

Role of the motor system in language knowledge

Iris Berent^{a,1}, Anna-Katharine Brem^{b,c}, Xu Zhao^a, Erica Seligson^c, Hong Pan^d, Jane Epstein^d, Emily Stern^d, Albert M. Galaburda^e, and Alvaro Pascual-Leone^{c,e}

^aDepartment of Psychology, Northeastern University, Boston, MA 02115; ^bDepartment of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom; ^cBerenson-Allen Center for Noninvasive Brain Stimulation, Beth-Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215; ^dFunctional Neuroimaging Laboratory, Departments of Radiology and Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02467; and ^eDepartment of Neurology, Division of Cognitive Neurology, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA 02215

Edited by Charles R. Gallistel, Rutgers University, Piscataway, NJ, and approved January 7, 2015 (received for review September 3, 2014)

All spoken languages express words by sound patterns, and certain patterns (e.g., *blog*) are systematically preferred to others (e.g., *lbog*). What principles account for such preferences: does the language system encode abstract rules banning syllables like *lbog*, or does their dislike reflect the increased motor demands associated with speech production? More generally, we ask whether linguistic knowledge is fully embodied or whether some linguistic principles could potentially be abstract. To address this question, here we gauge the sensitivity of English speakers to the putative universal syllable hierarchy (e.g., *blif* > *bnif* > *bdif* > *lbif*) while undergoing transcranial magnetic stimulation (TMS) over the cortical motor representation of the left orbicularis oris muscle. If syllable preferences reflect motor simulation, then worse-formed syllables (e.g., *lbif*) should (i) elicit more errors; (ii) engage more strongly motor brain areas; and (iii) elicit stronger effects of TMS on these motor regions. In line with the motor account, we found that repetitive TMS pulses impaired participants' global sensitivity to the number of syllables, and functional MRI confirmed that the cortical stimulation site was sensitive to the syllable hierarchy. Contrary to the motor account, however, ill-formed syllables were least likely to engage the lip sensorimotor area and they were least impaired by TMS. Results suggest that speech perception automatically triggers motor action, but this effect is not causally linked to the computation of linguistic structure. We conclude that the language and motor systems are intimately linked, yet distinct. Language is designed to optimize motor action, but its knowledge includes principles that are disembodied and potentially abstract.

embodiment | TMS | fMRI | phonology | language universals

Many animal species communicate using vocal patterns, and humans are no exception. Every hearing human community preferentially expresses words by oral patterns (1). Speech sounds, such as *d, o, g* give rise to contrasting patterns (e.g., *dog* vs. *god*), and certain speech patterns are systematically preferred to others. Syllables like *blog*, for instance, are more frequent across languages than *lbog* (2). Behavioral experiments further demonstrate similar preferences among individual speakers despite no experience with either syllable type (3–6).

Although such facts demonstrate that the sound patterns of language are systematically constrained, the nature of such constraints remains unknown. One explanation invokes universal linguistic constraints on the sound structure of language (7). However, in an alternative account, these patterns are thought to reflect motor, rather than linguistic, constraints caused by their embodiment in the motor system of speech (8–11). Indeed, the speech patterns that are attested across spoken languages are not arbitrary, and frequent patterns tend to optimize speech production (12). Such observations open up the possibility that the so-called “language universals” are action based. In this view, the encoding of a speech stimulus engages a motor articulatory network that simulates its production—the harder the motor simulation or production, the less preferred the stimulus (12, 13). In line with this possibility, past research has shown that the identification of speech sounds preferentially engages their specific

articulators. Labial speech sounds (e.g., *b*) engage the lip relative to the tongue motor area, whereas coronal speech sounds (e.g., *d*) produce the opposite pattern (14). Converging evidence is presented by studies using transcranial magnetic stimulation (TMS)—a noninvasive technique that induces focal cortical current via electro-magnetic induction to modulate (inhibit or facilitate) specific brain regions (15). Results show that TMS to the lip and tongue motor regions selectively affects the identification of corresponding speech sounds (16–20). However, although the motor system is demonstrably linked to the identification of isolated speech categories (e.g., of *b*), it is unclear whether it might further constrain their patterning into syllables (e.g., *b+la* vs. *l+ba*). Indeed, the extraction of speech categories and their patterning is attributed to two separate systems (phonetics vs. phonology) (21), with distinct computational properties and neurological implementations (22, 23). Whether motor action mediates the computation of syllable structure is unknown. Our investigation addresses this question.

In each trial, participants were presented with a single auditory stimulus—either a monosyllable or a disyllable (e.g., *blif* [blɪf] vs. *belif* [bəlɪf]) and were asked to indicate the number of syllables (i.e., one or two syllables) while receiving real or sham TMS (20-Hz bursts of four TMS pulses) over the motor representation of the left orbicularis oris muscle (OO). The effect of the TMS manipulation on speech production was gauged by a syllable elicitation task, administered before experiments 1 and 2 (Fig. 1), and its capacity to disrupt perception of labial sounds was supported by additional analyses (Fig. S1). To ensure that the stimulated motor region was relevant to the representation of

Significance

All spoken languages express words by sound patterns, and certain sound patterns (e.g., *blog*) are systematically preferred to others (e.g., *lbog*). However, whether those preferences reflect abstract rules or the motor difficulties associated with speech production is unknown. We address this question using transcranial magnetic stimulation and functional MRI. Results show that speech perception automatically engages the articulatory motor system, but linguistic preferences persist even when the motor system is disrupted. These findings suggest that, despite their intimate links, the language and motor systems are distinct. Language is designed to optimize motor action, but its knowledge includes principles that are disembodied and potentially abstract.

Author contributions: I.B., A.-K.B., H.P., E. Stern, A.M.G., and A.P.-L. designed research; A.-K.B., E. Seligson, H.P., and J.E. performed research; I.B., X.Z., H.P., J.E., and E. Stern analyzed data; and I.B., A.-K.B., H.P., J.E., E. Stern, A.M.G., and A.P.-L. wrote the paper.

Conflict of interest statement: A.P.-L. serves on the scientific advisory boards for Nexstim, Neuronix, Starlab Neuroscience, Neuroelectrics, Axilum Robotics, Magstim Inc., and Neosync and is listed as an inventor on several issued and pending patents on the real-time integration of transcranial magnetic stimulation with EEG and MRI.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: i.berent@neu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416851112/-DCSupplemental.

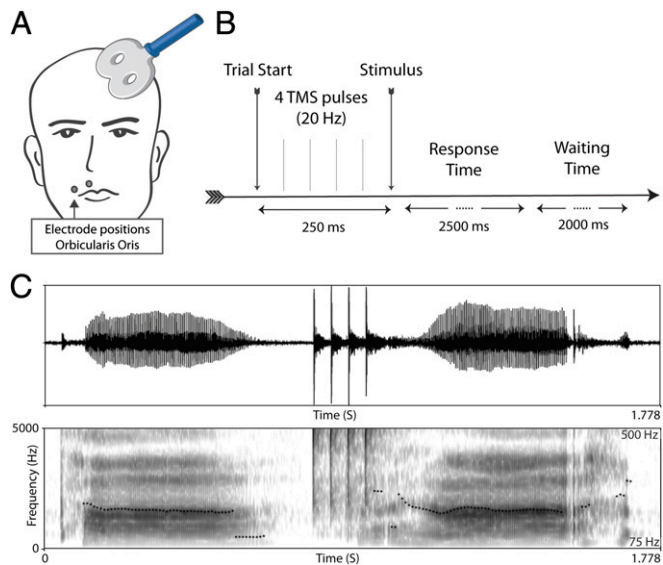


Fig. 1. The TMS procedure and its effect on speech production. During TMS trials, the coil was placed at a 45° angle, and electromyographic activity was recorded from the right first dorsal interosseus (FDI) and the right orbicularis oris (OO) muscle (A); the timing of repetitive TMS pulses relative in the experimental trial is depicted in B. C illustrates the effect of the TMS on speech production. The figure provides the waveform and spectrograms of two instances of the syllable *pa*. A first production is shown on the left, followed by four TMS pulses; the onset of the second *pa* syllable roughly coincides with the fourth pulse. TMS was associated with a marked perturbation in voicing and a delay in its onset (from 73 to 163 ms, for the first vs. second *pa* instance, respectively), along with a noticeable distortion in pitch (indicated by the dotted line).

syllable structure, we also confirmed its activation in a functional MRI (fMRI) experiment using the same task.

We compared responses to distinct syllable types, arrayed on a hierarchy based on their frequency across languages (3) and their structural well formedness (e.g., *blif*>*bnif*>*bdif*>*lbif*; for linguistic definition; *SI Text*) (24). Past behavioral research has shown that speakers are sensitive to the syllable hierarchy even when none of these syllable types is attested in their language (3–6). As the syllable becomes worse formed on the hierarchy, people tend to misidentify it as a disyllable (e.g., *lbif* becomes identified as *lebif*), and consequently, syllable count errors increase monotonically. Of interest is whether these errors reflect difficulties in the generation of a covert motor program. More generally, we ask whether people’s knowledge of their native language consists of analog patterns of speech action, or principles that are algebraic, disembodied, and abstract.

We predicted that if linguistic well-formedness preferences reflect motor simulation, then (i) as the syllable becomes worse formed, syllable count errors should increase (e.g., more errors to *lbif* relative to *blif*); (ii) speech motor areas of the brain should become more active on fMRI; and (iii) the effect of TMS on these motor regions should increase—the worse formed the syllable, the more likely it is to engage motor simulation, hence, the more susceptible to TMS. Furthermore, to the extent the dislike of ill-formed syllable reflects their excessive motor demands, then TMS should release such syllables from this burden and improve their identification. In contrast, if linguistic principles are disembodied and abstract (7, 21), then linguistic preferences, such as the syllable hierarchy, should persist irrespectively of TMS. Experiments 1 and 2 examine these predictions using TMS; experiment 3 uses an fMRI methodology.

Experiment 1

To determine whether sensitivity to the syllable hierarchy requires motor simulation, in experiment 1 we examined the effect of TMS (sham vs. real, bursts of four TMS pulses at 20 Hz) on syllable count of four types of monosyllables, arrayed according to the syllable hierarchy (e.g., *blif*>*bnif*>*bdif*>*lbif*). Items were presented in four blocks of trials, and participants were asked to discriminate those monosyllables (e.g., *lbif*) from their disyllabic (e.g., *lebif*) counterparts. The effect of TMS on sensitivity (d') was evaluated by means of 2 TMS/sham \times 4 syllable type \times 4 block order ANOVAs, conducted using both participants ($F1$) and items ($F2$) as random variables (separate results for monosyllables and disyllables are provided in Fig. S2). The TMS \times syllable type interaction was significant [$F1(3,24) = 3.27, P < 0.04$; $F2(3,69) = 5.08, P < 0.004$], and it was not further modulated by block order ($F < 1$).

Inspection of the means of the sham condition (Fig. 2) suggests that, as the syllable became worse formed, sensitivity (d') tended to decrease. Planned comparison indicated that syllables like *bnif* produced reliably higher sensitivity than *bdif* [$t1(24) = 5.70, P < 0.0001$; $t2(69) = 8.98, P < 0.0001$], which, in turn, produced higher sensitivity than *lbif*—the worst formed syllable on the hierarchy [$t1(24) = 6.19, P < 0.0001$; $t2(69) = 10.56, P < 0.0001$]. Responses to *blif*- and *bnif*-type items did not differ ($t < 1$). This finding confirms past research showing that English speakers are sensitive to the structure of syllables even if they are unattested in their language (3–6). Of interest is whether the effect of syllable structure is eliminated by TMS.

Results showed that bursts of four TMS pulses attenuated participants’ global sensitivity to the number of syllables. However, the effect of real TMS was confined to the better-formed syllables on the hierarchy—to syllables like *blif* and *bnif* ($P < 0.03$, Tukey HSD by participants and items), but it did not affect either *bdif* or *lbif* ($P > 0.6$, not significant). Moreover, participants remained sensitive to the syllable hierarchy despite TMS: sensitivity to the best-formed *blif*-type syllables was marginally higher than to *bnif*-type items [$t1(24) = 1.35, P < 0.20$; $t2(69) = 2.38, P < 0.03$]; *bnif*-type syllables produced reliably higher sensitivity than *bdif*-type syllables [$t1(24) = 2.31, P < 0.03$; $t2(69) = 4.25, P < 0.0001$], which, in turn, elicited better sensitivity than *lbif* [$t1(24) = 5.88, P < 0.0001$; $t2(69) = 10.07, P < 0.0001$]. These results suggest that English speakers are sensitive to the syllable hierarchy even when their lip motor area is disturbed by TMS.

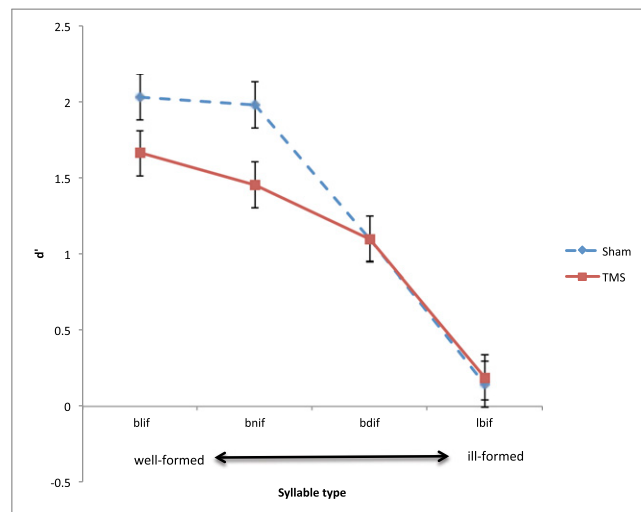


Fig. 2. The effect of TMS on participants’ sensitivity (d') to syllable structure. Error bars are 95% CIs for the difference between the means.

Experiment 2

To further assess the generality of these findings, in experiment 2, we gauged participants' sensitivity to another set of monosyllables that were all nasal-initial, unattested in English (e.g., *mlif* vs. *mdif*). Because these two syllable types are closely matched for their articulatory sequence (they both begin with lip-tongue sequences), this manipulation offers a more precise test of the role of the lip sensorimotor region in their perception.

To generate these monosyllables (e.g., *mlif*), we asked a native English speaker to produce their disyllabic counterparts (e.g., *melif*), and we gradually spliced out the vowel *e* in steady increments, resulting in a six-step continuum ranging from the monosyllable (e.g., *mlif*, step 1) to a disyllable (e.g., *melif*, step 6). The critical comparison concerns the structure of the monosyllables in step 1—sequences such as *mlif* or *mdif*. Syllables like *mdif* are worse-formed relative to *mlif*, and they are also more prone to misidentification as disyllables (e.g., *mdif*→*medif*) (5, 6). If sensitivity to syllable structure relies on motor simulation, then TMS should attenuate the identification of monosyllables (at step 1), and its effect should be stronger for ill-formed, *mdif*-type syllables relative to the better-formed *mlif*-ones.

The effect of bursts of four TMS pulses was evaluated by means of a 2 TMS/sham × 2 syllable type (*mlif/mdif*) × 6 vowel duration ANOVA over the proportion of disyllabic responses (arcsine transformed). Results yielded a reliable interaction of syllable type × vowel step [$F(5,35) = 6.28, P < 0.0005$], which was not further modulated by TMS ($F < 1$).*

Inspection of the means (Fig. 3) suggests that TMS impaired identification. Contrary to the motor simulation hypothesis, however, TMS attenuated the identification of the better-formed *mlif* continuum, an effect that was marginally significant [$t(7) = 2.18, P < 0.07$]. In contrast, TMS had no measurable effect on the identification of the ill-formed *mdif*-continuum [e.g., *mdif*; $t(7) = 1.20, P > 0.27$]. Moreover, regardless of TMS, participants remained highly sensitive to syllable structure, resulting in the systematic misidentification of ill-formed *mdif*-type syllables relative to their better-formed *mlif*-type counterparts (in steps 1–5, Tukey HSD, $P < 0.01$ for both TMS and sham). As expected, responses to disyllables (in step 6) did not differ ($P > 0.49$), as both are well formed. These results converge with the findings of experiment 1 to suggest that people's sensitivity to the syllable hierarchy persists even when motor simulation is suppressed.

Experiment 3

The persistent sensitivity to the syllable hierarchy, despite the disruption of the lip motor area, appears to challenge the action-based explanation for the syllable hierarchy. It is possible, however, that the syllable effect survived TMS because motor simulation was irrelevant to the identification of these stimuli. One possibility is that the lip motor area was not specifically engaged in motor planning—an explanation that appears unlikely, given that TMS disrupted the production of labial syllables (Fig. S1), that all syllables in experiments 1 and 2 included a labial speech sound, and that the items in experiment 2 were strictly matched for their articulatory sequence (i.e., all comprised of labial-coronal combinations). Alternatively, ill-formed syllables (e.g., *lbif*, *mdif*) might have escaped the effect of TMS because their articulatory program was just too difficult to simulate even under the sham condition (i.e., a floor effect).

*Because the materials included three item pairs, our statistical analyses only used participants as a random variable. To further ensure that the ANOVA outcomes are not the result of artifacts associated with binary data, we also submitted the data to a mixed effect logit model. In line with the ANOVA, the logit analysis yielded a reliable interaction of vowel by syllable type ($\beta = -0.29, SE = 0.11, Z = -2.67, P < 0.008$), which was not further modulated by TMS ($\beta = 0.08, SE = 0.12, Z \leq 0.7, P < 0.53$, not significant).

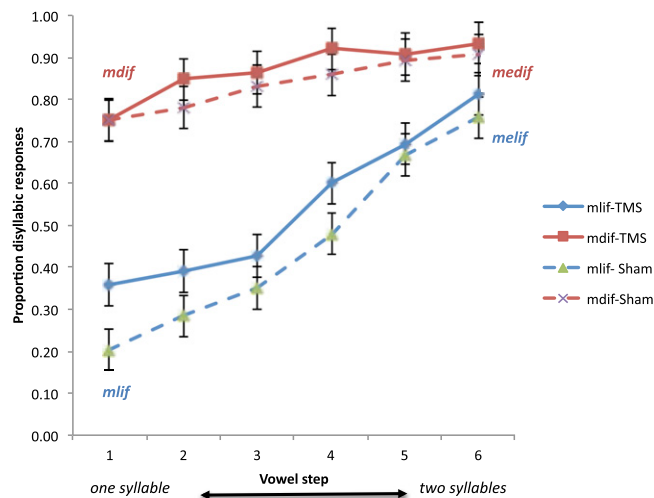


Fig. 3. The effect of TMS on the identification of nasal syllables. As the duration of the initial vowel (*e*) increased (e.g., from *mlif*, in step 1 to *melif*, in step 6), people were more likely to identify the stimulus as disyllabic. However, ill-formed monosyllables (e.g., *mdif*) were misidentified as disyllabic even in step 1, and this effect persisted even under TMS. Error bars are 95% CIs for the difference between the means.

If these explanations are correct, then the increased demands associated with the processing of ill-formed syllables should be evident in an increase in metabolic brain activity. To evaluate this possibility, in experiment 3 we used fMRI to gauge the involvement of lip motor area in performing syllable count with the same task and stimuli used in experiment 1. Of interest is whether the cortical motor region targeted by TMS in experiment 1 is activated by ill-formed syllables.

The fMRI experiment identified a sensorimotor lip area (14, 25) that was sensitive to the syllable hierarchy. This area matched the TMS stimulation site from experiments 1 and 2 (Fig. 4; for additional areas of activation, see ref. 26).

A 2 syllable (monosyllables vs. disyllables) × 4 type (large sonority rise, small rises, plateaus and falls e.g., *blif*, *bnif*, *bdif*, *lbif*, with a linear contrast of $[-3/4 -1/4 1/4 3/4]$) whole-brain voxel-wise ANCOVA of the BOLD signal yielded an interaction in sensorimotor cortex at Brodmann area 4, in the region of the lip area (25). This interaction was significant ($z = -3.02$).

The corresponding simple main effects of syllable type were significant in the ANCOVA for both monosyllables and disyllables (initial $P < 0.01$, uncorrected).

Contrary to the motor account, however, ill-formed syllables were associated with a decrease (rather than an increase) in the BOLD signal, whereas their disyllabic counterparts (e.g., *lebif*) tended to increase activation [possibly because these disyllables began with a sonorant sound, the perception of which engages the adjacent (25) larynx motor area].[†]

General Discussion

A large body of research indicates that the sound patterns of human languages are strictly and systematically constrained. Across languages, certain syllable types (e.g., *blog*) are overrepresented relative to others (e.g., *lbog*), and preferred syllables are more readily identified by individual speakers—both adults (14, 16–20) and infants (27). However, whether such preferences

[†]The increased BOLD response to disyllables is not an artifact of their greater length, as activation in Broca's area (left posterior BA 45) exhibited the opposite pattern (i.e., stronger activation to monosyllables, and stronger effects for ill-formed syllables compared with well-formed ones); for details see ref. 26.

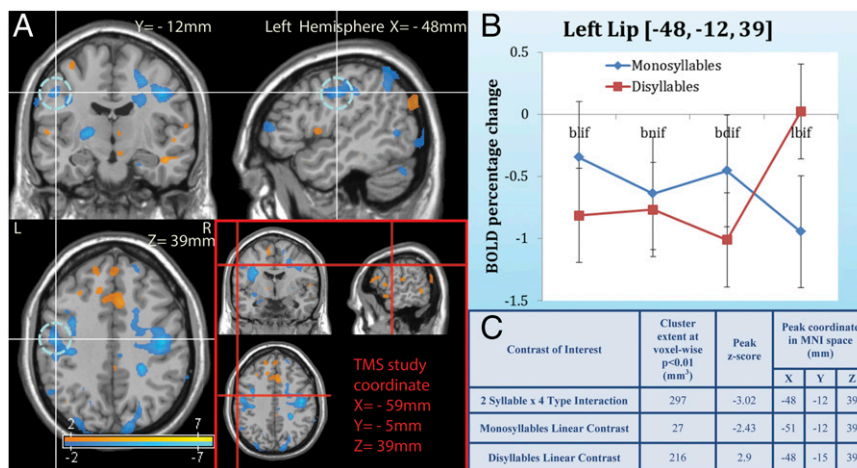


Fig. 4. The effect of the syllable hierarchy on the sensorimotor lip area. Syllable structure deactivated the lip sensorimotor area in the fMRI experiment [$P < 0.01$, uncorrected, at MNI cortical coordinate $(-48, -12, 39)$], and this region was adjacent to the cortical area projected from the coordinates of the cortical representation of the OO stimulation site in the TMS study [MNI cortical coordinates $(-59, -5, 39)$, shown by the red crosshatch in (A); these coordinates were averaged across all subjects]. Inspection of the BOLD signal (B) showed that as the syllable became worse formed, activation tended to decrease. The corresponding Z scores and coordinates are provided in C. Error bars are 95% CIs for the difference between the means. In this figure, we use an initial voxel-wise P value threshold at 0.05 for the purpose of visualization.

reflect abstract rules[‡] or the motor demands associated with speech production is unknown. Many previous studies have suggested that the identification of individual speech sounds (e.g., *b* vs. *p*) triggers motor simulation (14, 16–20, 28). These findings leave open the possibility that the combinatorial process that forms sound patterns (e.g., syllables such as *blog* vs. *lbog*) relies on algebraic rules, i.e., principles that are disembodied and abstract. To our knowledge, no previous research has examined this question.

Our research combines TMS and fMRI to examine this possibility. In line with the embodiment hypothesis, experiments 1–3 found that the identification of spoken syllables engages the articulatory motor system. These findings agree with previous studies (14, 16–20), demonstrating that speech perception automatically triggers action. However, our findings challenge the causal role of motor simulation in the computation of linguistic structure (29). Although the motor account predicts that ill-formed syllables should exert the greatest motor demands, our results show that ill-formed syllables are the least likely to engage the motor lip region and be impaired by TMS. These findings suggest that syllables like *lbif* are not disliked because of their online articulatory demands (i.e., greater motor demands → dispreference); rather, motor demands reflect their linguistic preference (dispreference → less simulation → fewer motor demands). Motor simulation is thus not the cause of linguistic preferences but its consequence.

Our results are limited inasmuch as the TMS procedure does not block motor simulation fully, and its application here was focused only on a single articulator—the lip. Such limitations notwithstanding, the most parsimonious explanation for the available evidence is that the motor system is preferentially engaged by well-formed structures (e.g., *blif*)—because they are well formed, easier to articulate, or similar to the syllables that exist in participants' language (i.e., English). The processing of cost of ill-formed syllables (e.g., *lbif*) must therefore arise from other sources—perhaps from the encoding of abstract linguistic restrictions on syllable structure that are shared across languages (7, 21).

[‡]We use rules to refer to algebraic operations over variables, applying to either inputs or outputs. This definition encompasses the narrower technical notions of both grammatical rules (i.e., operations on inputs) and grammatical constraints (i.e., operations over outputs), used in linguistics.

In line with this possibility, past research has shown that people's sensitivity to the syllable hierarchy is unlikely to result only from linguistic experience [similar results obtain in neonates (27) and in adult Korean speakers whose language lacks onset clusters altogether (4)].

It is also unlikely that the misidentification of ill-formed syllables only results from low-level auditory or phonetic difficulties that prevent people from accurately registering the phonetic form of auditory inputs such as *lbif*. First, our past research has shown that sensitivity to the syllable hierarchy is found with printed materials (i.e., in the absence of any auditory input) (5, 30). Conversely, demonstrable damage to the auditory/phonetic systems does not attenuate the sensitivity of dyslexic individuals to the syllable hierarchy (31). Other evidence against the auditory/phonetic explanation is provided by the fact that, when it comes to the phonetic categorization of individual sounds, it is the unfamiliar (i.e., nonnative) (28, 32–34) or degraded (i.e., masked by noise) (35) sounds that are most likely to engage the motor system. This result stands in stark contrast to our present findings, where motor simulation is selective to the most familiar syllable types (e.g., *blif*). Finally, the acoustic/phonetic account is directly countered by additional analyses of our own findings from experiment 1 (*SI Text*, Fig. S3, and Tables S1 and S2).

To gauge the contribution of low-level auditory/phonetic factors, we examined whether sensitivity to the syllable hierarchy can be captured by the duration and intensity of burst release—a phonetic property of stop consonants that has been linked to the misidentification of monosyllables (e.g., the misidentification of *bdif* and *bedif*) (36). Results showed that the burst modulated performance in the sham condition, but its effect was entirely eliminated by the TMS manipulation. In contrast, TMS did not eliminate the effect of the syllable hierarchy, and the effect of syllable type remained significant even after the properties of the burst are statistically controlled (Table S2). Although these findings do not rule out all acoustic or motor explanations, they are consistent with the possibility that the syllable hierarchy reflects linguistic principles that are abstract.

Our results illuminate the nuanced links between the sound pattern of language and its speech vessel. The motor system of speech production undoubtedly shapes language structure and mediates its online perception, but despite their inextricable bonds, these two systems appear to be distinct. Such a design is

only expected in view of the function of sound patterns in the language system as a whole. Indeed, the sound pattern of language has a double duty to follow (37). On the one hand, sound patterns must allow for efficient transmission of linguistic messages through the human body—preferentially, the oral articulatory system. At the same time, however, sound patterns must support the productivity of language—its capacity to form novel patterns by combining a small number of discrete elements. The reliance on abstract rules that are grounded in the speech system presents an adaptive solution that optimizes both pressures. How speech production shaped linguistic design in phylogeny and how it might constrain its acquisition in ontogeny warrant further study.

Materials and Methods

Methods for the TMS Experiments (Experiments 1 and 2).

Participants. We studied nine healthy, right-handed, native English speakers with normal hearing (mean age: 20.4 ± 1.59 y, six males) in experiment 1 and eight from the same group in experiment 2 (mean age: 20.6 ± 1.77 y, five males). Four additional subjects were excluded from the analysis, as their data were incomplete because of either programming errors ($n = 2$) or their withdrawal from the experiment ($n = 2$). All participants had acquired English before the age of 5 y and learned no other languages at home before the age of 10 y. Participants gave their written informed consent according to the procedures approved by the Institutional Review Boards of Beth Israel Medical Center and Northeastern University, and all study visits took place at the Berenson-Allen Center for Noninvasive Brain Stimulation at Beth Israel Deaconess Medical Center in Boston, MA. Handedness was assessed by the Edinburgh Handedness Inventory (38). Exclusion criteria included any psychiatric or neurological illnesses, metal in the brain, implanted medical devices, or intake of drugs listed as a potential hazard for the application of TMS (39).

Study design. In this single-blind, randomized cross-over study, participants received real or sham TMS in counterbalanced order on two study visits separated by at least 1 wk. During the experiment, participants were presented with novel auditory words and asked to indicate whether they included one syllable or two. Before each experiment, participants practiced the language tasks for 3 min. All stimuli were presented aurally through a computerized program and in-ear headphones. The stimulation target was determined by applying repetitive TMS pulses to the lip motor area (OO muscle). It was defined as the spot eliciting the largest TMS-induced motor evoked potentials (MEPs) in the OO muscle. Bursts of four biphasic TMS pulses (20 Hz) were triggered 50 ms after onset of each trial via key press by the participant. Fifty milliseconds after the fourth pulse, the auditory stimulus was presented, and participants were given 2.5-s response time followed by a rest period of 2 s (Fig. 1B). To ensure that the TMS manipulation affected articulation, we assessed speech production by an elicitation task, administered before experiments 1 and 2.

Elicitation task. During elicitation, participants were asked to produce 10 repetitions of a given syllable/syllable combination while undergoing TMS—either real or sham (the real/sham condition always matched the subsequent experimental session). Single syllables were *pa*, *ba*, *ta*, and *da*; the syllable combinations were *ba-pa* and *da-ta*. The effect of real TMS on this elicitation task was evident as illustrated in Fig. 1C. A formal analysis of the TMS on articulation was not done because (unlike the perceptual experiments), the application of TMS was not synchronized with the onset of production. The number of productions, the number of TMS applications, and their onsets were not controlled.

Experiment 1. The experiment presented four types of monosyllables beginning with a consonant cluster, along with their disyllabic counterparts. For the sake of expository convenience and comparison with our past research, we refer to those monosyllabic stimuli as *blif*, *bnif*, *bdif*, and *lbif* and to their disyllabic counterparts as *belif*, *benif*, *bedif*, and *lebif*; for the full list of stimuli, please see Table S3. These monosyllables differed on their well formedness across the syllable hierarchy (3, 24), ranging from the best formed (*blif*, type 1) to the worst formed (*lbif*, type 4). Most monosyllables (types 2–4) were unattested in English. Items were sampled from the materials used in past research (3, 4). They were arranged in quartets, matched for their rhyme (the final vowel and consonant), and were produced by a native Russian speaker (because all syllable types are possible in Russian, these stimuli can be produced naturally by speakers of this language). Disyllables contrasted with their matched monosyllables on the presence of a schwa between the two initial consonants (e.g., */blif/-bəlif/*). Each monosyllable type included 24 items, for a total of $24 \times 4 \times 2 = 192$ trials. The 24 quartets were divided into four blocks of 48 trials each. These

items were presented twice at two separate visits (randomized for order): once with real TMS and once with sham TMS (a total of 384 trials, eight blocks).

Experiment 2. The materials consisted of two types of continua used in our past research (6). Each such continuum ranged from a monosyllable (e.g., *mlif*) to a disyllable (e.g., *meliif*), by gradually increasing the duration of the schwa (*/ə/*) in six steps. In both continua, the monosyllabic end point was unattested in English, but in one continuum, the monosyllable was well formed across languages (e.g., *mlif*), whereas in the other, it was ill formed (e.g., *mdif*). These items were further matched for their articulatory sequence (labial-coronal; Table S4). The experiment included three such continua pairs, giving rise to 36 trials per block (3 pairs \times 6 steps). These blocks were repeated eight times (in counter balanced order): four times with real TMS, and four times with sham TMS (a total of 288 trials) on two separate visits.

EMG. EMG activity was recorded from the right first dorsal interosseus (FDI) and right OO muscles (Fig. 1A). The ground electrode was placed over the ulnar styloid process. EMG was band pass filtered (20–2,000 Hz) and digitized at a sampling rate of 5 kHz (Software: Spike2 V7; Device: Micro 1401; Cambridge Electronic Design Ltd.).

TMS. Neuronavigated TMS (in bursts of four biphasic pulses at 20 Hz) was applied using a figure-of-eight coil attached to a magnetic stimulator (Neuronavigation:Brainsight; Rogue Research; TMS: MagPro $\times 100$, coil: MCF-B65; MagVenture). Peak of motor cortical representation for OO was located on average at following MNI (40, 41) coordinates (experiment 1, *x/y/z*: $-68.97 \pm 9.02/-3.57 \pm 12.61/44.95 \pm 14.86$; experiment 2, *x/y/z*: $-70.23 \pm 6.09/-2.82 \pm 12.61/41.61 \pm 13.41$). The coil was placed tangentially on the head with the handle pointing backward at an angle of 45° (Fig. 1A). During sham stimulation, the edge of the coil was placed on the head. Resting motor threshold (RMT) for FDI and active motor threshold (AMT) for OO were assessed at the first study visit, and hotspots were marked on participants' anatomical MRI. These hotspots were used as stimulation targets during the experiments. If the first study visit involved sham stimulation, RMT and AMT were assessed again before the real TMS visit. RMT for FDI was defined as the minimum stimulation intensity required to elicit an MEP of at least 50 μ V in 5 of 10 trials. AMT for the OO muscle was defined as the minimal stimulation intensity required for eliciting an MEP of at least 200 μ V in 5 of 10 trials during isometric contraction. Participants practiced the maintenance of isometric contraction (pursing lips) with the help of visual feedback of ongoing muscle activity. They were then asked to maintain this activity level to assess AMT.

As the tolerance of single pulses is different from repetitive pulses, stimulation intensity was adjusted to participants' individual comfort level. Participants were initially given four consecutive pulses over the OO hotspot at 80% of AMT and intensity was gradually adapted in steps of 5% not exceeding 120% of AMT (experiment 1/2: mean % of stimulator output: $61.56 \pm 6.17/59.00 \pm 5.32$; mean % of AMT: $77.78 \pm 11.17/76.25 \pm 10.89$). Adverse effects of TMS were assessed with a questionnaire.

Methods for the fMRI Experiment (Experiment 3). Experiment 3 consisted of an interrogation of an existing data set (26) to address the role of the lip motor area in performing the syllable count task. The materials consisted of a subset of the materials from experiment 1, and the task was identical to experiment 1, but experiment 3 was administered to another group of participants, distinct from the one in experiments 1 and 2 ($n = 14$, $M = 22.57$ y, 10 females, right-handed native English speakers).

The NNL fMRI Hardware System (NordicNeuroLab) with E-Prime2.0 Professional software (Psychology Software Tools) was configured and programmed for sensory stimulus delivery and response recording that were synchronized with a Siemens MAGNETOM TIM Trio 3-T MRI scanner (VB17A; Siemens Medical Solutions), equipped with a standard 12-channel head coil. The fMRI experiment was conducted with a tailored scanning protocol with two anatomical image acquisitions and a series of fMRI runs using a modified gradient echo EPI sequence that allows one to insert periods of "silent" time in the pulse sequence (during which magnetization is maintained). The auditory stimuli were presented only during the predetermined silent gaps in the acquisition chain (42, 43) and synchronized with the auditory stimulus presentation via E-Prime and NNL fMRI Hardware System. In this event-related experimental design, each trial had only one silent volume (1.2 s, during which the short 1-s stimulus was delivered), followed by 10 volumes of acquisition (12 s, resting state, serving as the baseline).

The fMRI imaging data processing procedures were performed using laboratory optimized Statistical Parametric Mapping (SPM) software (44, 45), and a whole-brain voxel-wise multilevel random-effects model in an ANCOVA setting was estimated to detect activation and deactivation patterns in BOLD signal with particular focus on the effect of the syllable \times type interaction at the lip sensorimotor area. Additional information concerning the methods and analysis are detailed in ref. 26.

ACKNOWLEDGMENTS. We thank Robert Painter for assistance with the phonetic transcription of the materials. Work for this study was supported, in part, by grants from the Sidney R. Baer Jr. Foundation, the National Institutes of Health (R01 HD069776, R01 NS073601, R21 MH099196, R21

NS082870, R21 NS085491, and R21 HD07616), and Harvard Catalyst, the Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH, UL1RR025758).

- Hockett CF (1960) The origin of speech. *Sci Am* 203:89–96.
- Greenberg JH (1978) Some generalizations concerning initial and final consonant clusters. *Universals of Human Language*, eds Greenberg JH, Ferguson CA, Moravcsik EA (Stanford Univ Press, Stanford, CA), Vol 2, pp 243–279.
- Berent I, Steriade D, Lennertz T, Vaknin V (2007) What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104(3):591–630.
- Berent I, Lennertz T, Jun J, Moreno MA, Smolensky P (2008) Language universals in human brains. *Proc Natl Acad Sci USA* 105(14):5321–5325.
- Berent I, Lennertz T, Smolensky P, Vaknin V (2009) Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26(1):75–108.
- Berent I, Lennertz T, Balaban E (2012) Language universals and misidentification: A two-way street. *Lang Speech* 55(Pt 3):311–330.
- Prince A, Smolensky P (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar* (Blackwell Publishers, Malden, MA).
- Pulvermüller F, Fadiga L (2010) Active perception: Sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11(5):351–360.
- MacNeilage PF (2008) *The Origin of Speech* (Oxford Univ Press, Oxford, UK), p xi.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74(6):431–461.
- Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13(3):361–377.
- Mattingly IG (1981) Phonetic representation and speech synthesis by rule. *The Cognitive Representation of Speech*, eds Myers T, Laver J, Anderson J (North Holland, Amsterdam), pp 415–420.
- Wright R (2004) A review of perceptual cues and robustness. *Phonetically Based Phonology*, eds Steriade D, Kirchner R, Hayes B (Cambridge Univ Press, Cambridge, UK), pp 34–57.
- Pulvermüller F, et al. (2006) Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci USA* 103(20):7865–7870.
- Kobayashi M, Pascual-Leone A (2003) Transcranial magnetic stimulation in neurology. *Lancet Neurol* 2(3):145–156.
- Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur J Neurosci* 15(2):399–402.
- D'Ausilio A, et al. (2009) The motor somatotopy of speech perception. *Curr Biol* 19(5):381–385.
- Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29(31):9819–9825.
- Möttönen R, van de Ven GM, Watkins KE (2014) Attention fine-tunes auditory-motor processing of speech sounds. *J Neurosci* 34(11):4064–4069.
- D'Ausilio A, et al. (2014) Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philos Trans R Soc Lond B Biol Sci* 369(1644):20130418–20130418.
- Chomsky N, Halle M (1968) *The Sound Pattern of English* (Harper & Row, New York), pp xiv, 470.
- Caramazza A, Chialant D, Capasso R, Miceli G (2000) Separable processing of consonants and vowels. *Nature* 403(6768):428–430.
- Miozzo M, Buchwald A (2013) On the nature of sonority in spoken word production: Evidence from neuropsychology. *Cognition* 128(3):287–301.
- Clements GN (1990) The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, eds Kingston J, Beckman M (Cambridge Univ Press, Cambridge, UK), pp 282–333.
- Brown S, Ngan E, Liotti M (2008) A larynx area in the human motor cortex. *Cereb Cortex* 18(4):837–845.
- Berent I, et al. (2014) Language universals engage Broca's area. *PLoS ONE* 9(4):e95155.
- Gómez DM, et al. (2014) Language universals at birth. *Proc Natl Acad Sci USA* 111(16):5837–5341.
- Kuhl PK, Ramirez RR, Bosseler A, Lin J-FL, Imada T (2014) Infants' brain responses to speech suggest analysis by synthesis. *Proc Natl Acad Sci USA* 111(31):11238–11245.
- Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J Physiol Paris* 102(1-3):59–70.
- Berent I, Lennertz T (2010) Universal constraints on the sound structure of language: Phonological or acoustic? *J Exp Psychol Hum Percept Perform* 36(1):212–223.
- Berent I, Vaknin-Nusbaum V, Balaban E, Galaburda AM (2013) Phonological generalizations in dyslexia: The phonological grammar may not be impaired. *Cogn Neuropsychol* 30(5):285–310.
- Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* 33(1):316–325.
- Callan DE, et al. (2003) Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* 19(1):113–124.
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22(3):1182–1194.
- Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci USA* 111(19):7126–7131.
- Wilson C, Davidson L, Martin S (2014) Effects of acoustic-phonetic detail on cross-language speech production. *J Mem Lang* 77(0):1–24.
- Berent I (2013) *The Phonological Mind* (Cambridge Univ Press, Cambridge, UK).
- Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9(1):97–113.
- Rossi S, Hallett M, Rossini PM, Pascual-Leone A; Safety of TMS Consensus Group (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol* 120(12):2008–2039.
- Mazziotta J, et al. (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci* 356(1412):1293–1322.
- Mazziotta J, et al. (2001) A four-dimensional probabilistic atlas of the human brain. *J Am Med Assoc* 285(5):401–430.
- Schwarzbauer C, Davis MH, Rodd JM, Johnsrude I (2006) Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *Neuroimage* 29(3):774–782.
- Yang Y, et al. (2000) A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magn Reson Med* 43(2):185–190.
- Frackowiak RSJ (2004) *Human Brain Function* (Elsevier Academic Press, Amsterdam).
- Pan H, Epstein J, Silbersweig DA, Stern E (2011) New and emerging imaging techniques for mapping brain circuitry. *Brain Res Brain Res Rev* 67(1-2):226–251.