Investigating the Tone-Segment Asymmetry in Phonological Counting: A Learnability Experiment

Jian Cui[1], Hanna Shine[1], Youngah Do[2], Jesse Snedeker[1]

[1]*Harvard University*
[2]*University of Hong Kong*

1 Introduction

1.1 *Background* Tone systems have long been argued to have “properties that surpass segmental and metrical systems” (Hyman, 2011). Tone is different in the sense that it can do everything that segmental and metrical phonology can do, but not vice versa. Such a tone-segment asymmetry has been observed in terms of several phonological phenomena including unbounded circumambient high/mid tone plateauing (Stevick, 1969; Hyman, 1998; Yip, 2002; Hyman, 2011), first-last harmony (Bickmore, 2007; Rolle and Bickmore, 2022; Hyman, 2023) and phonological counting (Marlo et al., 2015; Paster, 2019). In phonology per se, we ask whether the inability of segments (or segmental features) stems from synchronic learning biases – i.e., whether the typology we observe corresponds to a specific learning profile – or whether it can only be attributed to an accidental diachronic gap.

The connection between learning biases and typological tendencies have been widely investigated for the domain of phonology, including epenthesis (Morley, 2018), velar palatalization (Wilson 2006), vowel harmony (e.g., Finley, 2008, 2012; Martin & Peperkamp, 2020; Martin & White, 2021; Huang, T., & Do, Y., 2023), consonant harmony (Koo & Cole, 2006; Conklin et al., 2023). In this paper, I focus on such an asymmetry in the context of phonological counting[1]. Typologically, tone and segment vary distinctly in whether they can be counted in the formalization of morpho-phonological rules. On the one hand, non-counting rules involving segments and tones are almost equally attested. On the other hand, counting rules exhibit a clear tone-segment asymmetry: while tone can be flexibly counted up to four, segments are never involved in such a non-local manner. For instance, in Kuria (E40), an H-tone is assigned to the initial mora of the verb to mark past tense, but it can also be assigned to the third or fourth mora to mark remote future and inceptive, respectively, as illustrated in (1).

(1) Kuria H-tone assignment (Marlo et al., 2015; Paster, 2019: 48)
    a. 1 n-to-o-hɔ́ɔ́tɔ́ɔ́t-ɛ́r-a Past
    FOC-1PL-PAST-reassure-APPL-FV
    ‘we reassured’

    b. 2 n-tɔ-ɔka-hoó̲tóót-ééy-e Past progressive
    FOC-1PL-PAST.HAB-reassure-APPL.PERF-FV
    ‘we have just been reassuring’

    c. 3 n-to-re-hɔɔtɔ̲́ɔ́t-ɛ́r-a Remote future

---

* Acknowledgement: We are thankful for helpful discussions with Kevin Ryan, Adam Albright, Nik Rolle, Adam Chong, Edward Flemming, Hailang Jiang, and Bingzi Yu. Also special thanks to Chelsea Tang for her invaluable help as a native Cantonese speaker with stimulus recording. Finally, we thank the *Mind, Brain, Behavior Initiative* at Harvard for the research grant that supported the project.

[1] In this paper, I use ‘phonological counting’ and ‘counting’ more as descriptive, theory-neutral terms for a non-local phenomenon that differentiates the behavior of tonal and segmental features.

Proceedings of *AMP 2025*

FOC-1PL-FUT-reassure-APPL-FV
‘we will reassure’

d. 4 to-ra-hɔɔtɔɔ́t-ɛ́r-a Inceptive
1PL-INCEPT-reassure-APPL-FV
‘we are about to reassure’

We also observe that segmental features play a role in the formalization of morpho-phonological rules: they are almost exclusively edge-oriented, with no counting patterns involved. For example, in the Gurage (Ethiopian Semitic) language Chaha, palatalization of the last root consonant marks the verbal agreement with a second person feminine singular subject, as shown in (2). This rule fails to apply to those roots where the final consonant is non-palatalizable.

(2) Palatalization as verbal agreement in Chaha (McCarthy, 1983)

2nd.m.sg	2nd.f.sg	
nəkəs	nəkəsʲ	‘bite’
nəmæd	nəmædʲ	‘love’
fəræx	fəræxʲ	‘be patient’

A similar example can be found in Tereno language (Bendor-Samuel, 1960), where 1st person and 3rd person possessive forms are distinguished by nasalization of the word-initial segment, which then undergo spreading up to the first stop or fricative in the word, e.g., ˈowuku ‘his house’ → ˈõw̃ũŋgu[2] ‘my house’.

The summary table below (though not exhaustive) illustrates a notable typological asymmetry regarding whether morpho-phonological rules allow the non-local counting pattern of tone and segment. The generalization we obtain here is that tonal features can be counted over tone-bearing units, whereas segmental features cannot be counted over their feature-bearing units. A necessary question to ask is why such an asymmetry exists.

	Tone	Segment
Counting	Attested • H-tone assignment to the third/fourth mora to mark inceptive in Kuria (E40; Marlo et al., 2015)	Non-attested
Non-counting	Attested • H-tone assignment to the first mora to mark past tense (E40; Marlo et al., 2015)	Attested • Palatalization of root-final consonant in Chaha verbal agreement (McCarthy, 1983)

Table 1. Typological profile of counting in tonal vs. segmental (morpho-)phonology

1.2 *Research question and hypotheses* In this study, we investigate whether the typological skew we observe between segment and tone could be potentially predicted by learners’ learnability biases. We adopted the Artificial Language Learning (ALL) paradigm and designed four artificial languages, each with a distinct pluralization rule. In these languages, singular-plural alternations are marked differently: 1) Language A: The plural form is realized by assigning a high tonal feature (i.e., H tone) to the first mora of the word, mimicking the past tense morphology in Kuria, as reviewed above; 2) Language B: The plural form is realized by assigning a nasal feature (i.e., [+nasal]) to the first consonant of the word, serving as the segmental counterpart of the non-counting tonal rule in Language A; 3) Language C: The plural form is marked by assigning an H tone to the third mora of the word, mimicking the inceptive morphology in Kuria; 4) Language D[3]: The plural form is marked by assigning a nasal feature to the third consonant of the

---

[2] During the spreading, there is an accompanying replacement process targeting the first stop/fricative in the word, where the stop *k* is replaced by a nasal consonant cluster *ŋg*.

[3] The rule designed here enforces a counting mechanism, which should be distinguished from general rules involving internal segmental changes (e.g., palatalization that requires identifying a suitable host (Kurisu, 2009)). To this end, all the words used in the stimuli contain only consonants that are appropriate hosts for a nasal feature**.**

word, representing a counting version of a segmental assignment rule. Participants were randomly assigned to learn one of the four rules and their learning performance were tested and analyzed.

Envisioning possible outcomes of the experiment, we made the following hypotheses.

Hypothesis 1 (Figure 1a). Three of the four rules that we are testing (non-counting segment, counting tone, non-counting tone) are attested in natural languages. One is not (counting segment). If this typological pattern has its roots in the relative difficulty of learning these rules then we would expect that the first three conditions would be easier than the fourth. This would result in an interaction between segment/tone and counting/noncounting. We might also expect to see a main effect of counting/non (with performance on non-local counting being lower) and a main effect of segment/tone (with segment being harder).

Hypothesis 2 (Figure 1b). The absence of segmental counting systems could be due to the relative difficulty of learning rules over segments (vs tones) and the relative difficulty of learning counting rules, without any need to posit a particular difficulty in learning rules that have both features. On this hypothesis we would expect: no interaction, main effect of counting/non (with performance on counting being lower) and a main effect of segment/tone (with segment being harder).

Hypothesis 3 (Figure 1c). The relative scarcity of rules requiring counting reflects learnability but there is no learnability difference between segmental and tonal rules. Note that this data pattern does not require that we give up the assumption that there could be a tight linkage between typology and learnability. The existence of a small set of languages allowing counting rules over tones (and no attested examples of a language allowing counting rules over segments) could simply be due to chance in either the languages that have been studied or the languages that have arisen – also known as diachronic gap. On this hypothesis we would expect: no interaction, main effect of counting/non (with performance on counting being lower) and no effect of segment/tone.

Hypothesis 4 (Figure 1d). There is no difference in learnability between any of these rules. This hypothesis would predict no effect of either variable and no interaction. If this data pattern is found, the question arises of whether differences between conditions might emerge in different learning contexts. For example, if performance is very high in all conditions, we might ask if there are differences when the level of exposure is lower and performance drops. Similarly, if performance is quite poor, we might ask if there are differences when exposure is higher and performance improves.

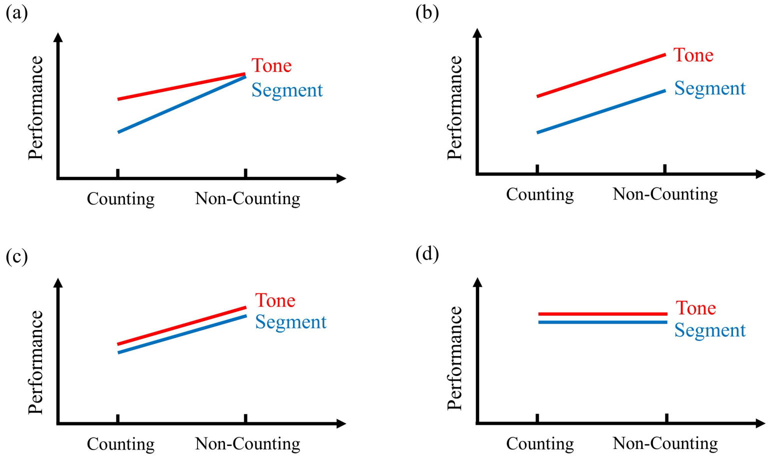

Figure 1. Visual illustration of the four hypotheses. Previewing the results of the current study, we observe a learning pattern that mirrors Figure 1a.

2 Methods

2.1 *Participants* 80 Cantonese-English bilingual adults were recruited through Prolific (N = 8) and a combination of The University of Hong Kong campus and personal networks (N = 72). Each condition had 20 participants, with 2 sourced from Prolific and 18 recruited offline. All participants were required to be at least 18 years old and native Cantonese speakers who are conversationally fluent in English. Participants gave informed consent and received 25HKD for completing the study. This study was approved by the Harvard University Institutional Review Board.

Cantonese speakers were chosen as the target population because the tonal stimuli designed for the experiment are level tones from the Cantonese tonal inventory, which they can easily perceive. Additionally, we specifically need Cantonese speakers with a good knowledge of English to ensure familiarity with the multisyllabic words used in the experiment.

To ensure data quality, we incorporated an attention check at the end of Test 1 phase. All participants passed the attention check and are therefore included. Additionally, we set a threshold for participants' performance on Test 1 -- only participants who obtain at least 80% accuracy at Test 1 can proceed directly to the second block of training and testing. Anyone who performs < 80% is prompted to redo Test 1 and then proceed. We did not follow our original pre-registration plan to exclude participants who did not achieve 80% correctness in Test 1 from the main analysis. The original aim of this criterion was to prevent potential ceiling effects among participants who received additional training. However, due to the difficulty in recruiting Cantonese native speakers and the absence of ceiling effects among these participants, we decided to include them in the main analysis. In the end, 80 participants met the criteria above and are all included in the analysis.

2.2 *Design and procedure* Participants completed, on their own computer, a forced-choice task administered through using the online experiment platform PCIbex (Zehr & Schwarz, 2018). The aim of the task was to learn how to realize the plural form of a known object in an artificial language.

Experiment 1 used a 2 x 2 between-subjects design with four conditions. The two factors were feature type (tone vs. nasal) and counting condition (counting vs. non-counting). For instance, for participants who were assigned to a tone counting condition, they will learn to assign a H-tone to the third mora to realize the plural morpheme. A summary of the four conditions is provided in Table 2.

	Tone	Nasal
Counting	H-assignment to the third mora	[+nasal]-assignment to the third consonant
Non-counting	H-assignment to the first mora	[+nasal]-assignment to the first consonant

Table 2. Summary of Experimental Conditions: Feature and Counting Factors

The experimental paradigm was composed of two phases which were each made up of a training component and a test component. In the training components participants were trained on a plural rule in the language involving a particular morpho-phonological change (in tone or segment). In the test components they were tested on their ability to generalize the learned rule to new singular-plural pairs of objects.

Each trial (both training and test) consisted of two images of a familiar object. The images of the objects were sourced from *unsplash.com* and no objects were repeated across the experiment. The image displayed on the left side of the screen showed a singular instance of the object while the image on the right showed a plural instance. There were 45 pairs of singular-plural object images and each participant saw a full set of them. Each training trial consisted of two audio descriptions of the images, one corresponding to the singular instance of the object and the other to the plural instance. Each test trial consisted of three audio labels of the images, one corresponding to the singular instance of the object and the other two were presented as potential descriptions of the plural instance.

Phase One In the training component of phase one (nine trials) participants first saw the singular object accompanied by the audio label. After playing the audio label the image of the plural object appeared accompanied by a corresponding audio label. After playing the audio label participants were prompted to continue to the next trial. Participants were given the option of replaying either audio label and no limit was placed on the number of plays. This process was repeated for each training trial. A demonstration of the training paradigm is shown in Table 3.

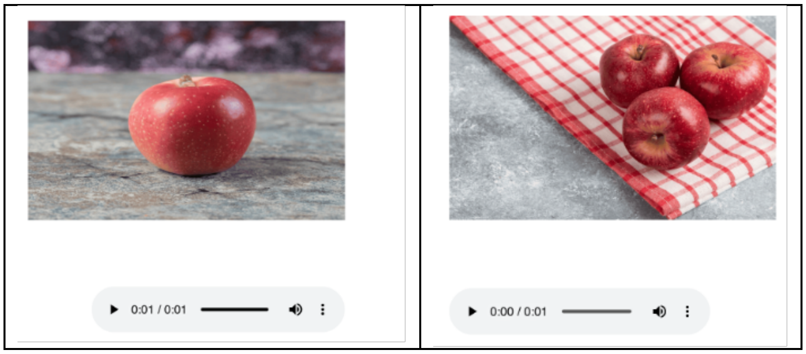

Table 3. Training paradigm: sequential presentation of singular/plural pictures and audio files.

In the test component of phase one (Test 1: nine trials) participants first saw the singular object accompanied by the audio label. After playing the audio label the image of the plural object appeared accompanied by a corresponding audio label. After playing the audio, a second audio label appeared and after playing the second audio participants were prompted to make a selection between the two. One audio attached to the plural object was the correct plural form based on the training and the other audio was either identical to the singular form audio (catch trials) or it was a plural form with a morphological change not featured in the training.

The order of appearance of the audios (correct and incorrect plural form) was based on a fixed list that was randomly created. Participants were given the option of replaying audio labels and no limit was placed on the number of plays. This process was repeated for every test trial. The test paradigm is illustrated in Table 4.

Table 4. Test paradigm: sequential presentation of singular/plural pictures and audio files

Given that some test trials in phase one used an audio that was identical to the singular form audio it was expected that participants should largely succeed during the test component and therefore participants were only able to proceed to phase two if they were at or above 80% accuracy. If participants were below this threshold they were prompted to redo phase one.

Phase Two The training component of Phase Two had the same setup as that of Phase One. The test component of Phase Two (Test 2) was identical to that of Phase One, except that it contained 18 trials in total and one critical change was made to the audio labels. Specifically, the correct audio option was contrasted with an audio label in which the trained morpho-phonological change was placed in the wrong position within the word. Following this, participants completed a brief-post task questionnaire consisting of short questions regarding their linguistic and musical background.

2.3 *Stimuli* The stimuli were created by pseudo-randomly combining segments from a subset of the Cantonese inventory (consonants: /p, k, t/, vowels: /i, ɛ, a, ɔ, u/, corresponding nasal set: /m, ŋ, n/, corresponding aspiration set: /pʰ, tʰ, kʰ/). Tonal stimuli involved include a high-level tone (55), mid-level tone (33) and a high rising tone (24). All the words in the stimulus set consist only of syllables without codas, and none of the multisyllabic stems are real words in Cantonese.

During the two training sessions and Test 1, participants were exposed to and tested on nine nonce words, including three three-syllable words, three four-syllable words, and three five-syllable words. In Test 2, participants were tested on 18 nonce words, consisting of six three-syllable words, six four-syllable words, and six five-syllable words. All singular forms carried a mid-level tone (33). The correct plural forms varied across conditions. In the Tone/Nasal-Counting condition, the correct plural form was marked by assigning a high-level tone (55) or a [+nasal] feature to the third mora or consonant of the word. In the Tone/Nasal-Non-counting condition, the high-level tone (55) or [+nasal] feature was instead assigned to the initial mora or consonant of the word.

Incorrect plural forms were designed differently for two tests. In Test 1, across all conditions, three of the incorrect forms were singular forms with no phonological alternation, while six were hypothetical plural forms that involved an untrained phonological change applied to the correct position in the word. In the tonal conditions, the incorrect phonological change was a high rising tone (24) rather than the expected high-level tone (55). In the nasal conditions, the incorrect change involved the feature [+spread glottis] instead of the expected [+nasal]. In Test 2, incorrect plural forms were created by assigning the target high-level tone or [+nasal] feature to the wrong position, specifically to the second mora or consonant rather than the intended location.

In total, we created 45 singular-formed words, 45*4 correct plural-formed words, 6*4 incorrect plural-formed words with wrong featural change and 18*4 incorrect plural-formed words with wrong positional change. All audio labels were recorded by a native Cantonese speaker and were recorded in one session in a sound-proof booth at the MIT Phonetics Lab.

3 Results

3.1 *Test 2* We first analyze the 18 test trials in Test 2, which were the crucial test trials in our design that examines participants' learnability of rules that involves a counting mechanism. Figure 2 illustrates the total number of correct responses on Test 2 for Experiment 1, grouped by feature type and counting condition. Participants in the *counting–nasal* condition achieved the lowest overall accuracy. In contrast, performance in both *non-counting* conditions and the counting-tone condition was noticeably higher, with the *non-counting–tone* condition showing the highest number of correct responses overall. The distribution also appears wider for the *counting–nasal* group, suggesting greater variability among participants in that condition.

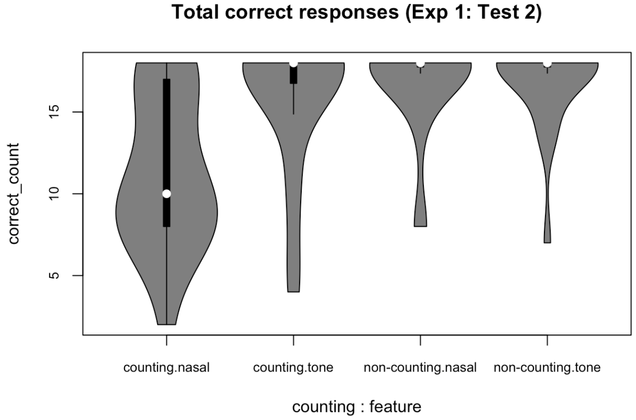

Figure 2. Total number of correct responses on Test 2 (Experiment 1), grouped by feature type (nasal vs. tone) and counting condition (counting vs. non-counting).

This result tentatively suggests that participants' ability to learn a nasal-related phonological rule, but not a tone-related rule, is influenced by the integration of a counting mechanism.

To confirm the observation above through statistical analysis, a series of generalized linear mixed-effects models, implemented in R (Bates et al., 2015), were built incrementally and compared to examine the effects of counting condition and feature type on the binary outcome variable Match (e.g., whether participants' choice match with the expected correct choice). In other words, instead of looking at the total correct responses on Test 2 trials previewed in Figure 1, the independent variable for the statistical analysis is predicted probability of getting each trail correct. Following the pre-registration, all models contained random intercepts for participants (coded as ID) and trials (coded as Trial) and they progressively increase in complexity, allowing for an assessment of whether adding predictors improves model fit.

We started with a Base Model. It only includes random effects ((1 | ID) + (1 | Trial)) but no fixed effects (i.e., no counting or feature predictors), assuming that variation in the dependent variable correct/incorrect is purely predicted by individual differences (ID) and trial effects. It serves as a null model (baseline reference) for later model comparison and selection. Following the base model, I fitted Model 1 (*match* ~ *counting* + (1 | ID) + (1 | Trial)), with the independent variable sum-coded. A significant effect of counting on accuracy is revealed: 1) the intercept ($\beta = 4.18$, $z = 6.76$, $p < .001$) represents the grand mean log-odds of a getting correct response across both counting conditions; 2) the negative coefficient for counting1 ($\beta = -1.18$, $z=-3.68$, $p < 0.001$) indicates that the counting condition leads to significantly lower accuracy compared to the overall mean. Model 2 (Match ~ Counting + Feature + (1 | ID) + (1 | Trial)) was designed to include feature as the other fixed effect and it found that both counting and feature significantly influence accuracy. The negative coefficient for counting1 ($\beta = -1.53$, $z = -4.30$, $p < .001$) indicates that the counting condition significantly reduces accuracy compared to the grand mean. Similarly, the negative coefficient for feature1 ($\beta = -0.85$, $z = -2.66$, $p < .01$) suggests that nasal features result in lower accuracy compared to the overall mean. Model 3 (Match ~ Counting*Feature + (1 | ID) + (1 | Trial)) that I built was to add additional interaction effects to Model 2. It not only confirms significant main effects of counting ($\beta = -1.46$, $z = -4.04$, $p < 0.001$) and feature ($\beta = -0.83$, $z = -2.6$, $p < 0.01$) but also indicates that a significant interaction term ($\beta = -0.78$, $z = -2.002$, $p < 0.05$), suggests that the negative effect of the counting condition is amplified when the feature is nasal, indicating an interaction between the two predictors (see Figure 3 for a straightforward visual review of the results).

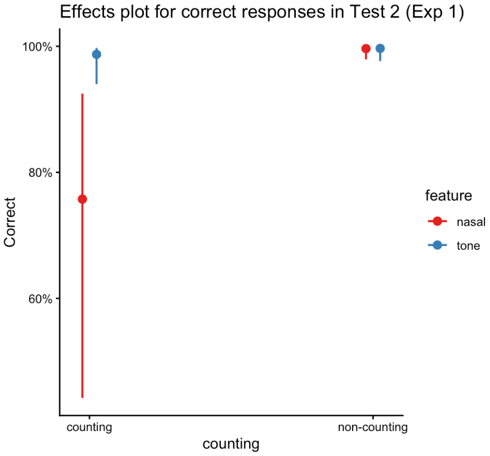

Figure 3. Effects plot for correct responses in Test 2 conditioned by Counting and by Feature

We further conducted a model comparison using the likelihood ratio test (ANOVA), which compare four progressively complex mixed-effects logistic models to assess the impact of counting and feature and their interaction on accuracy. The results showed that Model 1 (adding the predictor counting) significantly improves fit over the base model ($\chi^2(1) = 15.83$, $p < 0.001$), indicating that counting is a strong predictor of accuracy. Model 2 (adding the predictor *feature*) further improves fit ($\chi^2(1) = 7.37$, $p < 0.01$), suggesting that feature also plays a significant role. Model 3 (adding *counting * feature* interaction) results in a modest but significant improvement over Model 2 ($\chi^2(1) = 4.12$, $p < 0.05$), indicating that the effect of counting depends on feature.

The interaction effects between counting condition and feature type can be additionally confirmed by a pairwise analysis. The estimated marginal means (EMMeans) results indicate significant differences in the predicted probability of correct responses across conditions. Non-counting conditions for both tone and nasal features show near-ceiling accuracy (0.996 and 0.997, respectively), whereas the counting-nasal condition has the lowest accuracy (0.757) with the highest standard error (0.129). Counting-tone accuracy (0.987) is slightly lower than its non-counting counterpart but remains high. The contrast analysis reveals that the counting-nasal condition is significantly different from all other conditions, with notably lower accuracy compared to non-counting nasal (odds ratio = 0.0115, $p < 0.001$) and counting-tone (odds ratio =

0.0403, $p < 0.01$). No significant differences emerge between non-counting nasal and non-counting tone ($p = 0.9996$) or between counting-tone and non-counting tone ($p = 0.5562$).

3.1 *Test 1* Unlike Test 2, which specifically assessed participants' ability of learning counting, Test 1 consisted of nine trials using two types of forced-choice questions. Three trials evaluated whether participants could detect that pluralization in the language involved at least some phonological change. The remaining six trials tested whether participants could distinguish the specific trained pluralization change from an untrained one. In this experiment, participants were required to achieve at least 80% accuracy in Test 1 to directly proceed to Test 2. Those who did not meet this threshold were prompted to redo Test 1 before moving forward. Our analysis of the Test 1 phase focuses on participants' redo rates across four conditions, examining whether redo rates are influenced by feature type (tone vs. nasal) and counting condition (counting vs. non-counting).

A mixed-effects logistic regression model ($\text{Redo} \sim \text{feature} * \text{counting} + (1 \mid \text{ID})$) was used to examine the effects of counting condition (counting vs. non-counting) and feature type (tone vs. nasal) on the binary outcome variable Redo (Yes/No). The results indicate that feature type significantly influenced redo rates, with participants in the nasal condition being more likely to redo Test 1 compared to those in the tone condition ($\beta = 0.54$, $z = -2.10$, $p < .05$), suggesting that the Cantonese participants are more perceptually sensitive to tonal contrasts than nasal contrasts. In contrast, while it has a slightly lower redo rate, but non-counting condition is not significantly different from the counting condition ($\beta = 0.15$, $z = -0.61$, $p = 0.54$), which aligns with expectations since counting was not specifically tested in Test 1. The results additionally suggest that while learners may ultimately acquire tone and nasal rules to a similar extent in the non-counting condition, they did not learn the two rule types equally well, possibly constrained by certain perceptual factors/biases. Figure 4 illustrates how the Test 1 redo rate is affected by counting condition and by feature type.

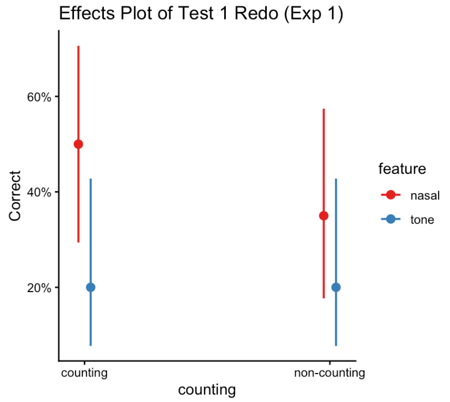

Figure 4. Effects plot of Redoing Test 1 conditioned by Counting and by Feature

4 Discussion

From the results above, we can identify two broad trends in how learners acquire patterns:

- Segmental feature assignment rules become particularly challenging in the counting condition, in contrast to corresponding tonal rules.
- Learners generally find tonal feature assignment rules easier to acquire than the corresponding segmental ones, regardless of whether counting is involved.

The first finding is consistent with the hypothesis that the typologically unattested segmental counting rule is more difficult to learn than the three attested rules (tone counting, tone non-counting, and segmental non-counting). This pattern of results aligns with the typological profile of tone versus segment in phonological counting, and is consistent with prior findings that typological asymmetries may be cued by learning biases (as reviewed in Section 1). In addition, referring back to the second generalization above, we found that although both tone and segmental non-counting rules can be learned to a similar extent, rules involving these two feature types are not learned equally well, with tonal rules being learned moderately better than segmental ones. This may further indicate that learners tend to acquire tone-based rules more easily than those involving segments, irrespective of the presence of counting. In what follows, I will first zoom in to discuss the potential factors that make tonal counting easier to learn and then zoom out to discuss how this line of learnability results inform the long-standing issue of tone-segment asymmetry.

4.1 *Zooming in: Why is tonal counting easier than segmental counting?* The Recent studies on phonological learning have increasingly focused on investigating the presence and characteristics of possible learning biases. Two such biases that have received significant attention in the literature and are relevant to the current study are substantive bias and structural bias. Substantive bias, rooted in phonetically based phonology (e.g., Hayes et al., 2004; Steriade, 2008; White, 2017; Glewwe, 2022), proposes that phonetic substance – such as knowledge related to articulation and perception – can be utilized in synchronic phonological learning4 (Wilson, 2006). Learners are argued to have an a priori preference for phonological patterns that align with phonetic principles. For example, patterns that require less articulatory effort or have lower perceptual cost are more readily learned. However, this type of bias is not absolute and may be overridden by linguistic input that does not conform to phonetic expectations.

Structural bias, also known as complexity bias, suggests that more complex patterns are harder to learn than simpler ones (Moreton & Pater, 2012a, 2012b). Here, the notion of complexity is typically defined in terms of phonological features. A phonological generalization may be considered more complex if it: 1) targets natural classes defined by a greater number of phonological features (e.g., Saffran & Thiessen, 2003; Skoruppa & Peperkamp, 2011), 2) requires more featural changes during the application of a phonological process (e.g., Skoruppa et al., 2011), or 3) involves a higher degree of featural contingency (e.g., Moreton, 2008).

On the one hand, I argue that structural bias may crucially contribute to a preference for tonal feature assignment rules over segmental ones. In the following, I consider the simplicity in tone from two perspectives: features and representational properties. First, in terms of features, a tonal feature assignment typically involves only one feature change during a phonological process, while a single segmental feature assignment often results in multiple changes (and corresponding repairs) in the phonological representation. For instance, assigning a nasal feature to a voiceless obstruent results in changes not only in nasality, but also in voicing, sonority etc. This featural consideration can be intuitively translated to a corresponding phonetic interpretation: manipulating a tonal feature typically involves modifying only a single acoustic dimension (i.e., F0), whereas changes to a segmental feature often result in alterations to additional acoustic correlates (e.g., various higher formants, duration, etc.). Tone is also simpler in terms of the size of its feature inventory: a tonal feature set are generally smaller than a segmental one, resulting in less complexity and fewer feature-based interactions compared to segmental features. Often times, one single tonal feature can suffice to specify one unique toneme and even when they interact with each other, tonal features may simply form complex contours without triggering any featural contradiction. On the other hand, one single segmental feature can define a natural class of segments, and we may often encounter featural incompatibilities during their interactions, potentially leading to repair strategies or non-application of the rule. Second, in terms of representational properties, one can also reason that tonal features operate on a structurally simpler tier, whereas segmental features are embedded within a more complex representational structure. While one might argue that a single segmental feature could, in principle, have its own representational tier – similar to how tone operates independently, we have a larger set of segmental features serving as defining primitives for phonemes in a given language. That's said, even if we posit a dedicated tier for a segmental feature, it would still operate alongside numerous other tiers, due to the inherent complexity of phonemic representation. Therefore, the intrinsic structural properties outlined above collectively favor the interpretation that segmental feature assignment rules are inherently more difficult than their tonal counterparts.

On the other hand, given the ceiling effects observed in participants' performance across the two tonal conditions, we may also consider the possibility that the relative difficulty in learning a segmental feature assignment rule stems from the greater perceptual salience of tonal features compared to consonantal ones (e.g., Liu et al., 2024). This interpretation aligns with a substantive bias, which can be understood through

---

[4] Biases rooted in phonetic substance are often examined in relation to diachrony and language transmission, where they are referred to as channel biases (e.g., Blevins, 2004, Moreton & Pater, 2012b). This perspective suggests that phonological systems can develop over time in ways that mirror phonetic tendencies, even if those phonetic influences are not active in the process of synchronic phonological learning. This type of bias is beyond the evaluation of the current study.

the phonetic principle that phonological processes tend to favor patterns that minimize perceptual effort – that is, more salient cues are easier to process than less salient ones. While it may be noted that learners performed similarly in the non-counting nasal condition and non-counting tonal conditions, suggesting that tonal and nasal features may be equally perceptible, this comparison is complicated by the fact that both conditions reached ceiling performance. As such, any perceptual advantage of tonal over segmental features may not be detectable in the current results[5] (see Figure 3). A follow-up study (Cui et al., in prep) based on this salience hypothesis tested whether the current results generalize to other segmental features – especially vowel features, which have been argued to be more essential than consonants for rule extraction (Bonatti, 2005; 2007) and perceptually more salient (Liu et al., 2024). The results show that learning performance for vowels is trending toward that for consonants. This further indicates that the observed learnability asymmetry reflects a broader, universal learning bias between tone and segment.

4.2 *Zooming out: How do the learnability results inform the issue of tone-segment asymmetry?*

Previous approaches to tone–segment asymmetry have largely appealed to either computational or representational accounts. Within the framework of formal language theory, it has been argued that tone is computationally more complex than segmental structure (Jardine, 2016), in that only tonal phonology extends beyond the class of weakly deterministic maps. This line of argument, however, has faced counterarguments. McCollum et al. (2020) provide an empirical challenge to Jardine's claim by examining unbounded circumambient processes in the segmental domain, arguing that segmental phonology requires the same level of expressive power as tonal phonology. Pater (2018) further questions the plausibility of treating formal language theory as a stand-alone alternative to Optimality Theory (or other grammatical frameworks), emphasizing that substantive bias is indispensable for understanding and constraining phonological typology.

Against the backdrop of these theoretical debates, the present study provides converging evidence that tone-segment asymmetry in typology is not illusory, as it is reflected in systematic synchronic learning biases observed in our experimental results. These learnability findings pose a potential challenge to purely computational accounts, raising the question of why phonological objects claimed to be more complex would nonetheless be easier to learn. At the same time, our results leave open a central unresolved issue: how best to formalize and phonologize the substantive biases that may underlie the observed learning asymmetries.

5 Conclusion

Tone-segment asymmetry has been a frequently discussed topic in phonology, with much theoretical work dedicated to characterizing the differences between the two. The present study provides the first experimental evidence, using an artificial learning paradigm, to investigate whether the typological asymmetry observed in phonological counting correlates with a learnability profile. The results show that the learnability contrast between tone and segment in counting conditions mirrors the typological asymmetry: tonal patterns are more easily learned than segmental ones. Future studies are needed to investigate whether similar learning asymmetries extend to other phenomena associated with tone-segment asymmetry, and to formalize the substantive biases that may contribute to observed learning and typological patterns.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, 67, 1-48.

Bendor-Samuel, J. T. (1960). Some problems of segmentation in the phonological analysis of Tereno. *Word*, 16(3),

---

[5] Another contributing factor that should not be overlooked is the perceptual salience of the word-initial position, which characterizes the non-counting conditions. The prominence of this edge position may have driven ceiling-level performance in both non-counting tone and segment conditions, thereby masking any observable differences in perceptual salience between tone and segment.

348-355.

Bickmore, L. (2007). *Cilungu phonology*. CSLI Publications.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6), 451-459.

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, 18(10), 924-925.

Conklin, J., Stork, E., Bastimar, R., He, X., & Conklin, J. T. (2024). Impacts of direction and morphological locus of control on learnability of sibilant harmony. *Laboratory Phonology*, 15(1).

Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass*, 6(5), 310-329.

Finley, S. (2008). *Formal and cognitive restrictions on vowel harmony* (Doctoral dissertation). Johns Hopkins University.

Finley, S. (2012). Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and cognitive processes*, 27(10), 1550-1562.

Glewwe, E. (2022). Substantive bias and the positional extension of major place contrasts. *Glossa: a journal of general linguistics*, 7(1).

Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically Based Phonology*. Cambridge University Press.

Huang, T., & Do, Y. (2023). Substantive bias and variation in the acquisition of vowel harmony. *Glossa: a journal of general linguistics*, 8(1).

Hyman, L. M. (1998). Positional prominence and the 'prosodic trough' in Yaka. *Phonology*, 15(1), 41-75.

Hyman, L. M. (2011). *Tone: Is it different?.* In The Handbook of Phonological Theory (pp. 197-239).

Hyman, L. M. (2023). Initial/Final Tone Agreement in Ekegusii (Bantu; Kenya). *Anuario del Seminario de Filología Vasca" Julio de Urquijo"*, 57(1-2), 441-456.

Kurisu, K. (2009). Palatalisability via feature compatibility. *Phonology*, *26*(3), 437-475.

Liu, Z., Chitoran, I., & Turco, G. (2024). Perceptual salience of tonal speech errors. In *Speech Prosody 2024* (pp. 427-431). ISCA.

Marlo, M. R., Mwita, L. C., & Paster, M. (2015). Problems in Kuria H tone assignment. *Natural Language & Linguistic Theory*, 33, 251-265.

Martin, A., & White, J. (2021). Vowel harmony and disharmony are not equivalent in learning. *Linguistic Inquiry*, 52(1), 227-239.

McCarthy, J. J. (1983). Consonantal morphology in the Chaha verb. In *Proceedings of the West Coast Conference on Formal Linguistics*, 2, 176-188.

McCollum, A. G., Baković, E., Mai, A., & Meinhardt, E. (2020). Unbounded circumambient patterns in segmental phonology. *Phonology*, 37(2), 215-255.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, 25(1), 83–127.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning. Part I: Structure. *Language and Linguistics Compass*, 6(11), 686–701.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning. Part II: Substance. *Language and Linguistics Compass*, 6(11), 702–718.

Morley, R. L. (2018). Is phonological consonant epenthesis possible? A series of artificial grammar learning experiments. *Phonology*, 35(4), 649-688.

Paster, M. (2019). Phonology counts. *Radical: A Journal of Phonology*, 1, 1-61.

Pater, J. (2018). Substance matters: a reply to Jardine (2016). *Phonology*, 35(1), 151-156.

Rolle, N., & Bickmore, L. (2022). Outward-sensitive phonologically-conditioned suppletive allomorphy vs. first-last tone harmony in Cilungu. *Morphology*, 32(2), 197-247.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494.

Skoruppa, K., & Peperkamp, S. (2011). Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science*, 35, 348–366.

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The role of phonetic distance in the acquisition of phonological alternations. In S. Lima, K. Mullin, & B. Smith (Eds.), *Proceedings of the 39th North Eastern Linguistics Conference* (pp. 717–729). Cascadilla Press.

Steriade, D. (2008). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–180). MIT Press.

Stevick, E. W. (1969). Pitch and duration in Ganda. *Journal of African Languages*, 8, 1-28.

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language*, 93(1), 1-36.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar

palatalization. *Cognitive science*, 30(5), 945-982.
Yip, M. (2002). *Tone*. Cambridge University Press.
Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832