



When focus shapes the flow: prosodic restructuring in Mandarin complex nominals

Anqi Xu¹, Yu-yin Hsu²

¹School of Humanities and Social Sciences, Harbin Institute of Technology, Shenzhen

²Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

xuanqi@hit.edu.cn, yu-yin.hsu@polyu.edu.hk

Abstract

While extensive studies have explored acoustic focus realization in Mandarin, little is known about how focus affects the prosodic phrasing of Mandarin complex nominals. This study examined how contrastive focus influences syllable duration of Mandarin numeral-classifier-noun phrases. Using a mini-dialogue paradigm, we elicited contrastive focus of different spans, alongside a baseline no-focus condition. Two production experiments revealed that when focus was placed on the numeral, the default prosodic grouping was disrupted, dissimilar to when focus encompasses the entire phrase or in neutral contexts, with tonal factors amplifying the reorganization. Our results indicate that prosodic organization in Mandarin is shaped by tones, morphosyntactic structures, focus marking and their interplay. Crucially, the results challenge rigid models of boundary phrasing and disyllabic footing, highlighting a multilevel interaction among phonetics, prosodic phrasing, and syntax in tonal languages.

Index Terms: Mandarin complex nominals, focus prosody, prosodic phrasing, duration, disyllabic footing

1. Introduction

Prosodic focus can be placed in a particular part of speech to highlight its importance. In English [1], Swedish [2], German [3] and many other languages, the on-focus words are marked by increased durations and fundamental frequency (f_0), followed by a sharp drop in these parameters. This phenomenon of duration and f_0 reduction after focus is referred to as post-focus compression (PFC). Mandarin, a tonal language, also exhibits on-focus enhancement and PFC, but its prosodic cues are more complex due to tone-specific effects on pitch and duration.

The lengthening of the on-focus words and shortening of the post-focus words hold in Mandarin [4], [5]. However, tone type influences the realization of focus. Only words with high tone (Tone 1) exhibit the typical raised on-focus f_0 and reduced post-focus f_0 , whereas other tonal conditions rely more on pitch expansion, rather than voice pitch. For example, the focused words with a falling tone (Tone 4) and a low tone (Tone 3) may even show decreased f_0 [4]. Moreover, the post-focus regions exhibit varied prosodic patterns, featuring compressed F0 range and occasionally reversed patterns. Some studies suggest that some tonal contexts may not necessarily trigger a f_0 range reduction in the post-focus words [6]; for instance, a post-focus rising tone (Tone 2) shows compression after a high tone but expansion after a low tone, while falling tone after focus shows expansion after a high tone but remains unchanged after a low tone. These findings suggest that Mandarin focus interacts intricately with its tonal system.

While focus prosody in Mandarin has been extensively studied, prosodic phrasing remains comparatively underexplored. Existing theoretical frameworks attribute phrasing preferences to rhythmic constraints, particularly the organization of syllables into disyllabic units. For instance, Duanmu [7] and Feng [8] posit that disyllabic footing serves as a default mechanism for achieving rhythmic balance, accounting for the predominance of disyllabic words and the systematic division of trisyllabic structures into “1+2” or “2+1” patterns, and four-syllabic structures into “2+2” patterns. However, their accounts diverge in the role of stress: Duanmu [7] links stress to left-prominent disyllabic feet, whereas Feng [8] treats disyllabic units as foundational prosodic constituents, independent of stress assignment.

In contrast to theoretical explorations of syntactic-prosodic mapping, empirical investigations into Mandarin phrasing have prioritized boundary marking over structural preferences. For example, a large-scale speech corpus analysis [9] revealed significant pre-boundary syllable lengthening, modulated by tonal properties: high and rising tones exhibited greater lengthening effects compared to low and falling tones. Similarly, Feng et al. [10] observed that neutral tones and rising tones amplify final-syllable duration in phrase-final positions. Critically, these studies focus primarily on phonetic correlates of phrasing reflected by duration and tones; whether and how morphosyntactic structures interact with prosodic organization remains to be examined.

Beyond tonal and boundary effects, syllable duration is systematically shaped by phrase length and internal morphosyntactic structure. Xu et al. [11] found that in noun phrases shorter than four syllables, final syllables are longest, initial syllables are slightly shorter, and medial syllables are the shortest. Additionally, the morphosyntactic structure of a four-syllable sequence affects its durational patterns: monosyllabic modifier + trisyllabic words' syllable-duration ranking follows the order 1st > 4th > 2nd > 3rd, whereas disyllabic word + disyllabic word follow 4th > 1st > 3rd > 2nd. These findings underscore duration as a robust phonetic cue for inferring internal morphosyntactic organization. Nevertheless, whether such durational hierarchies generalize to complex nominal phrases remains unresolved.

While prior work has extensively documented how prosodic focus operates in simple words [4], [5], [6], leaving unclear how focus interacts with the internal structure of complex phrases. Furthermore, theoretical frameworks for Mandarin prosodic phrasing [7], [8], though insightful, lack empirical validation. Importantly, emerging evidence suggests that durational patterns reflect morphosyntactic organization in Mandarin phrases [11], raising the possibility that syllable duration could serve as a key indicator for examining the interplay between focus and prosodic phrasing.

Thus, the goal of the study is to investigate whether focus affects the internal prosodic grouping of Mandarin numeral-classifier-noun phrases and whether tonal conditions interact with these effects. We will measure the duration of each syllable in the phrases expressing varying focus spans inside the phrase (i.e., different phrasal-internal focus structures) and tonal conditions. We hypothesize that focus reorganizes the default prosodic grouping of syllables in Mandarin complex nominal phrases, and that tonal conditions further modulate this prosodic reorganization.

2. Method

2.1. Stimuli

The target complex nominals in this study were numeral-classifier-noun phrases (NPs). In Experiment 1, the numeral was monosyllabic, while in Experiment 2, it was disyllabic, and their associated classifier and noun were all monosyllabic (see Table 1). Additionally, all syllables within each target phrase carried the same tone. To elicit natural contrastive focus, we employed a mini-dialogue paradigm, prompting focus on either the numeral (CNUM) or the entire phrase (CNP). A baseline condition (ODNP) was included, where all words conveyed old information, as shown in Table 2.

Table 1: *Examples of target items with different tonal conditions in Mandarin. Disyllabic numerals with Tone 2-Tone 2 combination are not available in the language.*

Tone	Pinyin & Characters	Glosses
T1N: High	sān zhī huā 三 枝 花	‘three [classifier for] flowers’
T1N: High	sān qiān zhī huā 三 千 枝 花	‘three thousand [classifier for] flowers’
T2N: Rising	shí tiáo yú 十 條 魚	‘ten [classifier for] fish’
T3S: Sandhied Low	wǔ wǎn jiǔ 五 碗 酒	‘five bowls of alcohol’
T3S: Sandhied Low	wǔ bǎi wǎn jiǔ 五 百 碗 酒	‘five hundred bowls of alcohol’
T4N: Falling	liù duì wà 六 对 袜	‘six pairs of socks’
T4N: Falling	liù wàn duì wà 六 万 对 袜	‘sixty thousand pairs of socks’

The numeral-classifier-noun phrases were embedded in carrier sentences of equal length. As shown in Table 2, for the ODNP condition, the whole phrase functioned as background information, whereas in the CNP and CNUM conditions, contrastive information was placed on the entire noun phrase and on the numeral alone, respectively (contrastive elements being corrected in the leading sentence are in bold). The target sentences were identical across conditions, differing only in the size and type of focus constituent prompted by the preceding context. Six target sentences were created for each tone condition.

Table 2: *Mini-dialogue illustrating three focus conditions with Mandarin numeral-classifier-noun phrases.*

Information	Leading sentence	Target sentence
Old NP (ODNP)	What happened to five bowls of liquor?	<i>Gloss:</i> <u>Five bowls of liquor</u> spilled all over the table.
Contrastive NP (CNP)	A glass of water spilled all over the table.	<i>Pinyin:</i> wǔ wǎn jiǔ sǎ le yī zhuō zi
Contrastive numeral (CNUM)	Two bowls of liquor spilled all over the table.	<i>Characters:</i> 五碗酒洒了一桌子

2.2. Participants

We invited native speakers of Putonghua from the northern provinces of China to participate in two production experiments. Each experiment included 14 speakers, ranging in age from 20 to 28 (mean = 23.5). None of the participants reported any history of speech or hearing problems.

2.3. Procedure

The production experiments were conducted in a sound-attenuated room. Pre-recorded stimulus questions (context) were played to the participants via headphones, after which they were asked to answer naturally using the target sentences. Each participant completed three practice trials. The context and target sentences were presented in randomized order. Speech was recorded in .wav format at a sampling rate of 44.1 kHz with 16-bit quantization. Prior to the experiment, participants completed a language-background questionnaire and signed an informed consent form.

2.4. Analysis

We annotated the recordings using Praat [12]. Syllable boundaries were determined using both visual (waveform and spectrogram) and auditory cues. We extracted the duration of each syllable in the target sentences for analysis. Linear mixed effects models were fitted to the data for each target syllable using *lme4* package [13] in R [14]. We began with a model that included random intercepts for Subjects and Items. Syllable Location, Focus, Tone, and their interaction were then added as potential fixed effects. P-values were obtained from likelihood-ratio tests comparing models with and without these fixed effects and their interactions, which were included only when they improved model fit ($\alpha = 0.05$). Post-hoc tests with Tukey correction for multiple comparisons were conducted using *emmeans* package [15].

3. Results

3.1. Experiment 1: NPs with a monosyllabic number

In Experiment 1, the duration ratio of each syllable relative to the entire complex nominal followed a consistent pattern: the initial and final syllables were longer than the medial syllable (Fig. 1). A significant main effect of Syllable Location was found ($\chi^2(2) = 627.54, p < .001$). Post-hoc comparisons revealed that the noun had the longest duration, followed by the numeral, while the classifier was the shortest ($p < .001$). The temporal pattern for phrases under contrastive focus on the whole phrase

(CNP) closely resembled that of NP expressing old information (ODNP) but differed significantly from phrases where focus was on the numeral (CNUM) (Fig. 2). Statistical analysis confirmed a significant interaction between Syllable Location and Focus Type ($\chi^2(6) = 221.62, p < .001$). Post-hoc tests further showed that in both ODNP and CNP conditions, the noun was significantly longer than the numeral ($p < .001$), whereas in the CNUM condition, the numeral and classifier did not differ significantly ($p = .998$).

We also found a significant interaction between Syllable Location and Tone ($\chi^2(9) = 290.7, p < .001$), indicating that tone influenced the duration ratio of each syllable within the phrase (Fig. 1). The durations of the noun with rising tone (T2N) and falling tone (T4N) are significantly longer than the other conditions ($p < .001$). Notably, for Tone 3, the numeral was the shortest among the three positions ($p < .05$), whereas for other tones, the general pattern remained consistent: Noun > Numeral > Classifier.

The three-way interaction between Syllable Location, Focus, and Tone was also significant ($\chi^2(18) = 40.38, p = .002$). As shown in Figure 2, the longest syllable within the entire phrase under corrective numeral focus (CNUM) varied across tonal conditions. For high tone without tone sandhi (T1N), the durational pattern remained consistent across all three focus types, with the numeral shorter than the noun under CNUM ($p < .001$). In contrast, under CNUM, the numeral became the longest syllable in T2N and T4N without tone sandhi ($p < .001$). Interestingly, for low tone with tone sandhi (T3S), the noun had significantly longer duration than the numeral ($p = .030$).

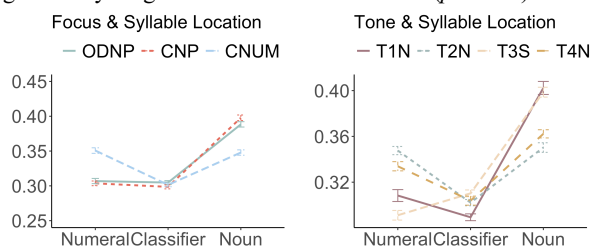


Figure 1: Interaction between Focus and Syllable Location (left)/ Tone and Syllable Location (right) on duration ratios of NP with a monosyllabic numeral.

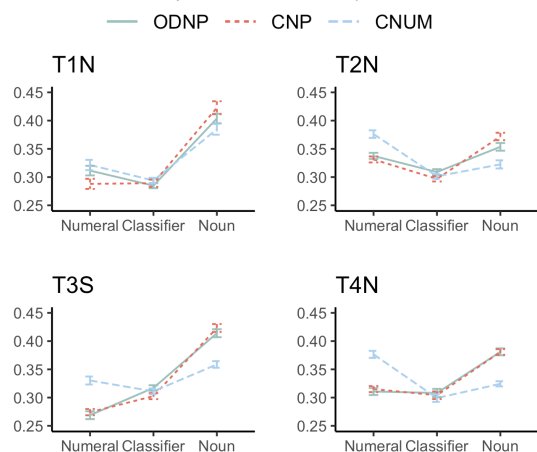


Figure 2: Three-way interaction between Syllable Location, Focus and Tone on duration ratios of syllables over NP with a monosyllabic numeral.

In summary, duration patterns under numeral focus (CNUM) were further influenced by tonal conditions, demonstrating how prosodic phrasing interacts with focus and tone in Mandarin complex nominals.

3.2. Experiment 2: NPs with a disyllabic number

In Experiment 2, the noun had the longest duration, followed by the first numeral, with the second numeral and the classifier exhibiting the shortest durations (Fig. 3). Statistical analysis revealed a significant main effect of Syllable Location on duration ($\chi^2(3) = 1355.1, p < .001$). Post-hoc tests showed that the difference between the second numeral and the classifier was not significant ($p = .935$), while all other contrasts were significant ($p < .001$).

As shown in Figure 3, the temporal pattern of phrases in ODNP and CNP differed from that in CNUM. A significant interaction was found between Syllable Location and Focus Condition ($\chi^2(8) = 185.36, p < .001$). As illustrated in Figure 4, the duration ratio of the first numeral in ODNP and CNP did not differ significantly ($p > .050$), whereas a pronounced difference was observed in the CNUM condition ($p < .001$). Additionally, the interaction between Syllable Location and Tone was significant ($\chi^2(8) = 209.15, p < .001$). The numeral with T3S was significantly shorter than those with other tones, while the T3S noun was lengthened compared to the other tonal conditions ($p < .050$).

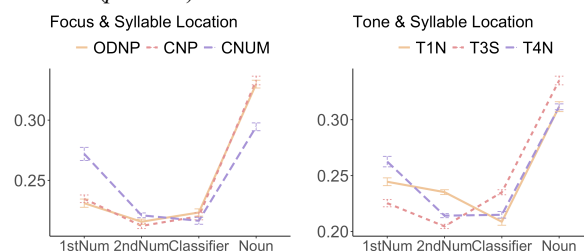


Figure 3: Interaction between Focus and Syllable Location (left)/ Tone and Syllable Location (right) on duration ratios of NP with a disyllabic numeral.

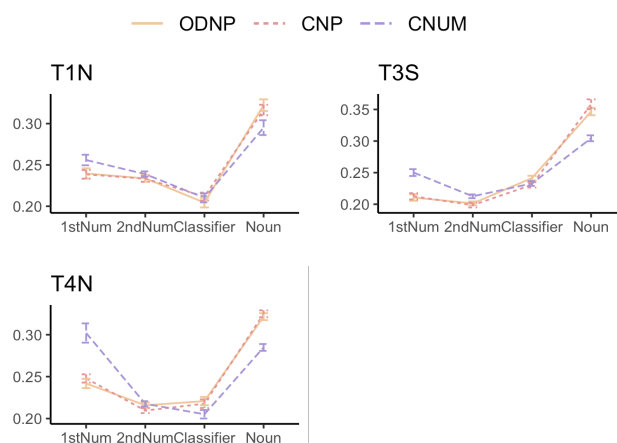


Figure 4: Three-way interaction between Syllable Location, Focus and Tone on duration ratios of syllables over the disyllabic numeral-classifier-noun phrases.

A significant three-way interaction was also found between Syllable Location, Focus Type, and Tone ($\chi^2(16) = 33.222, p = .007$). As shown in Fig.4, under corrective focus, the first

numerals with T1N and T3S remained shorter than the noun, consistent with other focus conditions ($p < .001$). However, under the T4N condition, the difference between the numeral and the noun under corrective focus was non-significant ($p = .533$). The results for complex nominals with disyllabic numerals were consistent with those for monosyllabic numerals. Contrastive focus triggered changes in the temporal pattern of syllables, with certain tonal conditions further amplifying these effects.

4. Discussion

This study examined the impact of focus on the internal prosodic structuring of Mandarin numeral-classifier-noun phrases. We measured the duration of each syllable in the complex nominals under three conditions of varying focus spans (ODNP, CNP, CNUM) in a mini-dialogue paradigm, with the numerals being either monosyllabic in Experiment 1 or disyllabic in Experiment 2. The results indicate that focus altered the temporal organization of syllables in the phrases, with some tonal conditions further amplifying these effects.

We have found that syllable duration followed a consistent pattern where the initial and final syllables were longer than the phrasal medial syllables, aligning well with previous acoustic studies on Mandarin complex words [11]. The phrases with old information (ODNP) and corrective focus on the whole NP (CNP) did not alter the temporal pattern. However, under corrective focus on the numeral (CNUM), the duration ratio of each syllable deviated from the expected patterns of trisyllabic compounds, presumably because of on-focus lengthening, a phenomenon well-documented in previous literature [4], [5]. This finding suggests a regrouping of syllables under corrective focus.

In accordance with previous findings concerning the tonal effect on phrasing, we also found that high tone and rising tone result in greater lengthening effect on the final syllable for monosyllabic numeral-classifier-noun condition [9], [10]. However, in disyllabic numeral phrases, the high (T1N) and rising tones (T2N) were not significantly longer than the low tone (T3S). This discrepancy may stem from differences in measurement methods, as our study analyzed relative syllable duration, whereas previous research reported absolute duration.

It is worth noting that focus and tone interact in shaping the temporal patterns of complex noun phrases. In monosyllabic numeral-classifier-noun phrases, rising and falling tones amplified the effect of corrective focus on durational modulation. Specifically, numerals with rising and falling tones were longer than the noun, contradicting the typical pattern where the final syllable of trisyllabic words is the longest [9], [10], [11]. Additionally, in NPs with a disyllabic numeral, the first numeral with a falling tone had a similar duration to the noun. Across both monosyllabic and disyllabic numeral conditions, the falling tone particularly enhanced focus-driven duration effects, highlighting its prominent role in prosodic modulation.

Furthermore, our results from the disyllabic-numeral condition suggest that internal phrasing does not necessarily adhere to disyllabic footing [7], [8]. Comparing Fig. 3 with Fig. 5 below, we observe that for phrases containing old information (ODNP) and those with corrective focus on the entire noun phrase (CNP), the internal phrasing follows an AB+CD compound structure, seemingly suggesting a preference for disyllabic footing in disyllabic numeral + classifier-noun

phrases. However, this pattern shifts when the focus is on the numeral (CNUM), where the durational pattern aligns more closely with an A+BCD structure. These findings demonstrate that focus can influence internal phrasing and, in some cases, override the disyllabic footing rule.

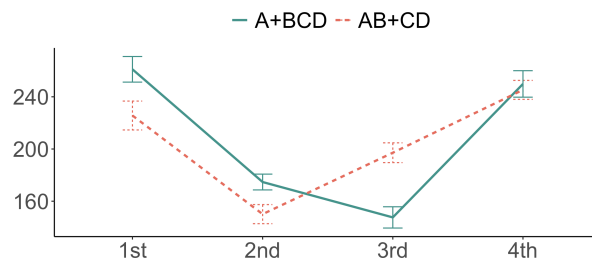


Figure 5: Duration (ms) of each syllable in four-syllabic phrases with two internal structures: $A(\text{modifier})+BCD(\text{noun})$ vs. $AB+CD$ (disyllabic words in parallel relationship). Adopted from Table 5 in [11].

A notable limitation of this study is the limited availability of comparable exemplars for trisyllabic versus four-syllabic phrasings. We primarily relied on Xu et al. [11] to explore potential internal phrasing patterns and their available dataset included only two morphosyntactic structures. Future research should extend this investigation to a broader range of structures to comprehensively assess whether there is a consistent tendency for disyllabic footing [7], [8].

Despite this limitation, our results reveal an evident influence of tone and focus on the modulation of temporal patterns of complex nominals. The study contributes to the broader understanding of how focus, tone, and morphosyntactic structure interact to shape prosodic phrasing in Mandarin, offering empirical support for a dynamic, rather than rigid, model of prosodic structuring of Mandarin.

5. Conclusions

This study investigated how contrastive focus affects the internal prosodic structure of Mandarin numeral-classifier-noun phrases. We found that when focus was placed on the numeral, the default prosodic grouping was disrupted, while focusing on the entire phrase preserved the original grouping. Particularly under certain tonal conditions, the reorganization can be intensified. These findings challenge conventional models of prosodic phrasing and support a more dynamic framework where prosodic boundaries are flexibly reconfigured by both information structure and tonal properties. By empirically demonstrating the interaction between morphosyntactic structure, tone, and focus, this study contributes a nuanced understanding of prosodic organization in tonal languages.

6. Acknowledgements

This research was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (RGC Project No. 15612922), and by the Hong Kong Polytechnic University through the Large Equipment Fund (1-BC7N). We would also like to thank the anonymous reviewers for their feedback and suggestions.

7. References

- [1] W. E. Cooper, S. J. Eady, and P. R. Mueller, “Acoustical aspects of contrastive stress in question–answer contexts,” *J Acoust Soc Am*, vol. 77, no. 6, pp. 2142–2156, Jun. 1985, doi: 10.1121/1.392372.
- [2] M. Heldner, “On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish,” *J Phon*, vol. 31, no. 1, pp. 39–62, Jan. 2003, doi: 10.1016/S0095-4470(02)00071-2.
- [3] C. Féry and F. Kügler, “Pitch accent scaling on given, new and focused constituents in German,” *J Phon*, vol. 36, no. 4, pp. 680–703, Oct. 2008, doi: 10.1016/j.wocn.2008.05.001.
- [4] Y. Xu, “Effects of tone and focus on the formation and alignment of f_0 contours,” *J Phon*, vol. 27, no. 1, pp. 55–105, Jan. 1999, doi: 10.1006/jpho.1999.0086.
- [5] Y. Chen and C. Gussenhoven, “Emphasis and tonal implementation in Standard Chinese,” *J Phon*, vol. 36, no. 4, pp. 724–746, Oct. 2008, doi: 10.1016/j.wocn.2008.06.003.
- [6] Y. Chen, “Post-focus F0 compression—Now you see it, now you don’t,” *J Phon*, vol. 38, no. 4, pp. 517–525, Oct. 2010, doi: 10.1016/j.wocn.2010.06.004.
- [7] S. Duanmu, “Zhongyin lilun he Hanyu de ci chang xuanze [Metrical theory and word length choices in Chinese],” *Zhongguo Yuwen*, vol. 4, no. 271, pp. 246–254, 1999.
- [8] S. Feng, “Lun Hayu De Ziran Yinbu[On the natural foot in Chinese],” *Zhongguo Yuwen [Chinese Linguistics]*, vol. 1, pp. 40–47, 1998.
- [9] C. Ni and W. Liu, “Durational Characteristics and Pitch Characteristics of the Prosodic Phrase in Mandarin Chinese,” in *2008 Chinese Conference on Pattern Recognition*, IEEE, Oct. 2008, pp. 1–5. doi: 10.1109/CCPR.2008.84.
- [10] Y. Feng, M. Chu, L. He, and S. Lv, “Statistical Analysis of Syllable Duration in Mandarin Speech [Han Yu Hua Yu Yin Jie Shi Chang Tong Ji Fen Xi],” in *Modern Phonetics in the New Century—Proceedings of the 5th National Conference on Modern Phonetics*, 2001.
- [11] Y. Xu and M. Wang, “Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin,” *J Phon*, vol. 37, no. 4, pp. 502–520, Oct. 2009, doi: 10.1016/j.wocn.2009.08.003.
- [12] P. Boersma and D. Weenink, “Praat: doing phonetics by computer (Version 4.5.)[Computer program],” *Retrieved from <http://www.praat.org/>*, vol. 5, no. 9/10, pp. 341–345, 2007.
- [13] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *J Stat Softw*, vol. 67, no. 1, p. 51, 2015, doi: 10.18637/jss.v067.i01.
- [14] R Development Core Team, “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>,” 2013.
- [15] S. R. Searle, F. M. Speed, and G. A. Milliken, “Population Marginal Means in the Linear Model: An Alternative to Least Squares Means,” *Am Stat*, vol. 34, no. 4, pp. 216–221, Nov. 1980, doi: 10.1080/00031305.1980.10483031.