



# Is Phonology Embodied? Evidence from Mechanical Stimulation

Iris Berent<sup>1</sup> · Melanie Platt<sup>1</sup>

Accepted: 25 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Across languages, certain syllables are systematically preferred to others (e.g., *plaf* > *ptaf*). Here, we examine whether these preferences arise from motor simulation. In the simulation account, ill-formed syllables (e.g., *ptaf*) are disliked because their motor plans are harder to simulate. Four experiments compared sensitivity to the syllable structure of labial- vs. corona-initial speech stimuli (e.g., *plaf* > *pnaf* > *ptaf* vs. *traf* > *tmaf* > *tpaf*); meanwhile, participants (English vs. Russian speakers) lightly bit on their lips or tongues. Results suggested that the perception of these stimuli was selectively modulated by motor stimulation (e.g., stimulating the tongue differentially affected sensitivity to labial vs. coronal stimuli). Remarkably, stimulation did not affect sensitivity to syllable structure. This dissociation suggests that some (e.g., phonetic) aspects of speech perception are reliant on motor simulation, hence, *embodied*; others (e.g., phonology), however, are possibly abstract. These conclusions speak to the role of embodiment in the language system, and the separation between phonology and phonetics, specifically.

**Keywords** Phonology · Embodiment · Abstraction · Motor simulation · Syllable structure · Sonority · Speech perception

## Introduction

Across languages, certain phonological patterns are systematically preferred to others. For example, syllables like *blog* are more frequent across languages than *lbog* (e.g., Greenberg, 1978) and they are easier for individual speakers to perceive; this is the case for adults (e.g., Berent et al., 2007; Davidson et al., 2015; Moreton, 2002; Pertz & Bever, 1975), children (e.g., Berent et al., 2011; Jarosz, 2017; Maïonchi-Pino et al., 2010, 2012, 2013; Ohala, 1999; Pertz & Bever, 1975) and neonates (Gómez et al., 2014). Why are certain syllable types systematically preferred

---

✉ Iris Berent  
i.berent@northeastern.edu

Melanie Platt  
me.platt@northeastern.edu

<sup>1</sup> Department of Psychology, Northeastern University, Boston, MA, USA

Generative linguistic accounts (e.g., McCarthy & Prince, 1998; Prince & Smolensky, 2004) attribute the sound patterns of language to abstract phonological principles (rules, constraints; hereafter, the *abstraction* hypothesis). But on an alternative *embodiment* explanation, the sound structure of language is determined by articulatory motor action during speech processing—either overt articulation (in production) or its covert simulation (in perception).

In considering this debate, one red herring ought to be removed upfront. By stating that knowledge of language is “abstract” (as opposed to “embodied”), one is *not* espousing some ethereal substance, distinct from the body. On the contrary, the “abstraction” view asserts that knowledge consists of a set of *physical* symbols that form part of the human body/brain (e.g., Fodor & Pylyshyn, 1988; Newell, 1980). Both sides, then, agree that knowledge of language forms part of the human brain. The debate is about *what* linguistic knowledge entails, and, correspondingly, *how* it is encoded in the human brain.

The abstraction view assumes that knowledge of language consists of abstract algebraic rules. The embodiment view, by contrast, denies that rules play a role. The embodiment view, then, does not merely assert that knowledge of language forms part of one’s body—this much is uncontroversial. Rather, it asserts that linguistic knowledge consists of the collection of one’s sensorimotor linguistic experiences (e.g., Barsalou, 2008; Barsalou et al., 2003; MacNeilage, 2008).

Thus, in the abstraction view, a stimulus like *lbog* is ill-formed because it violates the rules that govern syllable structure. *Lbog*, then, would remain ill-formed even if sensorimotor demands would be altered (e.g., by presenting the input in print, or by engaging the articulatory motor system). The “embodiment” alternative denies that such rules exist. In that view, *lbog* is disliked because such sequences are harder to articulate, and covert articulation plays a critical role in both perception and production (e.g., Blevins, 2004; Bybee & McClelland, 2005; MacNeilage, 2008). Phonological restrictions, then, present a test case for the role of embodiment and abstraction in cognition, generally (e.g., Barsalou, 2008; Glenberg et al., 2013; Leshinskaya & Caramazza, 2016; Mahon & Caramazza, 2008; Pulvermüller & Fadiga, 2010).

As we next show, a large body of literature suggests that motor simulation indeed constrains some aspects of speech perception. But whether the motor simulation can further account for phonological preferences is far from evident. The present research explores this question. In so doing, we seek to shed light on the role of abstraction and embodiment in cognition.

## The Role of Embodiment in Speech Perception: Phonetics or Phonology?

There is ample evidence to support the hypothesis that speech perception is embodied. And indeed, the sound patterns of language and its articulation are intimately linked (e.g., Browman & Goldstein, 1989; Gafos, 1999; Goldstein & Fowler, 2003; Hayes et al., 2004). Dispreferred syllables like *lbog* are typically ones that are harder to articulate (e.g., Mattingly, 1981; Wright, 2004) and harder to perceive (e.g., Blevins, 2004; Ohala & Kawasaki-Fukumori, 1990).

Further support for the embodiment position is provided by results suggesting that speech perception triggers articulatory *motor simulation*. One line of evidence is presented by the engagement of the *brain* motor system in speech perception. For example, the perception of labial sounds (e.g., *ba*) activates the lip motor area (e.g., Pulvermüller et al., 2006), and its disruption (by Transcranial Magnetic Stimulation, TMS) impairs the

perception of labial sounds more than coronal sounds (e.g., *da*; e.g., D'Ausilio et al., 2012; D'Ausilio et al., 2009; Fadiga et al., 2002; Möttonen, Rogers, & Watkins, 2014; Möttonen & Watkins, 2009). Likewise, the disruption of the lip motor area impairs the categorization of the *ba-da* continuum, but not non-labial controls (e.g., *ga-ka* Möttonen & Watkins, 2009; Smalle et al., 2014). Given that the brain area activated in perception also mediates lip action, its engagement in speech perception is likely due to its role in motor simulation.

In line with this possibility, a second body of work has shown that phonetic categorization is sensitive to *mechanical* stimulation. In these experiments, participants respond to speech sounds while they selectively engage one of their articulators (e.g., lip vs. tongue) mechanically. Results show that mechanical stimulation selectively modulates the perception of speech sounds.

For example, when six-month old infants restrain the movement of their tongue tip, they lose their sensitivity to nonnative sounds that engage this articulator (e.g., Bruderer et al., 2015). Mechanical stimulation further affects adult participants (for full review, see Berent, Platt, et al., 2020). For example, the identification of speech sounds is modulated by mechanical stimulation (e.g., Berent, Platt, et al., 2020; Ito et al., 2009; Nasir & Ostry, 2009), silent articulation (Sams et al., 2005), mouthing and imagery (e.g., Scott, Yeung, Gick, & Werker, 2013), the application of air puff (suggesting aspiration; Gick & Derrick, 2009), and imagery alone (Scott, 2013). Moreover, these effects of mechanical stimulation obtain even when the articulator in question is utterly irrelevant to response, and even when speech perception is gauged via perceptual sensitivity (*d'*; Berent, Platt, et al., 2020). For example, participants are demonstrably more sensitive to the voicing contrast of labial sounds that are ambiguous with respect to their voicing (e.g., *ba-pa*) when they lightly bite on their lips relative to the tongue; ambiguous coronal sounds (e.g., *da-ta*) show the opposite effect (i.e., sensitivity to voicing improves when participants bite on the tongue relative to the lips; Berent, Platt, et al., 2020). These results thus suggest that motor simulation affects perception, not merely response.

Nonetheless, the support for the embodiment view is limited by two considerations. First, most of the evidence for motor simulation obtains from studies of *phonetic* categorization of ambiguous speech sounds (e.g., voicing contrast /b/ and /p/). While these results are consistent with the possibility that *some* aspects of speech perception engage the motor system, they leave open the question of whether the contribution of motor simulation extends to the *phonological* level, including the computation of syllable structure.

A strong embodiment view asserts that it does: all aspects of speech perception are embodied, so there is indeed little reason to draw a distinction between phonetics and phonology. Generative linguistics, by contrast, typically assumes that speech is encoded at multiple levels of analysis, and that these levels could differ on their computational properties. The phonetic system extracts discrete features from the continuous speech stream, and phonetic representations are considered continuous and analog (e.g., Chomsky & Halle, 1968; Keating, 1988). Phonology, by contrast, is believed to rely on abstract, possibly, algebraic computations that are putatively disembodied (e.g., Berent, 2013; Chomsky & Halle, 1968; de Lacy, 2006; Prince & Smolensky, 1993/2004; Smolensky & Legendre, 2006).

In line with this possibility, linguists have long argued that phonological preferences can dissociate from phonetic pressures, inasmuch as language users sometimes prefer phonological structures that are phonetically suboptimal, presumably, because these structures satisfy abstract phonological principles (e.g., de Lacy, 2004, 2008; Hayes et al., 2004). Other arguments for abstraction are presented by the convergence of phonological principles across sensory modality—spoken and manual (e.g., Andan et al., 2018; Brentari,

2011; Sandler & Lillo-Martin, 2006). And indeed, speakers who are ASL-naive demonstrably project grammatical principles from their spoken language to ASL signs (e.g., Berent et al., 2016, 2021; Berent, Bat-El, et al., 2020). Such results suggest that some phonological principles are independent of language modality.

Taken together, these results support the possibility that phonetics and phonology are distinct. So even if phonetic categorization is in fact embodied, and reliant on motor simulation, the question remains whether these conclusions extend to *phonology*—a component that, at least by some accounts, relies, in part, on combinatorial principles that are disembodied and abstract.

To address this question, here, we examine the role of the motor system in the computation of syllable structure. To explain our manipulation, we briefly describe the restrictions on syllable structure. We next review past research on the role of motor simulation in this domain; these results lay the foundation for the present investigation.

## The Syllable Hierarchy

As noted, certain syllable types are systematically preferred across languages (e.g., Clements, 1990; Greenberg, 1978; Prince & Smolensky, 1993/2004; Smolensky, 2006; Steriade, 1988). For example, *plaf* is preferred to *pnaf*, which in turn, is preferred to *ptaf*; least preferred are syllables like *lpaf*, and together, these preferences suggest the following syllable hierarchy: *plaf* > *pnaf* > *ptaf* > *lpaf* (where > indicates a preference).

These preferences have been attributed to the phonological construct of sonority. Sonority is a scalar phonological property that correlates with the salience of phonological elements (e.g., Clements, 1990). 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).

To capture syllable structure, one can calculate the sonority cline of its onset by subtracting the sonority level of the first consonant from the sonority level of the second consonant ( $\Delta s$ ). As the sonority cline decreases, the syllable becomes progressively worse-formed, hence, dispreferred. For example, syllables like *pla* (with a large rise in sonority between the stop and the liquid,  $\Delta s=2$ ) are preferred to *pna* (small rise,  $\Delta s=1$ ), which, in turn, are preferred to the sonority plateau in *pta* ( $\Delta s=0$ ); least preferred, the sonority fall in *lpa* ( $\Delta s=-2$ ). Indeed, as the sonority cline decreases, the syllable becomes underrepresented across languages (e.g., Berent et al., 2007; Greenberg, 1978).

Past research has shown that the syllable hierarchy captures the behavior of individual speakers: the less preferred the syllable, the harder it is for speakers to accurately encode (in adults: e.g., Berent et al., 2007; Pertz & Bever, 1975; in children: Berent et al., 2011; Jarosz, 2017; Maionchi-Pino et al., 2012; Maionchi-Pino et al., 2010; Maionchi-Pino et al., 2013; Ohala, 1999; Pertz & Bever, 1975). For example, when asked to discriminate spoken monosyllables from their disyllabic counterparts (e.g., is *ptaf*=*petaf*?), speakers exhibit lower sensitivity (*d'*) and slower response times to ill-formed syllables (e.g., *ptaf*) compared to better-formed syllables (e.g., *pnaf*, *plaf*). Likewise, ill-formed syllables evoke stronger hemodynamic response in the brains of adults (on functional MRI, Berent et al., 2014) and neonates (using Near Infrared Spectroscopy, Gómez et al., 2014).

The finding that sensitivity to the syllable hierarchy obtains in neonates opens up the possibility that this knowledge does not arise only from the similarity of novel syllables to existing words (e.g., *pnaf* is preferred because it is analogized to *sniff*, Daland et al., 2011). Indeed, people exhibit these preferences even when most or all of these syllable

types do not occur in their languages—for adult speakers of English (e.g., Berent et al., 2009; Berent et al., 2007; Tamasi & Berent, 2015); French (e.g., Maïonchi-Pino, 2012); Hebrew (e.g., Bat-El, 2012); Polish (e.g., Jarosz, 2017); and Spanish (e.g., Berent et al., 2012), and even when their language arguably lacks onset clusters altogether (for Mandarin: Zhao & Berent, 2016; for Korean: Berent et al., 2008). And since sensitivity to the syllable hierarchy obtains even for printed words (e.g., Berent & Lennertz, 2010; Berent et al., 2009; Tamasi & Berent, 2015), there is reason to expect that the difficulty with ill-formed syllables does not arise solely for auditory/phonetic reasons (e.g., because English speakers cannot faithfully encode the phonetic form of *ptaf* from speech).

Whether speakers' sensitivity to the syllable hierarchy indeed arises by learning from experience, and whether it reflects auditory difficulties are important open questions *that we do not seek to settle here*. In the present research, our goal is strictly to explore whether sensitivity to the syllable hierarchy results from **motor simulation**. In this view, upon hearing *ptaf*, listeners (covertly) simulate its articulation by engaging their motor system, and since *ptaf* is presumably harder to articulate, such syllables are systematically misidentified. Our research examines this proposal; the role of experience and auditory pressures will not be examined here.

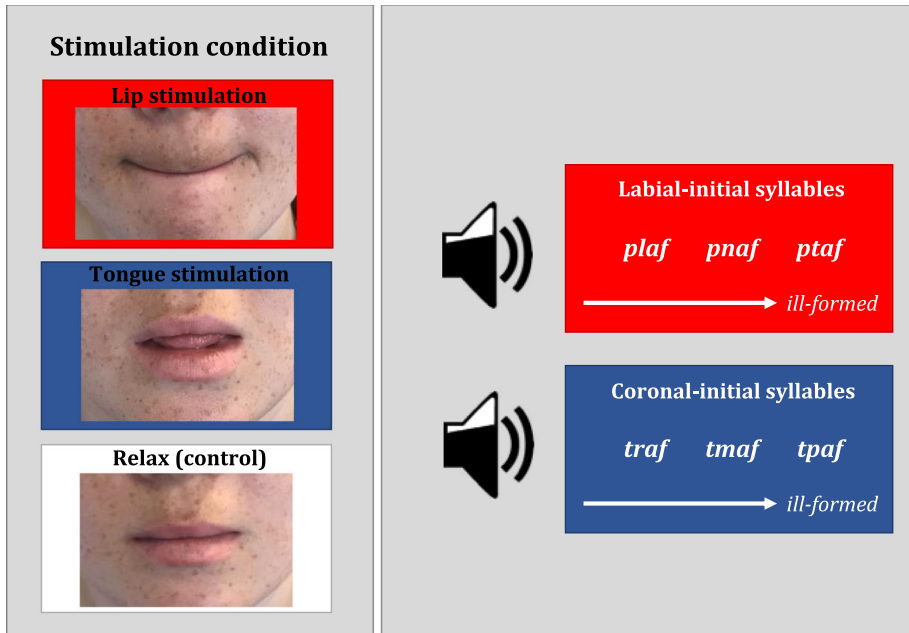
### Is Sensitivity to the Syllable Hierarchy Due to Motor Simulation?

Past research has evaluated whether motor simulation mediates sensitivity to syllable structure using both mechanical and transcranial stimulation. If sensitivity to syllable structure arises from motor simulation, then disrupting this process should attenuate sensitivity to the syllable hierarchy. This is because, in this embodiment account, ill-formed syllables like *ptaf* are hardest to perceive because they are hardest to simulate, and simulation, recall, is critical for speech perception. It thus follows that, if simulation is disrupted—either mechanically, or transcranially (by TMS)—then, paradoxically, the identification of ill-formed syllables should *improve*, and consequently, sensitivity to syllable structure (evidence by the slope of the identification function) should decline.

Past results do not support these predictions. When simulation was blocked mechanically, sensitivity to syllable structure was utterly unaffected (Zhao & Berent, 2018). Similarly, participants remained sensitive to the syllable hierarchy even when their brain motor system was stimulated by TMS (Berent et al., 2015). These results could suggest that the phonological computation of syllable structure is abstract, and thus, distinct from phonetic computations, where the effect of motor simulation have been amply demonstrated.

It is possible, however, that the lack of evidence for motor simulation only emerged because the manipulations we had used were incapable of detecting such effects. One concern is that these past manipulations were **non-selective**. Selectivity, is, indeed, the hallmark of motor simulation. Simulation is inferred when the perception of a given speech sound (e.g., labials) depends on the target of stimulation (e.g., lip vs. tongue). Demonstrating such effects thus requires that (a) The stimulation selectively targets distinct articulators; and (b) The speech stimuli under investigation contrast with respect to their reliance on those same articulators.

This, however, was not necessary the case in the past stimulation studies. The mechanical stimulation experiments simultaneously engaged both the lip and the tongue (Zhao & Berent, 2018). Similarly, past attempts to stimulate the lip motor area by TMS could have also stimulated the adjacent area controlling the tongue (Berent et al., 2015). Finally, the stimuli that were used in both studies did not systematically differ with respect to their



**Fig. 1** Experimental design

engagement of distinct articulators. Because these manipulations were non-selective, it is unclear whether they were capable of accurately gauging the role of motor simulation in the first place.

A second concern is that our results from English speakers might not have adequately reflected the true role of motor simulation because the syllable structures that were featured in these experiments were mostly unfamiliar to participants. The question thus arises whether the evidence for motor stimulation was stronger for speakers whose languages allow those structures. The present research addresses these two concerns.

## The Present Research

The following experiments systematically compare the effect of mechanical stimulation of two *articulators*—the lip and the tongue—on sensitivity to the syllable structure of two matched sets stimuli that systematically differ on their *place of articulation*—labial and coronal. Each such set consisted of three types of syllables, arrayed in terms of their well-formedness on the syllable hierarchy—large sonority rises ( $\Delta s = 2$ ), small rise ( $\Delta s = 1$ ), and sonority plateau ( $\Delta s = 0$ ). These syllables, in turn, further contrasted on their initial place of articulation. Labial-initial syllables consisted of items such as *plaf*, *pnaf*, and *ptaf*; the matched coronal counterparts were *traf*, *tmaf*, and *tpaf*.

Participants were asked to identify these labial- and coronal-initial stimuli while pressing lightly on either the lips or the tongue; a control condition instructed participants to simply relax (Fig. 1). This design allowed us to evaluate two distinct questions (a) Does the identification of these stimuli rely on motor simulation; and (b) Does motor simulation modulate sensitivity to the syllable hierarchy.

Consider, first, the role of motor simulation. Past research has shown that, when this same method of stimulation applies to phonetic categorization, it can produce the expected selective effects: biting on the lip had stronger effect on labials; biting on the tongue had stronger effect on coronals (Berent, Platt, et al., 2020). Given that (a) Our experimental materials feature contrasting speech sounds (i.e., labials vs. coronals), that (b) Speech perception demonstrably relies on motor simulation, and that (c) our method of choice can effectively detect its effect (Berent, Platt, et al., 2020), we expect to detect evidence for motor simulation in the present experiments. If participants rely on motor simulation, then their performance should be modulated by the congruence between the place of articulation of the stimulus (labial vs. coronal initial) and the stimulated articulator (lip vs. tongue). Statistically speaking, motor simulation should give rise to a *place* × *stimulation* interaction.

The critical question is whether motor stimulation would further affect sensitivity to syllable structure. In past research, sensitivity to syllable structure was gauged by examining its effect on perception: as the syllable becomes worse-formed, its identification typically declines. So if this sensitivity arises from motor simulation, then, as explained above, disrupting motor simulation should attenuate sensitivity to the syllable hierarchy. Specifically, we expect the slope of the effect of syllable type to decline when participants stimulate the congruent articulator (e.g., lip, for labial sounds) compared an incongruent articulator (e.g., lip, for coronal sounds). Statistically speaking, this effect should give rise to a three-way (stimulation × place × syllable type) interaction.

The embodiment account predicts that the syllable hierarchy arises from motor simulation, so stimulating the relevant articulators should selectively attenuate sensitivity to the syllable hierarchy. The abstraction hypothesis asserts that it won't. Note, however, that the abstraction hypothesis does not deny that motor simulation mediates certain components of speech perception, such as phonetic categorization. Accordingly, finding that the perception of spoken auditory stimuli relies on motor simulation would not falsify this hypothesis. Rather, what the abstraction hypothesis denies is that embodiment mediates the grammatical computation of language structure, such as syllable structure. The critical question, then, is whether motor stimulation plays a causal role in promoting the sensitivity to syllable structure; this should be evident by selective changes in the slope of the syllable hierarchy (i.e., as a stimulation × place × syllable type interaction). It is here where the predictions hypotheses contrast.

Experiments 1 & 3 evaluate these questions for *English* speakers, using two distinct procedures: Experiment 1 employs a syllable count task (e.g., *does plaf include one syllable or two?*); Experiment 3 examined this question using AX discrimination (*does plaf = pelaf?*). Since most of these syllable types are unattested in English, these experiments would allow us to evaluate whether motor simulation mediates sensitivity to novel linguistic stimuli, for which participants have had little articulatory experience.

It is possible, however, that the lack of experience of English speakers might minimize their engagement in motor articulation. To address this possibility, Experiments 2 & 4 apply the same manipulations on speakers of *Russian*—a language where these syllable types are attested. As such, our experiments not only allow us to evaluate the role of motor simulation in phonology but also examine whether it is modulated by linguistic experience.

## Experiment 1: English Speakers (Syllable Count)

Experiment 1 employed a syllable count task. Each trial featured a single stimulus—either a monosyllable (e.g., *plaf*) or a disyllable (e.g., *pelaf*). Participants were asked to indicate (via a button press) whether the stimulus included one syllable or two. Meanwhile, they were asked to bite lightly on either their lips, their tongue, or relax. These stimulation conditions were each administered across three distinct counterbalanced blocks, and each such block featured both labial and coronal initial stimuli.

Our experiment examined three questions (a) Are people sensitive to syllable structure; (b) Do people rely on motor simulation; and (c) Does motor simulation modulate sensitivity to syllable structure.

## Methods

### Participants

Participants were forty-eight native English speakers, students at Northeastern University. One additional participant was tested, but was not included in the analysis due to missing data resulting from a technical issue. Participants took part in the experiment either in partial fulfillment of course credit or remuneration (\$10). The procedures used in this and all subsequent experiments were approved by the Institutional Review Board at Northeastern University; all participants signed an informed consent form.

Sample size in this and all subsequent experiments was determined based on the previous experiments of Zhao and Berent (2018) and Berent, Platt, et al. (2020).

### Materials

The stimuli consisted of 72 pairs of auditory monosyllabic nonwords and their disyllabic counterparts (e.g., *plaf* and *pelaf*). Monosyllables had a CCVC structure (C = consonant, V = vowel). Disyllables contrasted with their monosyllabic counterparts on the pretonic schwa (e.g., *plaf* vs. *pelaf*).

The monosyllables consisted of three matched types arrayed according to the sonority profile of their onsets—either a large sonority rise, a small sonority rise, or a sonority plateau. Half of the triplets began with an initial labial stop consonant followed by a coronal (*plaf* > *pnaf* > *ptaf*) and the other half began with a coronal (alveolar) stop consonant followed by a labial (*traf* > *tmaf* > *tpaf*). The experiment featured a total of 48 such triplets (24 monosyllables and their 24 disyllabic counterparts), matched for their rimes.

All monosyllables (and their disyllabic counterparts) further began with a voiceless stop; syllables beginning with voiced stops were avoided because pilot work indicated that English participants confused these items with disyllables even when the monosyllable had an onset that was attested in English. We note that the onsets featured in this experiment (*pl/pn/pt* vs. *tr/tm/tp*) exhaust all possible combinations in English, as defined above. Altogether there were 144 auditory nonwords (2 place of articulation: labial vs. coronal) × 2 syllable count (monosyllabic and disyllabic) × 3 onset type (large sonority rise, small sonority rise, sonority plateau) × 12 items). These materials are provided in the Appendix.

The experiment also featured another set of 24 monosyllabic fillers with sonority falls and their disyllabic counterparts (a total of 48 items). These items were originally designed



to test the full sonority hierarchy. A phonetic inspection of these items, however, revealed phonetic cues that rendered these monosyllables indistinguishable from disyllables, and indeed, when these items were presented to Russian participants (in Experiment 2), most participants identified them as disyllables (for the proportion of monosyllabic response:  $M=0.48$ ,  $SD=0.34$ ; for the other three types:  $M=0.91$ ,  $SD=0.18$ ). We thus excluded these items from all analyses.

A female native Russian speaker recorded the stimuli. We chose a Russian speaker because the Russian language includes complex onsets of each of the four types used in our experiments, so all items could be produced naturally. The speaker read the words from printed lists presented in Cyrillic. Each word was in a phrasal context (e.g., X raz, “X *once*”) and were arranged in pairs of monosyllables and disyllables (e.g., *plaf* and *pelaf*), counterbalanced for order. The speaker was instructed to adopt flat intonation but otherwise read the words naturally.

The recordings were made using a Mac Pro (Apple, 2010) with OS X version 10.9.5 using PRAAT 6.0.05 (Boersma & Weenink, 2015) and an external AKG model C 460 B microphone with a foam cover (AKG Acoustics, Inc.) and prepared for the experiment using the same equipment and the RMS-Boersma script for PRAAT, version 6.0.05 (Boersma & Weenink, 2018). The recordings were next standardized so that all items were equated for their root-mean-square decibel level as well as the duration of silence preceding each item.

## Procedure

Participants initiated each trial by pressing the space bar; their response triggered the presentation of a single auditory stimulus—either a monosyllable or disyllable. Participants were instructed to indicate whether the stimulus had one syllable or two via a key press (1 = monosyllable; 2 = disyllable). Participants were encouraged to respond as quickly as possible. Slow responses (responses slower than 2500 ms) triggered a computer feedback message (“too slow!”). Prior to the experiment, participants were given a brief practice with English words (e.g., *please* vs. *police*).

Participants performed the syllable count under three stimulation conditions. In the lip stimulation condition, participants gently squeezed their lips between their teeth; in the tongue condition, they gently held their tongue between their teeth; in the control condition, participants held their mouths in a normal position. Participants were instructed to apply the stimulation throughout the block of trials, and maintain the non-stimulated articulator free to move. The performance of the stimulation was monitored by an experimenter, aided by mirrors, installed to reflect participants’ faces. Each such condition featured all 192 items (144 experimental items and 48 fillers).

Each stimulation condition was presented in a separate block (counterbalanced for order), featuring all stimuli (144 experimental items and 48 fillers). At the beginning of each such stimulation block, the computer presented participants with instructions on how to perform the stimulation procedure, along with a picture illustrating the stimulation; for reminder, a picture card was placed at the computer monitor. To combat fatigue, each block was divided into three parts (counterbalanced for Syllable  $\times$  Type  $\times$  Place), and participants were given 30 s breaks between those parts. Within each part, trial order was random.

Participants were tested in small groups of up to four participants at a time. Within such group, all participants were assigned to the same stimulation condition order.

The experiment was conducted using E-Prime 2.0 software (Psychology Software Tools, Inc., 2008) running on a Dell OptiPlex 775 (Dell, Inc., 2008) with Windows 7.0 Professional (Microsoft Corp., 2009).

## Results

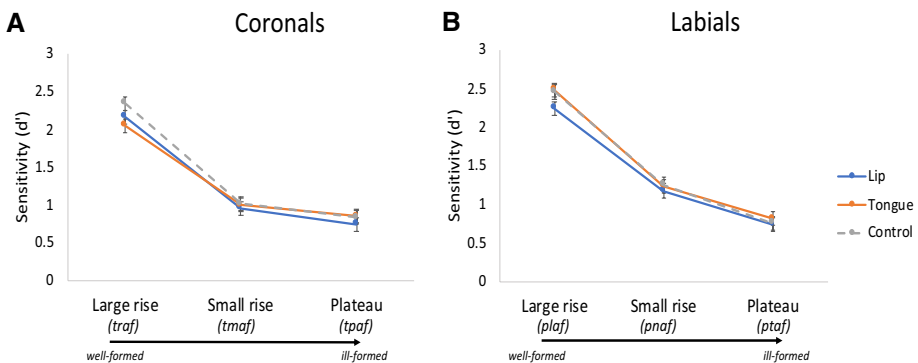
The analyses below examine the effect of syllable type, place of articulation, and stimulation on perceptual sensitivity ( $d'$ ). In light of the low sensitivity scores, response times were not evaluated. Since the set of items used in this experiment represents all possible instances of the structures of interest (i.e., English allows for no other labial-coronal combinations beginning with voiceless stops that satisfy the constraints defined in Methods), we conducted all analyses across participants only (for the data for this and all exp.'s: [https://osf.io/g39ta/?view\\_only%20=%204666d09c477d497eb6ff68a8eabdfa7f](https://osf.io/g39ta/?view_only%20=%204666d09c477d497eb6ff68a8eabdfa7f)).

These analyses address three questions (a) Are people sensitive to the syllable hierarchy (i.e., the main effect of syllable Type)? (b) Do people engage in motor simulation (i.e., the interaction of Stimulation with Place of articulation); and (c) Does motor simulation modulate sensitivity to the sonority hierarchy (i.e., the Simulation  $\times$  Place  $\times$  Type interaction)?

For simplicity, in this and all subsequent experiments, we focus on the statistical results pertinent to the three questions (a–c) above; all other significant results (e.g., the main effect of stimulation) are discussed in the SM. We first address these questions across all stimulation blocks of trials; to counter the effects of practice and fatigue, we next examine the first block of trials separately (i.e., the first stimulation condition presented to any given participant).

### All Blocks

Figure 2 plots the effect of syllable type on sensitivity to the syllable hierarchy; in this and all subsequent figures, error bars are 95% CI for the difference between the means.



**Fig. 2** Sensitivity to the syllable hierarchy for labial and coronal stimuli under motor stimulation

An inspection of the means suggests that, as the syllable became worse formed, perceptual sensitivity ( $d'$ ) declined. However, when considering performance across blocks there was no evidence that the effect of stimulation differed depending on the target's place of articulation.

A 3 Stimulation  $\times$  3 syllable Type  $\times$  2 Place of articulation repeated-measures ANOVA yielded a significant effect of Syllable Type ( $F(2,94)=129.83$ ,  $p<0.0001$ ,  $\eta_p^2=0.73$ ). Planned contrasts showed that the best formed syllables with large sonority rises (e.g., *plaf*) produced better sensitivity than the intermediate ones with smaller sonority rises (e.g., *pnaf*:  $t(94)=6.29$ ,  $p<0.0001$ ), which in turn, produced better sensitivity than the worse formed ones, with sonority plateaus (e.g., *ptaf*:  $t(94)=3.21$ ,  $p=0.002$ ). These results confirmed that participants were sensitive to syllable structure.

Was this sensitivity caused by motor simulation? If participants engaged in motor simulation, then we should have expected to find a reliable interaction of Stimulation  $\times$  Place. But an inspection of the means (Fig. 2) shows no evidence for such interaction, and this impression is confirmed by the ANOVA: the Stimulation  $\times$  Place interaction was not significant ( $F(2,94)=1.63$ ,  $p=0.20$ ,  $\eta_p^2=0.03$ ). Likewise, there was no evidence for the three-way interaction (Stimulation  $\times$  Place  $\times$  Type:  $F(4,188)=1.73$ ,  $p=0.14$ ,  $\eta_p^2=0.04$ ).

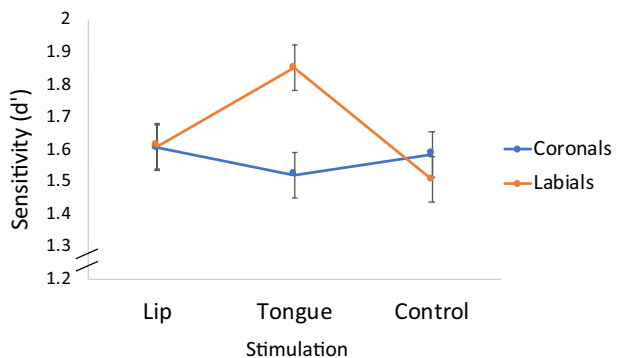
### First Block Only

Across blocks, we found no evidence for motor simulation. It is possible, however, that this null effect is due to fatigue or practice, which could have decreased the engagement in stimulation. To address this possibility, we further examined the same effects within the first block of trials separately. Here, Stimulation was manipulated between participants; Place of articulation and Type were manipulated within participants.

The ANOVA now yielded a Stimulation  $\times$  Place interaction ( $F(2,45)=4.29$ ,  $p=0.02$ ,  $\eta_p^2=0.16$ ). To explore this interaction further (see Fig. 3), we next compared the effect of lip and tongue stimulation directly (excluding the control condition). The Stimulation  $\times$  Place interaction remained significant ( $F(1,30)=4.46$ ,  $p=0.04$ ,  $\eta_p^2=0.13$ ).

Planned contrasts showed that pressing on the tongue increased sensitivity to labial- relative to coronal stimuli ( $t(38)=3.01$ ,  $p<0.004$ ); these two stimuli, however, did not differ when participants pressed on the lips ( $t<1$ ). Thus, perceptual sensitivity to labial stimuli selectively increased by stimulating the incongruent articulator (the tongue). This, however, was not the case for coronal stimuli (i.e., sensitivity to coronals did not improve by

**Fig. 3** The effect of motor stimulation on response to labials and coronals



stimulating the lips). As such, these results offer partial evidence that participants engaged in motor simulation.

Critically, motor simulation did not reliably modulate the effect of syllable Type. The ANOVA of the first block of trials showed no evidence for the three-way interaction, and this was the case regardless of whether stimulation was examined across all three conditions (lips, tongue, control;  $F < 1$ ) or only for the experimental conditions (lips and tongue;  $F < 1$ ).

In contrast, the main effect of Type was highly significant already in the first block ( $F(2,90)=127.01$ ,  $p < 0.0001$ ,  $\eta_p^2=0.74$ ). Planned contrasts confirmed that large sonority rises ( $M=2.47$ ) produced better sensitivity than small rises ( $M=1.33$ ,  $t(90)=6.32$ ,  $p < 0.0001$ ), which, in turn, resulted in better sensitivity than plateaus ( $M=1.05$ ,  $t(90)=2.95$ ,  $p=0.004$ ).

## Discussion

Experiment 1 asked whether sensitivity to the syllable hierarchy is due to motor simulation. To address this question, we first explored whether participants are sensitive to syllable structure. The results were loud and clear. The effect of syllable structure was robust: as syllable structure became worse-formed, perceptual sensitivity monotonically declined.

To determine whether participants engaged in motor simulation, we next compared the identification of labial- vs. coronal-initial syllables (e.g., *plaf* vs. *traf*) when participants pressed on either their lips or tongues. If participants engaged in motor simulation, then the engagement of a motor articulator (e.g., the lips) should differentially affect the perception of congruent speech sounds (e.g., labials) compared to incongruent ones (e.g., coronals). The results indeed showed evidence for simulation in the first block of trials, as the stimulation of the tongue yielded better perceptual sensitivity for labial-relative to coronal sounds. While these results were limited, inasmuch as they did not obtain across all blocks of trials, and they were only found for coronals, the findings do suggest that our experimental procedure can detect the effect of motor simulation.

Having shown participants engage in motor simulation, and that they are sensitive to syllable structure, the remaining crucial question is whether sensitivity to syllable structure arises from (failed) motor simulation. If it does, then sensitivity to syllable structure should be modulated by both the stimulation condition and stimulus place of articulation. Our results found no such interactions.

These results seem to support two conclusions. First, sensitivity to syllable structure is governed by abstract linguistic principles. Second, motor simulation might not play a major role in the computation of syllable structure. Each of these conclusions, however, is subject to some significant caveats.

First, the results from English speakers alone are insufficient to explicate the effect of syllable structure. Indeed, it is possible that this effect emerged not from linguistic knowledge, but from stimuli artifacts: perhaps the Russian talker who produced these stimuli articulated ill-formed monosyllables (e.g., *ptaf*) such that they are in fact phonetically indistinguishable from disyllables (e.g., *ptaf=petaf*). Accordingly, the “misidentification” of these items arose from the stimuli themselves, not from linguistic knowledge. Second, perhaps English speakers did not fully rely on motor simulation because of their limited familiarity with such syllable types. The results from English speakers could thus underestimate the true contribution of motor simulation in phonology.

To address these two concerns, Experiment 2 administered the same procedure to speakers of Russian—a language in which all syllable types are attested. If English stimuli misidentified ill-formed syllables because these stimuli are tainted, then Russian speakers should do the same. But if this misidentification is partly due to linguistic experience, then the performance of Russian speakers should far surpass their English counterparts.

The comparison with Russian speakers can further allow us to evaluate whether the contribution of motor simulation depends on linguistic experience. If the limited role of motor simulation in Experiment 1 is due to the unfamiliarity, then Russian speakers should show far stronger effects of motor simulation in Experiment 2.

## Experiment 2: Russian Speakers (Syllable Count)

### Methods

Thirty-six native Russian speakers took part in this experiment. To be considered a native Russian speaker, participants needed to have learned Russian before age 5 and not have learned any other languages before age 12. Participants were between the ages of 18–40 years.

The materials and methods were as in Experiment 1, except that the instructions were provided to participants in Russian in written format. The computer instructions were also provided in Russian. Translations from English to Russian were made by the same female native Russian speaker who recorded the stimuli. Participants were tested in the lab individually and monitored by a research assistant. Participants received either credit that partially fulfilled course requirements or remuneration (\$10).

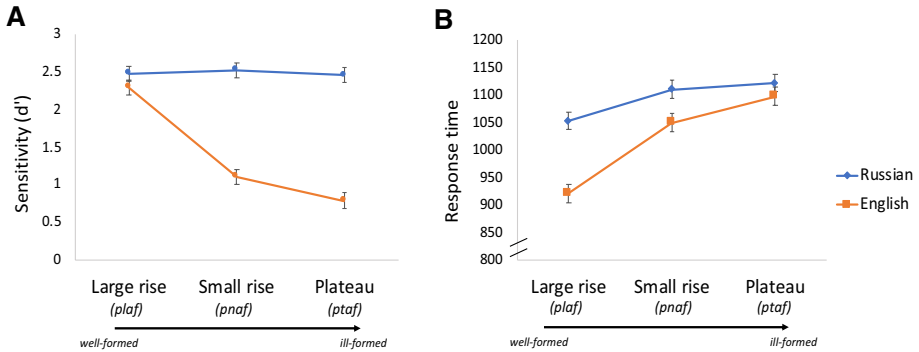
### Results

Since the Russian speakers showed far better sensitivity than English participants, in Experiment 2, we analyzed the results over both sensitivity ( $d'$ ) and response time to monosyllables. As in Experiment 1, we first analyzed the results across blocks; next we inspected performance in the first block of trials.

### All Blocks

Figure 4 plots the sensitivity ( $d'$ ) of Russian participants to syllable structure; for comparison, we plot their results along with the English participants in Experiment 1. An inspection of the means suggests that Russian speakers showed much better sensitivity, which was unaffected by syllable structure. However, for Russian participants, response time was fastest for the best-formed syllables with large sonority rise (e.g., *plaf*).

In the ANOVA (3 Stimulation  $\times$  2 Place of articulation  $\times$  3 syllable Type) on sensitivity ( $d'$ ), the main effect of syllable Type was not significant ( $F < 1$ ). The ANOVA of response time, however, did yield a reliable main effect of syllable Type ( $F(2,70) = 12.95$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.27$ ). Planned contrasts showed that the best formed syllables with large sonority rise produced faster responses than small rises ( $t(70) = 3.94$ ,  $p < 0.001$ ), which, in turn, did not differ from plateaus ( $t < 1$ ).



**Fig. 4** The effect of the syllable hierarchy on sensitivity ( $d'$ ) and response time for Russian and English speakers

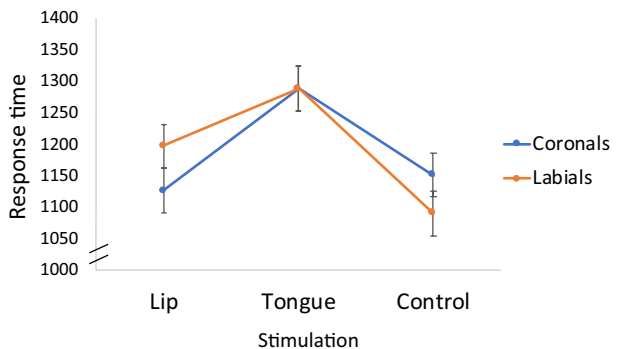
We next examined whether Russian participants engaged in motor simulation. If they did, then the effect of stimulation should interact with the stimulus place of articulation. And if motor simulation was further at the root of the effect of syllable structure, then a three-way interaction (Stimulation  $\times$  Place  $\times$  syllable Type) is expected. These interactions, however, were not found. There was no hint of a Stimulation  $\times$  Place interaction ( $d'$ :  $F < 1$ ; in RT:  $F(2,70) = 1.29$ ,  $p = 0.28$ ,  $\eta_p^2 = 0.04$ ). There was also no evidence for a three-way interaction (Stimulation  $\times$  Place  $\times$  syllable Type) in either the analysis of  $d'$  or response time (for  $d'$ :  $F < 1$ ; for response time:  $F(4,104) = 1.07$ ,  $p = 0.37$ ,  $\eta_p^2 = 0.03$ ).

### First Block Only

To address the possibility that motor simulation might be masked by fatigue/practice, we also applied these analyses to the first block of trials only.

We found a reliable Stimulation  $\times$  Place interaction in response time ( $F(2,33) = 3.53$ ,  $p = 0.04$ ,  $\eta_p^2 = 0.18$ ; in  $d'$ :  $F < 1$ ). Planned contrasts showed that, when participants pressed on their lips, they responded faster to coronals relative to labials ( $t(41) = 2.00$ ,  $p = 0.05$ ), but when participants pressed on their tongue, the advantage of coronals was eliminated ( $t < 1$ ; see Fig. 5).

**Fig. 5** The effect of motor stimulation on response time to monosyllables



The ANOVA of the first block also yielded a reliable effect of syllable Type in the analysis of response time ( $F(2,66)=3.61$ ,  $p=0.03$ ,  $\eta_p^2=0.10$ ; for  $d'$ :  $F < 1$ ). Planned contrasts showed that syllables with large rises produced faster responses than small rises ( $t(66)=2.15$ ,  $p=0.04$ ), which in turn, did not differ from plateaus ( $t < 1$ ). The three-way interaction was not significant (in  $d'$ :  $F(4,66)=1.01$ ,  $p=0.40$ ,  $\eta_p^2=0.06$ ; in RT:  $F(4,66)=1.96$ ,  $p=0.11$ ,  $\eta_p^2=0.11$ ).

## A Comparison with English Speakers

To determine whether the responses of Russian speakers differed from English participants, a final set of analyses compared the sensitivity ( $d'$ ) of the two groups via a 2 Language  $\times$  3 Stimulation  $\times$  2 Place of articulation  $\times$  3 syllable Type ANOVA. To explore the effect of language, here we report only the main effect of Language and its interaction with syllable Type and Stimulation (i.e., Language  $\times$  Stimulation  $\times$  Place; Language  $\times$  Stimulation  $\times$  Place  $\times$  Type). All other interactions with the language factor are described in the SM.

Only two such effects were reliable. There was a main effect of Language, as Russian speakers showed higher sensitivity ( $F(1,82)=91.63$ ,  $p < 0.0001$ ,  $\eta_p^2=0.53$ ). There was also a Language  $\times$  Type interaction ( $F(2,164)=74.42$ ,  $p < 0.0001$ ,  $\eta_p^2=0.48$ ).

We found no evidence that these differences were further modulated by motor stimulation, as the high-level interaction (Language  $\times$  Stimulation  $\times$  Place  $\times$  Type) was not significant ( $F(4,328)=1.56$ ,  $p=0.18$ ,  $\eta_p^2=0.02$ ). There was likewise no evidence that language modulated the effect of stimulation (for the Language  $\times$  Stimulation  $\times$  Place  $F(2,164)=1.40$ ,  $p=0.25$ ,  $\eta_p^2=0.02$ ).

Similar conclusions emerged in the analysis of the first block of trials only. The first block of trials yielded a reliable effect of Language ( $F(1,78)=36.25$ ,  $p < 0.0001$ ,  $\eta_p^2=0.32$ ) as well as a reliable Language  $\times$  Type interaction ( $F(2,156)=53.30$ ,  $p < 0.0001$ ,  $\eta_p^2=0.41$ ). The high-level interaction and Language  $\times$  Stimulation  $\times$  Place interaction were not significant ( $F < 1$ ;  $F(2,78)=2.06$ ,  $p=0.13$ ,  $\eta_p^2=0.05$ , respectively). As noted above, syllable structure modulated the sensitivity of English, but not Russian speakers.

## Discussion

Experiment 2 sought to shed light on the results of English speakers in Experiment 1. One concern was that the effect of syllable structure with these participants arose only from stimuli artifacts—the possibility that our monosyllabic stimuli were phonetically indistinguishable from disyllabic stimuli. The second concern was that the rather limited role of motor simulation with English participants was due to their lack of linguistic experience with these syllables.

The results do not support these possibilities. First, when the Russian speakers responded to the same stimuli, they showed far better perceptual sensitivity that was unaffected by syllable structure. This finding suggests that the effect of syllable structure in Experiment 1 is partly due to linguistic experience, rather than to stimuli artifacts alone. Second, despite their familiarity with syllable types, Russian speakers showed evidence for simulation only in the first block of trials, in line with the results for their English-speaking counterparts. In Part 2, we seek to extend these results to a second experimental procedure—AX discrimination.

### Experiment 3: English Speakers (AX Discrimination)

Experiment 3–4 extend the investigation of the motor system to a second procedure—AX discrimination. Participants were presented with the same set of stimuli as in Part 1. Here, however, each trial featured two successive stimuli, separated by 800 ms. Half of the trials presented identical (i.e., the same speech token, repeated), either monosyllables (e.g., *plaf-plaf*) or disyllables (e.g., *pelaf-pelaf*). The other half of the trials paired monosyllables with their disyllabic counterparts (e.g., *plaf-pelaf* or *pelaf-plaf*; counterbalanced for order). Participants were required to indicate whether or not the stimuli were identical. Critically, participants performed the task while they lightly bit on either their lips, their tongue, or relaxed (a control condition). And the stimuli, as noted, varied systematically in terms of their place of articulation and syllable structure.

The AX manipulation differs from the syllable count (in Experiments 1–2) in two important respects. First, the comparison of two stimuli imposes greater demands on working memory and attention; the greater load could potentially render the task more vulnerable to the demands of stimulation. Second, the comparison of monosyllables and disyllables encourages phonetic strategies—participants could differentiate the two items by searching for phonetic cues that discriminate them from each other (e.g., cues for the pre-tonic vowels), without encoding the items phonologically. Of interest is whether participants would nonetheless remain sensitive to syllable structure, and whether their sensitivity would be further modulated by motor stimulation. Experiment 3 evaluates this question with English speakers; Experiment 4 examines Russian participants.

## Methods

### Participants

Forty-eight participants took part in the experiment; they were native English speakers, students at Northeastern University. Ten additional participants were tested, but excluded from the analysis because they did not perform the stimulation task continuously. Participants took part in the experiment in partial fulfillment of a course credit or remuneration (\$10).

### Materials

The materials consisted of the same monosyllabic and disyllabic stimuli described in Experiment 1, except that here, these items were presented in pairs. Each such pair featured either two identical items (either two monosyllables or two disyllables) or a combination of a monosyllabic item with its disyllabic counterpart (e.g., *plaf-pelaf* or *pelaf-plaf*). These items were arranged in two lists (144 pairs, each), counterbalanced for Identity (identical or non-identical pairs), number of Syllables (monosyllables vs. disyllables), syllable Type (large rise, small rise, plateau), and Place of articulation (labial vs. coronal). Each participant was assigned to one such list.

The list presented to each participant appeared in all three stimulation conditions (lip, tongue and control), for a total of 432 trials (144 pairs×3 stimulation conditions). As in Experiments 1–2, each such stimulation block was further divided into three parts



(counterbalanced for place of articulation, syllable type, and identity), and these parts were separated by a brief break.

As in Experiments 1–2, the experiment also featured a total of 48 additional pairs involving sonority falls (144 additional trials across stimulation conditions); these items were excluded from all analyses because, as explained in Experiment 1, these stimuli could not be correctly identified even by native Russian speakers.

## Procedure

Participants initiated the trial by pressing the space bar. Their response triggered the presentation two successive auditory words. The first auditory word was presented for 500 ms, followed by a period of 800 ms of white noise, which, in turn, was followed by the second auditory word. Participants were asked to indicate whether the two items were identical by pressing the appropriate key (1 = identical; 2 = nonidentical). Slow responses (responses slower than 2500 ms) triggered a computer feedback message of “too slow!”. Prior to the experiment, participants were given a short practice featuring real words (e.g., *is please = police?*).

As in Experiment 1, at the beginning of each stimulation block, participants were instructed on how to perform the stimulation, aided by a picture; as a reminder, a picture card was also placed by the computer monitor. All participants were monitored by a research assistant, who ensured that the stimulation conditions were being completed as instructed. Mirrors were placed at each computer to allow for easier monitoring of the participants.

## Results and Discussion

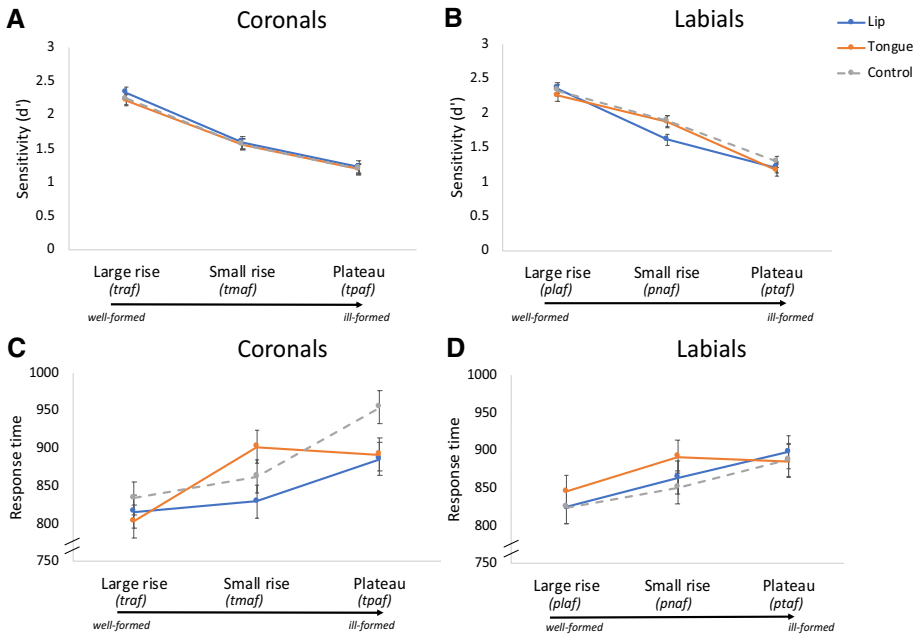
The following analyses examine the effect of syllable type and stimulation on sensitivity ( $d'$ ) and response time to nonidentical trials (e.g., *ptaf* vs. *petaf*).

The ANOVA found a reliable main effect of Type in both sensitivity ( $F(2,94) = 110.37$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.70$ ) and response time ( $F(2,66) = 30.02$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.48$ ). Figure 6 provides the means.

Planned contrasts showed that the better formed syllables with large rise yielded better sensitivity and faster response time relative to smaller rises ( $d'$ :  $t(94) = 8.71$ ,  $p < 0.0001$ ; RT:  $t(66) = 3.91$ ,  $p < 0.001$ ), which, in turn, produced better sensitivity and faster discrimination than plateaus ( $d'$ :  $t(94) = 8.71$ ,  $p < 0.0001$ ; RT:  $t(66) = 3.84$ ,  $p < 0.001$ ). As such, these results confirm that participants were sensitive to syllable structure.

However, we found no evidence for motor simulation. The Stimulation  $\times$  Place interaction was not significant (in  $d'$ :  $F(2,94) = 2.43$ ,  $p = 0.09$ ,  $\eta_p^2 = 0.05$ ; in RT:  $F(2,66) = 2.17$ ,  $p = 0.12$ ,  $\eta_p^2 = 0.06$ ). Likewise, there was no evidence of a three-way interaction (in  $d'$ :  $F(4,188) = 1.08$ ,  $p = 0.37$ ,  $\eta_p^2 = 0.02$ ; in RT:  $F < 1$ ). A separate analysis of the first block of trials found similar results: participants were sensitive to the syllable hierarchy, but there was no evidence for motor simulation (see SM).

Summarizing, English speakers were highly sensitive to syllable structure: as the syllable structure became worse formed, their perceptual sensitivity ( $d'$ ) declined, and their response time increased. However, we found no evidence that this sensitivity to syllable structure was affected by motor simulation. In fact, in the AX task, there was no evidence that participants relied on motor simulation, possibly because the AX task exacts steeper



**Fig. 6** The effect of motor stimulation on sensitivity to the syllable structure of labial- and coronal-initial stimuli

memory and attention demands than syllable count task. Since most syllable types were unfamiliar to English participants, this combination of task demands and inexperience could have prevented them from engaging in motor simulation.

To address this possibility, in Experiment 4, we administered the same task to speakers of Russian—a language where all syllable types are attested. We thus expect Russian speakers to show markedly better perceptual sensitivity relative to English speakers. Of interest is whether Russian speakers would be more likely to engage in motor simulation.

## Experiment 4: Russian Speakers (AX Discrimination)

### Methods

#### Participants

Forty-two native Russian speakers took part in this experiment. Native Russian speakers were defined described a in Experiment 2, and their ages were likewise between the 18–40 years. Six additional participants were tested, but their data were eliminated; four due to technical issues and two for failing to consistently apply the stimulation as instructed. Participants received either credit that partially fulfilled course requirements or remuneration (\$10).

## Materials

The experimental design and stimuli were as in Experiment 3, except that the instructions were written in Russian.

## Procedure

The first nineteen participants were tested in the lab individually, using the same E-Prime software as in Experiments 1–3, as well as the same procedures described therein. Due to the onset of the COVID-19 pandemic, the remaining participants were tested online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). The experiment was implemented in *Gorilla* to precisely follow the same design and timing specifications as in the E-Prime program, and the trials were further designed to look as similar as possible to the previous E-prime version.

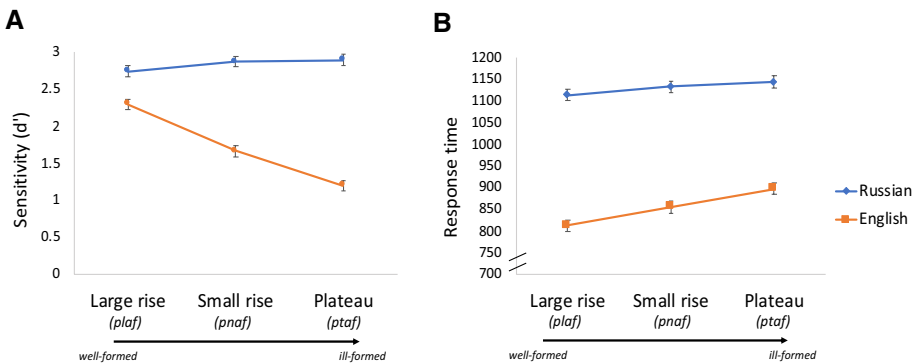
The online experiment was administered over a video call. A research assistant was present throughout the experiment to provide the instructions, answer any questions, and monitor the performance of the stimulation procedure. The procedure was likewise the same as in the E-Prime version (described in Experiment 3), except that, in the online version, participants were not provided with the picture card of the stimulation condition. This image, however, was displayed on their computer screen before each condition.

## Results and Discussion

We first report the results from Russian participants ( $d'$  and RT to nonidentical trials); we next compare their performance with English speakers (in Experiment 3).

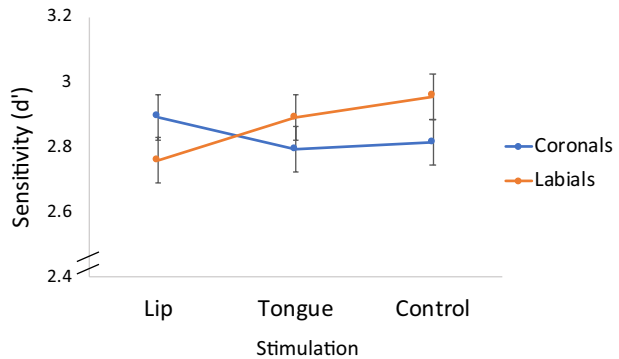
### All Blocks

An inspection of the means (Fig. 7) suggests that Russian participants showed better sensitivity than English speakers, and their performance was unaffected by syllable structure.



**Fig. 7** The effect of the syllable hierarchy on sensitivity ( $d'$ ) and response time for Russian and English speakers

**Fig. 8** The effect of motor stimulation and place of articulation on Russian participants



The ANOVA indeed found no effect of syllable Type in either sensitivity ( $F(2,82)=2.12$ ,  $p=0.13$ ,  $\eta_p^2=0.05$ ) or RT ( $F(2,80)=2.13$ ,  $p=0.13$ ,  $\eta_p^2=0.05$ ).

Did Russian participants engage in motor simulation? An inspection of the means (Fig. 8) suggested that when participants pressed on their lips (but not their tongues), the perceptual disadvantage of coronal- (relative to labial) stimuli was selectively reversed. The Stimulation  $\times$  Place interaction was indeed significant in the analysis of  $d'$  ( $F(2,82)=4.89$ ,  $p=0.01$ ,  $\eta_p^2=0.11$ ; for RT:  $F < 1$ ).

Planned contrasts showed that, in the control condition, coronals produced lower sensitivity than labials ( $t(82)=2.13$ ,  $p=0.04$ ); this however was not the case when participants pressed on the tongues ( $t(82)=1.46$ ,  $p=0.15$ ). And when participants pressed on their lips, coronals produced *better* perceptual sensitivity than labials ( $t(82)=1.99$ ,  $p=0.05$ ). There was no evidence for a three-way interaction ( $d$ :  $F(4,164)=1.48$ ,  $p=0.21$ ,  $\eta_p^2=0.03$ ; for RT:  $F < 1$ ).

A separate analysis of the first block of trials yielded similar conclusions: there was evidence that participants relied on motor simulation, but there was no evidence that simulation modulated their sensitivity to syllable structure (see SM).

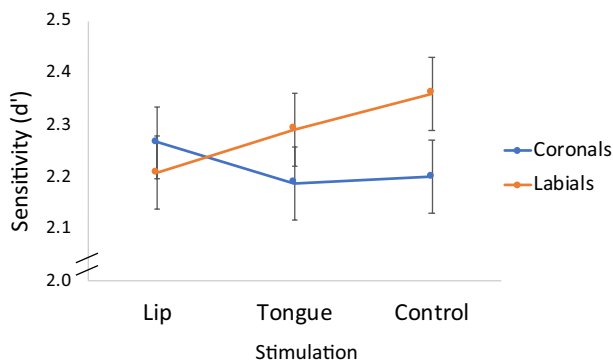
Summarizing, the results of Experiment 4 offered evidence that Russian speakers engaged in motor simulation, but, unlike English speakers, their performance was not reliably affected by syllable structure. A final set of analyses compared the performance of the two groups.

## A Comparison to English Speakers

To determine whether participants' sensitivity to syllable structure and their reliance on motor simulation was modulated by linguistic experience, we compared the performance of English and Russian speakers using a 2 Language  $\times$  3 syllable Type  $\times$  3 Stimulation  $\times$  2 Place ANOVA. As in previous experiments, here we primarily focus on the main effect of Language, and its interaction with syllable Type and motor stimulation (Language  $\times$  Stimulation  $\times$  Place; Language  $\times$  Stimulation  $\times$  Place  $\times$  Type); other effects involving the Language factor are discussed in the SM.

We found a significant effect main effect of Language, as Russian speakers exhibited better sensitivity, but their response time was slower ( $d'$ :  $F(1,88)=69.89$ ,  $p < 0.0001$ ,  $\eta_p^2=0.44$ ; RT:  $F(1,73)=31.96$ ,  $p < 0.0001$ ,  $\eta_p^2=0.30$ ). Because Russian participants were partly tested online, whereas English participants were tested in the lab, these differences in response time are difficult to interpret.

**Fig. 9** The effect of motor stimulation and place of articulation on (across language)



Critically, Language modulated the effect of syllable Type, and this interaction was significant in the analyses of  $d'$  ( $F(2,176)=80.23$ ,  $p<0.0001$ ,  $\eta_p^2=0.48$ ) and RT ( $F(2,146)=4.15$ ,  $p=0.02$ ,  $\eta_p^2=0.05$ ). As noted in the previous analyses, syllable type modulated the perceptual sensitivity of English, but not Russian speakers (see Fig. 7).

We found no evidence that these differences were further modulated by motor stimulation, as the high-level interaction (Language  $\times$  Syllable Type  $\times$  Stimulation  $\times$  Place) was not significant ( $d'$ :  $F(4,352)=2.09$ ,  $p=0.08$ ,  $\eta_p^2=0.02$ ; RT:  $F<1$ ). There was likewise no evidence that language modulated the effect of stimulation (for the Language  $\times$  Stimulation  $\times$  Place all  $F<1$  for  $d'$  and RT).

Nonetheless, when the responses of English and Russian participants were combined, we found a reliable Stimulation  $\times$  Place interaction in the analysis of  $d'$  ( $F(2,176)=7.39$ ,  $p<0.001$ ,  $\eta_p^2=0.08$ ). This interaction emerged because sensitivity to labial sounds was superior to coronals in the control ( $t(176)=3.78$ ,  $p<0.001$ ) and tongue ( $t(176)=2.47$ ,  $p=0.01$ ) conditions, but this advantage was eliminated when participants pressed on their lips ( $t(176)=1.36$ ,  $p=0.17$ ). Figure 9 provides the means.

Summarizing, the comparison of Experiments 3–4 confirms that the sensitivity of English speakers to the syllable hierarchy (in Experiment 3) is not due to stimuli artifacts: when Russian speakers were presented with the same stimuli, they showed no difficulty with ill-formed syllables. The difficulty of English speakers with ill-formed syllables is thus due to their knowledge of language.

We also found some evidence that Russian participants relied on motor simulation. While the separate analysis of English speakers found no evidence for simulation (in Experiment 3), when the two groups were combined, the evidence for simulation emerged, irrespective of language. While labial sounds produced better sensitivity when participants pressed on their tongues (as well as in the control condition), when participants pressed on their lips, this labial advantage was eliminated. Altogether, then, we found evidence that both groups possibly relied on motor simulation, and that English speakers were further sensitive to syllable structure. But there was no evidence that this sensitivity arose from motor simulation.

## General Discussion

It is well established that, across languages, certain syllable types are systematically preferred to others (e.g., *plaf* > *pnaf* > *ptaf*; Greenberg, 1978; see also Berent et al., 2007). It is also known that this syllable hierarchy is psychologically real, inasmuch as syllables that

are preferred across languages are also ones that individual speakers process more easily, even when those syllables are unattested in their language (e.g., Berent et al., 2007; Maïonchi-Pino et al., 2013; Pertz & Bever, 1975). Unknown, however, is the basis for this preference.

Here, we explored whether the syllable hierarchy arises from motor simulation. In this view, ill-formed syllables like *ptaf* are dispreferred because, upon hearing such speech sounds, listeners seek to simulate their articulatory production; since ill-formed syllables like *ptaf* are presumably harder to simulate, their perceptual identification declines. More generally, then, this account asserts that the syllable hierarchy arises not from abstract linguistic constraints. Rather, language structure, generally, and syllable structure, specifically, is **embodied** in action (e.g., Barsalou, 2008; Glenberg et al., 2013; Pulvermüller & Fadiga, 2010).

The present research thus asked (a) Whether participants are sensitive to syllable structure, and (b) Whether this sensitivity arises from motor simulation. We further evaluated the possibility that (c) Motor simulation and sensitivity to syllable structure are modulated by linguistic experience. To this end, we gauged evidence from motor simulation across two groups of participants, whose native languages differ markedly on the type of syllables they allow: English vs. Russian.

In four experiments, we compared the sensitivity of these two groups of speakers to the syllable structure of labial-initial (e.g., *plaf* > *pnaf* > *ptaf*) and coronal-initial stimuli (e.g., *traf* > *tmaf* > *tpaf*). Meanwhile, participants lightly pressed on either their lips, their tongues, or relaxed (in the control condition). Experiments 1–2 used syllable count (e.g., *does plaf have one syllable or two?*); Experiments 3–4 employed AX discrimination (e.g., *does plaf = pelaf?*).

Using this same procedures, past research has shown that participants rely on motor simulation in the **phonetic** categorization of speech sounds (Berent, Platt, et al., 2020). In these experiments, participants heard speech stimuli that were ambiguous with respect to their voicing—either labials (in between *balpa*) or coronals (in between *dalta*). Results showed that perceptual sensitivity ( $d'$ ) varied systematically, depending on whether participants stimulated an articulator that was congruent (e.g., lips, for labials) or incongruent (e.g., tongue, for labials) with the speech sounds. Here, we asked whether stimulation could likewise modulate **phonological** computations, specifically, the computation of syllable structure.

## Are Participants Sensitive to the Syllable Hierarchy?

To address this question, let us first consider whether participants—English and Russian participants—were sensitive to the syllable hierarchy. As expected, English speakers were highly sensitive to syllable structure: as syllable structure became worse-formed, perceptual sensitivity declined, and this was the case regardless of whether participants were asked to perform syllable count (in Experiment 1) or AX discrimination (in Experiment 3). The AX discrimination further found that worse-formed syllables elicited longer response times. In contrast, for Russian speakers (in Experiments 2 & 4), ill-formed syllables did not exert a similar toll on perceptual sensitivity, as these syllable types are all attested in their language.

These results make it clear that the difficulties of English speakers with ill-formed syllables (e.g., *ptaf*) are due to some knowledge, arising, in part, from linguistic experience. Whether this knowledge of syllable structure references sonority directly (e.g., Clements,

1990; de Lacy, 2007; Sheer, 2019), or whether it arises from other grammatical principles (e.g., Smolensky, 2006) cannot be settled here. Clearly, however, participants consult some linguistic knowledge, rather than only the acoustic properties of these stimuli.

### **Do Participants Rely on Motor Simulation?**

Was this difficulty, then, due to motor simulation? Before we can answer this question, we first need to ask whether participants relied on motor simulation. If they did, then the effect of stimulation should have been *selective*: pressing on the lips, for example, should have differentially affected the perception of labial- relative to coronal stimuli; pressing on the tongue should have produced the opposite effect.

Our four experiments yielded various indications that stimulation produced such selective effects. In the syllable count task, when English speakers pressed on their tongues, they showed better perceptual sensitivity to labials (relative to coronal) sounds; this however, was not the case when participants pressed on their lips. Russian speakers showed similar effects in response time: when Russian participants pressed on their lips, they responded faster to coronals- relative to labials; this, however, was not the case when they pressed on their tongues. Thus, in both cases, response to speech sounds was improved by stimulating the incongruent articulator. This evidence for simulation, however, only emerged in the first block of trials.

In the AX task, English speakers showed no evidence for motor simulation (in Experiment 3), possibly, due to its attention and memory demands. Russian speakers, however, did show such effects. Here (in Experiment 4), coronal sounds produced poorer perceptual sensitivity than labials sounds, and this was the case regardless of whether participants engaged their tongues or relaxed (in the control condition). But when Russian participants pressed on their lips, the disadvantage of coronal sounds was eliminated. Similar findings obtained in the combined analysis of English and Russian participants. Here, pressing on the tongue produced better sensitivity to labials (relative to coronals), but this was not the case when participants pressed on their lips (the congruent articulator). Thus, as in the syllable count task, perceptual sensitivity was higher when the stimulation engaged the incongruent- (relative to the congruent) articulator. We speculate that, because the congruent articulator is critical to simulation, its engagement disrupts speech perception; by comparison, then, incongruent stimulation is advantageous.

As noted, these results ought to be interpreted with caution, as the evidence for simulation did not emerge in every experiment and for both places of articulation. Nonetheless, these results suggest that mechanical motor stimulation can reliably modulate perceptual sensitivity of both English and Russian speakers. These results are in line with the possibility that the identification of speech sounds does rely on motor simulation, generally, and, at least in our experiments, these effects do not seem to systematically depend on linguistic experience.

### **Is Sensitivity to Syllable Structure Due to Motor Simulation?**

Having offered evidence that participants in our experiments did rely on motor simulation, we can now turn to the main question of whether motor simulation can explain the syllable hierarchy. We reasoned that, if ill-formed syllables (e.g., of the labial *ptaf*) are misidentified because they are harder to simulate, then once motor simulation is disrupted (e.g.,

by stimulating the lips), perceptual sensitivity to ill-formed syllables should improve in a selective manner (depending on their congruence between their place of articulation and the stimulated articulator), and consequently, sensitivity to the syllable hierarchy should be attenuated. Statistically, this should give rise to a three-way interaction (Stimulation  $\times$  Place  $\times$  syllable Type). Across our four experiments, this interaction was not reliable.

Null results, of course, could arise from many sources. In particular, we cannot rule out the possibility that this null effect emerged because our motor simulation procedure was insufficiently robust. As noted, however, in past research with phonetically ambiguous stimuli, this same procedure was shown to consistently affect perceptual sensitivity (Berent, Platt, et al., 2020); this is in line with the large body of literature showing that the phonetic categorization of ambiguous speech sounds is systematically modulated by both mechanical (e.g., Bruderer et al., 2015; Ito et al., 2009; Nasir & Ostry, 2009; Sams et al., 2005; Scott et al., 2013) and transcranial (e.g., D'Ausilio et al., 2012; D'Ausilio et al., 2009; Fadiga, Craighero, Buccino) stimulation. Yet in the present investigation, stimulation did not modulate sensitivity to syllable structure. This conclusion further agrees with past research that examined sensitivity to the syllable hierarchy using both mechanical (Zhao & Berent, 2018) and transcranial stimulation (Berent et al., 2015). In both cases, sensitivity to syllable structure remained robust despite the stimulation of the motor system. This convergence opens up the possibility that the grammatical computation of syllable structure are impervious to motor simulation.

## The Role of Motor Simulation in Speech Perception

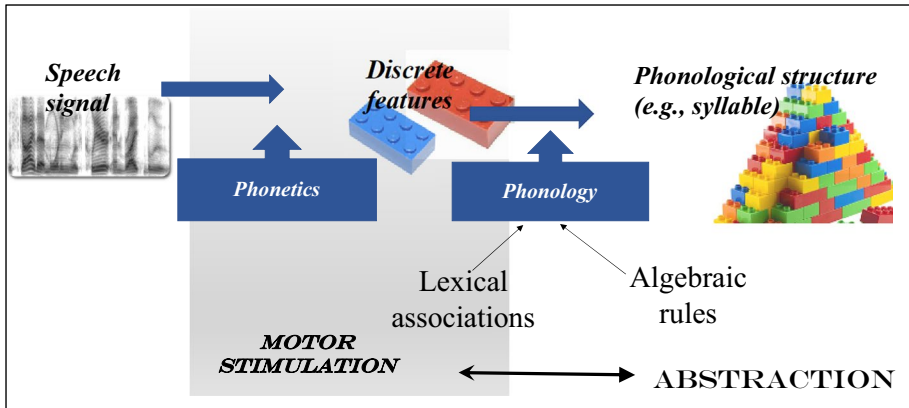
These conclusions, if correct, raise a puzzle. Given that motor simulation has been amply demonstrated to contribute to some aspects of speech perception, why does the computation of syllable structure seem unaffected by motor simulation? How can we accommodate the present results with the large literature on phonetic categorization, where the effect of motor stimulation is robust, and it arises in both mechanical and transcranial stimulation?

Our present results cannot unequivocally address this question. As noted, we cannot rule out the possibility that these differences could partly arise from methodological factors, such as stimulus clarity (i.e., ambiguous vs. unambiguous speech sounds), length, and task demands. Additionally, our conclusions with respect to the role of motor simulation in phonology are limited, insofar that they are based on a single case study (i.e., the syllable hierarchy). Thus, our present findings (that motor simulation does not mediate sensitivity to the syllable hierarchy) do not preclude the possibility that other phonological computations could rely on motor simulation (e.g., the place assimilation). These limitations notwithstanding, it is possible that these differences could also arise because phonetics and phonology represent different levels within the language system.

Indeed, speech perception entails multiple levels of analysis. The phonetic level transforms the continuous speech signal into discrete categories. The phonological level, by contrast, builds linguistic structure, informed, in part, by grammatical algebraic rules. We hypothesize that phonetic processing is strongly embodied, but the grammatical computation of phonological rules is disembodied and abstract (see Fig. 10).

This proposal nicely explains the present pattern of results. In this view, motor simulation in past experiments arose primarily at the phonetic level. Since the materials were spoken, and since spoken stimuli require phonetic categorization, participants must have engaged phonetic categorization—a process that is known to rely on motor simulation. But because the speech stimuli were unambiguous, the simulation demands might not have





**Fig. 10** The hypothesized role of motor simulation in speech perception. The processed hypothesized to rely on motor simulation are highlighted by the gray area

been as pronounced as with ambiguous speech sounds. These considerations explain why the effect of simulation in the present experiments is perhaps less robust than in the previous research on phonetic categorization.

Critically, we hypothesize that sensitivity to syllable structure relied primarily on abstract algebraic rules. Since, in the present proposal, the grammatical computation of syllable structure engages abstract computations that are impervious to motor simulation, this proposal explains why the effects of simulation did not modulate sensitivity to syllable structure. Altogether, then, this proposal explains why mechanical stimulation systematically modulates phonetic categorization, but not the (grammatical phonological) computation of syllable structure.

Is phonology, then, utterly impervious to motor simulation? We suspect not. Indeed, phonological preferences are demonstrably informed not only by abstract grammatical rules but also by lexical associations. For example, English speakers might encode by *plaf* (a large rise in sonority) by analogy to *plan*, for instance. And in the present proposal, these lexical associations could very well be informed by motor simulation (see Fig. 10). While this prediction was not borne out in the present findings, in past research, we found that the stimulation of the lip motor area by TMS selectively impaired perceptual sensitivity to syllables with large sonority rises (Berent et al., 2015). We would thus expect similar effects to emerge in mechanical stimulation as well. Critically, we predict a dissociation between the role of motor simulation in lexical association and phonetic categorization, on the one hand (where simulation should play a role), and the phonological grammar, on the other (where it shouldn't).

The possibility that grammatical phonological computations are abstract also agrees with previous research, showing that grammatical phonological computations can transcend linguistic modality. For example, in one line of research, we found that speakers with no command of a sign language spontaneously project the phonological and morphological principles from their spoken language to dynamic linguistic signs. Critically, responses to linguistic signs varied systematically, depending on the structure of speakers' spoken language (e.g., Berent et al., 2016, 2021; Berent, Bat-El, et al.,

2020). As such, these results suggest that the knowledge that participants invoked must have been at least partly disembodied and abstract.

Whether the computation of phonological structure from speech is partly disembodied is a question that awaits confirmation by further research. Our findings, however, suggest that a resolution of this matter might require a nuanced approach. Rather than asking “is speech perception embodied”, we advocate a research program that carefully evaluates the role of embodiment at distinct levels of analysis—phonetic categorization vs. grammatical computations. Thus, to advance the debate on the contribution of motor simulation to speech perception, the consideration of representational grain-size ought to play a role.

## Appendix

See Table 1.

**Table 1** Experimental materials

	Syllable type		
	Large rise	Small Rise	Plateau
Labial	pluf	pnuf	ptuf
	plaf	pnaf	ptaf
	plif	pnif	ptif
	plok	pnok	ptok
	plek	pnek	ptek
	plik	pnik	ptik
	plit	pnit	ptit
	plet	pnet	ptet
	plat	pnat	ptat
	plosh	pnosh	ptosh
	plish	pnish	ptish
	plesh	pnesh	ptesh
	Coronal	traf	tmaf
trof		tmof	tpof
truf		tmuf	tpuf
trak		tmak	tpak
trok		tmok	tpok
trik		tmik	tpik
trep		tmep	tpep
trap		tmap	tpap
trop		tmop	tpop
trush		tmush	tpush
tresh		tmesh	tpesh
trish		tmish	tpish

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10936-022-09871-x>.

**Acknowledgements** This research was supported by NSF Grant No. 1733984 (to IB/and Alvaro Pascual-Leone).

## Declarations

**Conflict of interest** All authors declare that they have no conflict of interest.

## References

- Andan, Q., Bat-El, O., Brentari, D., & Berent, I. (2018). Anchoring is amodal: Evidence from a signed language. *Cognition*, *180*, 279–283. <https://doi.org/10.1016/j.cognition.2018.07.016>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. <https://doi.org/10.1007/s00426-019-01269-0>
- Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. [https://doi.org/10.1016/S1364-6613\(02\)00029-3](https://doi.org/10.1016/S1364-6613(02)00029-3)
- Bat-El, O. (2012). The sonority dispersion principle in the acquisition of Hebrew word final codas. In S. Parker (Ed.), *The sonority controversy* (pp. 319–344). Mouton de Gruyter. <https://doi.org/10.1017/S0952675716000075>
- Berent, I. (2013). *The phonological mind*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139049610>
- Berent, I., Bat-El, O., Andan, Q., Brentari, D., & Vaknin-Nusbaum, V. (2021). Amodal phonology. *Journal of Linguistics*, *57*, 199–529. <https://doi.org/10.1017/S0022226720000298>
- Berent, I., Bat-El, O., Brentari, D., Dupuis, A., & Vaknin-Nusbaum, V. (2016). The double identity of linguistic doubling. *Proceedings of the National Academy of Sciences*, *113*(48), 13702–13707. <https://doi.org/10.1073/pnas.1613749113>
- Berent, I., Bat-El, O., Brentari, D., & Platt, M. (2020a). Knowledge of language transfers from speech to sign: Evidence from doubling. *Cognitive Science*, *44*, 1. <https://doi.org/10.1111/cogs.12809>
- Berent, I., Brem, A.-K., Zhao, X., Seligson, E., Pan, H., Epstein, J., Stern, E., Galaburda, A. M., & Pascual-Leone, A. (2015). Role of the motor system in language knowledge. *Proceedings of the National Academy of Sciences*, *112*, 1983–1988. <https://doi.org/10.1073/pnas.1416851112>
- Berent, I., Harder, K., & Lennertz, T. (2011). Phonological universals in early childhood: Evidence from sonority restrictions. *Language Acquisition*, *18*, 281–293. <https://doi.org/10.1080/10489223.2011.580676>
- Berent, I., & Lennertz, T. (2010). Universal constraints on the sound structure of language: Phonological or acoustic? *Journal of Experimental Psychology Human Perception & Performance*, *36*(1), 212–223. <https://doi.org/10.1037/a0017638>
- Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321–5325. <https://doi.org/10.1073/pnas.0801469105>
- Berent, I., Lennertz, T., & Rosselli, M. (2012). Universal linguistic pressures and their solutions: Evidence from Spanish. *The Mental Lexicon*, *13*(2), 275–305. <https://doi.org/10.1075/ml.7.3.02ber>
- Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, *26*(1), 75–108. <https://doi.org/10.1017/S0952675709001729>
- Berent, I., Pan, H., Zhao, X., Epstein, J., Bennett, M. L., Deshpande, V., Seethamraju, R. T., & Stern, E. (2014). Language universals engage Broca's area. *PLoS ONE*, *9*(4), e95155. <https://doi.org/10.1371/journal.pone.0095155>
- Berent, I., Platt, M., Theodore, R., Balaban, E., Fried, P., & Pascual-Leone, A. (2020b). Speech perception triggers articulatory action: Evidence from mechanical stimulation. *Frontiers in Communication*. <https://doi.org/10.3389/fcomm.2020.00034>

- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630. <https://doi.org/10.1016/j.cognition.2006.05.015>
- Blevins, J. (2004). *Evolutionary phonology*. Cambridge University Press.
- Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [computer program] (Version 5.406). Retrieved from <http://www.praat.org/>
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [computer program] (Version 6.0.37). Retrieved from <http://www.praat.org/>
- Brentari, D. (2011). Sign language phonology. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *Handbook of phonological theory*. Blackwells.
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251. <https://doi.org/10.1017/S0952675700001019>
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1508631112>
- Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review*, *22*(2–4), 381–410. <https://doi.org/10.1017/S0952675700001019>
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 282–333). Cambridge University Press.
- Daland, R., Hayes, B., Garellek, M., White, J., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, *28*, 197–234. <https://doi.org/10.1017/S0952675711000145>
- D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex A Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 882–887. <https://doi.org/10.1016/j.cortex.2011.05.017>
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology CB*, *19*(5), 381–385. <https://doi.org/10.1016/j.cub.2009.01.017>
- Davidson, L., Martin, S., & Wilson, C. (2015). Stabilizing the production of nonnative consonant clusters with acoustic variability. *The Journal of the Acoustical Society of America*, *137*(2), 856–872. <https://doi.org/10.1121/1.4906264>
- de Lacy, P. (2004). Markedness conflation in optimality theory. *Phonology*, *21*, 145–188. <https://doi.org/10.1017/S0952675704000193>
- de Lacy, P. (2006). *Markedness: Reduction and preservation in phonology*. Cambridge University Press.
- de Lacy, P. (2007). The interaction of tone, sonority, and prosodic structure. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 281–307). Cambridge University Press. <https://doi.org/10.1017/CBO9780511486371>
- de Lacy, P. (2008). Phonological evidence. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation* (pp. 43–77). Equinox Publications.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*, 399–402. <https://doi.org/10.1046/j.0953-816x.2001.01874.x>
- Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71. <https://doi.org/10.1016/j.cognition.2008.11.013>
- Gafos, A. I. (1999). *The articulatory basis of locality in phonology*. Garland Publishers.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502. <https://doi.org/10.1038/nature08572>
- Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 Years of cognitive psychology. *Perspectives on Psychological Science Journal of the Association for Psychological Science*, *8*(5), 573–585. <https://doi.org/10.1177/1745691613498098>
- Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (p. cm). Mouton de Gruyter. <https://doi.org/10.1515/9783110895094>
- Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespore, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, *111*(16), 5837–5841. <https://doi.org/10.1073/pnas.1318261111>

- Greenberg, J. H. (1978). Some generalizations concerning initial and final consonant clusters. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 2, pp. 243–279). Stanford University Press.
- Hayes, B., Kirchner, R. M., & Steriade, D. (2004). *Phonetically based phonology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486401>
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245. <https://doi.org/10.1073/pnas.0810063106>
- Jarosz, G. (2017). Defying the stimulus: Acquisition of complex onsets in Polish. *Phonology*, 34(2), 269–298. <https://doi.org/10.1017/S0952675717000148>
- Keating, P. A. (1988). The phonology-phonetics interface. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. I, pp. 281–302). Cambridge University Press.
- Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review*, 23(4), 991–1001. <https://doi.org/10.3758/s13423-015-0870-z>
- MacNeilage, P. F. (2008). *The origin of speech*. Oxford University Press.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology Paris*, 102(1–3), 59–70.
- Maïonchi-Pino, N. (2012). Phonological restriction knowledge in dyslexia: Universal or language-specific? In T. Wydell (Ed.), *Dyslexia: A comprehensive and international approach*. InTech. <https://doi.org/10.5772/11811>
- Maïonchi-Pino, N., de Cara, B., Écalle, J., & Magnan, A. (2012). Are French dyslexic children sensitive to consonant sonority in segmentation strategies?. Preliminary evidence from a letter detection task. *Research in Developmental Disabilities*, 33(1), 12–23. <https://doi.org/10.1016/j.ridd.2011.07.045>
- Maïonchi-Pino, N., Magnan, A., & Écalle, J. (2010). Syllable frequency effects in visual word recognition: Developmental approach in French children. *Journal of Applied Developmental Psychology*, 31(1), 70–82.
- Maïonchi-Pino, N., Taki, Y., Yokoyama, S., Magnan, A., Takahashi, K., Hashizume, H., & Écalle-Kawashima, J. R. (2013). Is the phonological deficit in developmental dyslexia related to impaired phonological representations and to universal phonological grammar? *Journal of Experimental Child Psychology*, 115(1), 53–73. <https://doi.org/10.1016/j.jecp.2012.10.006>
- Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 415–420). North Holland.
- McCarthy, J. J., & Prince, A. (1998). Prosodic morphology. In A. Spencer & A. M. Zeicky (Eds.), *Handbook of morphology* (pp. 283–305). Basil Blackwell.
- Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, 84(1), 55–71.
- Möttönen, R., Rogers, J., & Watkins, K. E. (2014). Stimulating the lip motor cortex with transcranial magnetic stimulation. *Journal of Visualized Experiments: Jove*, (88). <https://doi.org/10.3791/51665>
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience*, 29(31), 9819–9825. <https://doi.org/10.1523/JNEUROSCI.6018-08.2009>
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470. <https://doi.org/10.1073/pnas.0907032106>
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4(2), 135–183. [https://doi.org/10.1016/S0364-0213\(80\)80015-2](https://doi.org/10.1016/S0364-0213(80)80015-2)
- Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, 32, 397–421.
- Ohala, J. J., & Kawasaki-Fukumori, H. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Regional Meetings, Chicago Linguistic Society*, 2, 319–338.
- Pertz, D. L., & Bever, T. G. (1975). Sensitivity to phonological universals in children and adolescents. *Language*, 51(1), 149–162.
- Prince, A., & Smolensky, P. (1993/2004). *optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing: <https://doi.org/10.7282/T34M92MV>
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360. <https://doi.org/10.1038/nrn2811>

- Pulvermüller, F., Huss, M., Kherif, F., del Prado, M., Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865–7870. <https://doi.org/10.1073/pnas.0509989103>
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Research. Cognitive Brain Research*, 23(2–3), 429. <https://doi.org/10.1016/j.cogbrainres.2004.11.006>
- Sandler, W., & Lillo-Martin, D. C. (2006). *Sign language and linguistic universals*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139163910>
- Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, 24(9), 1824–1830. <https://doi.org/10.1177/0956797613478614>
- Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, 133(4), EL286. <https://doi.org/10.1121/1.4794932>
- Sheer, T. (2019). Sonority is diferent. *Studies in Polish Linguistics*, 1, 127–151. <https://doi.org/10.4467/23005920SPL.19.009.10989>
- Smalle, E. H. M., Rogers, J., & Möttönen, R. (2014). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex*. <https://doi.org/10.1093/cercor/bhu218>
- Smolensky, P. (2006). Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 2, pp. 27–160). MIT Press.
- Smolensky, P., & Legendre, G. (2006). Principles of integrated connectionist/symbolic cognitive architecture. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 1, pp. 63–97). MIT Press.
- Steriade, D. (1988). Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology*, 5, 37–155.
- Tamasi, K., & Berent, I. (2015). Sensitivity to phonological universals: The case of fricatives and stops. *Journal of Psycholinguistic Research*, 44(4), 59–81. <https://doi.org/10.1007/s10936-014-9289-3>
- Wright, R. (2004). A review of perceptual cues and robustness. In D. Steriade, R. Kirchner, & B. Hayes (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge University Press. <https://doi.org/10.1002/9780470756171.ch15>
- Zhao, X., & Berent, I. (2016). Universal restrictions on syllable structure: Evidence from Mandarin Chinese. *Journal of Psycholinguistic Research*, 45, 795–811. <https://doi.org/10.1007/s10936-015-9375-1>
- Zhao, X., & Berent, I. (2018). The basis of the syllable hierarchy: Articulatory pressures or universal phonological constraints? *Journal of Psycholinguistic Research*, 47(1), 29–64. <https://doi.org/10.1007/s10936-017-9510-2>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.