duration to cue coda stop voicing than older speakers [9]. These results may indicate that, in production, a trading relationship exists between vowel duration and glottalisation as cues to coda stop voicing. Furthermore, both younger and older speakers used the vowel duration cue most robustly for high vowels, which were less likely to occur in conjunction with glottalisation [9].

Recent research from American English suggests that glottalisation may hinder the perception of voiced codas. In an eye-tracking experiment, [13] found that listeners were slower to identify words with voiced codas (/b, d/) when they occurred in conjunction with glottalisation. Words ending in voiceless codas (/p, t/), on the other hand, showed no delay in identification when paired with glottalisation. [13] therefore suggest that listeners associate glottalisation with voiceless coda stops but not with voiced coda stops. However, [13] found that perception of voiceless stops was not generally improved by the presence of glottalisation. This suggests that other cues were able to successfully cue voicelessness, raising questions about the characteristics of the various cues and their weighting.

In this study we target two major cues (glottalisation and vowel duration) to examine how they are used individually and in combination in the perception of coda stop voicing in AusE. A preliminary analysis of younger listeners suggested that glottalisation may promote the percept of voicelessness, even in the presence of prolonged preceding vowel duration [14]. These initial findings are extended here to explore whether older listeners, who do not make as much use of glottalisation to cue coda voicelessness in production, similarly do not utilise glottalisation as much as young people in perception.

We are also interested in how glottalisation interacts with the preceding vowel duration cue. Inherently short vowels have been found to exhibit smaller coda voicing related durational differences compared to inherently long vowels [9, 15], which may suggest that preceding vowel duration is a less salient cue to coda voicing in inherently short vowel contexts. Correspondingly, younger listeners in [9] glottalised at higher rates in inherently short vowel contexts than in inherently long vowel contexts. Therefore, listeners may reliably be able to use vowel duration as a differentiating cue in inherently long vowel contexts whereas in inherently short vowel contexts glottalisation may confer some additional benefit.

Previous studies suggest links between perception and production, such that speaker/listeners who produce innovative variants are also sensitive to these in perception, whereas those with more conservative productions are more likely to use conservative cues [16, 17, 18] in perception. On the other hand, perception has been suggested to lead

10.21437/Interspeech.2018-1677

production in some sound changes [19, 20]; thus some listeners may be sensitive to a cue in perception without it being (fully) implemented in their production. In this case, it may be possible that older listeners are sensitive to glottalisation in perception despite employing it less in production.

Here we report on a perceptual task in which listeners were presented with audio stimuli co-varying in vowel duration and the presence of glottalisation. The results focus on the high vowel context to provide the optimal environment for analysing the utility of the vowel duration cue. [9] found that speakers utlised vowel duration to cue coda voicing more in high vowel contexts than in low vowel contexts. We examine here inherently long and short high vowels.

We hypothesise that:

- listeners will use duration to cue coda voicing, with longer vowel durations cueing voiced codas and shorter vowel durations cueing voiceless codas;
- the presence of glottalisation will promote the perception of coda voicelessness;
- preceding vowel duration will have a weaker effect on coda voicing for inherently short vowels, given the smaller vowel duration differences in production between voicing contexts for inherently short vowels than for inherently long vowels;
- if a trading relationship exists in perception between preceding vowel duration and glottalisation, the effect of glottalisation in promoting the percept of voicelessness will be stronger for inherently short vowels than for inherently long vowels, based on the finding in production that inherently long vowels make greater use of vowel duration differences between coda voicing contexts than inherently short vowels;
- the effect of glottalisation in promoting voicelessness will be stronger in younger listeners than in older listeners, given that in production younger speakers use glottalisation more whereas older speakers utilise preceding vowel duration more.

2. Methods

2.1. Participants

77 listeners took part in this perceptual task. Two groups of participants were recruited: a younger group, aged between 18 and 36 (n = 46; female: 38; male: 8; mean age: 21), and an older group, aged 50+ years (n = 31; female: 22; male: 9; mean age: 61). All participants were AusE native speakers, who were born in Australia (one participant was born overseas but came to Australia as a baby) and all had completed both primary and secondary education in Australia. All reported normal hearing.

2.2. Task

Participants took part in a binary forced choice word identification task in which minimal pairs differing in coda voicing (e.g. *bead/beat*) were orthographically displayed on an Apple Macbook computer monitor and audio stimuli were presented through Sennheiser HD 380 pro headphones. For each token, a fixation cross was displayed for 600ms in the centre of the screen. The orthographic representations of a minimal pair were then displayed, followed by the presentation of a single word audio stimulus after 500ms. For each stimulus item participants identified the word they heard by key press response. Once the participant responded, the orthographically displayed words disappeared and the next cycle began. Participants also took part in a production task immediately prior to the perception task, in which (among other items) all words included in the perceptual task were produced.

Each participant took part in a session in which 648 tokens were presented in two blocks. All tokens were of the form /bVC/, where /V/ was one of the vowels /iː, ɪ, ɐː, ɐ/. A subset of this data will be examined here, comprising a total of 4158 responses (54 responses per participant) for the two vowels: inherently long /iː/ and inherently short /ɪ/. Due to the length of the task, listeners were allocated either the inherently long vowel or inherently short vowel condition. Listeners were presented with randomised stimuli from a continuum where vowel duration increased in nine equally spaced steps as described below. For each continuum step, a non-glottalised token and a glottalised token were randomly presented. Three randomised repetitions of each item were included.

2.3. Stimuli

The stimuli for this task were created from naturally produced tokens, recorded in a sound treated recording studio in the Department of Linguistics at Macquarie University, using an AKG C535 EB microphone, Cooledit audio capture software via an M-Audio delta66 soundcard to a Pentium 4 PC at 44.1kHz sampling rate. The source tokens *bead* and *bid* were produced phrase medially in carrier sentences by a phonetically trained, non-rhotic female native AusE speaker, aged 25 years, with a modally voiced falling intonation pattern. These words were selected to enable a comparison between inherently long and inherently short vowels in the same phonetic context. The speaker also produced sustained instances of the vowels from these source tokens (i.e., /iː, ɪ/) with creaky voice.

The source tokens were manipulated in Praat [21] to remove cues to the voicing status of the final coda stop. As a first step, coda stop releases (including any aspiration) were excised and replaced with an ambiguous burst, taken from a low intensity voiced stop in an unstressed syllable. This burst was shown in our pilot experiments not to produce consistently either the percept of a voiced or a voiceless coda stop. The same ambiguous burst was used for both of the source tokens. In order to remove potential voicing in the closure, the closure periods in both source tokens were replaced with silence. In addition, in both source tokens the end of the vowel was truncated at the point immediately prior to F1 formant transition into the following consonant (i.e. to remove the F1 voicing cue). Intensity contours were then manipulated to ensure consistency of amplitude between source tokens.

Using the PSOLA function in Praat [21], nine equally spaced vowel duration steps were then created for each source token. The minimum and maximum durational values for each vowel within the /iː, ɪ/ continua were based on the mean values for young female speakers from Sydney in both the voiced and voiceless coda contexts, taken from [22] and reported in [9]. The shortest vowel duration step was two standard deviations shorter than the reported mean duration of the relevant vowel in the voiceless coda context, and the longest vowel duration step was two standard deviations greater than the reported mean duration of the vowel in the voiced coda context. This ensured that the first step was substantially shorter than the mean duration in the voiceless coda context, and that the ninth step was substantially longer than the mean duration in the voiced coda context. The coda closure duration for all of the steps in each continuum was the mean of the coda closure durations for that vowel (as reported in [9]) in both voiceless and voiced coda contexts combined. Table 1 lists the durations of the first and ninth step stimulus and the step size for each vowel examined here as well as the durations of the respective coda closure periods.

Table 1: *Vowel duration of shortest and longest steps per vowel continua.*

Vowel	**First step duration (ms)**	**Ninth step duration (ms)**	**Step size (ms)**	**Coda closure duration (ms)**
/iː/	104	340	29.5	82
/ɪ/	65	166	12.6	94

The F0 of each step was then manipulated to ensure consistency of pitch across tokens and continua and to ensure there was no F0 cue to coda voicing present. F0 at vowel onset was set at 265hz, and at vowel offset was set at 203hz, calculated according to the means for the speaker in voiced and voiceless coda contexts combined. Finally, the intensity of all tokens was scaled to 70 dB.

We then created a second set of these manipulated continua that were identical to the first set, and manually spliced glottalisation into the final portion of the vowels. For each continuum step, the final portion of the vowel was replaced with natural glottalisation taken from the matched sustained vowels produced with creaky voice as mentioned above. In line with proportions of glottalisation reported in [9], the final 25% of each long vowel step and the final 35% of each short vowel step was glottalised. Figure 1 illustrates the ninth step (i.e. longest vowel duration) of the *bead* continuum in both the non-glottalised and the glottalised conditions.

2.4. Statistical analysis

In order to analyse the effect of glottalisation and vowel duration on the perception of coda voicing we fitted a mixed effects logistic regression (GLMER) using [23] in [24]. The dependent variable was the response of voiced or voiceless coda stop. Fixed factors were vowel duration (step), condition (glottalised/non-glottalised), inherent vowel length (short/long), and age group (older/younger). We also included interaction terms for vowel duration by inherent vowel length, condition by inherent vowel length, and condition by age group. Participant was included as a random factor.

3. Results and Discussion

Figure 2 displays the proportion of voiced coda responses for the inherently short and inherently long vowel continua for older and younger listeners. As can be seen, for source *bead* and *bid* the proportion of voiced coda stop responses increases in line with an increase in vowel duration in both glottalised and non-glottalised conditions in each age group. Unsurprisingly, the results of the logistic mixed effects model showed a significant effect for vowel duration (Est.=-0.716; SE=0.032; z=-22.479; p< 0.0001), thereby confirming our first hypothesis that AusE listeners are sensitive to preceding vowel duration as a cue coda stop voicing: all else being equal, shorter preceding vowels cue perception of voiceless coda stops, whereas longer preceding vowels cue perception of voiced coda stops.

The model also showed a significant result for condition (Est.=2.166; SE=0.173; z=12.495; p<0.0001), demonstrating that the presence of glottalisation affects listeners' responses. As Figure 2 illustrates, in each continuum and for both age groups the vowel duration effect is reduced in the glottalised compared to the non-glottalised condition, with a lower proportion of voiced responses returned when glottalisation is present as opposed to when it is absent. This provides strong support for our second hypothesis that glottalisation would promote the perception of coda voicelessness suggesting that AusE listeners associate glottalisation with voiceless but not voiced codas, in line with [13].

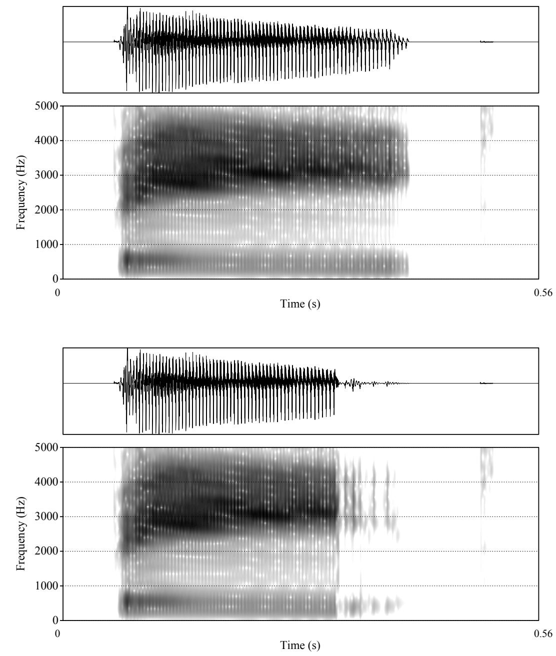

Figure 1: *Waveform and spectrogram of ninth step (longest vowel duration) stimuli in source 'bead' continuum. Upper panel shows non-glottalised condition; lower panel shows glottalised condition.*

We found a significant interaction between vowel duration and inherent vowel length (Est.=0.149; SE=0.043; z=3.479; p= 0.0005). Figure 2 shows that in the inherently short vowel context (source *bid*) the slope of the responses identified as voiced as vowel duration increases is more gradual compared to in the inherently long vowel (source *bead*) context. This is likely because inherently short vowels in AusE do not vary in duration between voiced and voiceless coda contexts to the same extent as inherently long vowels, and indicates that vowel duration is a weaker cue to coda voicing in the inherently short vowel context, in line with our third hypothesis. The model also showed a significant interaction between condition and inherent vowel length (Est.=0.855; SE=0.198; z=4.309; p<0.0001), with the presence of glottalisation exhibiting stronger promotion of voicelessness

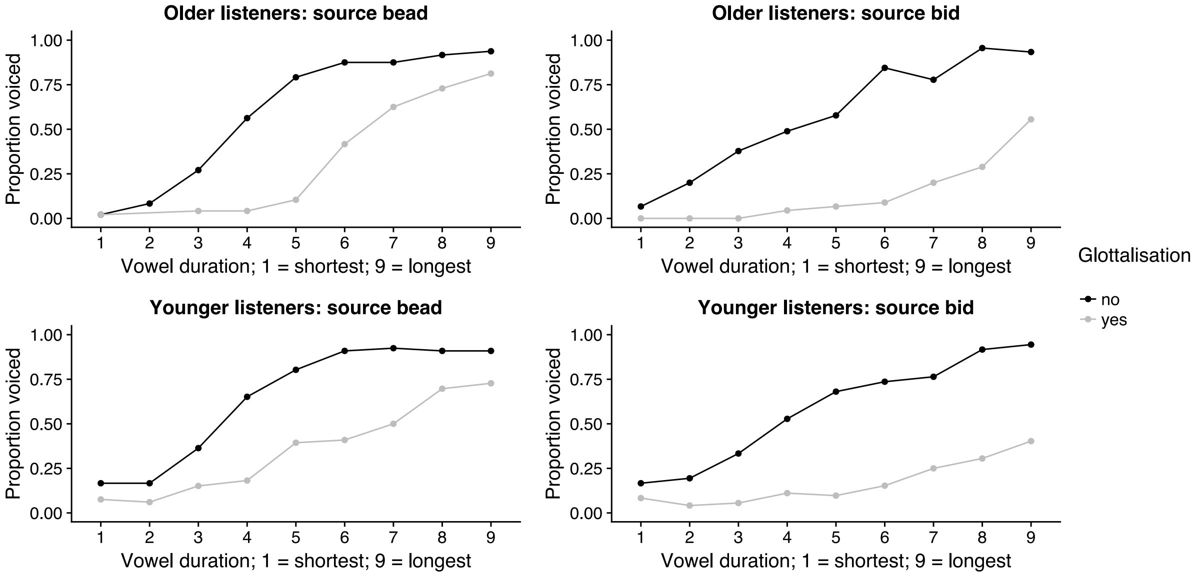

Figure 2: *Proportion of voiced coda responses for older and younger listeners in non-glottalised and glottalised conditions.*

in the inherently short vowel context compared to the inherently long vowel context. Post-hoc comparisons (with Tukey HSD correction) confirm that in the non-glottalised condition there is no significant effect of inherent vowel length (p=0.2114), whereas in the glottalised condition the difference between the vowel contexts is significant (p<0.0001). As Figure 2 illustrates, in the glottalised condition of the inherently long vowel continuum (source *bead*) the effect of vowel duration is such that the majority of tokens were selected as having voiced codas in the longest vowel duration steps, although this is nevertheless lower than in the non-glottalised condition. In contrast, in the short vowel continuum approximately half of the responses were for voiceless coda stops at even the longest vowel duration steps. This indicates that glottalisation has a stronger facilitative effect on the perception of voicelessness for inherently short vowels than for inherently long vowels, confirming our fourth hypothesis. Considering this result, it is possible that [13] may have found improved perception of voicelessness with glottalisation had they considered the potential impact of inherent vowel length in their study.

Finally, the model showed no significant interaction between condition and age group, suggesting that glottalisation has the same effect of promoting coda voicelessness for both older and younger listeners in this high vowel context. We hypothesised that younger listeners would utilise glottalisation more than older listeners in perception, as they have been shown to glottalise more and employ preceding vowel duration less than older speakers in production [9]. However, this hypothesis was not supported by the perception data. This suggests that older listeners are using glottalisation perceptually in the same way as younger listeners to cue coda voicelessness. A possible explanation for this apparent anomaly may lie in the fact that glottalisation appears to be a recent change to AusE, with the change being led by younger speakers [9]. As discussed above, some models of sound change suggest that changes in perception precede changes in production [19, 20]. With this is mind, it is possible that older AusE listeners are aware of glottalisation as a cue to coda voicelessness in perception, despite lagging behind the younger speakers in implementing this change into their own production. Younger listeners, on the other hand, are leading the change and hence display a more innovative utilisation of glottalisation that is observed both in perception and in production.

4. Conclusion

We found that in perception both younger and older AusE listeners used preceding vowel duration as a cue to the voicing of coda stops, but the addition of glottalisation had the effect of promoting the perception of coda voicelessness. We also found that the vowel duration cue was weaker for inherently short vowels compared to inherently long vowels, and that glottalisation had a stronger effect on the perception of coda voicing in inherently short vowel contexts. These results support a perceptual trading relationship between glottalisation and vowel duration as cues to coda stop voicing. The facilitative effect of glottalisation on the perception of voicelessness has been shown to affect both younger and older listeners in the same way, despite mismatches in production, raising questions about the relationship between perception and production in sound change. Future work will explore whether these findings apply to a greater range of contexts and speaker/listeners.

5. Acknowledgements

This work was partially supported by an Australian Government Research Training Program scholarship and an Endeavour Research Fellowship to the first author. We would like to thank Louise Ratko, Shuting Liu, Ulrich Reubold, Maria Paola Bissiri, and members of the MQ phonetics lab for their comments, suggestions and assistance, particularly regarding stimuli preparation.

6. References

[1] G. Docherty, *The timing of voicing in British English obstruents*. Berlin, Germany: Foris, 1992.

[2] L. Lisker, "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," *Language and Speech, vol. 29*, pp. 3–11, 1986.
[3] J. E. Flege and W. S. Brown, "The voicing contrast between /p/ and /b/ as a function of stress and position in utterance," *Journal of Phonetics, vol. 10*, pp. 334–345, 1982.
[4] M. K. Huffman, "Segmental and prosodic effects on coda glottalization," *Journal of Phonetics, vol. 33*, pp. 335–362, 2005.
[5] J. Kingston and R. L. Diehl, "Phonetic knowledge," *Language, vol. 70*, pp. 419–454, 1994.
[6] L. J. Raphael, "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," *Journal of the Acoustical Society of America, vol. 51*, pp. 1296–1303, 1972.
[7] J. Y. Song, K. Demuth, and S. Shattuck-Hufnagel, "The development of acoustic cues to coda contrasts in young children learning American English," *Journal of the Acoustical Society of America, vol. 131*, pp. 3036–3050, 2012.
[8] R. Wright, "A review of perceptual cues and cue robustness," in *Phonetically-based phonology,* B. Hayes, R. Kirchner, and D. Steriade, Eds. New York: Cambridge University Press, 2004, pp. 34–57.
[9] J. Penney, F. Cox, K. Miles, and S. Palethorpe, "Glottalisation as a cue to coda stop voicing in Australian English," *Journal of Phonetics*, *vol. 66*, pp. 161–184, 2018.
[10] O. B. Gordeeva and J. M. Scobbie, "A phonetically versatile contrast: Pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality," *Journal of the International Phonetic Association, vol. 43,* pp. 249–271, 2013.
[11] P. J. Roach, "Glottalization of English /p/, /t/, /k/ and /ʧ/ – a reexamination," *Journal of the International Phonetic Association*, *vol. 3*, pp. 10–21, 1973.
[12] J. Pierrehumbert, "Prosodic effects on glottal allophones," in *Vocal fold physiology: Voice quality control,* O. Fujimura and M. Hirano, Eds. San Diego: Singular, 1995, pp. 39–60.
[13] A. Chong and M. Garellek, "Online perception of glottalized coda stops in American English," *Laboratory Phonology*, vol. 9, pp.1–24, 2018.
[14] J. Penney, F. Cox, A. Szakay, "Weighting of vowel duration and glottalisation in the perception of coda stop voicing," presented at *Labphon 16, Lisbon, Portugal*, 2018.
[15] F. Cox and S. Palethorpe, "Timing differences in the VC rhyme of Standard Australian English and Lebanese Australian English," in *Proc. 17th ICPhS, Hong Kong*, 2011, pp. 528–531.
[16] A. W. Coetzee, P. S. Beddor, K. Shedden, W. Styler, and D Wissing, "Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis," *Journal of Phonetics, vol. 66*, pp. 185-216, 2018.
[17] J. Harrington, F. Kleber, and U. Reubold, "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," *Journal of the Acoustical Society of America, vol*. 123, pp. 2825–2835, 2008.
[18] P. S. Beddor, "The relation between language users' perception and production repertoires," in *Proc. 18th ICPhS, Glasgow, Scotland*, 2015.
[19] F. Kleber, J. Harrington, and U. Reubold, "The relationship between the perception and production of coarticulation during a sound change in progress," *Language and Speech, vol. 55*, pp. 383–405, 2012.
[20] J. J. Ohala, "*The listener as a source of sound change,"* in Papers from the Parasession on language and behavior, C.S. Masek, R.A. Hendrick, and M.F. Miller, Eds. Chicago: Chicago Linguistic Society, 1981, pp. 178–203.
[21] P. Boersma and D. Weenik, *Praat: Doing phonetics by computer*, version 5.4.09, available at http://www.praat.org, 2015.
[22] D. Burnham et al., "Building an audio-visual corpus of Australian English: Large corpus collection with an economical portable and replicable black box," in *Proc. INTERSPEECH 2011, Florence, Italy,* 2011, pp. 841–844.
[23] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software, vol. 67*, pp. 1-48, 2015.
[24] R Core Team (2016). *R: A language and environment for statistical computing*, version 3.3.1, available at https://www.R-project.org/.