Open Library of Humanities

On the learnability of level-based and unit-based tonal OCP generalizations: An artificial grammar learning study

Tsung-Ying Chen, National Tsing Hua University, TW, chen.ty@mx.nthu.edu.tw

In previous research of tonal phonology, contour tones were assumed to be phonologically presented not only as a sequence of level tones but also as a single unit. Obligatory Contour Principle (OCP) constraints, by extension, were thought to prohibit identical level tones associated to adjacent terminal nodes (OCP-Terminal) and identical adjacent tonal units (OCP-Unit). This proposal has nevertheless been challenged theoretically and empirically. This study seeks to offer new experimental evidence to help solve the debate by comparing the learnability of the two OCP generalizations in an artificial grammar learning paradigm. In Exp I, we exposed disyllabic tonal patterns conforming either to OCP-Terminal or to OCP-Unit to two target groups of learners, who were then tested if they could extend the hidden generalization to their auditory acceptability judgment of novel disyllabic items. Exp II had the same exposure phase but required learners to produce novel disyllabic items with tonal combinations of their choice. In both experiments, the two target groups demonstrated signs of learning target OCP generalizations from different channels. Crucially, learners implicitly acquired the OCP-Terminal generalization but only learned the OCP-Unit pattern as explicit knowledge. These findings led us to conclude that OCP-Unit may not be involved in implicit and automatized phonological computation.

1 Introduction

The last 60 years have seen advancement in our understanding of tonal phonology following the transition from structural (e.g., Wang 1967; Woo 1969) to generative and autosegmental approaches (Goldsmith 1979), which were later incorporated into a constraint-based framework (e.g., Myers 1997; Chen 2000; Yip 2002). However, the debate over whether level tones and contour tones can be *both* phonologically represented as a single constituent has never truly settled.[1]

In a tonal geometry that we refer to as a level-based system, level tones are always independent phonological entities. Contour tones are thus assumed to consist of two level tones with separate root nodes autosegmentally linked to a tone-bearing unit (TBU), such as the rising tone LH presented in **Figure 1a**. In a unit-based system adopted from Yip (1989) in **Figure 1b**,[2] level tones and contour tones are both represented as single constituents, whose terminal level tones are governed by one single root node.[3]
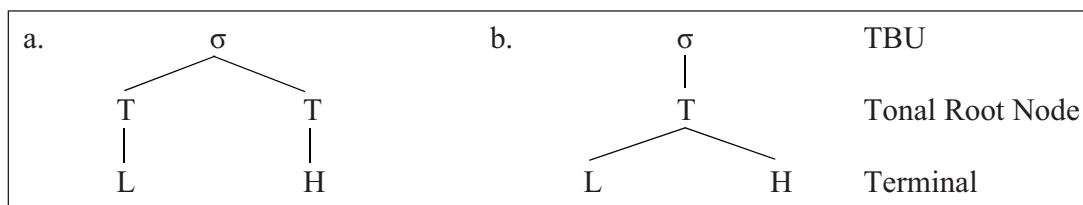
Figure 1: An autosegmental representation of LH as (a) a cluster of level tones associated to separate root nodes and (b) a constituent with one root node governing all terminal level tones.

With the level-based and unit-based views, we could assume two distinct types of Obligatory Contour Principle (OCP; Leben 1973; McCarthy 1986) that prohibits identical adjacent tonal entities:[4] One bans same adjacent terminal level tones, or *OCP-Terminal*, while the other forbids adjacent root nodes to have the identical terminal level tones in the same linear order, or *OCP-Unit*, as defined in (1). In a level-based tonal model, the two OCP constraints predict the same tonal gaps (e.g., *H-H, *H-HL, *L-LH) since each tonal unit is equivalent to a terminal level tone.

---

1 Similar debates over the phonological representation of segmental combinations persist as well. Readers are referred to Lin (2011) and Berns (2016) for an overview of cluster-based vs. unit-based views on affricates, and an attempt to explain the representational nature of nasal-stop sequences could be found in Downing (2005).

2 Note that the root node and the terminal nodes in Yip's (1989) model are in fact specified with the feature [upper] and [raised] respectively. In the current review, we use H and L consistently for a more direct comparison between the level-based and unit-based models.

3 We chose not to discuss tonal units represented with single features such as [rise] and [fall] in linear feature matrices (e.g., Wang 1967) or with an additional autosegmental tier that separates a contour node from a register node (e.g., Bao 1990; 1999). The former would be more restricted than non-linear approaches in terms of explaining tonal patterns, and the latter is computationally equivalent to Yip's unit-based model (Oakden 2020).

4 For non-dissimilatory tonal processes predicted to be possible in level-based and unit-based tonal models, see Yip (1989), Duanmu (1994), Chen (2000, §2), and Chen (2010).

Thus, in a level-based tonal model, it is necessary to assume only OCP-Terminal as part of tonal phonology. However, the distinction between the two OCP constraints becomes non-trivial in a unit-based model, in which same contour tones (e.g., *HL-HL, *LH-LH) violate only OCP-Unit by having identical terminal level tones linked to adjacent root nodes.

(1)   OCP-Terminal vs. OCP-Unit
    a. OCP-Terminal: Adjacent terminal level tones are forbidden.
    b. OCP-Unit: Adjacent root nodes with same terminal tonal specifications are forbidden.

The controversies with the unit-based view predicting two partially overlapping tonal dissimilatory processes are twofold (see also §2). First, tonal dissimilatory patterns assumed to be best explained in a unit-based model using OCP-Unit are usually subject to an alternative, level-based analysis. Furthermore, a possible tone language in a unit-based model that only obeys OCP-Unit remains unattested. These uncertainties have led many theorists to question the necessity to posit OCP-Unit given its limited explanatory power and many potential theoretical problems it has raised. The debates over the nature of tonal OCP and by extension, the phonological representation of tones, have nevertheless ended with a stalemate as the discussion hardly goes beyond analyses of impressionistically transcribed dictionary data. Experimental studies are thus the key to shed new light on the issue and ultimately explain the nature of phonological tones (Zhang 2010; 2014).

In the current study, we investigated experimentally if OCP-Unit in a unit-based model is part of unconscious and automatized phonological knowledge by studying the implicit learnability of OCP-Unit patterns in two artificial grammar learning (AGL) experiments (Reber, 1967). Our research hypothesis is that patterns conforming to OCP-Terminal is an implicitly learnable phonological generalization as they have been observed in typologically diverse tone languages. Then, if a unit-based representation of tones is also computed at an abstract level of phonology and OCP-Unit is part of tonal phonology, OCP-Unit should be equally learnable as an implicit generalization. In the two experiments, we first exposed naïve learners to auditory input that strictly obeyed either OCP-Terminal or OCP-Unit in a brief training session. We then asked the learners to judge or produce novel items to test if they had acquired the target abstract knowledge and could extend it successfully. Contrary to the null hypothesis that both OCP generalizations are implicitly learnable, our experimental results indicated that OCP-Unit played at best a marginal role in the learners' test performance. The current study and its findings are expected to not only breathe new life into the discussion of the fundamental issues in tonal phonology, but also supply a rare example of tonal AGL experiments with methodological improvements.

In the rest of the article, we will continue to provide a more in-depth literature review in §2, elaborate on the experimental design and the results of the two AGL experiments respectively in §3 and §4, and finally explore residue issues in a general discussion in §5.

2. Tonal OCP: From theoretical debates to experimental investigation

In tonal phonology, the evidence of OCP-Terminal has been firmly established as many languages appear to avoid juxtaposing same level tones in the output. Previous studies have also found OCP-Terminal to target either only H sequences (e.g., Clark 1989; Mmusi 1992; Aranovich 1994; Myers 1997; Downing 2003; Jenks & Rose 2011) or only L sequences (e.g., OCP-L; Clark 1990; Daly & Hyman 2007; Salffner 2010). It is thus plausible to further divide OCP-Terminal into OCP-H and OCP-L. OCP-Terminal provides a simple account of interactions between level and contour tones that result in tonal alternations in surface representations, as demonstrated in the data of Luba (Hyman 2007: 11–12) in (2). In (2a), the input LH-H has to be changed to L-H as the two underlyingly adjacent terminal Hs would violate OCP-Terminal in a faithful output. Likewise in (2b), the rising tone LH in the input is simplified as H in the output so there exists no surface L sequence that would violate OCP-Terminal. Finally, in (2c), when the rising tone LH is adjacent to a terminal level tone that could create an H or L sequence on both ends, the rising contour is simplified as L, rather than H. This output preference implies that an L sequence is more tolerable than an H sequence in Luba (Hyman 2007: 11–12), which could further support the independence of OCP-H and OCP-L.[5]

(2)   Tonal dissimilation in Luba; bolded tones are deleted in the output
        a.   L**H** + H → L-H
        b.   L + **L**H + L → L-H-L
        c.   L + L**H** + H → L-L-H

Contrary to the abundant evidence of OCP-Terminal found in typologically diverse languages, the trace of OCP-Unit has been largely observed in Chinese tone languages, which are renowned for their rich contour tone system and complex tone sandhi patterns. The Tianjin dialect is among the instances that has been frequently discussed in the literature.[6] Chen's (2000) review of Tianjin tone sandhi includes four primary disyllabic patterns listed in (3).[7] The tonal epenthesis and deletion in (3a–b) could eliminate the underlying L sequence and avoid the potential OCP-Terminal violation.[8] The contour tone simplification process in (3c–d) are necessary presumably

---

5   In a constraint-based framework (e.g., Prince & Smolensky 2004), this stronger bias against adjacent Hs could be captured with the ranking OCP-H » OCP-L » MAX-T.

6   See also Lin (2011; 2019), Hsiao (2015), and Wee (2019) for the operation of OCP-Terminal in other Chinese tone languages. In these studies, OCP-Terminal and OCP-Unit are commonly referred to as OCP-t (t = terminal node) and OCP-T (T = tonal root node) respectively.

7   It is important to note that substantial diachronic changes in Tianjin tone sandhi have been found in more recent phonetic studies, including Zhang & Liu (2011; 2016), Li & Chen (2016), and Li et al. (2019), perhaps due to a close language contact with Standard Chinese. Nevertheless, one could still argue that the unit-based analysis applies to the original tone sandhi patterns before the onset of the diachronic changes.

8   We are aware of the perceptual basis of (3a) (i.e., tonal absorption; Hyman 2007: 12). However, the perceptual force may coincide with an abstract, symbolic identity restriction (e.g., Berent et al. 2002; Berent 2013), which could altogether facilitate the learning of the tonal alternation.

because two identical and adjacent contour tones in the output would violate OCP-Unit (e.g., Lin 2008; Wee 2015). OCP-Terminal is also at play to determine the output of (3c–d), which never includes an H or L sequence. Wee (2004) claimed to discover two additional tone sandhi processes (4) related to OCP-Terminal, in which an underlying H sequence is avoided in the output (cf. Zhang & Liu 2011). In sum, Tianjin tone sandhi seems to support the operation of both OCP-Terminal and OCP-Unit in tonal phonology.

(3)   Four primary Tianjin tone sandhi patterns (Chen 2000: 105–106)
        a.   L + L → LH-L
        b.   HL + L → H-L
        c.   LH + LH → H-LH
        d.   HL + HL → L-HL

(4)   Two extra Tianjin tone sandhi processes from Wee (2004)
        a.   LH + H → L-H
        b.   LH + HL → L-HL

Although introducing OCP-Unit into a unit-based tonal model may help complete an analysis of dissimilatory tonal patterns, it also raises theoretical issues. Crucially, a unit-based analysis is frequently subject to an alternative analysis motivated by phonetic naturalness without an arbitrary and complex grammar construction. For instance, dissimilation between contour tones in (3c–d), with fewer tonal ups and downs in the outputs, could simply be the consequence of reducing overall articulatory complexity (e.g., Hyman 2007: 16–18; Hyman & Vanbik 2004). Another case where an analysis is in fact complicated by the inclusion of OCP-Unit could be seen in Wee's (2015) constraint-based analysis of Tianjin tone sandhi (**Table 1**). The goal of the analysis is to solve a dilemma caused by the surface tonal sequence HL-LH; if the avoidance of adjacent Ls in (3a–b) is due to a top-ranked OCP-Terminal, it is not possible to explain why HL-LH does not undergo tone sandhi. Wee thus proposed a ranking in which OCP-Terminal is ranked below faithfulness constraints and analyzed (3a) as an alternation driven by a top-ranked OCP-Unit.[9] The HL-LH sequence, without violating OCP-Unit, is now acceptable because as single constituents, the two adjacent contour tones are not identical at the root level. The only pattern left unexplained is (3b), since the low-ranked OCP-Terminal itself cannot be held accountable for the derivation HL + L → H-L. To tackle this issue, Wee proposed a top-ranked local conjunction constraint that binds Head-Tone-Complexity (HTC) in (5) and OCP-Terminal (i.e., HTC & OCP-Terminal); with this constraint, the underlying tonal sequence in (3b) HL + L cannot surface faithfully since the output not only has adjacent Ls

9   Note that the diachronic changes of Tianjin tone sandhi documented in recent studies (see fn.7) do not necessarily undermine Wee's OCP account of the tonal alternations. For example, Zhang & Liu (2011) found that the output of L + L is H-L rather than LH-L, but the analysis that OCP-Unit drives the alternation of L + L remains valid.

(i.e., OCP-Terminal violation) but also a more complex head tone HL in the right-headed tonal sequence (i.e., HTC violation).

(5) Head-Tone-Complexity (HTC): Non-head tones must not be more complex than head tones. Tonal complexity hierarchy = Rising » Falling » High » Low

Input: /L+L/	OCP-Unit	HTC & OCP-Terminal	DEP/MAX	OCP-Terminal
L-(L)	*!			
☞LH-(L)				
Input:/HL+L/	OCP-Unit	HTC & OCP-Terminal	DEP/MAX	OCP-Terminal
HL-(L)		*!		
☞H-(L)				
Input: /HL+LH/	OCP-Unit	HTC & OCP-Terminal	DEP/MAX	OCP-Terminal
☞HL-(LH)				
H-(LH)			*!	

Table 1: An excerpt of Wee's (2015) analysis of Tianjin tone sandhi; () = prosodic head.

Concerns over local conjunction aside (e.g., McCarthy 2002: 18–19), a phonetically based analysis without OCP-Unit in **Table 2** is possible. First, one could analyze the alternation in (3a-b) with a top-ranked OCP-Terminal that forces an output to avoid any surface H or L sequence. The preservation of the underlying HL+LH sequence in the output could be explained by a lower articulatory complexity of HL-LH compared to that of other output candidates avoiding adjacent Ls. For instance, if the initial HL in HL+LH is reduced to H as in the output H-LH, the articulatory gestures must be adjusted rapidly from the offset of the initial H to the onset of the following LH to leave enough time to fully realize the rising contour. By contrast, the output in (3a) (i.e., HL+L → H-L) does not require the same acute gestural change; a simple low-pitch target to be realized after the initial H could be reached even with a slower gestural transition. This difference in articulatory complexity can be captured by a markedness constraint NoAcuteChange (NAC) defined in (6), which is violated by H-LH but not H-L.[10]

(6) NoAcuteChange (NAC): Acute changes in the articulatory gesture are prohibited.

---

10 The tonal alternation (4b) (i.e., LH+HL → L-HL) could be attributable to a constraint that bans non-domain-final rising tones (e.g., Zhang 2002; 2004; et seq.), which is productive in modern Tianjin (e.g., Zhang & Liu 2016).

Input: /L+L/	OCP-Terminal	NAC	DEP/MAX
L-L	*!		
☞LH-L			
Input: /HL+L/	OCP-Terminal	NAC	DEP/MAX
HL-L	*!		
☞H-L			
Input: /HL+LH/	OCP-Terminal	NAC	DEP/MAX
☞-LH	*		
H-LH		*	*!

Table 2: A possible phonetically-based reanalysis of Wee's (2015) analysis in Table 1.

Our goal is not to offer a full reanalysis of any particular tone sandhi pattern but to highlight the conflicts in the analyses of tonal dissimilation arising from opposite theoretical perspectives. Without new evidence, the analytical debates can hardly see a major breakthrough.

Following the above review, one may turn to typological survey for a new lead, as a grammar hypothesis generates predictions on possible languages. In a minimal constraint grammar assuming a unit-based representation of tones with OCP-Unit, OCP-Terminal, and a faithfulness constraint (Faith), there are three crucial rankings in (7) and each corresponds to a possible tone language. With a top-ranked Faith in (7a), a tone language would not demonstrate tonal dissimilation at all, such as Cantonese (Yue-Hashimoto 1972; Yip 2002: 174–178). A top-ranked OCP-Terminal in (7b) predicts a language with dissimilation only between adjacent terminal level tones, as is the case with Luba reviewed earlier in this section. Finally, the ranking (7c) predicts a tone language in which surface tonal sequences never include identical tonal constituents next to each other. According to Wee (2019: 167–168), however, this type of tone languages is yet to be discovered,[11] a gap that renders the unit-based view on tonal dissimilation questionable.

(7)　A crucial factorial typology with Faith, OCP-Unit, and OCP-Terminal
　　a.　Faith » {OCP-Unit, OCP-Terminal}
　　b.　OCP-Terminal » Faith » OCP-Unit
　　c.　OCP-Unit » Faith » OCP-Terminal

11　Wee (2019: 167–168) claimed that the Boshan dialect of Chinese (Qian 1993; Chen 2000: 165) may belong to the language type. However, the analysis of Boshan tone sandhi is also inevitably complicated by the diachronic development of tones and could not fully endorse a top-ranked OCP-Unit.

In sum, while analytical debates may revive and typological evidence will mount, an experimental test would also be useful for tapping into the psychological reality of unit-based tonal dissimilation and OCP-Unit. Thus, we put forth an experimental investigation in the current study focusing on the learnability of the grammar (7b) and (7c) in the unit-based tonal model. If tones could dissimilate as single constituents, both OCP-Terminal and OCP-Unit should be *synchronically* available at the level of abstract phonological computation to acquire the grammatical generalizations in (7b) and (7c). Our prediction is that (7b) with a top-ranked OCP-Terminal, supported by typological evidence, is a computable and learnable phonological generalization. The primary research question would be whether the grammar (7c) with a top-ranked OCP-Unit is equally learnable.

In this study, we attempted to answer the research question in AGL experiments (Reber 1967; et seq.). The experimental paradigm has been popularized to test if linguistic regularities hidden in learning input could be acquired as abstract generalizations and extended to novel forms, and whether these generalizations are acquired implicitly (i.e., without learners' awareness). A widely adopted AGL experimental design is to present minimally contrasting learning input to separate learner groups to investigate the *relative* learnability of distinct linguistic patterns.[12] In recent decades, this AGL paradigm has been extended to examine hypotheses regarding intrinsic inductive biases for phonological regularities that are computable, formally simpler, formally/ phonetically more natural, or typologically attested or more common (Pycha et al. 2003; Wilson 2003; 2006; Seidl & Buckley 2005; Peperkamp et al. 2006; Moreton 2008; Finley & Badecker 2009; Carpenter 2010; Finley 2012; 2015; 2017; Gallagher 2013; White 2014; White & Sundra 2014; Hayes & White 2015; Lai 2015; Martin & Peperkamp 2020; among many others). To our best knowledge, Kao (2017) and Chen (2020) have been so far the only two studies that investigated the relative learnability of tonal generalizations in the AGL paradigm.[13] In Kao's (2017) study, inductive biases were found for a contour formation rule that preserves the linear order of underlying tonal sequences and a tone retention rule that preserves an underlying L sequence. Chen's (2020) experimental results indicated a bias in favor of a phonetically natural tonal constraint over a phonetically unnatural one.

In our AGL study, we created two minimally contrasting disyllabic tonal patterns with four tones (H, L, HL, and LH) to compare the relative learnability of the two target OCP generalizations. In the OCP-Terminal language, surface tonal patterns were created without any sequence including adjacent identical terminal level tones. By contrast, disyllabic combinations in the OCP-Unit language never included pairs of identical level or contour tones. If both OCP-Unit and OCP-Terminal are part of learners' unit-based phonological grammar of tones, learners exposed to either language

---

12 For additional methodological advantages of this paradigm and its comparison with other AGL paradigms, see Hamrick & Sachs (2018).

13 Wang & Saffran (2014) and Caldwell-Harris et al. (2015) also investigated the learning of tonal combinations in AGL experiments, although inductive bias was not their main research focus.

should be able to converge on the respective target OCP generalization and extend the acquired generalization to novel test items. Unlike previous AGL experiments testing various inductive biases, we also included necessary awareness measures in our experimental design to disentangle the effects of implicit (or unconscious) learning from explicit (or conscious) learning (see also Moreton & Pertsova (2016)). Awareness measures are crucial since the two OCP generalizations are assumed to be part of online, automatized, and unconscious phonological knowledge, which is assumed to guide children in early language acquisition.[14] Thus, our experimental results should not only indicate a successful learning of the OCP generalizations, but also reveal the implicit nature of the acquired knowledge. In particular, for adult learners recruited for our study, AGL would be practically similar to L2 learning, in which explicit learning is highly influential (e.g., Krashen 1982; DeKeyser 2003; Ellis 2005; Ionin et al. 2009; Morgan-Short et al. 2012; Lichtman 2013; Hulstijn 2015). It is thus important not to misinterpret the outcome of explicit learning as evidence of implicit phonological knowledge, and the effects of explicit learning must be properly isolated. Ultimately, our research question could be formulated as (8), which could be answered with experimental results from two AGL experiments discussed in the rest of this article.

(8)   Are OCP-Terminal and OCP-Unit equally learnable as an implicit phonological generalization?

3 AGL experiment I with an acceptability judgment task

In Exp I, we tested if learners could acquire the target OCP generalizations implicitly after being briefly exposed to training input and extend these abstract generalizations to their auditory acceptability judgment of novel test items.

3.1 Participants

A total of 90 L1 speakers of Taiwan Mandarin enrolled as an undergraduate or graduate student at National Tsing Hua University in Taiwan were recruited for Exp I. None of them reported any learning or hearing impairment or majored in a field related to linguistics. The participants were randomly assigned to one of the three learner groups (Unit: male = 15, female = 15; Terminal: male = 12, female = 18; Control: male = 17, female = 13). The age of these 90 participants ranged from 20 to 34 years old (sd = 2.72) and the age means did not differ significantly across the three learner groups.[15] All the 90 participants reached an accuracy rate of 90% or higher in a random memory recognition task administered during the training session (see §3.3) and completed both training and test sessions in Exp I. All participants were paid 100 NTD for their participation.

---

14   See, for example, Demuth (1993) and Kappa & Papoutsi (2019), for the acquisition of OCP in child phonology.

15   10. Unit vs. Terminal: $t(58)$ = −0.158, p = .875; Unit vs. Control: $t(58)$ = −0.436, p = .665; Terminal vs. Control: $t(58)$ = −0.316, p = .753.

3.2 Materials

As briefly explained in §2, we created an artificial language composed of disyllabic sequences and manipulate tonal patterns in the training input for each of the three groups of learners. The consonants, vowels, and tones were all selected from Taiwan Mandarin, the L1 of the participants. The disyllabic sequences were combinations of simple CV syllables [pi, pu, pa, tʰi, tʰu, tʰa, ki, ku, ka, ni, nu, na, mi, mu, ma], in which C1/C2 and V1/V2 must not repeat (e.g., [pinu] and [kuma]; cf. *[ki**ni**] and *[tʰutʰa]). This step would have helped lower the chance of directing learners' attention incorrectly to any phonological generalization potentially related to consonant and vowel harmony. As a result, 120 disyllabic training items were generated for Exp I, which were associated with di-tonal patterns composed of H, LH, L, and HL. Note that the four tones corresponded to the tonal labels T1, T2, T3, and T4 in Taiwan Mandarin respectively. In addition, T3 was viewed as a low-level tone L, rather than as a dipping tone MLH as in most analyses of Standard Mandarin phonology, which coincided with the finding in Huang's (2017) phonetic investigation. Treating T3 as L also helped eliminate a potential perceptual confusion between the dipping tone MLH and the rising tone LH (e.g., Huang 2001; Fon et al. 2004; Liu & Samuel 2004) in our training input, which could undermine the learning of target tonal patterns.

For the Terminal group, eight di-tonal patterns without violating the target constraint (**Table 3**) were distributed pseudo-randomly across the 120 training items (i.e., 15 tokens for each of the eight tonal patterns). For the Unit group (**Table 4**), in addition to the four di-tonal patterns violating OCP-Unit (i.e., *H-H, *LH-LH, *L-L, and *HL-HL), another four di-tonal patterns were also left out from the training input as *accidental gaps* (H-HL, LH-L, L-H, and HL-LH). This step assured that both target groups processed the same number of di-tonal patterns during the training session (i.e., eight di-tonal types). The accidental gaps were carefully selected to avoid an asymmetrical distribution of the four tones (i.e., each tone appeared in each position twice in the eight available tonal patterns). Furthermore, four out of the eight available di-tonal patterns still served as the positive evidence against OCP-Terminal (e.g., LH-H and HL-L).

	H	**LH**	**L**	**HL**
H		H-LH	H-L	
LH		LH-LH	LH-L	
L	L-H			L-HL
HL	HL-H			HL-HL

Table 3: Di-tonal combinations without violating OCP-Terminal.

	H	LH	L	HL
H		H-LH	H-L	
LH	LH-H			LH-HL
L		L-LH		L-HL
HL	HL-H		HL-L	

Table 4: Di-tonal combinations in the training input of the OCP-Unit language.

The above difference between the two target languages may raise methodological concerns, such as whether the training input in the two target learning conditions was *qualitatively* matched. It might be in fact more challenging for the Unit group to acquire the target OCP generalization as the learners missed part of the acceptable input types. However, we assumed that the accidental gaps at best played a marginal role in the learning of an implicit generalization with a top-ranked OCP-Unit. In natural language acquisition contexts, learners do not have to be or may not have the chance to be exposed to all positive evidence for rejecting non-target linguistic generalizations (i.e., poverty of the stimulus; Chomsky 1980; et seq.). Crucially, in the training input for both target groups, there was no exception to the respective target OCP generalization, a design that should have made the two learning settings comparable. Even in a very unlikely case that the accidental gaps in the training input for the Unit group contained additional learnable segmental or tonal patterns, these patterns should have been learned independently from unit-based tonal dissimilation; that is, learners would have implicitly rejected test items with systematic patterns hidden in accidental gaps as well as tonal sequences violating OCP-Unit. Accordingly, we chose to prioritize the control of the number of input tokens to avoid between-group differences arising from the processing of two *quantitatively* distinct training sets.

To further minimize the gap between the two sets of training input, two additional steps in our stimulus design were needed. First, di-tonal patterns that were present in the training input for *both* target groups were always paired with the same disyllabic sequences. Second, di-tonal patterns different only in the initial or final tone were also paired with the same disyllabic sequences (e.g., Unit: [pi^{L}ku^{LH}] vs. Terminal: [pi^{L}ku^H]). Since the two sets of training input included four contrasting tonal patterns, half of the 120 training items were minimally different in their tonal sequence across the two target groups (see Appendix A). The 120 training items were not a homophone or a pseudo-homophone of a disyllabic lexical item in Taiwan Mandarin (or Taiwanese Southern Min, which might be the home language of some participants as well).

We also included a Control group exposed to random combinations of the four tones to measure the baseline performance with training stimuli that did not comply either OCP

constraint. This design was essential to verify if the learning performance of target learners was merely an artifact of applying irrelevant grammatical or extragrammatical knowledge. For the Control group, the 120 training items were associated to all possible tonal combinations except L-L, which is an apparent violation of participants' L1 phonotactics (i.e., T3 Sandhi; L+L → LH+L). Each of the 15 tonal combinations was associated with four disyllabic input tokens (i.e., 4 × 15 = 120) to have a fully balanced tonal distribution. Care was also taken to reduce the variation in the training input across the three learner groups. If a training item was paired with the same tonal sequence for the two target groups, the same pair was used in the Control condition, too. If a training item differed in its tonal combination across the two target learning conditions, the same segmental combination in the Control condition was paired with one of the two tonal sequences used in the target conditions (e.g., Terminal: [tʰuᴸmiᴴ]; Unit: [tʰuᴸmiᴸᴴ]; Control: [tʰuᴸmiᴴ]). When it was not possible to strictly follow the above two principles while maintaining a fully balanced distribution, the bottom line was to have at least one identical tone in the same position for the same training item across the three learning conditions (e.g., Terminal: [tʰuᴸmaᴴᴸ]; Unit = [tʰuᴸmaᴴᴸ]; Control = [tʰuᴸᴴmaᴴᴸ]). Without positive evidence against either target OCP generalization, we anticipated the Control group to randomly accept or reject test items in the acceptability judgment task.

For the auditory acceptability judgment task, we created one different set of 75 disyllabic test items from 12 monosyllables [fi, fu, fa, si, su, sa, xi, xu, xa, li, lu, la] whose consonants were not used in the training input. Duplicated onset consonants were included in three practice items (i.e., [fifu], [susa], and [lali]) but none of the remaining 72 test items repeated onset consonants or nucleus vowels in both syllables (e.g., [fisu]; cf. *[**fif**a] or *[**xifi**]) for a consistent stimulus design across the two sessions. All three groups were presented with this fixed set of 75 test items, which are listed in Appendix B.

All training and test items were recorded by the author speaking Taiwan Mandarin as his L1 (see Appendix C for a detailed explanation of recording and processing the auditory stimuli).

3.3 Procedure

The experiment was administered in a quiet room using PsychoPy v3.1.5 (Peirce et al. 2019) on a laptop/desktop computer with the output volume adjusted to a comfortable level. Participants were randomly assigned to one of the three learner groups and instructed to complete a learning task (i.e., the training session). In this learning task, learners were told to listen to words of a minor Chinese dialect and to try their best memorizing these words. Participants were also informed in advance that their learning performance would be assessed during the learning task and in a test session. The design of the auditory acceptability judgment task was not revealed to the participants prior to their completion of the training phase.

On each training trial, an eye fixation cross appeared at the center of the computer screen for 500 ms, and then a randomly ordered training item was immediately presented auditorily to the participants via a Musical Fidelity® MF-100 headphone. The training session proceeded automatically to the next trial three seconds after the offset of each auditory stimulus. To assess whether participants were attentive to the auditory inputs and followed the instruction to memorize them, a memory recognition task was administered at random intervals. After listening to two to five input tokens, participants were presented with an auditory input and were instructed to judge if the input matched the last input token they had heard immediately before the task. Chance was equal for the test input to be the token from the last training trial or a randomly selected training item. Participants responded by pressing S (Yes) or L (No) on the keyboard without being pressured by a time limit. The training session resumed automatically after a valid response, but no feedback on response accuracy was provided. At the end of the training session, the experimental software presented the accuracy rate of the memory recognition task on the computer screen. Participants who scored an accuracy rate of 90% or higher were qualified to participate in the test session. All of the 90 participants passed the threshold.

Before the test session, the 90 participants were explicitly told that some rules were hidden in the training items, and the test session was designed to assess if they had acquired these rules. We then explained to our participants that none of the auditory inputs in the test session was used in the training session, and they had to judge whether the test items conformed to the hidden rules. Participants were encouraged to rely on their intuition and respond spontaneously, especially if they were unsure about the target rules. We chose to explicitly discuss hidden rules in our instructions without specifying the target patterns since it was helpful for participants to fully understand their task and perform the test with similar strategies. In our pilot experiments, we avoided referring to hidden rules but asked our participants to judge novel test items based on whether the test items "sounded like or unlike the minor Chinese language in the training phase". A few participants nevertheless misunderstood the instruction and took the test as another memory recognition task; they simply accepted those they believed to be included in the training session even if it was emphasized that the test items never appeared in the training session. Thus, while our instruction could have either made participants explicitly aware of the target tonal patterns or biased participants to rely more on their explicit knowledge, it seemed necessary for participants to fully understand the test procedure. This was also the reason why an awareness measure was included in the test session, as we will explain below.

After the test session started, each test trial began with an eye fixation displayed at the center of the computer screen for 500 ms, which was followed by the auditory presentation of a randomly selected test item. After the offset of the auditory input, participants had four seconds to judge whether the test item conformed to the hidden rules by pressing the S key (Yes) or the L key (No) on the keyboard as quickly as possible. With a valid response, participants were then

asked if they were confident with their own judgment by pressing either the S key (Yes) or the L key (No) within 10 seconds. The lack of a positive correlation between learners' subjective confidence ratings and their correct responses could be equal to the lack of awareness of target linguistic generalizations (i.e., *zero correlation criterion*; Dienes 2007; Graham & Williams 2018); that is, the target linguistic patterns are acquired as implicit and unconscious knowledge. Alternatively, if a strong correlation is found, learners may be explicitly aware of the target knowledge. If participants did not provide their acceptability judgment in time, the test session proceeded to the next trial without asking participants for their confidence rating.

The test sessions began with three practice trials (see §3.2 for the materials) to familiarize participants with the procedure. The accuracy rate of participants' responses was calculated and presented to all participants in PsychoPy at the end of a test session. Responses were coded as *correct* if participants of the target groups accepted test items conforming to the target OCP generalization and rejected those violating the generalization. However, since the Control group was exposed to random di-tonal patterns in the training session, we did not set any fixed correct response for this group. Alternatively, *for the debriefing purpose only*, responses for the Control group were randomly coded as correct or incorrect in PsychoPy. Accordingly, the Control group was mostly presented with an accuracy rate around 50% when the test session ended. We then provided supplementary information regarding their learning performance should they have any question. The two sessions in Exp I together took 30 minutes on average to complete.

3.4 Results

The acceptability judgment task elicited a total of 6,345 valid responses in non-practice trials from 90 participants.[16] The number of trials without a valid response for the three groups was 33 (1.5%) for the Terminal group, 46 (2.1%) for the Unit group, and 56 (2.6%) for the Control group. The mean reaction time of valid acceptability judgment responses within each learner group was 1,491 ms (Terminal; $sd = 820$), 1,521 ms (Unit; $sd = 821$), and 1,599 ms (Control; $sd = 857$).[17] The data set was analyzed using mixed-effects logistic regression using the *lme4* package (Bates et al. 2020) in R 4.0.4 (R Core Team 2021), in which the binary dependent variable (Accept) was coded as 0 (reject) and 1 (accept). We chose to analyze the acceptance patterns since AGL learners in acceptability judgment tasks may be less likely to correctly reject than to correctly accept novel items (e.g., Chen 2020); the high accuracy rate of accepting novel items may be canceled by the low accuracy rate of rejecting novel items, which leads to null

---

16 The raw result data and R codes for all analyses in Exp I and II are available at https://osf.io/zt8jh/.

17 Unpaired two-sample *t*-tests assuming an equal variance suggested a significant difference in reaction time between each of the two target groups and the Control group (Terminal vs. Control: $t(4229) = -4.18, p < .001$; Unit vs. Control: $t(4216) = -3.02, p = .003$) but a non-significant difference between the two target groups ($t(4239) = 1.19, p = .236$).

results. In addition, it was just not possible to code and analyze correct responses for the Control group (see §3.3). The fixed predictors included Group with ternary levels (Control vs. Terminal vs. Unit) and OCP-H, OCP-L, and OCP-Unit with binary levels representing constraint violation (Yes vs. No). The predictors were Helmert-coded to compare a level to other levels of the same predictor, rather than merely to a reference level.[18] This was crucial for us to test performance differences between all learner groups in our regression modeling. We also divided OCP-Terminal into OCP-H and OCP-L since the two constraints were found to operate independently across languages (see §2) and their rankings could thus be acquired separately as well. Two-way interactions between Group and the three OCP variables were also included, which were critical to answer our main research question. We did not include the interactions between the OCP variables since the study was not designed to investigate the ganging effect of OCP violations. By-subject and by-item random slopes and intercepts were also included in our search of an appropriate regression model.

We first followed Barr et al. (2013) to begin with a 'maximal' model in (8) including all random intercepts and slopes into consideration, with the BOBYQA optimizer (Powell 2009) configured to have a maximum of 100,000 iterations. When the model failed to converge, we gradually removed random effects with a zero or near-zero variation until the model converged. This second step of model simplification, as opposed to model selection, was recommended in Mastuschek et al. (2017) to construct a model that could best balance the chance of Type I and Type II errors. The simplified model (9) that converged is summarized in **Table 5**, and the three crucial two-way interactions between Group and OCP are visualized in **Figure 2**.

(8) Accept ~ Group + OCP-H + OCP-L + OCP-Unit + Group × (OCP-H + OCP-L + OCP-Unit) + (1 + OCP-H + OCP-L + OCP-Unit | Participant) + (1 + Group | Item)

(9) Accept ~ Group + OCP-H + OCP-L + OCP-Unit + Group × (OCP-H + OCP-L + OCP-Unit) + (1 + OCP-L + OCP-Unit | Participant) + (1 | Item)

The above analysis suggested a significantly lower acceptance rate for the Terminal group than for the Control group ($\beta = -0.264$, $SE = 0.088$, $z = -3$, $p = .003$) with no significant difference between the Unit group and the other two groups *combined* ($\beta = -0.037$, $SE = 0.051$, $z = -0.734$, $p = .463$). In addition, there existed no across-the-board OCP effect. Multiple significant two-way interactions between Group and OCP were also discovered in our mixed-effects modeling with implications for the learning of the OCP generalizations. The left panel in **Figure 2** illustrated

---

18 Helmert coding was generated using the R base function contr.helm(), which compares a factor level to all previous levels, which may be alternatively viewed as *reversed* Helmert coding. The order of levels in Helmert coding in R is alphabetical by default. Thus, the base level is Control for Group and Yes for the three OCP predictors.

	β	SE	z	p
Intercept	−0.074	0.117	−0.364	.526
GroupTerminal	−0.264	0.088	−3	.003 *
GroupUnit	−0.037	0.051	−0.734	.463
OCPHYes	−0.056	0.07	−0.803	.422
OCPLYes	−0.033	0.081	−0.405	.686
OCPUnitYes	−0.131	0.081	−1.619	.106
GroupTerminal:OCPHYes	−0.252	0.04	−6.269	<.001 *
GroupUnit:OCPHYes	−0.014	0.023	−0.628	.53
GroupTerminal:OCPLYes	−0.127	0.053	−2.391	.017 *
GroupUnit:OCPLYes	0.101	0.031	3.312	<.001 *
GroupTerminal:OCPUnitYes	−0.001	0.052	−0.016	.988
GroupUnit:OCPUnitYes	−0.059	0.03	−1.953	.051

Table 5: The mixed-effects modeling of the acceptability judgment data; * = $p < .05$.
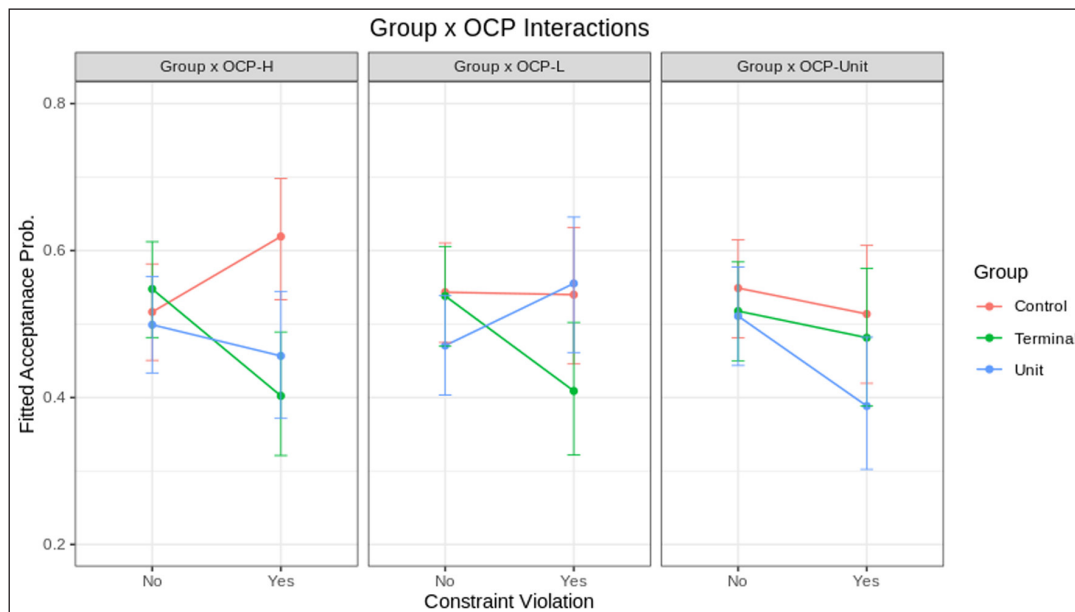
Figure 2: Two-way Group $\times$ OCP interactions from the mixed-effects model.

a difference in responses from the Control group and the Terminal/Unit group to test items with or without a violation of OCP-H. For the Control group, the learners were more inclined to accept the test items violating OCP-H, whereas the learners of the Terminal/Unit group were less inclined to accept the same set of the test items. This difference corresponded to the significant interaction term GroupTerminal:OCPHYes indicating a significantly more negative OCP-H effect for the Terminal group than for the Control group. The nonsignificant interaction term GroupUnit:OCPHYes suggests that the OCP-H effect was not significantly more negative for the Unit group than for the Control/Terminal group combined. The middle panel in **Figure 2** indicated three distinct OCP-L effects for each of the three learner groups. First, OCP-L violation did not affect the judgments of the Control group. Second, the Terminal group, if compared to the Control group, was more likely to reject the test items violating OCP-L. This difference was reflected in the significant interaction term GroupTerminal:OCPLYes with a more negative slope for the Terminal group than for the Control group. The Unit group demonstrated a *positive* effect of OCP-L than the Control/Terminal group combined, which was represented by the significant interaction term GroupUnit:OCPLYes. In the rightmost panel in **Figure 2**, there were two similar slopes for the Control/Terminal group suggesting only a mild negative effect of OCP-Unit violation. The slope for the Unit group stood out with a more drastic decline in the acceptance probability for test items violating OCP-Unit. However, the difference between the Unit group and the Control/Terminal group combined was nonsignificant (GroupUnit:OCPUnitYes: β = –0.059, SE = 0.03, z = –1.953, p = .051).

A reviewer asked if syllables that are phonotactically illicit in Taiwan Mandarin (i.e., [fi], [si], and [xi]) in the test items had any effect on participants' judgments, given the possibility that more attention might be drawn to them. Accordingly, we separated our entire data set into two subsets depending on the inclusion of a phonotactically illicit syllable. Both subsets were submitted to the model (9), which converged. For the subset *without* illicit syllables, the same model indicated the same significant and nonsignificant main effects and interactions in **Table 5**. The modeling of the subset *with* illicit syllables presented two minor differences. The significant interaction GroupTerminal:OCPLYes in **Table 5** became marginal (β = –0.112, SE = 0.063, z = –1.761, p = .08). In addition, the nonsignificant interaction GroupTerminal:OCPUnitYes in the previous analysis turned out to be significant (β = –0.141, SE = 0.069, z = –2.036, p = 0.042). The former might be related to the overall weaker effect of OCP-L as we will discuss below, and the latter only revealed another remarkable difference between the Terminal and Control groups. In general, the results of these additional analyses were generally congruent with the grand analysis, and the impact of including phonotactically illegal syllables in the test items seemed minimal.

As hypothesized in §2, the patterns we found in the above analysis may reflect the outcomes of implicit *and* explicit learning, as the latter would be influential for adult learners. To examine possible effects of explicit knowledge on participants' test performance, we analyzed the

correlation between response accuracy and confidence with the data of the two target groups. For the Terminal group, responses were coded as *correct* if the learners accepted test items complying with OCP-H and OCP-L and rejected those violating the two OCP constraints. For the Unit group, responses were coded as *correct* only when the learners rejected test items with a tonal sequence violating OCP-Unit and accepted others without identical tonal constituents. The response accuracy served as a dependent variable regressed against the Helmert-coded categorical independent variables Group (Terminal vs. Unit) and Confidence (No vs. Yes) and the two-way Group × Confidence interaction. As in the previous mixed-effects modeling, we started by building a maximal model taking by-subject and by-item random intercepts and slopes into consideration as in (10). The model nevertheless failed to converge and was reduced to (11) summarized in **Table 6**. The model summary indicated that a higher level of confidence predicted a higher probability of correct responses, and this correlation did not vary significantly between the two target groups. In conclusion, the learning performance of the two groups of learners was indeed considerably influenced by explicitly learned knowledge. We also regressed confidence level against the binary factor representing the inclusion of phonotactically illicit syllables (Yes vs. No) in a separate logistic mixed-effects modeling. The results showed a nonsignificant effect of illicit syllables ($\beta = 0.079$, $SE = 0.067$, $z = 1.185$, $p = .236$).

(10)   Accuracy ~ Group × Confidence + (1 + Confidence | Participant) + (1 + Group × Confidence | Item)

(11)   Accuracy ~ Group × Confidence + (1 + Confidence | Participant) + (1 + Group | Item)

	β	SE	Z	p
Intercept	3.55	0.615	5.771	<.001 *
GroupUnit	–0.026	0.235	–0.111	.911
ConfidenceYes	0.183	0.079	2.33	.02 *
GroupUnit:ConfidenceYes	–0.098	0.079	–1.237	.216

Table 6: The mixed-effects modeling testing the zero-correlation hypothesis; * = $p < .05$.

With a strong influence of explicit learning, it is of question which OCP generalizations discovered in our first grand analysis were learned with a more implicit nature. We thus turned to analyze a data subset including only *unconfident* judgments of all three learner groups. Acquired OCP generalizations uncovered in this analysis could thus be viewed as implicit and unconscious

knowledge. This subset included a total of 2,345 trials (36.2% of the entire data set) and were initially submitted to the model in (9) for a direct comparison of significant main effects and interactions. However, as the model failed to converge due to a smaller data set, we were forced to further remove the by-subject random slope of OCP-Unit and simplify the model into (12). The summary of the modeling statistics is provided in **Table 7** and visualized in **Figure 3**.

(12)   Accept ~ Group + OCP-H + OCP-L + OCP-Unit + Group × (OCP-H + OCP-L + OCP-Unit) + (1 + OCP-L | Participant) + (1 | Item)

	β	SE	z	p
Intercept	–0.312	0.116	–2.695	.007 *
GroupTerminal	–0.137	0.123	–1.107	.268
GroupUnit	–0.02	0.067	–0.301	.763
OCPHYes	–0.126	0.067	–1.876	.061
OCPLYes	0.039	0.076	0.513	.608
OCPUnitYes	–0.055	0.076	–0.731	.465
GroupTerminal:OCPHYes	–0.24	0.067	–3.585	<.001 *
GroupUnit:OCPHYes	0.003	0.036	0.069	.945
GroupTerminal:OCPLYes	–0.063	0.076	–0.826	.409
GroupUnit:OCPLYes	0.007	0.043	0.17	.865
GroupTerminal:OCPUnitYes	–0.018	0.077	–0.227	.82
GroupUnit:OCPUnitYes	–0.035	0.041	–0.851	.395

Table 7: The mixed-effects modeling of the unconfident judgment data; * = $p < .05$.

In the current analysis of unconfident responses, we still found a stronger OCP-H effect for the Terminal group than for the Control group (the leftmost panel in **Figure 3**) as indicated by the significant interaction term GroupTerminal:OCPHYes (β = –0.24, *SE* = 0.067, *z* = –3.585, *p* = .003). The slope for the Unit group was not significantly more negative than the Control/Terminal group combined. No other significant between-group difference was discovered, however; the between-group variation in the OCP-L effect was neutralized in the current analysis (the middle panel in **Figure 3**), and the difference in the OCP-Unit effect on participants' responses also shrunk (the rightmost panel in **Figure 3**).
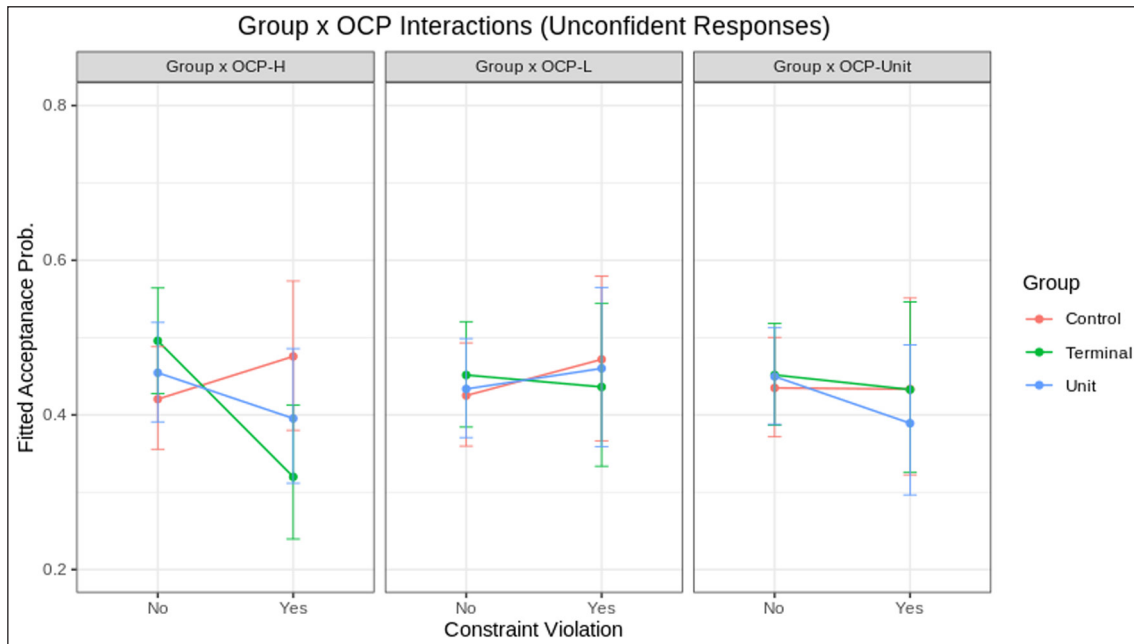
Figure 3: Two-way Group × OCP interactions from the mixed-effects model of unconfident judgments.

3.5 Discussion

In Exp I, we asked learners to judge the acceptability of test items with variable di-tonal sequences after listening to input tokens manipulated to present distinct tonal distributions. Our primary findings were twofold. In our grand analysis, we found that the Terminal group, unlike the Control and Unit groups, demonstrated a bias *against* test items violating OCP-Terminal (i.e., OCP-H and OCP-L), whereas the Unit group showed only a marginal negative effect of OCP-Unit and an unexpected negative effect of OCP-H. This difference led to an interim conclusion that the level-based and unit-based generalizations are not equally learnable. In a follow-up analysis, we focused on unconfident responses that would reflect the implicit nature of an acquired OCP generalization. We still found a remarkable between-group difference in the OCP-H effect, which was significantly more negative for the Terminal group. Other between-group differences with respect to the effects of OCP-L and OCP-Unit violation were neutralized. Notably, the decrease in the fitted acceptability rate driven by OCP-H violation was greater for the Terminal group in the follow-up analysis (17.6%) than in the grand analysis (14.5%). This finding led us to conclude that only a level-based OCP generalization could be potentially acquired as implicit phonological knowledge, which could be easily generalized with abstract identity avoidance (cf. Berent et al. 2002; Berent 2013).

There is still room for alternative interpretations of our findings, which reside largely in the perception of tones. In Exp I, we were not able to incorporate a true poverty-of-the-stimulus design

(Wilson 2006) by testing our participants' learning performance with a set of tones that were not used in the training session. The participants could have always explicitly compared the tonal sequences of the test items to those of the training items when making their acceptability judgments. For the two target groups, test items with a new tonal sequence might have stood out from other test stimuli and raised participants' awareness of target OCP generalizations. This potentially strong dependence on the explicit process may have overshadowed intrinsically weaker effects of OCP-L and OCP-Unit even if they were indeed learnable implicit grammatical generalizations. For example, if adjacent Ls are intrinsically favored over adjacent Hs in phonological grammar (see §5.1), it would take longer to acquire a robust OCP-L generalization, and the OCP-L effect would be masked by the strong influence of the explicit perceptual comparison.

The similarities and differences in the OCP-H effect may have also reflected the influences of pure tonal perception. First, unlike the two target groups, the Control group demonstrated a bias *toward* test items violating OCP-H. This may be due to an asymmetry in the perception of adjacent Hs and Ls, as a sustained high pitch may help direct listeners' attention to the speech (e.g., Evans 2015: 3);[19] that is, since adjacent Hs were perceptually salient in the training input, it was easier for the Control group to explicitly notice that adjacent Hs in the test items also appeared in the training input and consequently accept these items. This was evident with a greater *increase* in the fitted acceptance rate driven by OCP-H violation for the Control group in the grand analysis (10.3%) than in the analysis of unconfident responses (5.5%). The same perceptual bias may also be the primary cause of a reduced negative OCP-H effect for the two target groups; both target groups demonstrated as a smaller *decrease* in the acceptance rate caused by OCP-H violation in the grand analysis (Terminal: 14.5%; Unit: 4.3%) than in the follow-up analysis of unconfident responses (Terminal: 17.6%; Unit: 6%). Perhaps, when learners were more attentive to a tonal sequence including adjacent Hs sequence in test items, they confidently (but falsely) believed that the sequence appeared in the training items.

In addition, while confidence rating is a heuristic for identifying the knowledge source of participants' auditory acceptability judgment, it may not be a robust awareness measure due to its subjective nature. Maie & DeKeyser (2020) warned that participants may be biased to respond with a low confidence level even if an explicit, conscious knowledge is recruited for their judgment (e.g., phonetic memory). Or, as a reviewer suggested, it could be the application of implicit knowledge that have unconsciously raised the subjective confident level. Thus, in order to gain more conclusive evidence of the implicit learnability of the two OCP generalizations, an additional experiment is required to incorporate a more objective awareness measure and an assessment of learning performance that is less dependent on perception.

---

19 However, as suggested by a reviewer, adjacent Ls could also be perceptually more salient than adjacent Hs since vowel duration is longer with a low tone (e.g., Gandour 1977).

4 AGL experiment II with a production task

In Exp II, learning performance was assessed in a *production* task that required participants to produce each *visually presented* test item with a free di-tonal combination, with or without violating the hidden tonal patterns (see §4.3). This experimental design would force participants to actively apply learned phonological generalizations rather than to perform perceptual comparisons. Learners' awareness of acquired target OCP generalizations was measured based on whether participants could consciously produce outputs violating the hidden generalizations in real time. The output per se could thus serve as a more objective indicator of awareness than self-reported confidence ratings.

4.1 Participants

Another group of 17 female and 28 male participants enrolled as an undergraduate or graduate student at National Tsing Hua University in Taiwan were recruited for Exp II. They had little linguistic training, spoke Taiwan Mandarin as their native language, and aged between 20 and 29 years old (*mean* = 21.6, *sd* = 1.97). The 45 participants neither participated in Exp I nor reported any hearing or learning impairment at the time of the study. Two participants were excluded from result analyses for monotonic tonal patterns used in their production (see §4.4). The rest of 43 participants passed the accuracy threshold of 90% in the memory recognition task during the training session and then completed the test session. Participants of Exp II were paid 150 NTD for their participation.

4.2 Materials

The three sets of 120 disyllabic training items for each of the three learner groups were identical to those in Exp I. A different set of 34 disyllabic sequences (two practice items plus 32 target test items) were created specifically for the production task in the test session in Exp II. They were combinations of CV monosyllables (/sɑ/, /su/, /lɑ/, /li/, /lu/, /fɑ/, /fu/, /xɑ/, and /xu/) that do not violate segmental phonotactics in Taiwan Mandarin and were therefore pronounceable for our participants. The 34 CVCV sequences were then converted into Zhuyin Fuhao (an onset-rime orthographic system used specifically in Taiwan) visually presented to the participants as a target of their production. In each orthographic string, tonal labels were replaced with a question mark (e.g., ㄌㄧ？ㄏㄚ？ = /li?xɑ?/) to remind our participants to freely combine any of the four Taiwan Mandarin tones in their production. The 32 target test items were specifically designed not to be near-homophones of disyllabic lexical words in Taiwan Mandarin to minimize the interference of the L1 lexicon (see, for example, Chen (2020)). The 32 target test items were then divided into two subsets of 16 test items for two test conditions (see §4.3), and specific attention was paid to carefully control the distribution of CV combinations by syllable position in both test conditions. All 34 test items are listed in Appendix D.

4.3 Procedure

The instruction for the training session was identical for both Exp I and Exp II, and the same random memory recognition task was embedded in the training phase (see §3.3 for details).

After completing the training session, participants were told that the training input was created following some hidden rules, and the goal of the test session was to investigate if they had acquired these rules. The test session was divided into two blocks, which was designed to incorporate the *inclusion-exclusion* task (Curran 2001; Destrebecqz & Cleeremans 2001; Chan & Leung 2014) as an objective awareness measure. The first block was the *inclusion* condition, in which participants would see disyllabic orthographic sequences one by one without tonal information, and they had to *follow* the hidden rules to freely combine lexical tones in Taiwan Mandarin to read each test item aloud.[20] Put differently, participants were asked to produce di-tonal outputs that can be *included* in acceptable surface forms generated by acquired target knowledge. A good performance in this condition could be ascribed to both implicit and explicit knowledge; the learners may be explicitly aware of and thus consciously follow a target generalization to produce acceptable outputs, or simply produce acceptable forms spontaneously without effortful and conscious knowledge or memory retrieval. The second block was the *exclusion* condition, and the participants' task was to produce test items with a di-tonal combination that participants thought to be *incompatible* with the hidden rules. To succeed in this condition, participants must be *explicitly* aware of an acquired target generalization to possibly avoid producing acceptable outputs. With this design, the nature of an acquired generalization lies in the difference and similarity in participants' performance between the two conditions. If a target generalization is acquired as implicit knowledge, learners would have no conscious access to the generalization and could not violate it voluntarily in the *exclusion* condition. Implicit learners are thus more likely to generate outputs conforming to a target generalization in *both* test conditions. Alternatively, if learners are aware of a target generalization, they would have full control of when to apply the explicitly acquired knowledge. Explicit learners are thus expected to produce more outputs violating the target generalization in the *exclusion* condition than in the *inclusion* condition. In Exp II, the inclusion condition always preceded the exclusion condition to prevent explicit processing from interfering with the test performance in the inclusion condition (e.g., Lichtman 2013).

Before the first block, participants were given only the instruction on the *inclusion* task. We encouraged the participants to follow their intuition and use as many di-tonal combinations in their production as possible. This would allow us to investigate a broad phonological generalization learned by the participants, rather than a partial preference for a few di-tonal patterns. Participants were also advised to produce T3 in Taiwan Mandarin as L rather than MLH for methodological benefits; a low-toned T3 not only matched the f0 contour used to create the training input but

---

20 See Chan & Leung (2014) for a similar experimental design in their study of the implicit learning of L2 Spanish stress. We will also discuss possible influences of explicitly referring to Taiwan Mandarin tones in §4.5.

also made it easier for us to transcribe the difference between T2 and T3 in their production (see §3.2). Only after participants completed the inclusion condition, the instruction on the second block was given, which was to ask participants to produce another set of orthographic sequences by trying to violate the hidden rules. As in the inclusion condition, we requested the participants to use diverse di-tonal patterns in their production following their intuition.

In both test conditions, each test trial started with an eye-fixation cross at the center of the computer screen lasting for 500 ms, which was followed by the visual presentation of a randomly selected target orthographic sequence. After the onset of the visual presentation of a target sequence, participants had four seconds to read the sequence aloud with a di-tonal pattern of their choice before the test session proceeded automatically to the next trial. When each test condition began, the participants completed two practice trials to be familiarized with the procedure before proceeding to their production of 16 target test items.

Both training and test sessions in Exp II were administered using PsychoPy v3.1.5 on a desktop computer in a quiet room. Training items were presented auditorily via a Musical Fidelity® MF-100 headphone, and the production of test items was recorded at a sample rate of 44,100 Hz in PsychoPy via an Audio-Technica® MB 3k microphone mounted to a stand and connected directly to the desktop computer. The training and test sessions took an estimate of 20 minutes in total. The experimental design as well as the accurate response patterns were debriefed upon participants' requests after the end of the entire test session.

4.4 Results

Di-tonal patterns of the recordings collected in the test session were transcribed by two research assistants speaking Taiwan Mandarin as their L1, and the consistency rate in their transcriptions reached 90.5%. When there were disagreements, a third research assistant decided which transcription was more accurate. Among all $32 \times 45 = 1{,}440$ tokens, we excluded 133 tokens (9.2%) that were ultimately judged to show hesitation, unnatural interruption, or incompleteness. We then further removed four rare instances of L-L from the data set, which were all produced in the *exclusion* condition, perhaps as a deliberate attempt to produce an output extremely distant from training items. In the remaining 1,303 tokens, the participants on average produced 10.6 di-tonal *types* ($sd = 2.28$) of all 15 possible di-tonal types excluding L-L. Two participants (Terminal: 1; Unit: 1) produced only six or fewer di-tonal patterns (i.e., $\leq mean - 2 \times sd$) and were thus excluded from the analyses below; a limited use of di-tonal patterns would not be informative in our study of the learnability of the broad tonal generalizations. These screening processes generated a subset of 1,245 production tokens for our result analyses. This subset was then sorted by learner group (Control vs. Terminal vs. Unit), di-tone type, OCP violation (OCP-H, OCP-L, and OCP-Unit), and test condition (inclusion vs. exclusion) as a production corpus analyzed using Poisson regression. Since the sorted data set included zero counts, we followed

Myers (2012) to add one token to each count to avoid poor data fit caused by excessive zeros in Poisson regression modeling (e.g., He et al., 2017).

The production corpus with adjusted counts was first submitted to mixed-effects Poisson regression taking token counts as the dependent variable and Group, Condition, OCP-H, OCP-L, and OCP-Unit as Helmert-coded fixed variables. Three-way interactions between Group, Condition, and each of the three OCP variables were also included for an assessment of between-group and between-condition variation in the OCP effects on the learners' production. By-subject and by-item random effects were included in an initial modeling attempt. We took the procedures adopted in the analyses of Exp I results to begin with a maximal mixed-effects model in (12), which failed to converge. Later attempts in searching for a mixed-effects model to include any random effect were not successful, perhaps owing to a small data set. As a last resort, we fitted the results to a generalized Poisson regression model without random effects in (13), which is summarized in **Table 8**.

(12)   Counts $\sim$ Group + OCP-H + OCP-L + OCP-Unit + Group $\times$ (OCP-H + OCP-L + OCP-Unit) + (1 + OCP-H + OCP-L + OCP-Unit | Participant) + (1 + Group | Item)

(13)   Counts $\sim$ Group $\times$ Condition $\times$ (OCP-H + OCP-L + OCP-Unit)

From the modeling results, we would first highlight the *negative* main OCP effects on the production of the test items: The participants generally avoided output tokens violating OCP-H ($\beta = -0.347$, $SE = 0.034$, $z = -10.082$, $p < .001$), OCP-L ($\beta = -0.494$, $SE = 0.047$, $z = -10.509$, $p < .001$), or OCP-Unit ($\beta = -0.707$, $SE = 0.059$, $z = -12.07$, $p < .001$). These negative main OCP effects also varied considerably by Group and Condition, which are visualized in **Figure 4**, **Figure 5**, and **Figure 6**.

In **Figure 4**, we can observe the negative OCP-H effect across the two test conditions except for the Control group, which showed a *positive* effect of OCP-H in the *inclusion* condition. The difference was partially reflected in the significant three-way interaction term GroupTerminal:CondInclude:OCPHYes ($\beta = -0.243$, $SE = 0.041$, $z = -5.924$, $p < .001$) showing a more negative OCP-H effect for the Terminal group than for the Control group. Another significant interaction term GroupUnit:CondInclude:OCPHYes ($\beta = -0.056$, $SE = 0.025$, $z = -2.239$, $p = .025$) suggested that the negative OCP-H effect was stronger in the inclusion condition for the Unit group than for the other two groups combined. This could be ascribed to the opposite OCP-H effects for the Control/Terminal group, which canceled each other in this comparison between Unit and Control/Terminal. To further test if the negative OCP-H effect differed significantly across the two test conditions between the two target groups, a post-hoc pairwise comparison excluding the Control group was conducted with the model in (13). The three-way interaction term GroupUnit:CondInclude:OCPHYes was found to be nonsignificant ($\beta = 0.038$, $SE = 0.044$, $z = 0.86$, $p = .39$). Thus, the negative OCP-H effect in the inclusion condition was comparable for the two target groups.

	β	SE	z	p
Intercept	0.186	0.073	2.544	.011 *
GroupTerminal	−0.428	0.074	−5.882	<.001 **
GroupUnit	−0.119	0.06	−1.994	.046 *
CondInclude	−0.379	0.073	−5.167	<.001 **
OCPHYes	−0.347	0.034	−10.082	<.001 **
OCPLYes	−0.494	0.047	−10.509	<.001 **
OCPUnitYes	−0.707	0.059	−12.07	<.001 **
GroupTerminal:CondInclude	−0.405	0.074	−5.503	<.001 **
GroupUnit:CondInclude	−0.157	0.06	−2.628	.009 **
GroupTerminal:OCPHYes	−0.178	0.041	−4.345	<.001 *
GroupUnit:OCPHYes	−0.02	0.025	−0.785	.433
GroupTerminal:OCPLYes	−0.348	0.062	−5.582	<.001 **
GroupUnit:OCPLYes	0.102	0.03	3.388	<.001 **
GroupTerminal:OCPUnitYes	−0.074	0.044	−1.671	.095
GroupUnit:OCPUnitYes	−0.214	0.053	−4.052	<.001 **
CondInclude:OCPHYes	−0.017	0.034	−0.492	.622
CondInclude:OCPLYes	−0.162	0.047	−3.44	<.001 **
CondInclude:OCPUnitYes	−0.374	0.059	−6.391	<.001 **
GroupTerminal:CondInclude:OCPHYes	−0.243	0.041	−5.924	<.001 **
GroupUnit:CondInclude:OCPHYes	−0.056	0.025	−2.239	.025 *
GroupTerminal:CondInclude:OCPLYes	−0.212	0.062	−3.399	<.001 **
GroupUnit:CondInclude:OCPLYes	0.057	0.03	1.9	.057
GroupTerminal:CondInclude:OCPUnitYes	−0.254	0.044	−5.782	<.001 **
GroupUnit:CondInclude:OCPUnitYes	−0.222	0.053	−4.206	<.001 **

Table 8: The Poisson regression modeling of the production counts; $* = p < .05$; $** = p < .01$

To summarize, the Control group, like the two target groups, avoided violating OCP-H in the exclusion condition but showed a preference for di-tonal patterns violating OCP-H in the inclusion condition. In addition, the two target groups did not differ significantly across the two test conditions in terms of their avoidance of OCP-H violation.
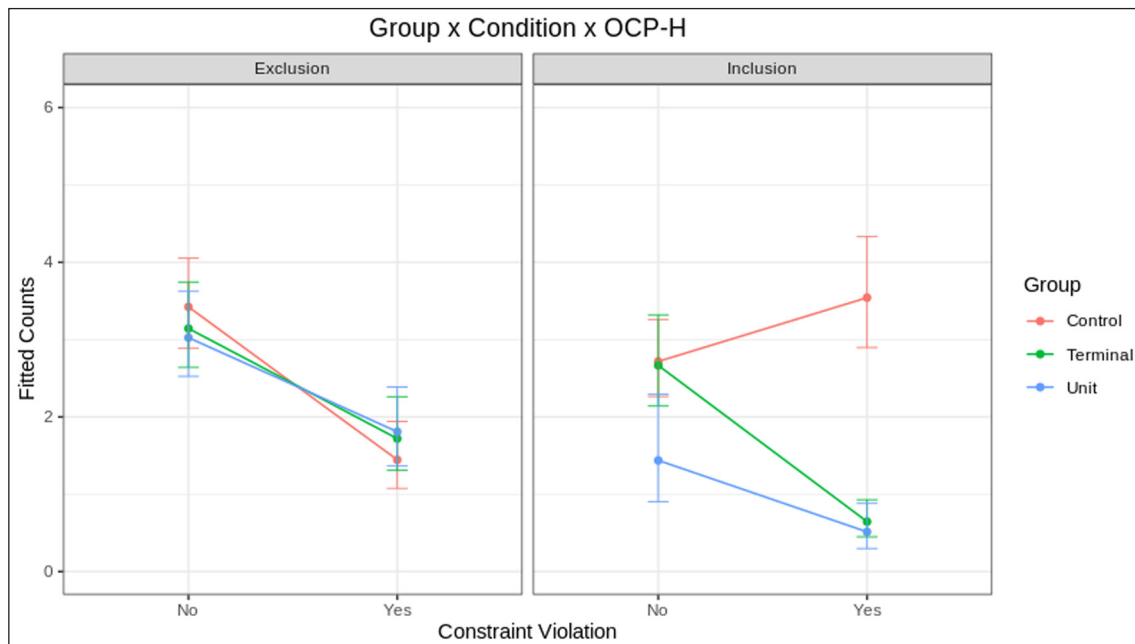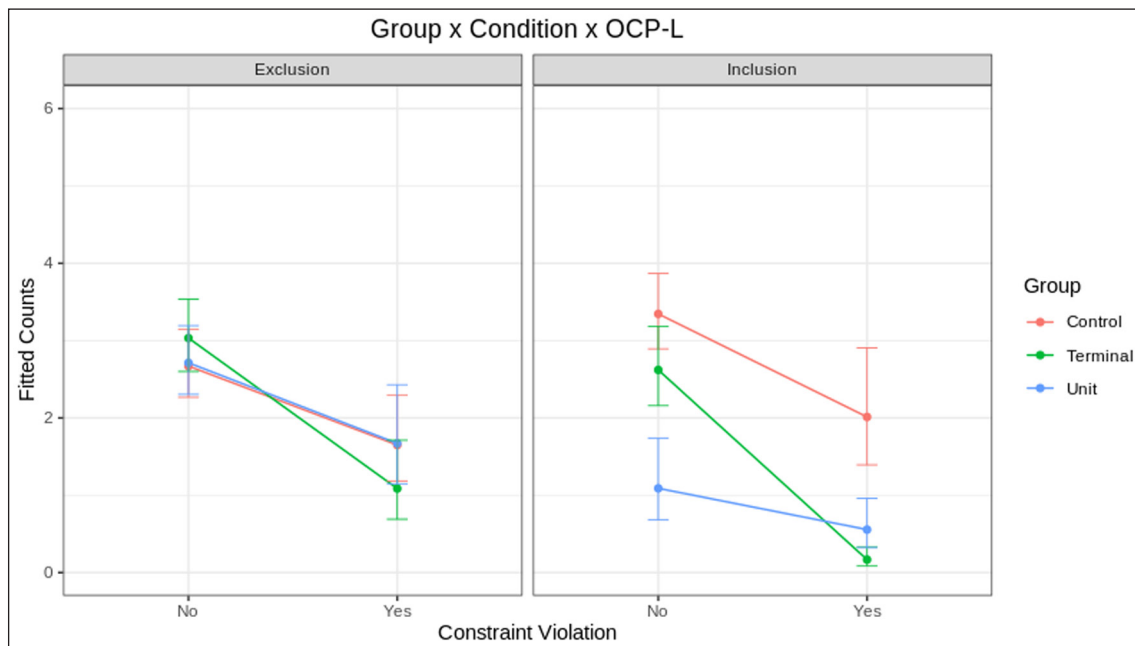

Figure 4: The Group $\times$ Condition $\times$ OCP-H interaction from the Poisson regression model.

The negative OCP-L effects presented in **Figure 5** were more consistent across the three learner groups, but with a steeper slope only for the Terminal group. The difference was partially reflected in the significant two-way interaction term GroupTerminal:OCPLYes (β = –0.349, SE = 0.062, z = –5.582, p < .001) in **Table 8** comparing the Terminal group to the Control group. Another significant two-way interaction term GroupUnit:OCPLYes (β = 0.102, SE = 0.03, z = 3.388, p < .001) also indicated a *less* negative OCP-L effect for the Unit group than for the other two groups combined. Furthermore, the negative OCP-L effect was even stronger for the Terminal group than the Control group in the inclusion condition (i.e., GroupTerminal:CondInclude:OCPLYes: β = –0.212, SE = 0.062, z = –3.399, p < .001). The same negative effect was attenuated for the Unit group in the inclusion condition if compared to the other two groups altogether, although the difference did not reach statistical significance (GroupUnit:CondInclude:OCPLYes: β = 0.057, SE = 0.03, z = 1.9, p = .057). However, the post-hoc pairwise comparison between the two target groups revealed a significant three-way interaction term GroupUnit:CondInclude:OCPLYes (β = 0.192, SE = 0.062, z = 3.098, p =

.002). This could be the supporting evidence for a *less* negative OCP-L effect for the Unit group than for the Terminal group in the inclusion condition.

Considering all the individual findings here, we can reach the conclusion that the Terminal group was more inclined to avoid OCP-L violation across the board in their production, particularly in the inclusion condition, whereas the Unit group resembled the Control group with their production outputs less affected by OCP-L violation in both test conditions.

Figure 5: The Group $\times$ Condition $\times$ OCP-L interaction from the Poisson regression model.

Finally, the negative OCP-Unit effect in **Figure 6** also varied by Group and Condition to a different extent. Crucially, the overall negative effect was not significantly different between the Control group and the Terminal group (GroupTerminal:OCPUnitYes: $\beta = -0.074$, $SE = 0.044$, $z = -1.671$, $p = .095$) but was significantly different between the Unit group and the other two groups combined (GroupUnit:OCPUnitYes: $\beta = -0.214$, $SE = 0.053$, $z = -4.052$, $p < .001$). The negative effect was nevertheless stronger in the inclusion condition for the Terminal group than for the Control group (GroupTerminal:CondInclusion:OCPUnitYes: $\beta = -0.254$, $SE = 0.044$, $z = -5.782$, $p < .001$) and stronger for the Unit group than for the other two groups combined (GroupUnit:CondInclusion:OCPUnitYes: $\beta = -0.222$, $SE = 0.053$, $z = -4.206$, $p < .001$). The same post-hoc comparison excluding the Control group pointed to a stronger negative OCP-Unit effect for the Unit group in the inclusion condition than for the Terminal group (GroupUnit:CondInclusion:OCPUnitYes: $\beta = -0.206$, $SE = 0.083$, $z = -2.48$, $p = .013$).

Altogether, the experimental results of the production task suggested an overall negative OCP-Unit effect, which was greatest for the Unit group in the inclusion condition. The negative effect, however, was also reduced to the greatest extent for the Unit group in the exclusion condition. In other words, the Unit group produced output without violating OCP-Unit more frequently than the Terminal group did in the inclusion condition. The Unit group nevertheless also tried harder than the Terminal group to avoid violating OCP-Unit in the exclusion condition. Unlike the two target groups, the Control group consistently avoided outputs with OCP-Unit violation across the two conditions.
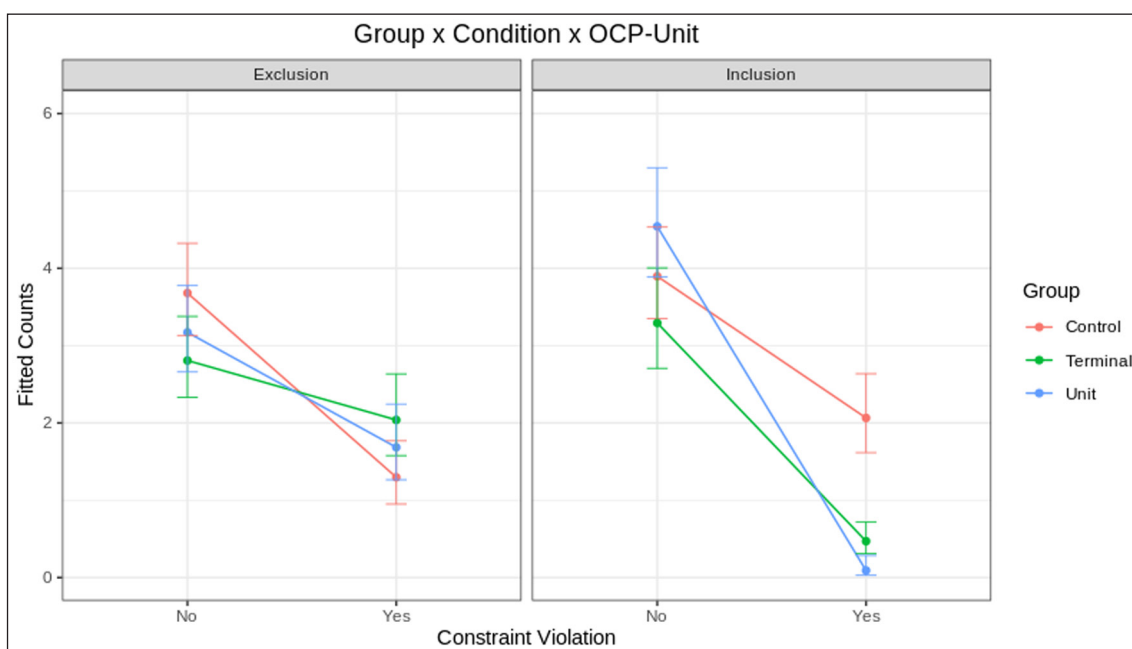
Figure 6: The Group × Condition × OCP-Unit interaction from the Poisson regression model.

4.5 Discussion

In Exp II, we investigated the learnability of the three OCP generalizations and assessed the learning performance in a production task without any auditory interference to free learners from the influences of acoustic similarities between training and test items. The learning performance was assessed in an inclusion-exclusion design aiming to objectively measure the qualitative differences in acquired phonological knowledge.

The results indicated that the OCP-H generalization was learned in an opposite way for the Control group and the two target groups (**Figure 4**); whereas the Terminal group and the Unit group seemed to hold themselves from producing outputs with adjacent Hs, the Control group by and large preferred to include adjacent Hs in their production of test items. Accordingly, we can

conclude that the OCP-H generalization was learned by both target groups but not by the Control group. Crucially, as the negative effect of OCP-H was comparable for both target groups across the two test sessions, it may be safe to conclude that the learners of the two target groups could not consciously avoid violating OCP-H in the exclusion condition. Thus, it could be plausibly assumed that both target groups acquired an implicit OCP-H generalization.

Our analysis of the OCP-L effects distinguished the Terminal group from the other two groups (**Figure 5**), as a stronger negative OCP-L effect was consistent across the two test conditions for the Terminal group. Since the learners of the Terminal group constantly produced outputs violating OCP-L in the exclusion condition, we could argue that OCP-L was acquired as implicit knowledge for the Terminal group. The weaker negative OCP-L effect did not differ for the Unit and the Control groups across the two test conditions. The Unit group thus demonstrated no sign of learning the OCP-L generalization.

Next, all three groups were sensitive to OCP-Unit violation in different ways (**Figure 6**). Outputs violating OCP-Unit were generally avoided by the Control group across the two conditions. However, since the negative OCP-Unit effect was stronger for the two target groups in the *inclusion* condition but substantially weaker in the *exclusion* condition, the OCP-Unit generalization might have been acquired as explicit knowledge. Crucially, the between-condition difference in the negative OCP-Unit effect was more salient for the Unit group than for the Terminal group, which might suggest the development of a more explicit OCP-Unit generalization for the Unit group. In our retrospective interview, two participants from each of the two target groups were able to verbalize the rule accurately (e.g., "identical tones should not be put together"), which partially supported the explicit nature of the acquired OCP-Unit generalization.

One additional subset of exceptions in Exp II merits further examination, namely the four rare output tokens with a L-L sequence against the participants' L1 tonal phonotactics. With these tokens, one might question whether the inclusion-exclusion task could really tap into implicit knowledge if the L1 knowledge, which is generally assumed to be implicit, could be violated in online production. After inspecting these four tokens, we found that they all occurred in the *exclusion* condition; two were from the same participant in the Terminal group, and two participants of the Control group respectively contributed one token. Since the exclusion condition required participants' deliberate effort to product outputs that they believed to be different from the training input, the learners in the task should have recruited their explicit knowledge of the tonal gaps in the training input, rather than their L1 phonology. It may thus not be surprising that this conscious process external to participants' L1 phonology could have allowed them to produce outputs violating their L1 tonal phonotactics. The occasion was nevertheless extremely rare, as the automatized L1 phonology would still restrict the learners' production to a great extent. Put differently, the four L-L tokens only support the connection between the exclusion

condition and the application of consciously acquired target knowledge, rather than undermine the conclusions regarding implicitly learned generalizations.

All in all, with Exp II, we provided supporting evidence of the Terminal group learning both OCP-H and OCP-L as an implicit phonological generalization. Alternatively, the learners of the Unit group learned OCP-Unit as an explicit generalization and unexpectedly demonstrated the implicit learning of the OCP-H generalization.

5 General discussion

In this study, we designed two AGL experiments to test if level-based and unit-based tonal dissimilation are equally learnable as an implicit phonological generalization by exposing three groups of learners to auditory input with distinct tonal patterns. In Exp I, we tested the learning performance with a set of novel test items in an acceptability judgment task. In Exp II, learners were asked to produce novel test items with free combinations of tones. The primary findings for the two target groups Terminal and Unit are summarized in **Table 9**. In both AGL experiments, we consistently found evidence of the learning of OCP-H and OCP-L as an implicit phonological generalization for both target groups, and OCP-Unit was at best learned as explicit knowledge. Accordingly, we would conclude that level-based and unit-based OCP patterns are *not* two equally learnable implicit phonological generalizations, and learners are expected to favor the implicit learning of the OCP-Terminal generalization. In §5.1, we will discuss the possible sources of the learned OCP generalizations and alternative explanations. We then focus on the theoretical implications of our findings for the proposal of contour tone unit in §5.2. Limitations and directions for future studies are summarized in §5.3.

	Group = Terminal	**Group = Unit**
Exp I Acceptability Judgment	OCP-H (implicit)	OCP-H (implicit)
Exp II Production	OCP-H and OCP-L (implicit) OCP-Unit (explicit)	OCP-H (implicit) OCP-Unit (explicit)

Table 9: A summary of the acquired OCP generalization by group in Exp I and Exp II.

5.1 Possible sources of the learned OCP and anti-OCP generalizations

Two possible primary sources of the OCP generalizations acquired implicitly are explored in this section, namely grammar induction (e.g., the acquisition of target constraint ranking) or domain-general statistical learning (e.g., Conway & Christiansen 2006; Walk & Conway 2015). We argue

that implicit statistical learning cannot be held solely accountable for the implicit learning of OCP generalizations demonstrated by the two target groups in our experimental findings for two reasons. First, if the OCP-H patterns were implicitly acquired by the Unit group via statistical learning of underrepresented tonal patterns due to the absence of H-H and H-HL from the training input, the OCP-Unit patterns should have been acquired as an implicit generalization as well. Second, statistical learning may not have an explanation for why only OCP-H was acquired implicitly by the Terminal group in Exp I since both adjacent Hs and Ls were completely absent from the training input.

Alternatively, the implicit bias against adjacent Hs across the board for the two target groups could be explained in a grammatical account, such as the cumulative effects of phonological markedness in a weight-based constraint model (e.g., Harmonic Grammar; Legendre et al. 1990). In previous research of phonological tones, H was claimed to be universally more marked than L (e.g., Maddieson 1978: 342–343; Pulleyblank 1986: 125–127; cf. Hyman 2010). This asymmetry might be the foundation of why H, rather than L, is deleted in Luba when OCP-H and OCP-L are violated on both ends of LH (see §2). It also helps account for the higher learnability of the contour simplification rule retaining L instead of H in the output in Kao's (2017) AGL study reviewed in §2. Following this line of reasoning, it is plausible to assume that the markedness constraint *H may inherently have a higher constraint weight than *L. If this is the case, the sum of violated constraint weights would be intrinsically higher for OCP-H and *H than for OCP-L and *L, which gives rise to a stronger bias against di-tonal patterns with adjacent Hs. We will leave this hypothesis to be tested in future works.[21]

One may also wonder if L1 phonology plays any role in shaping the implicit knowledge of OCP-H and OCP-L above, given our use of L1 tones in the training input and an explicit reference to these L1 tones in the instruction on the production task in Exp II. As we discussed earlier, learning an AGL is essentially equivalent to learning an L2 for adult participants, in which the transfer of L1 knowledge might be inevitable. For instance, in Chen's (2020) AGL study, a significant effect of L1 phonological neighborhood density on learners' performance was discovered for all participant groups. However, as Chen (2020) still found a substantial difference in the learnability of target tonal phonotactics, the author concluded that the L1 interference may only lead to an underestimation of the difference in learners' performance, rather than contribute directly to the difference. Likewise, in the current study, we do not rule out any possible L1 interference, but we assume that this interference was not a determining factor that resulted in the significant differences in test performance across the three learner groups. To explore potential L1 interferences, we could examine the performance of the Control group exposed only to randomly distributed di-tonal sequences without consistent tonal gaps;

---

21 Readers are also referred to Breiss (2020) for an AGL study of cumulative effects on phonotactic learning.

any bias demonstrated by the Control group could be partially ascribed to L1 transfer. For instance, in both AGL experiments, the Control group consistently demonstrated an implicit *anti*-OCP-H generalization, which might be resulted from the transfer of L1 statistical knowledge in addition to the perceptual salience of adjacent Hs (see §3.5). One possibility is that there is higher type frequency of di-tonal combinations with adjacent Hs in Taiwan Mandarin, and this statistical trend in the learners' L1 lexicon was transferred to their judgment and production in our AGL experiments. To test this hypothesis, we extracted the first 500 top-frequency disyllabic word entries from the Academic Sinica Spoken Corpus of Taiwan Mandarin (Tseng 2005) and calculated the number of words with a surface di-tonal pattern violating either OCP-H and OCP-L. Among the 500 entries, we found 150 violating OCP-H (56 = LH-HL; 22 = LH-H) and only 74 violating OCP-L (30 = HL-LH; 22 = HL-L), an asymmetry that seems to be in line with our speculation. Assuming the L1 influence is true, the implicit OCP-H generalization acquired by both target groups further confirms that learners could work against the transferred L1 lexical trend to learn novel linguistic patterns in AGL settings.

Unlike the implicit OCP generalizations discussed above, the OCP-Unit generalization was acquired as explicit knowledge by all three learner groups in Exp II. This explicit knowledge may have arisen from the learners' metalinguistic knowledge of the tonal labels in their L1, which then directed their awareness to the lack of di-tonal sequences with identical tones in the training input. In our experiments, we conducted a retrospective interview with the participants after they completed the test session to ask them to verbally report the patterns hidden in the training input. Many of them used the tonal labels such as *first tone* and *second tone* to refer to the four tones in Taiwan Mandarin, which they had learned from formal language education in Taiwan. Similar descriptions were also used by the participants in our Exp I with no explicit reference to Taiwan Mandarin in the instructions. Some participants even referred to *ping* (平) and *zhe* (仄) labels associated with the historical origins of Chinese tones, presumably because of their familiarity with Chinese literature. Since tones are transcribed as units with the metalinguistic tonal labels, it may not be surprising that the metalinguistic knowledge of L1 tones, which is explicitly "committed to memory through practice and rehearsal" (e.g., Hulstijn 2005; Andringa & Rebuschat 2015: 188), is also largely unit-based.

5.2 Contour tone unit

The evidence collected in this study runs counter to an implicit and perhaps also grammar-based OCP generalization that operates on the root node of single tonal constituents, which implies the redundancy of a unit-based tonal representation. The origin of the debate over whether tones are phonologically represented as single constituents dates to Pike's (1948) pioneering research, which distinguished *level-pitch register* from *gliding-pitch* tonal systems. Pike also stated that gliding is "the basic tonemic unit" in the latter (1948: 8) and "must be treated as unitary

tonemes and cannot be broken down into end points" (1948: 10). This distinction was further extended to Pike's another influential statement that "many of the languages of China appear to have systems somewhat like this [gliding-pitch tonal system]" (1948: 9; texts in the brackets are supplied by the author). It has led many researchers to view tones, contours in particular, in Chinese languages as single phonological units (cf. Woo 1969; Duanmu 1994; Evans 2008). Although our study did not investigate all tonal processes based on the unit-based model, we have presented new experimental evidence for rethinking the necessity of positing a unit-based representation of tones at the level of phonological computation. That said, whether adjacent terminal level tones could be projected into a single phonological unit at a higher prosodic level (e.g., Pierrehumbert & Beckman 1988: §6.5.1; Yip 1989: 163–164) awaits further investigation.

There also exists possibilities that terminal level tones could be encoded as a unit in speech processing or speech planning that does not involve the computation of abstract phonological identity. For instance, Xu & Wang's (2001) phonetic study showed that contour tones on a syllable are produced with a unitary tonal target, which is approximated differently from the same contour formed by a sequence of level tones across adjacent syllables. They considered this finding to be the evidence of phonological features like [fall] and [rise] for contour tones (cf. Wang 1967). However, it might not be surprising that a tonal contour in the same speech planning frame (e.g., syllable in Mandarin Chinese) is produced as a whole because its phonetic spell-out relies on the gradual adjustment of the same set of articulators (e.g., the vocal cords), which might take place beyond the level of phonological computation (cf. Articulatory Phonology; Browman & Goldstein 1992; et seq.). Likewise, findings from other phonetic studies suggested that pitch contour is a more salient perceptual cue than pitch height in certain tone languages (e.g., Gandour 1981; 1983; Fon et al. 2004; Xu et al. 2006), which could be viewed as the processing of tonal contours as a unit. Perceptual salience is nevertheless not direct evidence for a unit-like phonological representation of tones; further decomposition of tones into discrete units may still occur in phonological computation after a phonetic f0 contour is mapped to a tonal category.

Arguments favoring a unit-based representation of tones also come from studies of speech errors at the suprasegmental level. In Wan & Jaeger's (1998) corpus study of speech errors in Taiwan Mandarin, errors with tonal substitution were found to always involve whole-tone substitution regardless of whether the target tone is level or contour. This observation was argued as the evidence of a unit-based representation of contour tones; if contour tones are merely represented as a sequence of independent level tones, we would have observed errors that involve the replacement of only one of the level tones or the splitting of a contour tone into separate level tones, at least in Taiwan Mandarin. Nevertheless, since whole-tone substitution frequently results in homophones in Taiwan Mandarin due to a small inventory of syllable-tone combinations (e.g., [fən$^{\text{L}}$ xɔŋ$^{\text{LH}}$] 'pink' vs [fən$^{\text{H}}$ xɔŋ$^{\text{LH}}$] 'dividend'), whole-tone substitution is barely distinguishable from *whole-word* substitution. Speech errors might thus have no implication on the phonological

representation and computation of level and contour tones. Even if more solid evidence of whole-tone substitution is available,[22] it could still be that contour tones are encoded and accessed as chunks (i.e., *proximate unit* in the sense of Chen et al. (2002) and O'Seaghdha et al. (2010)) for the ease of phonological encoding and more rapid access of syllable-tone combinations in Taiwan Mandarin. In the end, we might not be able to extrapolate from speech errors during the encoding process to the abstract representation of tones at the phonological level.

5.3 Limitations and directions for future research

Before closing, we would like to acknowledge a few limitations on our study that will have to be addressed in future works to draw more insights on the learnability and the implicit nature of OCP generalizations as well as the phonological representation of tones. In particular, some may remain dubious as to whether the differences in learnability found in AGL studies can truly reflect learners' linguistic competence. After all, the difficulty of learning a target pattern does not necessarily entail the incompatibility between the pattern and the hardwired implicit linguistic knowledge. It could be the case that the target pattern is compatible with the implicit knowledge but is also intrinsically more complex and less transparent. We have partially resolved this ambiguity by demonstrating separate channels for acquiring the two respective OCP generalizations (i.e., implicit/automatized/unconscious vs. explicit/effortful/conscious learning), but more convincing conclusions could be reached with the following methodological improvements.

First, while we tested our learners with a different set of stimuli in both AGL experiments, a true poverty-of-the-stimulus design (e.g., Wilson 2006) was not possible and same tones were used to created training and test items. In other words, we were not able to investigate whether an acquired phonological generalization could be extended to different members of the same phonological class. This was in part due to a considerably small number of natural phonological tonal contrasts (Yip 2002: §2.4–2.5) that are available for creating tonal patterns with different members of the same phonological classes. On top of it, tone languages hardly exploit all these tonal contrasts, making it nearly impossible to recruit speakers with all these tonal contrasts in their native language in a tonal AGL study. Training participants to learn all these tonal contrasts may be a viable option, but the process may be time-consuming and could provide extra hints that spoil the purpose of our AGL experiments. Our remedy was thus to incorporate awareness measures into our experimental design so grammar learning could still be identified even if the same phonological tones were explicitly *reused* from the phonological memory. Of

---

22 In their study of tonal errors in Cantonese, Alderete et al. (2019: 11) proposed that the most unambiguous evidence for whole-tone substitution could be found in combinations between contour tones and syllables ending in [-p], [-t], or [-k]. As these syllable-tone combinations are phonotactic illicit in Cantonese, they cannot be instances of whole-word substitution. With a loose restriction on syllable-tone combinations in Taiwan Mandarin, such evidence of whole-tone substitution as an independent process may not be available at all.

course, future works may need to adopt additional objective awareness measures (e.g., Suzuki 2017; Maie & DeKeyser 2020) to validate and replicate the implicit and explicit learning of OCP generalizations observed in this study.

Second, a better comparison of the learnability of OCP generalizations with qualitatively similar training inputs will be necessary. In our experiments, the accidental gaps incorporated into the training input for the Unit group could have accidentally established a more complicated learning setting. The end result could have been a negative effect on learning performance that was not able to be identified with our experimental design or result analyses. One possible change is to limit the comparison to the OCP-H and OCP-Unit generalizations for an equal number of systematic gaps and accidental gaps to be included in the training input.

Third, a follow-up study may have to tease apart the learnability driven by phonetical naturalness from that motivated by phonological computability. In our study, we removed di-tonal sequences HL-L, LH-H, HL-LH, and LH-HL from the training input for the Terminal group. The absence of HL-L and LH-H may have triggered the learning of a constraint phonetically based on tonal absorption (e.g., Hyman 2007) in addition to the OCP generalizations implied in these gaps. To better estimate the net effect of OCP, the influence of tonal absorption should be subtracted from the learning performance. This could be done with another Absorption group exposed to training input only with di-tonal gaps linked to the learning of tonal absorption. The similarity in the learning performance of the Absorption and Terminal groups would thus be informative in terms of evaluating the net effect of OCP-Terminal violation.

Finally, the results in our Exp II could not be analyzed with mixed-effects modeling for more conservative estimates due to a lower number of participants per group. Accordingly, the significant main effects and interactions, albeit largely consistent with those found in Exp I, still warrant replication in large-scale studies.

6 Concluding remarks

In the current study, we revisited a long-standing debate over the necessity to analyze tonal OCP in a unit-based model that represents level and contour tones as single constituents. As a rare attempt to experimentally investigate the issue in two AGL experiments, we found no supporting evidence of a learnable, unit-based implicit generalization of OCP-Unit. Consequently, it may become more cautious to assume a unit-based representation of phonological tones. This conclusion was made based on the experimental designs that covered both perception and production with methodological improvements taking learners' awareness into account. The findings of this study are expected to initiate a more extensive examination of the evidence in experimental works, which are desperately needed to continue the exploration of the phonological knowledge of tones.

Additional file

The additional file for this article can be found as follows:

- **Appendix.** Appendices A to D. DOI: https://doi.org/10.16995/glossa.5795.s1

Ethics and consent

This study has been approved by the Research Ethics Committee of National Tsing Hua University in Taiwan (REC ID: 10610HS071). Informed consents were properly obtained from all participants.

Acknowledgements

We are indebted to the Associate Editor, Juliet Stanton, for her kind assistance and comments throughout the review process. We also highly appreciate the constructive criticisms from four anonymous reviewers, Carlos Gussenhover, James Myers, and the audience at NINJAP 2019 and AMP 2020 that have helped improve the earlier drafts. Finally, we would like to give credit to our assistants Ssu-Han Chang, Han-Chun Lin, Yi-Shan Lin, Wei-Hsin Lo, and Tzu-Hsuan Tseng, who have contributed to different aspects of the study. The usual disclaimer applies.

Funding information

The study is funded by the Ministry of Science and Technology, Taiwan (107-2410-H-007-002-MY2; 108-2410-H-007-030-MY3).

Competing interests

The author has no competing interests to declare.

References

Alderete, John & Chan, Queenie & Yeung, H. Henny. 2019. Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition* 191. 103952. DOI: https://doi.org/10.1016/j.cognition.2019.04.021

Andringa, Sible & Rebuschat, Patrick. 2015. New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition* 37(02). 185–196. DOI: https://doi.org/10.1017/S027226311500008X

Aranovich, Raul. 1994. The Tone System of Acatlan Mixtec and Some Exceptions to the OCP. *Linguistic Notes from La Jolla* 17. 3–26.

Bao, Zhiming. 1990. *On the nature of tone*. Cambridge, MA: MIT dissertation.

Bao, Zhiming. 1999. *The structure of tone*. Oxford, UK: Oxford University Press.

Barr, Dale J. & Levy, Roger & Scheepers, Christoph & Tily, Harry J. 2013. Random effects structure for confirmatory hypothesis testing : Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, Doug & Bolker, Ben & Maechler, Martin & Walker, Steven. 2020. lme4: Linear mixed-effect models using S4 classes. v1.1-26. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Berent, Iris. 2013. The phonological mind. *Trends in Cognitive Sciences* 17(7). 319–27. DOI: https://doi.org/10.1016/j.tics.2013.05.004

Berent, Iris & Marcus, Gary F. & Shimron, Joseph & Gafos, Adamantios I. 2002. The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition* 83(2). 113–139. DOI: https://doi.org/10.1016/S0010-0277(01)00167-6

Berns, Janine. 2016. The Phonological Representation of Affricates. *Language and Linguistics Compass* 10(3). 142–156. DOI: https://doi.org/10.1111/lnc3.12179

Breiss, Canaan. 2020. Constraint cumulativity in phonotactics: evidence from artificial grammar learning studies. *Phonology* 37(4). 551–576. DOI: https://doi.org/10.1017/S0952675720000275

Browman, Catherine P. & Goldstein, Louis. 1992. Articulatory phonology: an overview. *Haskins Laboratories Status Report on Speech Research* 111/112. 23–42.

Caldwell-Harris, Catherin L. & Lancaster, Alia & Ladd, D. Robert & Dediu, Dan & Christiansen, Morten H. 2015. Factors influecing sensitivity to lexical tone in an artificial language: Implications for second language learning. *Studies in Second Language Acquisition* 37(2). 335–357. DOI: https://doi.org/10.1017/S0272263114000849

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27(03). 345–392. DOI: https://doi.org/10.1017/S0952675710000199

Chan, Ricky K.W. & Leung, Jenny H.C. 2014. Implicit learning of L2 word stress regularities. *Second Language Research* 30(4). 463–484. DOI: https://doi.org/10.1177/0267658313510169

Chen, Jenn-Yeu & Chen, Train-Min & Dell, Gary S. 2002. Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task. *Journal of Memory and Language* 46(4). 751–781. DOI: https://doi.org/10.1006/jmla.2001.2825

Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486364

Chen, Tsung-Ying. 2010. Some remarks on Contour Tone Units. *Journal of East Asian Linguistics* 19(2). 103–135. DOI: https://doi.org/10.1007/s10831-010-9057-9

Chen, Tsung-Ying. 2020. An inductive learning bias toward phonetically driven tonal phonotactics. *Language Acquisition* 27(3). 331–361. DOI: https://doi.org/10.1080/10489223.2020.1769630

Chomsky, Noam. 1980. Principles and parameters in syntactic theory. In Hornstein, Norbert & Lightfoot, David (eds.), *Explanation in linguistics: The logical problem of language acquisition*, 32–75. London, UK: Longman.

Clark, Mary M. 1989. OCP Effects in Zulu. *Linguistic Analysis* 19(1–2). 59–76.

Clark, Mary M. 1990. *The Tonal System of Igbo*. Dordrecht: Foris. DOI: https://doi.org/10.1515/9783110869095

Conway, Christopher M. & Christiansen, Morten H. 2006. Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science* 17(10). 905–912. DOI: https://doi.org/10.1111/j.1467-9280.2006.01801.x

Curran, Tim. 2001. Implicit learning revealed by the method of opposition. *Trends in Cognitive Sciences* 5(12). 503–504. DOI: https://doi.org/10.1016/S1364-6613(00)01791-5

Daly, John P. & Hyman, Larry M. 2007, July 17. On the representation of tone in Peñoles Mixtec. *International Journal of American Linguistics*. University of Chicago Press. DOI: https://doi.org/10.1086/519057

Dekeyser, Robert M. 2003. Implicit and explicit learning. In Doughty, Catherine J. & Long, Michael H. (eds.), *Handbook of Second Language Acquisition*, 313–348. Oxford, UK: Blackwell.

Demuth, Katherine. 1993. Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language* 20(2). 275–301. DOI: https://doi.org/10.1017/S030500090000828X

Destrebecqz, Arnaud & Cleeremans, Axel. 2001. Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review* 8(2). 343–350. DOI: https://doi.org/10.3758/BF03196171

Dienes, Zoltán. 2007. Subjective measures of unconscious knowledge. *Progress in Brain Research* 168. 49–64. DOI: https://doi.org/10.1016/S0079-6123(07)68005-4

Downing, Laura J. 2003. Compounding and tonal non-transfer in Bantu languages. *Phonology* 20(1). 1–42. DOI: https://doi.org/10.1017/S0952675703004457

Downing, Laura J. 2005. On the ambiguous segmental status of nasals in homorganic NC sequences. In van Oostendorp, Marc & van de Weijer, Jeroen (eds.), *The Internal Organization of Phonological Segments*, 183–216. Berlin, Boston: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110890402.183

Duanmu, San. 1994. Against contour tone units. *Linguistic Inquiry* 25(4). 555–608.

Ellis, Rod. 2005. Measuring implicit and explicit knowlegde of a second language: A psychometric study. *Studies in Second Language Acquisition* 27(2). 141–172. DOI: https://doi.org/10.1017/S0272263105050096

Evans, Jonathan P. 2008. 'African' Tone in the Sinosphere. *Language and Linguistics* 9(3). 463–490.

Evans, Jonathan P. 2015. High is not just the opposite of Low. *Journal of Phonetics* 51. 1–5. DOI: https://doi.org/10.1016/j.wocn.2015.05.001

Finley, Sara. 2012. Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes* 27(10). 1550–1562. DOI: https://doi.org/10.1080/01690965.2012.660168

Finley, Sara. 2015. Consequences of monotonicity: Representations and learnability. *Theoretical Linguistics* 41(1–2). 69–78. DOI: https://doi.org/10.1515/tl-2015-0003

Finley, Sara. 2017. Learning metathesis: Evidence for syllable structure constraints. *Journal of Memory and Language* 92. 142–157. DOI: https://doi.org/10.1016/j.jml.2016.06.005

Finley, Sara & Badecker, William. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61(3). 423–437. DOI: https://doi.org/10.1016/j.jml.2009.05.002

Fon, Janice & Chiang, Wen-Yu & Cheung, Hintat. 2004. Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32(2). 249–280.

Gallagher, Gillian. 2013. Learning the identity effect as an artificial language: bias and generalisation. *Phonology* 30(2). 253–295. DOI: https://doi.org/10.1017/S0952675713000134

Gandour, Jack. 1977. On the Interaction between Tone and Vowel Length: Evidence from Thai Dialects. *Phonetica* 34(1). 54–65. DOI: https://doi.org/10.1159/000259869

Gandour, Jack. 1981. Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics* 9(1). 20–36.

Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11(3). 149–175. DOI: https://doi.org/10.1016/S0095-4470(19)30813-7

Goldsmith, John A. 1979. *Autosegmental Phonology*. Cambridge, MA: MIT dissertation.

Graham, Calbert R. & Williams, John N. 2018. Implicit learning of Latin stress regularities. *Studies in Second Language Acquisition* 40(1). 3–29. DOI: https://doi.org/10.1017/S0272263116000371

Hamrick, Phillip & Sachs, Rebecca. 2018. Establishing Evidence of Learning in Experiments Employing Artificial Linguistic Systems. *Studies in Second Language Acquisition* 40(1). 153–169. DOI: https://doi.org/10.1017/S0272263116000474

Hayes, Bruce & White, James. 2015. Saltation and the P-map. *Phonology* 32(2). 267–302. DOI: https://doi.org/10.1017/S0952675715000159

He, Hua & Zhang, Hui & Ye, Peng & Tang, Wan. 2017. A test of inflated zeros for Poisson regression models. *Statistical Methods in Medical Research* 28(4). 1157–1169. DOI: https://doi.org/10.1177/0962280217749991

Hsiao, Y. E. 2015. Rethinking OCP Effects on Tone Sandhi. *Language and Linguistics* 16(6). 927–945. DOI: https://doi.org/10.1177/1606822X15602616

Huang, Karen. 2017. From pitch contour variation to tone change. *International Journal of Chinese Linguistics* 4(2). 273–307. DOI: https://doi.org/10.1075/ijchl.16016.hua

Huang, Tsan. 2001. The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics* 55. 23–42.

Hulstijn, Jan H. 2005. Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition* 27(2). 129–140. DOI: https://doi.org/10.1017/S0272263105050084

Hulstijn, Jan H. 2015. Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In Rebuschat, Patrick (ed.), *Implicit and Explicit Learning of Languages*, 25–46. London, UK: John Benjamins. DOI: https://doi.org/10.1075/sibil.48.02hul

Hyman, Larry M. 2007. Universals of tone rules: 30 years later. In Riad, Tomas & Gussenhoven, Carlos (eds.), *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*, 1–34. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110207569.1

Hyman, Larry M. 2010. Markedness and the Phonological Typology of Two-height Tone Systems. *UC Berkeley PhonLab Annual Report* 6. 283–296. DOI: https://doi.org/10.5070/P779B3P9D8

Hyman, Larry M. & Kenneth Vanbik. 2004. Directional rule application and output problems in Hakha Lai tone. *Language and Linguistics* 5(4). 821–861.

Ionin, Tania & Zubrizarreta, María Luisa & Philippov, Vadim. 2009. Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition* 12(3). 337–361. DOI: https://doi.org/10.1017/S1366728909990149

Jenks, Peter & Rose, Sharon. 2011. High tone in Moro: Effects of prosodic categories and morphological domains. *Natural Language and Linguistic Theory* 29(1). 211–250. DOI: https://doi.org/10.1007/s11049-011-9120-x

Kao, Sophia. 2017. *Phonological Learning Bias in Tone Patterns*. New York, NY: State University of New York at Stony Brook dissertation.

Kappa, Ioanna & Papoutsi, Marieta. 2019. OCP factors governing the realization of [Obstruent + Sonorant] clusters in child Greek: A case study. In Guijarro-Fuentes, Pedro & Suarez-Gomez, Cristina (eds.), *Proceedings of GALA 2017: Language Acquisition and Development*, 437–450. Cambridge Scholars Publishing.

Krashen, Stephen. 1982. *Principles and practice in second language learning and acquisition*. Oxford, UK: Pergamon.

Lai, Regine. 2015. Learnable vs. Unlearnable Harmony Patterns. *Linguistic Inquiry* 46(3). 425–451. DOI: https://doi.org/10.1162/LING_a_00188

Leben, William R. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT dissertation.

Legendre, Geraldine & Miyata, Yoshiro & Smolensky, Paul. 1990. *Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An application*. Boulder, CO.

Li, Qian & Chen, Yiya. 2016. An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics* 54. DOI: https://doi.org/10.1016/j.wocn.2015.10.002

Li, Qian & Chen, Yiya & Xiong, Ziyu. 2019. Tianjin Mandarin. *Journal of the International Phonetic Association* 49(1). 109–128. DOI: https://doi.org/10.1017/S0025100317000287

Lichtman, Karen. 2013. Developmental Comparisons of Implicit and Explicit Language Learning. *Language Acquisition* 20(2). 93–108. DOI: https://doi.org/10.1080/10489223.2013.766740

Lin, Hui-shan. 2008. Variable directional applications in Tianjin tone sandhi. *Journal of East Asian Linguistics* 17(3). 181–226. DOI: https://doi.org/10.1007/s10831-008-9024-x

Lin, Hui-shan. 2011. Sequential and Tonal Markedness in Dongshi Hakka Tone Sandhi. *Language and Linguistics* 12(2). 313–357.

Lin, Hui-shan. 2019. Tonal (non-)transfer in Kunming Reduplication. *Journal of East Asian Linguistics* 28(1). 55–105. DOI: https://doi.org/10.1007/s10831-019-09190-8

Lin, Yen-Hwei. 2011. Affricates. In van Oostendorp, Marc & Ewen, Colin J. & Hume, Elizabeth V. & Rice, Keren (eds.), *The Blackwell companion to phonology, Vol. I, General issues and segmental phonology*, 367–390. Oxford, UK: Blackwell.

Liu, Siyun & Samuel, Arthur G. 2004. Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech* 47(2). 109–138. DOI: https://doi.org/10.1177/00238309040470020101

Maddieson, Ian. 1978. Universals of tone. In Greenberg, Joseph & Ferguson, Charles A. & Moravcsik, Edith A. (eds.), *Universals of Human Language*, 337–356. Stanford, CA: Stanford University Press.

Maie, Ryo & Dekeyser, Robert M. 2020. Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition* 42(2). 359–382. DOI: https://doi.org/10.1017/S0272263119000615

Martin, Alexander & Peperkamp, Sharon. 2020. Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony and disharmony. *Phonology* 37(1). 65–90. DOI: https://doi.org/10.1017/S0952675720000044

McCarthy, John J. 1986. OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17(2). 207–263.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511613333

Mmusi, Sheila Onkaetse. 1992. OCP Violations in Setswana: Evidence for Redefining the OCP? *Studies in the Linguistic Sciences* 22(1). 123–142.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127. DOI: https://doi.org/10.1017/S0952675708001413

Moreton, Elliott & Pertsova, Katya. 2016. Implicit and Explicit Processes in Phonotactic Learning. In Scott, Jennifer & Waughtal, Deb (eds.), *Proceedings of the 40th annual Boston University Conference on Language Development*, 277–290. Somerville, MA: Casadilla Press.

Morgan-Short, Kara & Steinhauer, Karsten & Sanz, Cristina & Ullman, Michael T. 2012. Explicit and Implicit Second Language Training Differentially Affect the Achievement of Native-like Brain Activation Patterns. *Journal of Cognitive Neuroscience* 24(4). 933–947. DOI: https://doi.org/10.1162/jocn_a_00119

Myers, James. 2012. Testing phonological grammar with lexical data. In Myers, James (ed.), *In search of grammar: Empirical methods in linguistics*, 141–176. Taipei, Taiwan: Academia Sinica.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language & Linguistic Theory* 15(4). 847–892. DOI: https://doi.org/10.1023/A:1005875608905

Oakden, Chris. 2020. Notational equivalence in tonal geometry. *Phonology* 37(2). 257–296. DOI: https://doi.org/10.1017/S0952675720000123

O'Seaghdha, Padraig G. & Chen, Jenn-Yeu & Chen, Train-Min. 2010. Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition* 115(2). 282–302. DOI: https://doi.org/10.1016/j.cognition.2010.01.001

Peirce, Jonathan & Gray, Jeremy R. & Simpson, Sol & MacAskill, Michael & Höchenberger, Richard & Sogo, Hiroyuki & Kastman, Erik & Lindeløv, Jonas Kristoffer. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203. DOI: https://doi.org/10.3758/s13428-018-01193-y

Peperkamp, Sharon & Le Calvez, Rozenn & Nadal, Jean-Pierre & Dupoux, Emmanuel. 2006. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101(3). B31-41. DOI: https://doi.org/10.1016/j.cognition.2005.10.006

Pierrehumbert, Janet B. & Beckman, Mary E. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pike, Kenneth L. 1948. *Tone Languages*. Ann Arbor: University of Michigan Press.

Powell, Michael J.D. 2009. *The BOBYQA algorithm for bound constrained optimization without derivatives*. Cambridge, UK. Department of Applied Mathematics and Theoretical Physics, Cambridge University.

Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory: constraint interaction in generative grammar*. Oxford, UK: Blackwell. DOI: https://doi.org/10.1002/9780470759400

Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel. DOI: https://doi.org/10.1007/978-94-009-4550-0

Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted. 2003. Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. In Garding, Gina & Tsujimura, Mimu (eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 104–114. Somerville, MA: Casadilla Press.

Qian, Zhengyi. 1993. *Boshan Fangyan Yanjiu [A Study of the Boshan Dialect]*. Beijing: Shehui Kexue Wenxian Chubanshe.

R Core Team. 2021. R: a language and environment for statistical computing. v4.0.4. Retrieved from http://www.r-project.org/

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6(6). 855–863. DOI: https://doi.org/10.1016/S0022-5371(67)80149-X

Salffner, Sophie. 2010. *Tone in the phonology, lexicon and grammar of Ikaan*. SOAS, London, UK: University of London dissertation.

Seidl, Amanda & Buckley, Eugene. 2005. On the Learning of Arbitrary Phonological Rules. *Language Learning and Development* 1(3–4). 289–316. DOI: https://doi.org/10.1080/15475441.2005.9671950

Suzuki, Yuichi. 2017. Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics* 38(5). 1229–1261. DOI: https://doi.org/10.1017/S014271641700011X

Tseng, Shu-chuan. 2005. Contracted Syllables in Mandarin: Evidence from Spontaneous Conversations. *Language and Linguistics* 6(1). 153–180.

Walk, Anne M. & Conway, Christopher M. 2015. Implicit statistical learning and language acquisition: Experience-dependent constraints on learning. In Rebuschat, Patrick (ed.), *Implicit*

and Explicit Learning of Languages 2, 191–212. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/sibil.48.09wal

Wan, I-ping & Jaeger, Jeri. 1998. Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15(3). 417–461. DOI: https://doi.org/10.1017/S0952675799003668

Wang, Tianlin & Saffran, Jenny R. 2014. Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology* 5(AUG). 953. DOI: https://doi.org/10.3389/fpsyg.2014.00953

Wang, William S. Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33(2). 93–105. DOI: https://doi.org/10.1086/464946

Wee, Lian-Hee. 2004. *Inter-tier Correspondence Theory*. New Jersey, NJ: Rutgers University dissertation.

Wee, Lian-Hee. 2015. Prominence from Complexity: Capturing Tianjin Ditonal Patterns. *Language and Linguistics* 16(6). 891–926. DOI: https://doi.org/10.1177/1606822X15602614

Wee, Lian-Hee. 2019. *Phonological Tone*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/9781316410912

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130(1). 96–115. DOI: https://doi.org/10.1016/j.cognition.2013.09.008

White, James & Sundara, Megha. 2014. Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition* 133(1). 85–90. DOI: https://doi.org/10.1016/j.cognition.2014.05.020

Wilson, Colin. 2003. Experimental Investigation of Phonological Naturalness. In Garding, Eva & Tsujimura, Mimu (eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 101–114. Somerville, MA: Casadilla Press.

Wilson, Colin. 2006. Learning Phonology With Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30(5). 945–982. DOI: https://doi.org/10.1207/s15516709cog0000_89

Woo, Nancy. 1969. *Prosody and Phonology*. Cambridge, MA: MIT dissertation.

Xu, Yi & Wang, Q. Emily. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33(4). 319–337. DOI: https://doi.org/10.1016/S0167-6393(00)00063-7

Xu, Yisheng & Gandour, Jackson T. & Francis, Alexander L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America* 120(2). 1063–1074. DOI: https://doi.org/10.1121/1.2213572

Yip, Moira. 1989. Contour tones. *Phonology* 6(1). 149–174. DOI: https://doi.org/10.1017/S095267570000097X

Yip, Moira. 2002. *Tone*. Cambridge, UK: Cambridge University Press.

Yue-Hashimato, Oi-kan. 1972. *Phonology of Cantonese, Volume 1*. Cambridge, UK: Cambridge University Press.

Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution*. New York: Routledge.

Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. In Hayes, Bruce & Steriade, Donca & Kirchner, Robert (eds.), *Phonetically based Phonology*, 157–190. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486401.006

Zhang, Jie. 2010. Issues in the Analysis of Chinese Tone. *Language and Linguistics Compass* 4(12). 1137–1153. DOI: https://doi.org/10.1111/j.1749-818X.2010.00259.x

Zhang, Jie. 2014. Tones, Tonal Phonology, and Tone Sandhi. In Huang, James C.-T., Li, Audrey Y.-H. & Simpson, Andrew (eds.), *The Handbook of Chinese Linguistics*, 443–464. Oxford:UK: Johns Wiley & Sons. DOI: https://doi.org/10.1002/9781118584552.ch17

Zhang, Jie & Liu, Jiang. 2011. Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica* 68(3). 161–191. DOI: https://doi.org/10.1159/000333387

Zhang, Jie & Liu, Jiang. 2016. The productivity of variable disyllabic tone sandhi in Tianjin Chinese. *Journal of East Asian Linguistics* 25(1). 1–35. DOI: https://doi.org/10.1007/s10831-015-9135-0