

# Supporting Information

Berent et al. 10.1073/pnas.1416851112

## SI Text

In what follows, we first define the syllable hierarchy. We next report the results of several auxiliary analyses designed to address alternative explanations for the findings.

**The Definition of the Syllable Hierarchy.** Across languages, certain syllable types (e.g., *blif*) are preferred to others (e.g., *lbif*). Linguistic research captures these preferences by appealing to the sonority profile of the consonant sequence occurring at the syllable's onset.

Sonority is a scalar phonological property that correlates with the loudness of segments. The most sonorous (i.e., salient) consonants are glides [e.g., *w*, *y*; with a sonority (*s*) of 4], followed by liquids (e.g., *l*, *r*; *s* = 3), nasals (e.g., *m*, *n*, *s* = 2), and obstruents (e.g., *b*, *p*; *s* = 1). Sonority, in turn, determines the well formedness of the syllable.

Syllables favor large sonority clines in their onsets—the larger the sonority cline, the better formed the syllable (1). For example, syllables like *bla* (with a large rise in sonority between the stop (*s* = 1) and the liquid (*s* = 3,  $\Delta s$  = 2) are preferred to *bna* (small rise,  $\Delta s$  = 1), which, in turn, are preferred to the sonority plateau in *bda* ( $\Delta s$  = 0); least preferred is the sonority fall in *lba* ( $\Delta s$  = -2). Indeed, as sonority distance decreases, the syllable becomes underrepresented across languages (2, 3).

Other sonority constraints have been noted in the syllable's final consonants (its coda), as well as in syllable contact (i.e., the co-occurrence of consonants across syllables). In both cases, small sonority clines are preferred (1). For example, disyllabic sequences like *al.ba* are preferred to *ab.la* (4). Such observations are significant because they suggest that the restrictions on consonant co-occurrence strictly depend on their structural role in the syllable.

However, although the notion of sonority can adequately capture the restrictions on syllable structure, it is unclear whether the grammar directly encodes restrictions on sonority (1) or whether these effects emerge from more basic grammatical constraints on feature co-occurrence (5).

Our interest here only concerns the question of embodiment vs. abstraction (i.e., whether the grammar includes abstract restrictions on syllable structure)—the precise formal account of these facts falls beyond the scope of our inquiry, and we will not discuss it further.

**Alternative Explanations for the Results.** The following sections examine several alternative explanations for the results. We begin by investigating whether our TMS manipulation disrupted the lip motor area and whether this effect impaired the identification of labial sounds, specifically. We next reexamine the effect of TMS in experiment 1 and compare responses to monosyllables versus disyllables. Our final set of analyses investigates the possibility that sensitivity to the syllable hierarchy is governed by the phonetic/acoustic properties of our materials rather than their phonological structure.

**The efficacy of our TMS manipulation.** In view of the resilience of the syllable hierarchy to TMS, one might worry that our manipulation failed to disrupt the lip's motor activity. As noted in *Materials and Methods*, however, the stimulation hotspot was identified by measuring the amount of TMS-induced muscle activity (motor evoked potentials (MEPs)) in the OO muscle (*Materials and Methods* and Fig. 1A). It was defined as the spot eliciting the largest TMS-induced MEPs in the OO muscle at minimal stimulation intensity for the whole experiment. Although the stimulation was induced cortically (as measured through the MEPs),

its effects were furthermore evident peripherally, as the ipsilateral peripheral nerves can also be stimulated and lead to facial twitches. A similar protocol was used in previous research, and it has been shown to impair the identification of labial sounds (6, 7).

To further test the functional efficacy of our TMS manipulation, we thus examined whether the disruption of the OO by TMS specifically impaired the perception of labial sounds in our present experiment. To this end, we reanalyzed the results of experiment 1 to compare the effect of TMS on labial-initial stimuli (e.g., *plik*) with those that do not begin with a labial consonant (e.g., *twaf*). We reasoned that if motor simulation mediates syllable count, then its disruption by TMS should have a stronger effect on labial-initial syllables. To ensure that our evaluation is not confounded by syllable type or the rhyme, we limited the analysis to stimuli that were matched for those properties. Our items (Appendix 1) included a total of two matched triplets that began with a labial stop (*plik*, *pnik*, *pkik* and *praf*, *pnaf*, *ptaf*); the sonorant-initial members (*lik*, *rpf*) were not included, as their initial phoneme was not labial. We next assessed the effect of TMS on sensitivity (*d'*) to labial-initial vs. nonlabial initial items in experiment 1 (all items in experiment 2 were labial-initial, so no such analyses are possible there). The means are provided in Fig. S1.

A 2 (labial/nonlabial)  $\times$  2 (TMS/sham)  $\times$  3 (syllable type) ANOVA yielded a marginally significant TMS  $\times$  labial interaction [ $F(1,8) = 3.81$ ,  $P < 0.09$ ;  $F(2,22) = 3.89$ ,  $P < 0.07$ ], and this effect was not further modulated by syllable type [ $F(2,16) = 2.18$ ,  $P < 0.15$ ;  $F(2,44) = 1.23$ ,  $P < 0.31$ , not significant].

Tukey HSD tests showed that TMS reduced sensitivity to labial-initial items ( $P < 0.02$ , by participants and items), but it did not reliably affect responses to nonlabial items (by participants:  $P < 0.15$ , by items:  $P < 0.004$ ). In addition, labial-initial items produced higher sensitivity than controls in the sham condition ( $P < 0.02$ , by participants and items), but this labial advantage was eliminated under the TMS manipulation ( $P > 0.41$ , not significant).

Summarizing, these findings demonstrate that the behavioral syllable count task was mediated by motor simulation that engaged the lip motor representation, and this simulation was further disrupted by our TMS manipulation. These results must be taken with caution, as the critical interaction was not fully significant (mostly likely, due to small number of labial-initial items). Nonetheless, the findings suggest that our TMS manipulation specifically disrupted the lip articulator.

**The effect of TMS on monosyllable.** Another concern is that the effect of TMS on participants' sensitivity in experiment 1 is driven not by the hierarchy of monosyllables, but rather by some artifact of their disyllabic counterparts. To counter this possibility, we next assessed the effect of TMS on the identification of monosyllables, specifically (Fig. S2).

A 2 TMS  $\times$  4 syllable type ANOVA on response accuracy to monosyllables (arcsine transformed) yielded an interaction, significant by items and marginally so across participants [ $F(3,24) = 2.59$ ,  $P < 0.08$ ;  $F(2,69) = 6.63$ ,  $P < 0.001$ ]. Planned contrasts confirmed that TMS only impaired the better-formed syllables. Specifically, the effect of TMS was significant for *bnif* [ $t(9) = 2.52$ ,  $P < 0.02$ ;  $t(24) = 3.71$ ,  $P < 0.005$ ] and marginally so for *blif*-type [ $t(9) = 1.88$ ,  $P < 0.08$ ;  $t(24) = 3.85$ ,  $P < 0.0005$ ]. In contrast, TMS did not reliably modulate responses to *bdif*- or *lbif*-type syllables (all  $t < 1$ ). The results from monosyllables thus mirror the conclusions from our omnibus analyses of overall sensitivity (*d'*). In both analyses, the effect of TMS was selective to syllable types that are better formed on the sonority hierarchy.

**The Role of Phonetic Factors.** The analyses reported thus far show that the misidentification of ill-formed syllables is not due to motor action alone. These results, however, do not necessarily demonstrate that people rely on abstract linguistic principles. Indeed, misidentification could have occurred because people failed to register the acoustic phonetic properties of the auditory stimuli. In this view, the phonetic form assigned by English speakers to inputs such as *lbif* consists of [lɒbif] (rather than the intended [lbif]). Critically, misidentification is caused not by the violation of abstract phonological constraints, but rather by “innocent misperception” (8) that occurs at lower levels of analysis—auditory or phonetic.

In what follows, we address two classes of phonetic explanations for the results. The first class attributes misidentification to the presence of phonetic properties that are specific to Russian—the native language of the talker who produced the stimuli used in experiment 1. We next consider the possibility that misidentification is due to the misinterpretation of cues that define stop consonants more broadly.

**The role of Russian-specific cues.** Experiment 1 presents our English-speaking participants with a double challenge. Not only are these syllables nonnative to English, but they are delivered by a talker who is a native speaker of Russian. Misidentification, then, could occur because participants failed to parse the phonetic characteristics of the Russian phonemes as intended by that talker. In what follows, we consider the role of three such characteristics.

We proceed in two steps. We first examine whether those cues do, in fact, elicit misidentification by investigating their effect in the sham condition. To ensure that these factors do not confound the TMS manipulation, we next proceed to either remove all items that carry that feature or explicitly gauge its effect on our TMS manipulation. The results make it clear that these phonetic factors cannot capture the effect of the syllable hierarchy or the TMS manipulation.

#### **The role of phonetic cues in the sham condition.**

- The velar-*ɫ*. Several of the monosyllables with falling sonority (e.g., *lbif*) began with the velar *ɫ* ([ɫ]). Because this velar allophone often occurs in English nuclei and codas [æpt] “apple,” its presence in the onset could have signaled disyllabicity for reasons that are entirely phonetic.

To test this possibility, we examined the effect of the initial sonorant on the sensitivity (*d'*) to sonority falls by means of a 2 sonorant type (*ɫ*- vs. non-*ɫ*-initial) × 4 block ANOVA (in this and all subsequent analyses, we use items as the random factor). The results yielded no effects involving the sonorant factor (all  $F < 1.51$ ,  $P > .23$ )\*.

- The Russian trill. Another artifactual explanation for the results is presented by the possibility that the Russian trill [r], occurring in some of our items, could have been misinterpreted by English speakers as an obstruent-like segment, rather than as a sonorant. Had such misidentifications occurred, then the sonority cline in obstruent-[r] combinations would have been deflated. Accordingly, the advantage of sonority rises over falls should have been attenuated for pairs such as *brif* vs. *rbif*. To test this possibility, we examined the effect of the sonorant status (trills vs. nontrills) on sensitivity to syllables with large sonority rises vs. falls (the two categories where trills were included) using a 2 trill (trill/nontrill) × 2 syllable type × 4 block ANOVA. Results yielded a reliable interaction of trill by syllable type [ $F_2(1,22) = 9.61$ ,  $P < 0.002$ ]. However, contrary to the phonetic explanation above, trills actually resulted in a somewhat higher sensitivity than nontrills (Tukey HSD,  $P < 0.07$ ).

Moreover, the advantage of sonority rises (e.g., *blif*) over falls (e.g., *lbif*) obtained irrespective of whether the syllable included a trill ( $M = 1.30$  vs.  $M = -0.02$ , for large rises vs. falls) or a nontrill sonorant ( $M = 1.02$  vs.  $0.15$ , for large rises vs. falls; Tukey HSD, both  $P < 0.0002$ ).

- Palatalization. A third phonetic concern is that palatalized consonants (e.g., *pn<sup>ɨ</sup>ik*) could have been misidentified as consonant clusters (as [pnjik]). Because English generally bans three-consonantal onsets, these palatalized onsets would be rendered particularly ill formed. Crucially, these effects would occur for reasons that are not directly related to the syllable hierarchy.

However, our past research (9) counters this possibility. In that experiment, English speakers were asked to transcribe the present auditory materials (with small sonority rises, plateaus, and falls; large rises were not included). None of the written responses indicated the insertion of a glide.

We further examined the effect of palatalization in the present experiment by means of a 2 palatalization (palatalized vs. nonpalatalized onsets) × 4 syllable type × 4 block ANOVA. This analysis yielded a reliable main effect of palatalization [ $F_2(1,88) = 9.76$ ,  $P < 0.003$ ], which did not further interact with syllable type ( $F < 1$ ). Contrary to the predictions of the phonetic explanation, palatalized onsets ( $M = 1.12$ ) actually yielded significantly higher sensitivity than their nonpalatalized counterparts ( $M = 0.68$ ; probably because the high tongue position in palatalized consonants is inconsistent with the perception of a schwa, so participants were less likely to entertain the disyllabic misperception).

**The effect of Russian-specific cues on the TMS manipulation.** The previous analyses of the sham condition lend no support for the possibility that the effect of the syllable hierarchy is only due to phonetic artifacts. To ensure that these factors do not taint our TMS manipulation, we next reassessed the effect of TMS while either removing or controlling for these factors. To this end, we removed from the analyses all quartets including (i) palatalized items and (ii) sonority falls that begin with a velar *ɫ* (a total of 9 excluded quartets, resulting in 15 remaining quartets). Because our materials matched onsets with large sonority rises and falls for the occurrence of a trill, we were able to control this cue factorially, by means of a 2 trill (trill/no trill) × 2 TMS × 4 syllable type × 4 block ANOVA.

The TMS × syllable type interaction did not reach significance in this small subset of items [ $F(3,39) = 2.07$ ,  $P < 0.12$ ], nor was it modulated by the trill factor ( $F_1 < 1$ ). Nonetheless, the overall pattern of results was in full agreement with the omnibus analysis (Fig. S3). TMS only attenuated sensitivity to better-formed onsets [for *blif*:  $t(39) = 1.68$ ,  $P < 0.06$  one-tailed; for *bnif*:  $t(39) = 2.26$ ,  $P < 0.03$ ; for *bdif*:  $t < 1$ ; for *lbif*:  $t(39) = 1.29$ ,  $P < 0.21$ , not significant]. Moreover, despite the administration of TMS, *bnif*-type syllables yielded reliably higher sensitivity relative to *bdif* [ $t(39) = 4.97$ ,  $P < 0.0001$ ], which in turn yielded better sensitivity to *lbif* [ $t(39) = 6.38$ ,  $P < 0.0001$ ]. Similar preferences also obtained in the sham condition [for *bnif* vs. *bdif*:  $t(39) = 6.38$ ,  $P < 0.0001$ ; for *bdif* vs. *lbif*:  $t(39) = 8.52$ ,  $P < 0.0001$ ]. Syllables like *blif* did not reliably differ from *bnif* in either the TMS or sham conditions (both  $t < 1$ ).

**The role of stop burst release.** The previous section examined the possibility that the effect of the syllable hierarchy results from the misinterpretation of phonetic cues that are characteristics of Russian phonetics. However, hearers are also known to misanalyze phonetic cues that are familiar from their native language. Past research has shown that hearers sometimes misanalyze nonnative monosyllables that begin with stop consonants (e.g., *bdif*) as disyllabic (e.g., as *bedif*). These errors often occur because the burst release associated with the initial stop (e.g., *b*) is misinterpreted as evidence for a vowel (10–12). The misidentification

\*One might wonder whether the phonetic properties of the velarized-*ɫ* might further account for the effect of syllable type in experiment 2. Note, however, that this phonetic account predicts greater difficulty with sonority rises relative to falls. Our results are clearly inconsistent with this possibility.

of ill-formed syllables in our experiments could thus result from uncontrolled differences in the salience of the burst associated with the initial stop consonant. The following section examined this possibility.

We proceeded in two steps. First, we measured the duration and intensity of the burst release of the three types of stop-initial items in experiment 1 (e.g., *blif* > *bnif* > *bdif*). The fourth type (e.g., *lbif*) was not included, as these items do not begin with a stop. For the same reason, we did not conduct these analyses in experiment 2, as all items begin with a sonorant.

An inspection of the means (Table S1) showed that the salience of the burst did not vary monotonically along the syllable hierarchy. An ANOVA found no significant difference between the three syllable types with respect to burst duration ( $F < 1$ ) or intensity [ $F(2,46) = 2.20, P < 0.13$ , not significant].

We next examined whether the salience of the burst modulated sensitivity ( $d'$ ) in the syllable count task. Specifically, we asked two questions: (i) do the phonetic properties of the burst uniquely modulate performance in the syllable count task; and (ii) does syllable type exert an independent effect even after the properties of the burst are controlled.

We addressed these questions in a series of regression analyses. In these analyses, we entered syllable type and the salience of the burst (its duration and intensity) as predictors in two steps. To examine the unique effect of the burst, we first forced syllable type into the model, whereas the burst salience was entered as the last predictor; to examine the effect of the syllable type, we next reversed the order of the predictors. These analyses were conducted for both the real TMS and the sham conditions.

We found that the burst uniquely captured 6.9% of the variance in the sham condition (Table S2), and its effect was highly

significant. Longer bursts were associated with lower sensitivity [ $d'$ ;  $r = -0.17, t(70) = 1.44, P < 0.07$ , one-tailed], whereas no significant correlation was found with respect to intensity [ $r = 0.12, t(70) = 1$ , not significant]. This result suggests that longer bursts promote the misperception of a schwa, and consequently, some monosyllables are misidentified as disyllables. TMS, however, eliminated the sensitivity to this phonetic cue, as the unique effect of the burst was no longer significant in the TMS condition.

To determine whether the burst can subsume the effect of the syllable hierarchy, we next reversed the order of the predictors. When syllable type was entered as the last predictor, its effect was significant in both the sham and TMS conditions, even after controlling for the phonetic effect of burst (in the first step). It should be noted, however, that the portion of the unique variance captured by syllable type was smaller in the TMS (9.9%) relative to the sham condition (30.5%). Nonetheless, TMS did not eliminate the unique contribution of syllable type. Thus, although syllable type exerted a significant effect irrespective of TMS, the phonetic effect of the burst was entirely eliminated by the TMS manipulation.

Taken as a whole, these results confirm that syllable count is mediated by acoustic/phonetic factors, such as the salience of the burst. This result is only expected, given that the identification of auditory materials inevitably hinges on acoustic and phonetic analyses. Crucially, these low-level effects do not survive the effect of TMS, and they do not account for the effect of the syllable hierarchy. These conclusions are consistent with the possibility that the effect of syllable structure reflects the computation of abstract phonological structure that is inexplicable by low level acoustic/phonetic properties.

- Clements GN (1990) The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, eds Kingston J, Beckman M (Cambridge Univ Press, Cambridge, UK), pp 282–333.
- Greenberg JH (1978) Some generalizations concerning initial and final consonant clusters. *Universals of Human Language*, eds Greenberg JH, Ferguson CA, Moravcsik EA (Stanford Univ Press, Stanford, CA), Vol 2, pp 243–279.
- Berent I, Steriade D, Lennertz T, Vaknin V (2007) What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104(3):591–630.
- Zec D (2007) The syllable. *The Cambridge Handbook of Phonology*, ed de Lacy P (Cambridge Univ Press, Cambridge, UK), pp 161–194.
- Smolensky P (2006) Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, eds Smolensky P, Legendre G (MIT Press, Cambridge, MA), Vol 2 pp 27–160.
- Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29(31):9819–9825.
- Smalle EHM, Rogers J, Möttönen R (2014) Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation [published online ahead of print October 1, 2014]. *Cereb Cortex*, 10.1093/cercor/bhu218.
- Blevins J (2007) Interpreting misperception: Beauty is in the ear of the beholder. *Experimental Approaches to Phonology*, eds Sole M-J, Speeter Beddor P, Ohala M (Oxford Univ Press, Oxford, UK), pp 144–154.
- Berent I, Lennertz T, Rosselli M (2012) Universal phonological restrictions and language-specific repairs: Evidence from Spanish. *The Mental Lexicon* 13(2):275–305.
- Wilson C, Davidson L (2013) Bayesian analysis of non-native cluster production. *NELS 40: Proceedings of the 40th Annual Meeting of the North East Linguistic Society*, eds Seda K, Moore-Cantwell C, Staubs R (Massachusetts Institute of Technology, Cambridge, MA), Vol 2, pp 265–278.
- Wilson C, Davidson L, Martin S (2014) Effects of acoustic-phonetic detail on cross-language speech production. *J Mem Lang* 77(0):1–24.
- Kang Y (2003) Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20(2):219–273.





