THE RELATIONSHIP BETWEEN VOWEL CHANGE AND NASAL LOSS IN THE CHENGDU DIALECT OF MANDARIN CHINESE: EVIDENCE FROM RT-MRI

Sishi Liao, Phil Hoole, Jonathan Harrington

Institute for Phonetics and Speech Processing (IPS), LMU Munich
sishi.liao | hoole | jmh@phonetik.uni-muenchen.de

ABSTRACT

This study aims to examine whether the pre-nasal vowel change that commonly occurs with the phonologization of nasalization, such as the raising and possibly fronting (e.g. Latin 'manus' to French /mɛ̃/), extends beyond the language family and can be generalized to where the nasal trace is lost (e.g. /an/-rime with nasal coda loss in the Chengdu dialect). To do so, real-time MRI was applied to obtain vocal tract aperture signals in the palatal and hyperpharyngeal areas in different rimes (/a, an, i, o/) from four L1 Chengdu speakers. The distances of /a, an/ to anchors /i, o/ were compared in a transformed articulatory space, demonstrating that the vowel in /an/-rime is raised and fronted compared to that in /a/-rime. This concomitant vowel shift leads us to conclude that the universal pre-nasal vowel shift does extend to the nasal loss cases where the VN has become an oral vowel.

Keywords: Chengdu dialect; sound change; vowel raising; vowel fronting; /an/-rime

1. INTRODUCTION

In a vowel-nasal (VN) sequence, the degree and temporal extent of nasalization on the preceding vowel can vary due to anticipatory nasalization [1]; this coarticulatory variation can evolve into sound changes [2], [3] that ultimately result in the reduction of the nasal consonant [4]. As a result, the VN sequence can turn into either a nasal vowel Ṽ [5] or a raised plain oral vowel V̞ [6].

Phonologization of nasalization or nasal loss is commonly accompanied by the pre-nasal vowel height shift, where the vowel is alleged to be higher for low vowels and lower for high vowels, leading to vowel height centralization [7]. This nasal-induced shift in vowel height is thought to result from the intra-speaker interplay between perception and production of speech. Specifically, when low vowels are nasalized, they are perceived as phonetically higher [8], which can lead to tongue dorsum raising [9] during its production.

This nasal-induced vowel height shift is a well-documented phenomenon in phonetics [10] and phonology [11]–[13] but has not been extensively investigated in the VN → V̞ cases [6]. One example of such cases exists in the Chengdu dialect of Mandarin Chinese [14], in which the nasal consonant of /(V)an/-rime has been lost, and the whole rime has become oral. Considering that sound changes often involve the reweighting of multiple cues [15], an unresolved issue is whether the loss of the nasal consonant is associated with changes in vowel quality.

Vowel height/backness are often estimated from the first two formant frequencies in studies of dialect variation [16] and sound changes [17], [18]. However, the relationship between formants and the articulatory data can be non-linear and sometimes ambiguous [19], [20]. To overcome this limitation, researchers have increasingly turned to advanced technology to investigate the configuration of the vocal tract. Techniques such as ultrasound [21], electromagnetic articulography (EMA) [22], and real-time MRI [23]–[25] can provide clearer visual representations of the vocal tract and its movement, allowing for more accurate interpretation of vowel quality in terms of physiological tongue position.

Many studies [24], [26] on vowel and consonant quality go well beyond examining values at specific moments by investigating the dynamic shape of the signal trajectory, using methods such as functional principal component analysis (FPCA) and discrete cosine transform (DCT) [27], [28]. For example, [29] examined the /u/-fronting in the 3D space composed by the first three DCT coefficients of the F2 and compared the relative Euclidean distance of the /u/ to two anchors /a, i/ between younger and older speakers. Another study [30] analyzed /s/-retraction in the sibilant of /str/ by examining the first spectral moment (M_1) trajectory; they compared the first 3 DCT coefficients using the normalized orthogonal projection onto the line connecting the canonical /s/ and /ʃ/ for each speaker.

This study compares the vocal tract constriction in the palatal and hyperpharyngeal regions (relating to tongue height and backness) between the vowels in the historically nasalized /an/-rime, which has become oral [14], and those in /a/-rime followed by a nasal. The aim is to investigate whether there are changes in vowel quality accompanying the nasal loss in this dialect [14], similar to the vowel height shift in cases where nasality has been preserved [2].

2. METHODOLOGY

2.1. Participants and experiment settings

This study recruited four first-language speakers of the Chengdu dialect (2 females, with an average age of 25.75 years) with no self-reported speech-language disorders. Prompt texts and videos explaining the experiment procedures were sent to the participants in advance, who also signed the consent forms prior to the experiment.

The experiment took place at the Max Planck Institute for Multidisciplinary Sciences in Göttingen, Germany. Real-time MRI data were collected with a 3T MRI system, with the participant in a supine position in the machine and the head placed in a head coil. The monitor showed images of the randomized prompt texts, which were reflected through a set of mirrors to the participants' vision. The texts were split into blocks of PowerPoint slides, with each block consisting of about 16 slides and lasting about one minute; any block with mistakenly pronounced tokens was repeated.

2.2. Speech materials

The speech material included four types of sentences differing in the target segments. Each target segment consisted of a CV_1NV_2 sequence, where C is an onset consonant /p, p^h, f/, V_1 is a nucleus vowel /a, i, o/ with four tones, N is an alveolar nasal /n/, and V_2 is /a/ if V_1N is tautosyllabic, or /ai/ if V_1N is heterosyllabic. The meaning of the sentence was either "she is called <X> aunty" or "she is called <X> grandma", both of which have the same syntactic structure and convey similar semantic meanings. An example set of sentences with C being /p/, and V_1 varying in /a, i, o/ are shown with characters and respective pronunciations in IPA in **Table 1**.

A total of 301 tokens were analyzed, excluding all the mispronounced ones. The target interval in each utterance was segmented and labeled using Praat [32]. In the $CV_1N.V_2$ cases, where the nasal coda was claimed to be lost [14] (realized as $CV_1.V_2$), the target interval began at the voice onset of V_1 and ended at the V_1-V_2 transition; in the $CV_1.NV_2$ cases, the interval also began at the voice onset of V_1 and ended at the transition from V_1 to N, where the nasal damped the intensity and caused the formation of anti-resonance. The reason for /i, o/ being the anchors is that they form the most peripheral contrast in height and backness in the vowel inventory of this dialect, so that /a, an/-rimes could be compared more accurately in both dimensions.

Target segments		IPA of the sentences				
V_1(N)-rime		她	叫	**X**	阿	姨
$CV_1(N).V_2$	pa(n).V_2	t^ha	tɕiɑʊ	**pa(n)**	**a**	i
V_1-rime		她	叫	**X**	奶	奶
$CV_1.NV_2$ anchors: /i, o/	pa.nV_2	t^ha	tɕiɑʊ	**pa**	**naɪ**	naɪ
	pi.nV_2	t^ha	tɕiɑʊ	**pi**	**naɪ**	naɪ
	po.nV_2	t^ha	tɕiɑʊ	**po**	**naɪ**	naɪ

Table 1: An example set of sentences (with onset being /p/) in IPA, with target segments in bold and target vowels underlined. The nasal consonant is presumed to be lost in the /an/-rime words.

2.3. Real-time MRI and vocal tract aperture

The real-time MRI data were reconstructed by means of [33] with a frame rate of 50 fps. The MR images were recorded synchronously with noise-suppressed audio so that the time stamps derived from the audio segmentation could be mapped onto the image-derived time series signals.

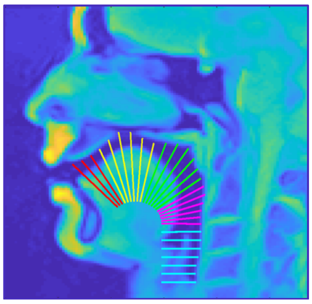

Figure 1: An example of the grid lines used to measure the aperture along the vocal tract (red: alveolar, yellow: palatal, green: velar, purple: hyperpharyngeal, blue: hypopharyngeal)

The analysis of the vocal tract configuration was based on the methods in [24]. The first step of processing the MR images was image registration, which eliminated the influence of any possible slight head movement. Then 28 grid lines were mapped onto the MR image, 20 of which were semi-polar lines extending from the alveolar ridge to the pharynx, 8 of which were horizontal lines from the pharynx to the glottis. The 28 gridlines were divided into five articulatory regions: alveolar, palatal, velar, hyperpharyngeal, and hypopharyngeal. For each articulatory region, the measure of vocal tract constriction was derived based on the average pixel intensity along the gridlines in each region. This measurement in the respective region increases when there is more high-intensity tongue tissue and decreases when there is more low-intensity air; see [25] for more details. For each speaker, the value of each articulatory signal was normalized to [0, 1] based on the overall minimum and maximum occurring in the whole corpus.

2.4. Discrete Cosine Transform

In each articulatory region, the derived vocal tract displacement signal within the vowel interval was resampled to 1000 data points using the function in the `fdasrsf` package [34] in Python. The first DCT coefficient (*k0*) for each resampled curve was then calculated with the function (1):

$$(1) \qquad k_0 = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} x(n)$$

To establish a transformed articulatory space, we used the *k0*_hyperpharyngeal and *k0*_palatal values from each vowel interval as the coordinates on the x- and y-axis, respectively. In this two-dimensional space, a token with a more raised and fronted tongue position would locate closer to /i/, and a token with a lowered and more retracted tongue position would locate closer to /o/.

2.5. Orthogonal projection ratio and statistics

To determine the vowel quality differences between /a/-rime and /an/-rime for each speaker, the orthogonal projection ratio (*op* ratio) was calculated with /i, o/ as anchors in the following manner. The centroids of the two anchors were determined on the <*k0*_hyperpharyngeal, *k0*_palatal> space and were then connected. The *op* ratio onto this line was then calculated for each token; the two centroids of /i, o/ have an *op* ratio of ±1. For any token, a positive *op* ratio denotes proximity towards /i/, a negative *op* ratio denotes proximity towards /o/, and an *op* ratio = 0 means this token is of equal distance towards /i/ and /o/.

A linear mixed-effect model was applied to the *op* ratios using the `lmer` function in the `lmerTest` package [35] from `R` to test the proximity of the /a, an/ tokens to the anchors /i, o/. The *op* ratio of each /a, an/ observation was set as the response, the `RIME` as the fixed factor, and the `SPEAKER`, `ONSET`, and `TONE` as the random factors. The model was then applied with ANOVA to test the difference between /a, an/ in this transformed articulatory space.

3. RESULTS

3.1. Movement trajectories and DCT coefficients

The degree of tongue height/backness during the interval of /a, an, o, i/, denoted by the articulatory signal in the palatal/hyperpharyngeal region, was resampled and plotted as a function of time in **Figure 2**. The left panel shows the tongue displacement in the palatal region: the larger the value, the higher the tongue position. The right panel shows the tongue displacement in the hyperpharyngeal region: the larger the value, the more retracted the tongue.

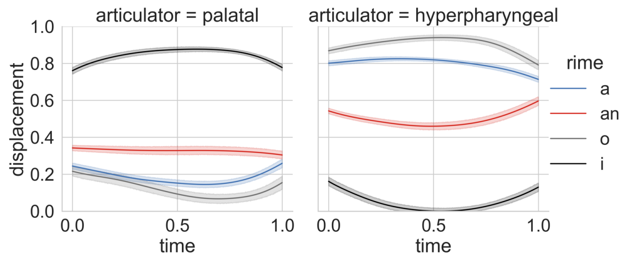

Figure 2: The tongue displacement as a function of time. The larger the value, the higher (left) / retracted (right) the tongue.

In the palatal region, the overall tongue displacement of the vowel in /an/-rime was larger than that in /a/ rime, showing a higher tongue position for /an/ than /a/. In the hyperpharyngeal region, the overall tongue displacement of the vowel in /an/-rime was less than that in /a/-rime, showing a fronter tongue position for /an/ than /a/.

To compare the degrees of tongue height and backness between /a, an/-rimes for each speaker, the DCT coefficient *k0* for both the palatal and hyperpharyngeal regions were plotted in a 2D space in **Figure 3**. Each point represents one utterance token, with the DCT coefficient *k0* of the hyperpharyngeal signal on the x-axis, and the *k0* of the palatal signal on the y-axis.

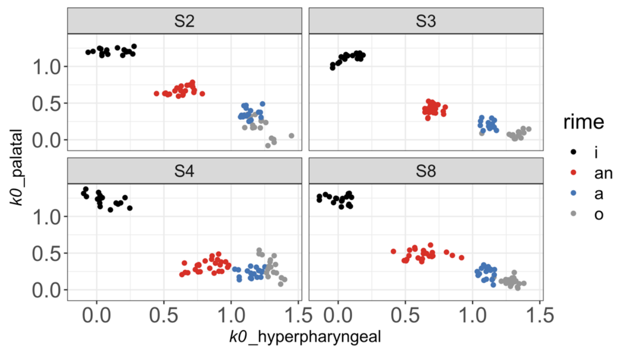

Figure 3: The DCT *k0* space (hyperpharyngeal × palatal) of each utterance token from each speaker, in which speaker S2 and S4 were female.

3.2. Orthogonal projection

The means of /i/- and /o/-rimes in the DCT *k0*_hyperpharyngeal × *k0*_palatal space were calculated and centered at ±1 in the orthogonal projection ratio plot, respectively colored in black and gray. In this transformed articulatory space, the /an/-rime was closer to

/i/ and farther from /o/, compared to the /a/-rime, denoting a more raised and fronted tongue position for /an/-rime. See **Figure 4**.

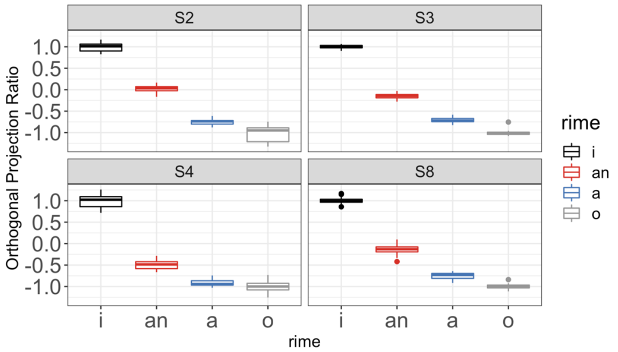

Figure 4: The orthogonal projection ratio in DCT *k0*_hyperpharyngeal × *k0*_palatal space.

3.3. Statistics

If vowels in /an/-rime words in the Chengdu dialect were produced with a more raised and fronted tongue position than that in /a/-rime words, then their *op* ratios should differ in the DCT *k0* space. The random factor `TONE` caused a lack of convergence and was thus not included in the model. The output of the model (**Table 2**) with 162 observations showed that the average *op* ratio was 0.59 higher (SE = 0.07, 95% CI = [0.37, 0.81]) for the vowels in /an/-rime than that in /a/-rime.

	Est.	Std. E	df	t val.	Pr(>|t|)
(Inter-cept)	-0.78	0.05	3.02	-17.11	0.000 ***
rime.an	0.59	0.07	3.00	8.11	0.004 **

Table 2. The results for the `lmer` model: *op* ratio ~ rime + (rime | speaker) + (1 | onset).

The results of the lmer model showed that there was a significant difference in the *op* ratio between /a/- and /an/-rime ($F[1, 3.00] = 65.75$, $p < 0.01$).

4. DISCUSSION

The result of this study shows that in the Chengdu dialect of Mandarin Chinese, the vowels in /an/-rimes are pronounced with a significantly higher and fronter tongue position than that in /a/-rime in a transformed articulatory space. This finding of the concomitant vowel raising and fronting with the nasal coda loss in /an/-rime [14] (VN → V̩ˑ) is consistent with other evidence showing vowel height centralization accompanying synchronic vowel nasalization [7].

This sound change does not seem to be derived from language contact because, in the more dominant common language – standard Mandarin, the nasal coda in /an/-rimes has been maintained. It would be necessary for a future investigation on the interaction between vowel quality change and nasal coarticulation, specifically in other Mandarin varieties that may have varying degrees of nasality in the /(V)an/-rimes. Further investigations are also needed to establish the diachronic sequence of vowel raising and nasal loss, i.e., whether the changes in the present study correspond to VN→ V̩ˑN→V̩ˑ or VN→Ṽ→ V̩ˑ.

5. ACKNOWLEDGMENT

The first author is supported by the China Scholarship Council (grant No. 202008440473). This study is supported by the InterAccent project, which receives funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 742289). Website: https://www.phonetik.uni-muenchen.de/Forschung/interaccent/interAccent.html

6. REFERENCES

[1] P. S. Beddor, "A coarticulatory path to sound change," *Language (Baltim).*, vol. 85, no. 4, 2009, doi: 10.1353/lan.0.0165.

[2] J. J. Ohala, "The listener as a source of sound change: An update," in *The Initiation of Sound Change*, Maria-Josep Solé and Daniel Recasens, Eds. John Benjamins, 2012, pp. 21–36.

[3] J. J. Ohala and M. Ohala, "The phonetics of nasal phonology: Theorems and data," *Nasals, Nasalization, and the Velum*, vol. 5, pp. 225–249, 1993, doi: 10.1016/B978-0-12-360380-7.50013-2.

[4] J. J. Ohala and M. G. Busa, "Nasal loss before voiceless fricatives: a perceptually-based sound change," *Riv. Linguist.*, vol. 7, pp. 125–144, 1995.

[5] B. L. Rochet, "The formation and evolution of the French nasal vowels," *Form. Evol. French nasal vowels*, Dec. 1976, doi: 10.1515/9783111328287.

[6] M. Cresci, "VN > V in Camuno: An alternative historical pathway to nasal loss," *Ital. J. Linguist.*, vol. 31, no. 1, pp. 61–92, 2019, doi: 10.26346/1120-2726-132.

[7] P. S. Beddor, "Phonological and Phonetic Effects of Nasalization on Vowel Height," Indiana University, 1983.

[8] R. A. Krakow, P. S. Beddor, L. M. Goldstein, and C. A. Fowler, "Coarticulatory influences on the perceived height of nasal vowels," *J. Acoust. Soc. Am.*, vol. 83, no. 3, p. 1146, Aug. 1988, doi: 10.1121/1.396059.

[9] P. S. Beddor, R. A. Krakow, and L. M. Goldstein, "Perceptual Constraints and Phonological Change: A Study of Nasal Vowel Height," *Phonol. Yearb.*, vol. 3, pp. 197–217, 1986.

[10] T. Arai, "Comparing Tongue Positions of Vowels in Oral and Nasal Contexts," in *INTERSPEECH*, 2005, pp. 1033–1036.

[11] R. Shosted, C. Carignan, and P. Rong, "Managing

the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi," *J. Acoust. Soc. Am.*, vol. 131, no. 1, p. 455, Jan. 2012, doi: 10.1121/1.3665998.
[12] C. Carignan, "Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels," *J. Phon.*, vol. 63, pp. 87–105, 2017, doi: 10.1016/J.WOCN.2017.04.005.
[13] N. Almeida, S. Silva, C. Cunha, and A. Teixeira, "Data-Driven Analysis of European Portuguese Nasal Vowel Dynamics in Bilabial Contexts," *Appl. Sci.*, vol. 12, no. 9, p. 4601, 2022, doi: 10.3390/APP12094601.
[14] S. Liao *et al.*, "Nasal Coda Loss in the Chengdu Dialect of Mandarin: Evidence from RT-MRI," *INTERSPEECH*, pp. 1347–1351, 2022, doi: 10.21437/INTERSPEECH.2022-10775.
[15] J. Kuang and A. Cui, "Relative cue weighting in production and perception of an ongoing sound change in Southern Yi," *J. Phon.*, vol. 71, pp. 194–214, Nov. 2018, doi: 10.1016/J.WOCN.2018.09.002.
[16] J. Riverin-Coutleé, E. Kapia, C. Cunha, and J. Harrington, "Vowels in urban and rural Albanian: The case of the Southern Gheg dialect," *Phonetica*, vol. 79, no. 5, pp. 459–512, Oct. 2022, doi: /10.1515/phon-2022-2025.
[17] J. Harrington, "Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts," *Lab. Phonol.*, no. 9, pp. 125–143, 2007.
[18] J. Riverin-Coutlée and J. Harrington, "Phonetic change over the career: a case study," *Linguist. Vanguard*, 2022, doi: 10.1515/LINGVAN-2021-0122.
[19] P. Hoole, "Articulatory-acoustic Relations in German Vowels," in *14th International Congress of the Phonetic Sciences*, 1999, pp. 2153–2156.
[20] A. B. Ximenes, J. A. Shaw, and C. Carignan, "A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English," *J. Acoust. Soc. Am.*, vol. 142, no. 1, p. 363, Jul. 2017, doi: 10.1121/1.4991346.
[21] C. Carignan, "Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels," *J. Acoust. Soc. Am.*, vol. 143, no. 5, p. 2588, May 2018, doi: 10.1121/1.5034760.
[22] T. Rebernik *et al.*, "A review of data collection practices using electromagnetic articulography," *Lab. Phonol.*, vol. 12, no. 1, pp. 1–42, Mar. 2021, doi: 10.5334/LABPHON.237.
[23] C. Carignan *et al.*, "Planting the seed for sound change: Evidence from real-time mri of velum kinematics in German," *Language (Baltim).*, vol. 97, no. 2, pp. 333–364, 2021, doi: 10.1353/LAN.2021.0020.
[24] C. Carignan *et al.*, "Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI," *Lab. Phonol.*, vol. 11, no. 1, pp. 1–26, Mar. 2020, doi: 10.5334/LABPHON.214.
[25] E. Kunay *et al.*, "Vowel height and velum position in German: Insights from a real-time magnetic resonance imaging study," *J. Acoust. Soc. Am.*, vol. 152, no. 6, p. 3483, Dec. 2022, doi: 10.1121/10.0016366.
[26] M. Pouplier, J. Cederbaum, P. Hoole, S. Marin, and S. Greven, "Mixed modeling for irregularly sampled and correlated functional data: Speech science applications," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 935–946, 2017, doi: 10.1121/1.4998555.
[27] M. Gubian, F. Torreira, and L. Boves, "Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts," *J. Phon.*, vol. 49, pp. 16–40, Mar. 2015, doi: 10.1016/J.WOCN.2014.10.001.
[28] J. Harrington and F. Schiel, "/u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change," *Language (Baltim).*, vol. 93, no. 2, pp. 414–445, 2017, doi: 10.1353/LAN.2017.0019.
[29] J. Harrington, F. Kleber, and U. Reubold, "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 2825–2835, 2008, doi: 10.1121/1.2897042.
[30] M. Stevens, J. Harrington, F. Schiel, M. Stevens, J. Harrington, and F. Schiel, "Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English," *Glossa a J. Gen. Linguist.*, vol. 4, no. 1, 2019, doi: 10.5334/GJGL.620.
[31] C. Carignan, R. Shosted, C. Shih, and P. Rong, "Compensatory articulation in American English nasalized vowels," *J. Phon.*, vol. 39, no. 4, pp. 668–682, 2011, doi: 10.1016/J.WOCN.2011.07.005.
[32] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2022. [Online]. Available: http://www.praat.org/
[33] M. Uecker, S. Zhang, D. Voit, A. Karaus, K. D. Merboldt, and J. Frahm, "Real-time MRI at a resolution of 20 ms," *NMR Biomed.*, vol. 23, no. 8, pp. 986–994, 2010, doi: 10.1002/NBM.1585.
[34] J. D. Tucker, W. Wu, and A. Srivastava, "Computational Statistics and Data Analysis Generative models for functional data using phase and amplitude separation," *Comput. Stat. Data Anal.*, vol. 61, pp. 50–66, 2013, doi: 10.1016/j.csda.2012.12.001.
[35] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *J. Stat. Softw.*, vol. 82, no. 13, pp. 1–26, 2017, doi: 10.18637/JSS.V082.I13.